OKANAGAN Université Paul Sabatier (Toulouse III)

INSTITUT de MATHEMATIQUES de TOULOUSE

Habilitation à Diriger des Recherches

Section 26 : Mathématiques Appliquées et Applications des Mathématiques
Spécialité : Analyse Convexe, Optimisation, Calcul Scientifique

Contributions à l'Analyse Convexe Computationnelle (AC^2).

Contributions to Computer-Aided Convex Analysis (CA^2)

présentée le 3 juin 2008 à

l'Université Paul Sabatier de Toulouse

par

Yves LUCET

Docteur de l'Université Paul Sabatier de Toulouse (France)
Professeur Assistant à l'Université de Colombie Britannique (Canada)

devant le jury composé de :

G. CARLIER	Professeur à l'Université de Paris-Dauphine (Paris IX)	(Rapporteur)
P. L. COMBETTES	Professeur à l'Université P. et M. Curie (Paris VI)	(Rapporteur)
R. CORREA	Professeur à l'Université du Chili à Santiago du Chili	(Rapporteur)
J.-B. HIRIART-URRUTY	Professeur à l'Université Paul Sabatier (Toulouse III)	(Directeur)
J.-B. LASSERRE	Directeur de recherche au CNRS, LAAS de Toulouse	(Examinateur)
D. NOLL	Professeur à l'Université Paul Sabatier (Toulouse III)	Examinateur
M. THÉRA	Professeur à l'Université de Limoges	(Examinateur)
J. YE	Professeur à l'Université de Victoria, Colombie Britannique (Canada)	(Examinatrice)

Remerciements

Mes remerciements vont tout d'abord à Jean-Baptiste Hiriart-Urruty, Professeur à l'Université Paul Sabatier, pour m'avoir suggéré de présenter cette Habilitation à Diriger des Recherches (HDR) et m'avoir encouragé tout au long de ce travail. J'ai particulièrement apprécié notre travail en commun sur la transformée de Legendre paramétrée et sa suggestion d'écrire mon article de revue des applications de l'analyse convexe computationnelle. Ses conseils ont profondément marqué ma carrière d'enseignant-chercheur.

Je remercie particulièrement mon collègue et collaborateur, Professeur Heinz Bauschke de l'Université de Colombie Britannique, pour m'avoir accueilli dans son cercle de collaborateurs et suggéré de travailler ensemble sur l'étude de l'opérateur proximal. Ses encouragements ont grandement amélioré la qualité de mes travaux. Le Professeur Jon Borwein occupe une place spéciale dans mes remerciements pour m'avoir encouragé à étudier l'analyse convexe computationnelle : mes travaux auraient été bien différents si je ne n'avais pas eu la chance de le rencontrer. Je lui suis aussi très reconnaissant pour son accueil chaleureux lors de mon arrivée au Canada. Merci à mes autres co-auteurs : Shawn Wang, Boris Mordukhovich, Jane Ye, Rafal Goebel pour m'avoir donné la chance de travailler ensemble sur des problèmes fascinants.

Ma gratitude va aussi aux rapporteurs de ce travail, les Professeurs G. Carlier, P. Combettes et R. Correa pour avoir pris le temps de lire en profondeur mes travaux et pour leurs suggestions pertinentes. Merci au Professeur D. Noll pour avoir accepté de présider ce jury. J'ai particulièrement apprécié sa présence, ainsi que celle de J.-B. Lasserre, Directeur de recherche au CNRS LAAS de Toulouse, qui reviennent 11 ans après ma thèse de doctorat examiner à nouveau mes travaux. Merci enfin aux Professeurs J. Ye et M. Théra d'avoir participé à ce jury.

Mes remerciements vont aussi à tous les étudiants qui ont accepté de travailler sous ma supervision tout au long de ces années.

Merci aux membres du groupe OCANA de constituer un environnement de travail unique et si agréable. Merci à mes collègues du département d'informatique de l'université de Colombie Britannique du campus de Kelowna ainsi qu'à mes collègues de l'unité 4 de la faculté d'arts et de sciences pour constituer un environnement de travail si agréable.

Finalement ce travail et ma carrière n'auraient pas pu voir le jour sans le soutien indéfectible de ma femme Patricia Lasserre, Professeur à l'Université de Colombie Britannique. Sa présence dans le même département m'a permis de me consacrer pleinement à mon travail d'enseignant-chercheur. Merci plus particulièrement pour ses nombreux conseils et ses encouragements durant toutes ces années.

Table des matières

1 Curriculum Vitae **7**
1.1 Activité professionnelle actuelle 7
1.2 Diplômes 7
1.3 Expérience professionnelle antérieure 8
1.4 Activité en matière d'enseignement 8
1.5 Activité antérieure en enseignement 8

2 Synthèse des travaux de recherche présentés **9**
2.1 Analyse convexe computationnelle 9
2.1.1 Calcul paramétrique de la transformée de Legendre-Fenchel et application au calcul de l'enveloppe de Moreau [5] 9
2.1.2 Un algorithme linéaire basé sur la transformée de Legendre en temps linéaire pour le calcul de la transformée en distance Euclidienne [6] 10
2.1.3 Nouveaux algorithmes séquentiels pour le calcul exact de la transformée en distance Euclidienne [8] 10
2.1.4 Calcul rapide de l'enveloppe de Moreau : Algorithmes numériques [7] 10
2.1.5 Le modèle linéaire-quadratique par morceaux pour l'analyse convexe computationnelle [10] 11
2.1.6 De quelle forme est votre conjuguée ? Une revue de l'analyse convexe computationnelle et de ces applications [9] 11
2.2 Analyse convexe 13
2.2.1 Ensembles et fonctions convexes compactement epi-Lipschitz [4] 13
2.2.2 Comment transformer de façon continue une fonction convexe en une autre fonction convexe [2] ? 13
2.2.3 Méthodes intrinsèques primales-duales pour les primitives d'opérateurs cycliquement monotones [3] 14
2.2.4 Propriétés de l'opérateur de moyenne proximale [1] 14
2.3 Analyse non lisse 15
2.3.1 Analyse de sensibilité pour la fonction valeur de problèmes d'optimisation avec des contraintes de type inégalités variationnelles [11] . 15

3 Formation et encadrement **17**
3.1 Thèses 17

3.2 Comités de thèses et encadrement d'étudiants 17

4 Diffusion et transfert de connaissances 19

4.1 Bourses et contrats de recherche 19

4.1.1 Financements externes 19

4.1.2 Financements internes 19

4.1.3 Autres financements 20

4.2 Articles publiés dans des revues à comité de lecture 20

4.3 Actes de conférences avec comité de lecture 21

4.4 Articles soumis pour publication (dans des revues à comité de lecture) . . . 21

4.5 Logiciels 21

4.6 Conférences et présentations 21

4.7 Rédaction de rapports de lecture-arbitrage (*referee reports*) 22

4.8 Responsabilités administratives 23

4.9 Sociétés savantes 23

5 Exemplaire des onze publications présentées (travaux de recherche) 25

Chapitre 1

Curriculum Vitae

Yves Lucet

Assistant Professor

Optimisation, Analyse Convexe, Calcul Scientifique

9112 Buchanan Rd
Coldstream BC V1B 3B7 Canada
Tél. : +1 250 545 6062
Canadien et Français
Né à Toulouse le 23 novembre 1970

Computer Science
I. K. Barber School of Arts & Sciences #4
The University of British Columbia
3333 University Way,
Kelowna BC V1V 1V7 Canada
Tél. : +1 250 807 9505
yves.lucet@ubc.ca
http ://people.ok.ubc.ca/ylucet/

1.1 Activité professionnelle actuelle

Assistant Professor (depuis juillet 2005) à I. K. Barber School of Arts & Sciences, The University of British Columbia Okanagan, Kelowna BC, Canada.

1.2 Diplômes

Doctorat en Optimisation et Analyse Convexe, Université Paul Sabatier, Toulouse, 1997. Thèse : *La transformée de Legendre-Fenchel et la convexifiée d'une fonction : Algorithmes rapides de calcul, analyse et régularité du second ordre.*
Directeur : Jean-Baptiste Hiriart-Urruty.

Diplôme d'études approfondies en Mathématiques appliquées, Université Paul Sabatier, Toulouse, 1993. Directeur : Jean-Baptiste Hiriart-Urruty.

Maîtrise en Mathématiques, Université Paul Sabatier, Toulouse, 1992

Licence en Mathématiques, Université Paul Sabatier, Toulouse, 1991

Diplôme d'Études Universitaires Générales (DEUG) mention sciences, section sciences des structures et de la matière, Université Paul Sabatier, Toulouse, 1990

1.3 Expérience professionnelle antérieure

Août 2002 à juin 2005 : Assistant Professor à Okanagan University College, Kelowna, BC Canada.

Septembre 1999 à juillet 2002 : Consultant en informatique et en optimisation, Salmon Arm BC, Canada.

Mai 1999 à août 1999 : Boursier post-doctoral au Centre for experimental and constructive Mathematics, Simon Fraser University, Canada, avec Jon Borwein.

Juin 1998 à avril 1999 : Boursier post-doctoral au Department of mathematics and statistics, University of Victoria, Canada, avec Jane Ye.

Janvier 1998 à mai 1998 : Boursier post-doctoral au Department of mathematical and statistical sciences, University of Alberta, Canada, avec René Poliquin.

Septembre 1997 à décembre 1997 : Boursier post-doctoral au Centre for experimental and constructive Mathematics, Simon Fraser University, avec Jon Borwein.

Septembre 1993 à août 1997 : Moniteur de l'enseignement supérieur et boursier de l'enseignement supérieur et de la recherche à l'Université Paul Sabatier.

1.4 Activité en matière d'enseignement

Depuis mon arrivée à l'Université de Colombie Britannique (juillet 2005), j'ai enseigné des cours d'informatique de la deuxième à la quatrième année en analyse algorithmique (Analysis of Algorithms), optimisation numérique (cours et TP ; Numerical Optimization), le cours d'introduction aux structures discrètes (Introduction to Discrete Structures), les deux cours en ingénierie du logiciel (cours, TD et TP ; Software Engineering et Software Engineering Project), et un cours d'analyse convexe (directed studies in Computer Science). J'ai aussi enseigné des cours doctoraux (graduate courses) en optimisation numérique et analyse convexe.

1.5 Activité antérieure en enseignement

À Okanagan University College j'ai enseigné de la première à la quatrième année les cours suivants : ingénierie du logiciel (Software Engineering et Software Engineering Project), programmation visuelle (Visual Programming II), XML (Topics in Computer Science XML), un cours spécialisé en Mathématiques (Directed Studies in Mathematics), deux cours spécialisés en informatique (Directed Studies in Computer Science), le cours de logique digitale (Digital Logic and Microcomputer Hardware), le cours de projet en informatique (Projects in Computer Science), le cours de graphisme (Computer Graphics) et le cours sur les systèmes d'exploitations (Technical Aspects of Operating Systems).

J'ai enseigné les deux premiers cours d'analyse en mathématiques (cours et TP ; Calculus I et Calculus II) à l'Université de Victoria et le cours d'algèbre linéaire (Linear Algebra) à l'Université d'Alberta. Lors de mon contrat de moniteur à l'Université Paul Sabatier, j'ai enseigné les TD et TP des cours d'analyse numérique de deuxième année et de licence de mathématique.

Chapitre 2

Synthèse des travaux de recherche présentés

Chacun des onze articles que nous présentons ici appartient à l'une des trois catégories suivantes : l'analyse non lisse, l'analyse convexe et l'analyse convexe computationnelle. Ces catégories sont clairement disjointes, elles sont même emboîtées : la dernière étant un cas particulier de la seconde, qui est elle-même un cas particulier de la première.

L'analyse convexe se concentre sur l'étude des fonctions et ensembles convexes et en particulier des problèmes d'optimisation avec données convexes. Elle étend les résultats classiques en calcul différentiel en remplaçant l'hypothèse de différentiabilité par la convexité. L'analyse non lisse généralise ces résultats en remplaçant l'hypothèse (globale) de convexité par des propriètés (locales) de régularité. Enfin l'analyse convexe computationnelle comprend le calcul numérique et symbolique des objets (opérateurs, fonctions et ensembles) fondamentaux de l'analyse convexe. Le terme "computationnel" vient du latin computus (pour calcul) qui a aussi donné le nom comput.

2.1 Analyse convexe computationnelle

Nos travaux en analyse convexe computationnelle sont contenus dans cinq articles [5, 7, 8, 9, 10] et ont fait l'objet d'une conférence avec actes [6]. Nous avons étudié les algorithmes rapides de calcul, proposé de nouveaux algorithmes et de nouveaux modèles, étudié leurs applications pour le calcul de la transformée en distance et recensé les applications de ces algorithmes dans des domaines divers dont l'analyse d'images et les équations aux dérivées partielles.

2.1.1 Calcul paramétrique de la transformée de Legendre-Fenchel et application au calcul de l'enveloppe de Moreau [5]

Nous présentons un nouvel algorithme de calcul numérique de la transformée de Legendre-Fenchel (conjuguée) pour les fonctions d'une variable réelle. L'algorithme est basé sur le calcul des points exposés du graphe de la conjuguée et sur la fonction dite perspective fermée. En utilisant les outils de l'analyse convexe, nous montrons que la paramétrisation $s = f'(x)$, $f^*(s) = sx - f(x)$ permet de reconstruire le graphe de la conjuguée sauf pour les parties qui demandent une interpolation linéaire.

L'algorithme est étendu au calcul de l'enveloppe de Moreau en utilisant la relation directe entre les deux transformées. Dans ce cas, l'interpolation linéaire est remplacée par une interpolation quadratique. Nous présentons aussi des comparaisons numériques avec des algorithmes de calcul rapide de l'enveloppe de Moreau. Tous les algorithmes comparés ont une complexité linéaire mais l'algorithme paramétrique jouit d'une plus petite constante de proportionnalité. Une implémentation en Scilab de l'algorithme montre une accélération par un facteur 400 par rapport aux algorithmes existants.

Les avantages de l'algorithme paramétré sont sa simplicité et sa rapidité. Ses inconvénients comprennent sa limitation aux fonctions convexes d'une variable et le besoin d'interpoler le résultat.

2.1.2 Un algorithme linéaire basé sur la transformée de Legendre en temps linéaire pour le calcul de la transformée en distance Euclidienne [6]

Dans ces actes de conférence nous expliquons comment calculer la transformée en distance Euclidienne en appliquant un algorithme de transformation rapide, la transformée de Legendre en temps linéaire (LLT ; Linear time Legendre transform). À partir d'une image binaire, nous transformons les données avec une pénalisation infinie pour calculer l'enveloppe de Moreau en appliquant l'algorithme LLT. L'algorithme résultant a une complexité linéaire et est trivialement parallélisable. Nous montrons ensuite comment la même approche peut être étendue à des images non binaires.

2.1.3 Nouveaux algorithmes séquentiels pour le calcul exact de la transformée en distance Euclidienne [8]

Le problème du calcul de la transformée en distance d'une image binaire a fait l'objet de plusieurs algorithmes en analyse d'images. Nous le considérons comme un cas particulier du calcul de l'enveloppe de Moreau. Les algorithmes de calcul rapide permettent alors son calcul en temps linéaire. Nous comparons les temps de calculs des différents algorithmes rapides dans ce contexte. Les résultats illustrent l'intérêt du compromis entre la généralité d'un algorithme comme le Linear-time Legendre transform (LLT) qui ne demandent pas d'hypothèse de convexité et l'algorithme Non Expansive Prox (NEP) qui atteint une plus grande rapidité mais ne s'applique qu'aux données convexes.

2.1.4 Calcul rapide de l'enveloppe de Moreau : Algorithmes numériques [7]

Nous récapitulons les algorithmes rapides pour le calcul de l'enveloppe de Moreau et présentons un nouvel algorithme, baptisé NEP (Non-Expansive Prox), basé sur la propriété de contraction de l'application proximale. Le temps de calcul des algorithmes est comparé numériquement et donne l'avantage à l'algorithme NEP. La rapidité de ce dernier est contre balancé par sa restriction aux fonctions convexes.

Les différents algorithmes sont implémentés dans une boîte à outils, publié avec l'article, sous le logiciel de calcul mathématique Scilab. Cette boîte à outils comprend une aide en ligne, des démonstrations des différents algorithmes ainsi qu'une batterie de tests. L'article détaille les différentes fonctions disponibles ainsi que des illustrations des transformées rapides calculables par le logiciel.

2.1.5 Le modèle linéaire-quadratique par morceaux pour l'analyse convexe computationnelle [10]

Un nouveau modèle pour l'analyse convexe computationnelle est introduit. Les algorithmes linéaires-quadratiques par morceaux utilisent les propriétés des fonctions linéaires-quadratiques par morceaux pour produire un calcul stable et robuste des transformées utilisées en analyse convexe. Leur propriété fondamentale est que la transformée de toute fonction linéaire-quadratique par morceaux est aussi linéaire-quadratique par morceaux, et ceci pour les quatre opérations fondamentales de l'analyse convexe : l'addition, la multiplication par un scalaire, la conjugaison de Legendre-Fenchel et la régularisation par enveloppe de Moreau.

Cette propriété de fermeture permet d'améliorer significativement les résultats de convergence connus précédemment dans le cadre des algorithmes rapides. Les nouveaux algorithmes résultants sont de complexité linéaire, donc aussi efficaces que les algorithmes rapides, et améliorent aussi la modélisation du domaine. Le nouveau modèle permet un calcul exact (symbolique) pour les fonctions linéaires-quadratiques par morceaux. Ainsi, nous pouvons calculer symboliquement les transformées de fonctions indicatrices, linéaires ou quadratiques définies par morceaux et toutes combinaisons de ces fonctions y compris celles qui prennent la valeur $+\infty$ à certains points (fonctions à valeurs étendues). La richesse de cette classe permet d'illustrer bien des résultats de l'analyse convexe et de rechercher des contre-exemples. Ces algorithmes sont qualifiés de symboliques-numériques car ils sont exacts pour les fonctions linéaires-quadratiques par morceaux. L'approximation numérique de toute autre fonction est calculée avant l'invocation des algorithmes, ce qui permet de séparer clairement l'étape d'approximation numérique de l'étape de calcul de la transformée proprement dit. Le modèle permet aussi une meilleure approximation car il manipule directement des fonctions d'ordre deux. Il permet ainsi l'utilisation d'algorithmes d'interpolation quadratiques.

Dans ce travail, les algorithmes PLQ sont limités aux fonctions convexes (à valeurs étendues, $+\infty$) d'une variable (l'extension à des fonctions non convexes d'une variable a fait l'objet d'une partie du mémoire de mastère de M. Trienis soutenu en décembre 2007).

2.1.6 De quelle forme est votre conjuguée ? Une revue de l'analyse convexe computationnelle et de ces applications [9]

Cet article constitue une revue de l'analyse convexe computationnelle et présente une large variété d'applications. Dans une première partie, nous rappelons les définitions des transformées fondamentales de l'analyse convexe. Nous présentons ensuite les algorithmes pour calculer ces dernières : Algorithmes de calcul symbolique, algorithmes rapides et algorithmes linéaires-quadratiques par morceaux. La deuxième partie de l'article présente un ensemble de problèmes bénéficiant de ces algorithmes.

Le problème d'intégration des fonctions convexes, étant donné un échantillon fini de leurs sous-gradients est résolu par l'opérateur de moyenne proximale qui est calculable rapidement par les algorithmes linéaires-quadratiques par morceaux. Les flots dans les réseaux séries-parallèles sont modélisables par une suite d'inf-convolutions de fonctions. Leur calcul peut donc être effectué par des algorithmes rapides. Le problème de la transition de phase en thermodynamique requiert le calcul d'inf-convolutions de fonctions, alors que l'étude de l'équilibre thermodynamique demande le calcul de l'enveloppe convexe d'une fonction. Les réseaux électriques (ainsi que les systèmes mécaniques comprenant des ressorts en série ou en parallèles) sont aussi modélisés par des inf-convolutions de fonctions. Tous ces problèmes bénéficient des algorithmes efficaces de l'analyse convexe computationnelle. De même, la navigation d'un robot a été effectuée en calculant la transformée de pente (une généralisation de la conjuguée de Legendre-Fenchel) ou par le calcul de la transformée en distance (un cas particulier de l'enveloppe de Moreau).

Dans une deuxième sous-partie, nous considérons les applications en analyse d'images, vision par ordinateur et morphologie mathématique. Ceci constitue un large domaine d'application des algorithmes d'analyse convexe computationnelle. Certains de ces algorithmes ont d'ailleurs été développés spécifiquement pour ces applications. Le troisième large domaine d'application est constitué par (certaines) équations aux dérivées partielles. Les formules de Lax et Hopf permettent de calculer des solutions des équations de Hamilton-Jacobi par des algorithmes d'analyse convexe computationnelle, alors que les schémas numériques de résolution d'équations aux dérivées partielles permettent de calculer certaines transformées comme l'enveloppe convexe. Le dernier large domaine d'application de notre article est constitué par les algèbres max-plus, les réseaux de communications et l'analyse multi-fractale. Dans tous ces domaines, le calcul efficace des transformées de l'analyse convexe facilite la résolution des problèmes. Le cas des algèbres max-plus est particulièrement intéressant : la modélisation de systèmes discrets demande l'extension des résultats classiques à des algèbres exotiques, souvent max-plus ou min-plus.

2.2 Analyse convexe

Quatre articles traitent de problèmes en analyse convexe [1, 2, 3, 4]. L'opérateur de moyenne proximale a été à la base de trois de ces travaux. Ses propriétés d'homotopie, d'héritage de régularité et de conjugaison en ont fait un outil remarquable pour les problèmes en analyse convexe.

2.2.1 Ensembles et fonctions convexes compactement epi-Lipschitz [4]

Nous donnons ici plusieurs caractérisations de la propriété dite de compacité epi-Lipschitz pour les ensembles convexes fermés. Nous démontrons en particulier qu'un ensemble convexe fermé est compactement epi-Lipschitz si, et seulement si, il a un intérieur relatif non vide et génère un sous espace fermé de codimension finie. Nous établissons aussi le fait que ces ensembles ont des frontières non dégénérées au sens suivant : les points sur la frontière d'ensembles compactement epi-Lipschitz sont des points d'appui non dégénérés.

Après une discussion sur les extensions possibles au cas non convexe, nous établissons des propriétés analogues pour les fonctions convexes propres semi-continues inférieurement. Les outils utilisés sont les théorèmes de séparation, la conjuguée de Legendre-Fenchel et l'opérateur d'inf-convolution. Finalement, nous donnons une version du théorème de dualité de Fenchel pour les fonctions convexes semi-continues inférieurement et compactement epi-Lipschitz.

2.2.2 Comment transformer de façon continue une fonction convexe en une autre fonction convexe [2] ?

L'origine de ce travail repose d'une part sur l'étude de la moyenne proximale, un opérateur introduit précédemment dans le contexte de l'étude d'opérateurs monotones, et d'autre part sur l'étude de l'existence d'une homotopie pour l'ensemble des fonctions convexes : comment peut-on construire une homotopie entre deux fonctions convexes ? Il est clair que la moyenne arithmétique n'est pas adapté au cadre des fonctions convexes qui peuvent prendre la valeur $+\infty$.

En utilisant les outils de l'analyse convexe, nous étudions ladite moyenne proximale et établissons plusieurs de ses propriétés-clefs. En particulier, en utilisant le fait que la transformée de Legendre-Fenchel de la moyenne proximale de deux fonctions est la moyenne proximale de ses transformées de Legendre-Fenchel, nous établissons la propriété d'homotopie de la moyenne proximale ainsi que ses propriétés héritées de stricte convexité ou de différentiabilité. Plusieurs exemples illustrant les résultats pour le cas de fonctions indicatrices, linéaires et quadratiques sont détaillés. Les résultats sont appliqués à la construction de normes héritant de "bonnes" propriétés.

L'article fournit une introduction à la moyenne proximale, accessible à des lecteurs n'ayant pas forcément de connaissance préalable en analyse convexe, et constitue le point de départ des deux travaux suivants [1, 3].

2.2.3 Méthodes intrinsèques primales-duales pour les primitives d'opérateurs cycliquement monotones [3]

Nous présentons une méthode constructive pour générer des primitives de fonctions respectant la symétrie par rapport à la dualité convexe. Étant donné un échantillon fini $A = \{(s, x)\}$ de sous-gradients d'une fonction convexe, il existe une infinité de primitives convexes de cette fonction : fonctions convexes dont le graphe du sous-différentiel contient A. Contrairement au calcul de primitives de fonctions continu, ces fonctions ne sont pas uniques à une constante près. Rockafellar a posé la question suivante en 2005 lors de conférences à Borovets (Bulgarie) et à Banff (Canada) : Peut-on construire une primitive de façon symétrique par rapport à la dualité convexe ? En d'autres termes, si on dispose d'une méthode pour créer une primitive et on l'applique à l'ensemble A et à l'ensemble A^{-1}, trouve-t-on des fonctions conjuguées ?

Nous répondons positivement à la question posée. Notre construction est basée sur les fonctions de Rockafellar et de Fitzpatrick pour les opérateurs monotones, couplées avec l'opérateur de moyenne proximale. Nous introduisons un opérateur appelé "ancêtre commun" qui explicite les relations entre les fonctions de Rockafellar et de Fitzpatrick. Bien que ces dernières soient des primitives, elles ne possèdent pas les propriétés recherchées. Pour cela, nous utilisons l'opérateur de moyenne proximale pour générer une méthode primale-duale symétrique à partir de toute méthode générant des primitives.

2.2.4 Propriétés de l'opérateur de moyenne proximale [1]

Nous étudions systématiquement les propriétés de l'opérateur de moyenne proximale en analyse convexe : domaine, conjuguée de Legendre-Fenchel, sous-différentiel et épicontinuité. Plusieurs reformulations de cet opérateur sont présentées, en particulier en fonction de l'enveloppe de Moreau (régularisée de Moreau-Yosida). Le calcul du domaine affine un de nos résultats précédents [3], alors que le calcul du sous-différentiel permet de déduire des résultats de régularité et de stricte convexité. Nous étudions aussi les propriétés de convergence et d'épiconvergence de l'opérateur de moyenne proximale. Comme conséquence de ces résultats, nous déduisons que la moyenne arithmétique et la moyenne épigraphique sont en homotopie.

2.3 Analyse non lisse

2.3.1 Analyse de sensibilité pour la fonction valeur de problèmes d'optimisation avec des contraintes de type inégalités variationnelles [11]

Ce travail se situe en analyse non lisse. Nous démontrons plusieurs résultats pour l'analyse de sensibilité de problèmes d'optimisation comprenant des inégalités de type contraintes variationnelles. Les outils utilisés sont le sous-différentiel limite singulier et les codérivées au sens de Mordukhovich.

Plus précisément, nous considérons un problème de minimisation avec des contraintes d'inégalités variationnelles. Ces problèmes sont aussi appelés problèmes d'optimisation avec contraintes d'équilibres. Nous étudions la fonction valeur associée à la perturbation de contraintes d'égalités, d'inégalités et d'inégalités variationnelles. Nous démontrons, sous certaines conditions de croissance, que la fonction valeur est semi-continue inférieurement et que ses sous-différentiels réguliers et singuliers sont contenus dans l'opposé des ensembles de multiplicateurs associés aux codérivés du problème original. Ces résultats permettent de déduire la régularité de la fonction valeur dans le cas où l'ensemble des multiplicateurs singuliers contient seulement le vecteur zéro. Sous des hypothèses plus fortes, ces inégalités permettent de déduire la différentiabilité de la fonction valeur.

Nous comparons aussi plusieurs ensembles de multiplicateurs et donnons un exemple pour lequel les multiplicateurs associés aux codérivés sont les plus appropriés. Les résultats sont ensuite appliqués aux problèmes d'optimisation avec contraintes à deux niveaux.

Une erreur dans l'énoncé du résultat principal nous a conduit à la publication d'un rectificatif [11 bis] qui introduit une hypothèse additionnelle requise par notre résultat principal.

Bibliographie

1. H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang. The proximal average : Basic theory. *SIAM J. Optim.*, Mar. 2008. Accepté pour publication.
2. H. H. Bauschke, Y. Lucet, and M. Trienis. How to transform one convex function continuously into another. *SIAM Review*, 50(1) :115–132, 2008.
3. H. H. Bauschke, Y. Lucet, and X. Wang. Primal-dual symmetric intrinsic methods for finding antiderivatives of cyclically monotone operators. *SIAM J. Control Optim.*, 46(6) :2031–2051, 2007.
4. J. Borwein, Y. Lucet, and B. Mordukhovich. Compactly epi-lipschitzian convex sets and functions in normed spaces. *J. Convex Anal.*, 7(2) :375–393, 2000.
5. J.-B. Hiriart-Urruty and Y. Lucet. Parametric computation of the Legendre–Fenchel conjugate. *J. Convex Anal.*, 14(3) :657–666, Aug. 2007.
6. Y. Lucet. A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform. In *Proceedings of the Second Canadian Conference on Computer and Robot Vision*, 262–267, 2005. IEEE Computer Society Press.
7. Y. Lucet. Fast Moreau envelope computation I : Numerical algorithms. *Numer. Algorithms*, 43(3) :235–249, Nov. 2006.
8. Y. Lucet. New sequential exact Euclidean distance transform algorithms based on convex analysis. *Image and Vision Computing*, Dec. 2006. Accepté pour publication.
9. Y. Lucet. What shape is your conjugate ? A survey of computational convex analysis and its applications. Manuscrit soumis en 2007.
10. Y. Lucet, H. H. Bauschke, and M. Trienis. The piecewise linear-quadratic model for computational convex analysis. *Comput. Optim. Appl.*, July 2006. Accepté pour publication.
11. Y. Lucet and J. Ye. Sensitivity analysis for the value function for optimization problems with variational inequalities constraints. *SIAM J. Control Optim.*, 40(3) :699–723, Sept. 2002.

11 bis. Y. Lucet and J. Ye. Erratum : Sensitivity analysis of the value function for optimization problems with variational inequality constraints. *SIAM J. Control Optim.*, 41(4) :1315–1319, Dec. 2002.

Chapitre 3

Formation et encadrement

3.1 Thèses

– Directeur de thèse de Mike Trienis en mastère en sciences (M.Sc.) interdisciplinaires (interdisciplinary graduate studies), spécialité optimisation, à University of British Columbia, de mai 2006 à décembre 2007. Soutenu avec succès le 14 décembre 2007. Sujet : Analyse convexe computationnelle : Des déformations continues aux primitives convexes finies.
– Directeur de thèse de Valentin Koch en Honneur (Honours) en informatique à University of British Columbia, depuis septembre 2007. Sujet : Optimisation globale pour la thérapie par rayons à intensités modulées.
– Directeur de thèse de Scott Fazackerley en Honneur (Honours) en informatique à University of British Columbia, depuis septembre 2007. Sujet : Algorithmes linéaires quadratiques par morceaux en deux dimensions.
– Codirecteur de thèse de Bryce Cutt en Honneur (Honours) en informatique à University of British Columbia, de septembre 2006 à avril 2007 (codirecteur : Patricia Lasserre). Sujet : Segmentation du champ de traitement par radiation dans les images de portail doubles (Segmentation of Radiation Treatment Field in Dual Portal Images).
– Codirecteur de thèse de Jeff Dicker en Honneur (Honours) en informatique à University of British Columbia, de septembre 2005 à avril 2006 (codirecteur : Patricia Lasserre). Sujet : Une implémentation des méthodes de marche rapide et des méthodes de niveaux (Fast Marching Methods and Level Set Methods : An Implementation).

3.2 Comités de thèses et encadrement d'étudiants

– Membre du comité consultatif (advisory committee) pour la thèse de mastère en sciences interdisciplinaires (M.Sc. Interdisciplinary Graduate Studies) de Bryce Cutt depuis septembre 2007.
– Rapporteur et membre du comité de thèse pour Liangjin Yao en mastère en sciences interdisciplinaires (M.Sc. Interdisciplinary Graduate Studies) soutenue avec succès le 14 décembre 2007.

- Membre du comité consultatif (advisory committee) pour la thèse de mastère en sciences interdisciplinaires (M.Sc. Interdisciplinary Graduate Studies) de Mélissa Lavallée depuis septembre 2006.
- Encadrement de Valentin Koch en tant qu'assistant de recherche à University of British Columbia de mai 2007 à août 2007. Financé par une bourse de recherche de premier cycle du conseil de recherche en sciences naturelles et en génie du Canada. Sujet : Optimisation globale pour la thérapie par rayons à intensités modulées.
- Encadrement de Scott Fazackerley en tant qu'assistant de recherche à University of British Columbia de mai 2007 à août 2007. Financé par une bourse de recherche de premier cycle du conseil de recherche en sciences naturelles et en génie du Canada. Sujet : Analyse convexe computationnelle.
- Encadrement de Valentin Koch en tant qu'assistant de recherche à University of British Columbia de mai 2006 à août 2006. Sujet : Optimisation pour le traitement par rayons à intensités modulées.
- Encadrement de Jeff Dicker en tant qu'assistant de recherche à University of British Columbia de mai 2005 à août 2005. Financé par une bourse de recherche de premier cycle du conseil de recherche en sciences naturelles et en génie du Canada. Sujet : Analyse convexe computationnelle.
- Encadrement de Andrew Tonner en tant qu'assistant de recherche à University of British Columbia de mai 2005 à août 2005. Sujet : Modèle de calcul de dosimétrie en traitement par rayons à intensités modulées.
- Encadrement de François Stelluti en tant qu'assistant de recherche à University of British Columbia de mai 2004 à août 2004. Sujet : Étude du traitement par rayons à intensités modulées
- Responsable de mémoire de Mike Trienis en Directed Studies in Computer Science, 2006. Sujet : Analyse convexe computationnelle.
- Co-responsable de mémoire de Jim Moffat en Directed Studies in Computer Science, 2004. Sujet : Traitement d'image en thérapie par rayons.
- Co-responsable de mémoire de Kin Hang Lau en Directed Studies in Mathematics, 2004. Sujet : Modèle élastique pour le problème de planification du traitement en thérapie par rayons.

Chapitre 4

Diffusion et transfert de connaissances

4.1 Bourses et contrats de recherche

4.1.1 Financements externes

– Responsable d'un financement en informatique de $75 000, de 2008 à 2012. Sujet : Analyse convexe assistée par ordinateur et optimisation.
– Responsable d'un financement en informatique de $33 000, de 2005 à 2008. Sujet : Analyse convexe computationnelle.
– Co-responsable d'un financement de la fondation canadienne du cancer du sein, de $65 000, avec H. Bauschke, P. Lasserre, P. Gill et R. Rasika (responsable principal), de 2007 à 2009. Sujet : Analyse quantitative de la densité mammographique du sein dans des mammogrammes digitaux.
– Responsable d'un financement de la fondation canadienne pour l'innovation, de $188 000, pour l'établissement du laboratoire d'analyse convexe assisté par ordinateur à University of British Columbia Okanagan, Kelowna BC Canada. Financement obtenu en 2007, date d'ouverture du laboratoire : 2010.

4.1.2 Financements internes

– Responsable d'un financement de $5,000 pour le projet "From applications to technology transfer of computational convex analysis", 2008.
– Responsable d'un financement de $5,000 pour le projet "numerical algorithms for primal-dual symmetric reconstruction of convex functions", 2007.
– Responsable d'un financement de $2,000 du Pacific Institute for the Mathematical Sciences pour le projet "Image Interpolation Through Computational Convex Analysis", 2007.
– Responsable d'un financement de $5,000 pour le projet "Efficient Dose Calculation for Intensity Modulated Radiotherapy Treatment Planning Optimization", 2006.
– Responsable d'un financement de $5,000 pour le projet "Fast Computational Convex Analysis Algorithms with applications to image processing and Intensity Modulated

Radiation Therapy (IMRT) treatment planning optimization", 2005.
- Responsable d'un financement de $5,000 pour le projet "Aperture-based optimization for Intensity Modulated Radiotherapy Treatment (IMRT)", 2004.
- Responsable d'un financement de $5,000 pour le projet "Aperture-based optimization for Intensity Modulated Radiotherapy Treatment (IMRT)", 2003.

4.1.3 Autres financements

- Bourse de chercheur post-doctoral du Pacific Institute for the Mathematical Sciences de $20,000 de 1997 à 1999.
- Boursier du Ministère de l'Enseignement Supérieur et de la Recherche de $60,000 de 1993 à 1997.
- Moniteur de l'enseignement supérieur de 1993 à 1997.
- Bourse de DEA en 1992-1993.
- Bourse de voyage de l'Université Paul Sabatier de $1000 en 1994.

4.2 Articles publiés dans des revues à comité de lecture

1. H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang. The proximal average : Basic theory. *SIAM J. Optim.*, Mar. 2008. Accepté pour publication.

2. H. H. Bauschke, Y. Lucet, and M. Trienis. How to transform one convex function continuously into another. *SIAM Review*, 50(1) :115–132, 2008.

3. H. H. Bauschke, Y. Lucet, and X. Wang. Primal-dual symmetric intrinsic methods for finding antiderivatives of cyclically monotone operators. *SIAM J. Control Optim.*, 46(6) :2031–2051, 2007.

4. J. Borwein, Y. Lucet, and B. Mordukhovich. Compactly epi-lipschitzian convex sets and functions in normed spaces. *J. Convex Anal.*, 7(2) :375–393, 2000.

5. J.-B. Hiriart-Urruty and Y. Lucet. Parametric computation of the Legendre–Fenchel conjugate. *J. Convex Anal.*, 14(3) :657–666, Aug. 2007.

6. Y. Lucet. Fast Moreau envelope computation I : Numerical algorithms. *Numer. Algorithms*, 43(3) :235–249, Nov. 2006.

7. Y. Lucet. New sequential exact Euclidean distance transform algorithms based on convex analysis. *Image and Vision Computing*, Dec. 2006. Accepté pour publication.

8. Y. Lucet, H. H. Bauschke, and M. Trienis. The piecewise linear-quadratic model for computational convex analysis. *Comput. Optim. Appl.*, July 2006. Accepté pour publication.

9. Y. Lucet and J. Ye. Sensitivity analysis for the value function for optimization problems with variational inequalities constraints. *SIAM J. Control Optim.*, 40(3) :699–723, Sept. 2002.

9 bis. Y. Lucet and J. Ye. Erratum : Sensitivity analysis of the value function for optimization problems with variational inequality constraints. *SIAM J. Control Optim.*, 41(4) :1315–1319, Dec. 2002.

4.3 Actes de conférences avec comité de lecture

1. Y. Lucet. A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform. In *Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV 2005)*, pages 262–267, Victoria BC, Mai 2005. IEEE Computer Society Press.

4.4 Articles soumis pour publication (dans des revues à comité de lecture)

1. Y. Lucet. What shape is your conjugate ? A survey of computational convex analysis and its applications. Manuscrit soumis en 2007.

4.5 Logiciels

Une boîte à outils, appelée CCA, composée d'algorithmes rapides et d'algorithmes linéaires-quadratiques par morceaux pour le calcul de la conjuguée de Legendre-Fenchel et de l'enveloppe de Moreau a été développée sous le logiciel mathématique Scilab. Plusieurs versions ont été publiées. Certaines fonctions sont décrites dans les travaux [7, 10]. Cette boîte à outils a été instrumentale pour la formulation de conjectures, la découverte de contre-exemples, le test d'algorithmes et la génération d'exemples lors de la rédaction des articles [1, 2, 3, 5, 8, 9, 6].

La version 1.5.2 est disponible. Outre l'implémentation des divers algorithmes, elle contient une aide en ligne, des démonstrations des différentes commandes et une batterie de tests. Le code a été évalué par l'éditeur de logiciels lors de la soumission de l'article [7].

4.6 Conférences et présentations

1. Hybrid Symbolic-Numeric Algorithms for Computational Convex Analysis (invité). Second International Conference on Continuous Optimization + Modeling and Optimization : Theory and Applications, Hamilton, Ontario, Canada. Août 2007.

1 bis. Organisateur (Invité) de la session Computational Convex Analysis lors de la Second International Conference on Continuous Optimization + Modeling and Optimization : Theory and Applications, Hamilton, Ontario, Canada. Août, 2007.

2. Primal-Dual Reconstruction of a Convex Function (invité). West Coast Optimization Meeting, University of Washington, Seattle, Washington, États-Unis. Mai 2006.

3. Organisateur de la session Symbolic and Numerical Computations for Convex Analysis lors du 6th International Congress of Industrial and Applied Mathematics (ICIAM), Suisse. Juillet 2007.

3 bis. Hybrid Symbolic-Numeric Algorithms for Computational Convex Analysis. 6th International Congress of Industrial and Applied Mathematics (ICIAM), Suisse. Juillet 2007.

4. The Piecewise Linear-Quadratic Model for Computational Convex Analysis. 2nd International Conference on Algorithmic Operations Research (AlgOR 2007), Surrey, British Columbia, Canada. Janvier 2007.

5. Adaptative Greedy Schemes for Computational Convex Analysis. SIAM Conference on Discrete Mathematics, Victoria, British Columbia, Canada. Juin 2006.

6. Fast Algorithms for the Legendre Conjugate and the Moreau Envelope with Applications. OCANA Seminar Series, University of British Columbia, Kelowna, British Columbia, Canada. Decembre 2005.

7. Fast Algorithms for the Fenchel Transform. Canada-Chile meeting on the mathematics of economic geography and natural resource management, Banff, Alberta, Canada. Novembre 2005.

8. Fast Moreau Envelope Algorithms and Applications. CMS Winter 2005 Meeting, Victoria, British Columbia, Canada. Novembre 2005.

9. Co-présenté avec Patricia Lasserre. Defining a System Dynamics Model for the University Undergraduate Learning Environment. From the Inside Out Learning Conference, Kelowna, British Columbia, Canada. Février 2005.

10. Co-organisateur de la session MS34 Symbolic and Numerical Computations for Convex Optimization lors de la Sixth SIAM conference on optimization, Atlanta, Georgia, États-Unis. Mai 1999.

10 bis. Fast Fenchel Transform and Fast Moreau-Yosida Approximate. Sixth SIAM conference on optimization, Atlanta, Georgia, États-Unis. Mai 1999.

4.7 Rédaction de rapports de lecture-arbitrage (*referee reports*)

- SIAM Journal on Optimization : 2 articles (depuis 2006).
- Journal of Convex Analysis : 2006.
- Image and Vision Computing : 2006.
- Algorithms for Operation Research : 2006.
- International Symposium on Symbolic and Algebraic Computation : 2006.
- Third International Conference on Education and Information Systems : Technologies and Applications : 3 articles (2005).
- Second Canadian Conference on Computer and Robot Vision : 2 articles (2005).
- Western Canadian Conference on Computer Education 2004 : 2 articles (2004).

4.8 Responsabilités administratives

- Co-organisateur du West Coast Optimization Meeting, Kelowna BC, Canada. Octobre 2007.
- Co-organisateur des séminaires du groupe OCANA (Optimization, Convex Analysis and Nonsmooth Analysis) depuis 2005.
- Responsable de la section optimisation du Interdisciplinary Graduate Program pour le campus de l'Okanagan de l'Université de Colombie Britannique depuis 2006.
- Membre du conseil consultatif des programmes doctoraux du campus de l'Okanagan de l'Université de Colombie Britannique depuis 2007.
- Membre du comité doctoral de l'unité de mathématiques du campus de l'Okanagan de l'Université de Colombie Britannique depuis 2007.
- Collaborateur avec l'agence contre le cancer de Colombie Britannique (Canada) depuis 2004.

4.9 Sociétés savantes

Membre de :

- Society for Industrial and Applied Mathematics (SIAM) et de l'Activity Group on Optimization.
- Association for Computing Machinery (ACM).
- Mathematical Programming Society (MPS).
- Canadian Mathematical Society (CMS).

Chapitre 5

Exemplaire des onze publications présentées (travaux de recherche)

The Proximal Average: Basic Theory

Heinz H. Bauschke*, Rafal Goebel†, Yves Lucet‡ and Xianfu Wang§

January 27, 2008 (revised version; originally submitted on April 6, 2007)

Abstract

The recently introduced proximal average of two convex functions is a convex function with many useful properties. In this paper, we introduce and systematically study the proximal average for finitely many convex functions. The basic properties of the proximal average with respect to the standard convex-analytical notions (domain, Fenchel conjugate, subdifferential, proximal mapping, epi-continuity, and others) are provided and illustrated by several examples.

2000 Mathematics Subject Classification:
Primary 90C25; Secondary 26A51, 26B25, 26E60, 46C05, 47H05, 52A41.

Keywords: Arithmetic average, arithmetic mean, convex analysis, convex function, epi-convergence, epigraphical average, epi-topology, essential smoothness, essential strict convexity, Fenchel conjugate, harmonic mean, Legendre function, proximal average, proximal mapping, subdifferential operator.

1 Overview

Let f_1 and f_2 be two functions that are convex, lower semicontinuous, and proper, and let λ_1 and λ_2 be strictly positive real numbers adding up to 1. How can we average the two functions f_1 and f_2 with respect to the weights λ_1 and λ_2 in a useful way? Perhaps the first approach is to consider the *arithmetic average* $\lambda_1 f_1 + \lambda_2 f_2$. However, functions in convex analysis are allowed to take on

*Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.

†Mathematics, University of Washington, Box 354350, Seattle, Washington 98195-4350, USA. E-mail: goebel@math.washington.edu.

‡Computer Science, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: yves.lucet@ubc.ca.

§Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: shawn.wang@ubc.ca.

the value $+\infty$, for example to model constraints in optimization problems. Thus, the arithmetic average can turn out to be $+\infty$ everywhere and then carries little information about f_1 and f_2; this happens whenever f_1 and f_2 are nowhere both finite. How could we possibly average such functions? A second thought may suggest to construct the *epigraphical average* $\lambda_1 \star f_1 \mathbin{\text{⩲}} \lambda_2 \star f_2$ obtained by forming a convex combination of the epigraphs of f_1 and f_2. Unfortunately, if the functions f_1 and f_2 lack coercivity, then the epigraphical average fails to be helpful: for instance, if f_1 and f_2 are two distinct linear functions, then their epigraphical average is identically equal to $-\infty$, and hence of little use. The *proximal average*, first introduced in [6] in the context of fixed point theory and recently studied in [4, 5, 7, 10, 15] from various viewpoints, avoids the mentioned difficulties and possesses numerous properties that are attractive to Convex Analysts.

The aim of this paper is to provide the basic theory of the proximal average. In addition, we extend it to more than two functions and we allow for an additional positive parameter. For the reader's convenience and the sake of completeness, the presentation of the theory is largely self-contained. It is shown that the proximal average has many desirable properties in terms of its domain, Fenchel conjugate, Moreau envelope, proximal mapping, subdifferential, epi-continuity, and other convex-analytical notions. Moreover, the arithmetic and epigraphical averages turn out to be limits of the proximal average, as the parameter tends to 0 and $+\infty$, respectively. Various examples illustrate our results. An interesting topic for future research is the extension to series and integrals.

The rest of this paper is organized as follows. Section 2 collects the notation used throughout this paper, and in Section 3 we collect and present results that simplify later proofs. The proximal average is introduced in Section 4 where also its domain is characterized. In Section 5, we present one very useful result (Theorem 5.1) which states that the Fenchel conjugate of the proximal average is the proximal average of the Fenchel conjugates. An important consequence of this result is that the proximal average is convex, lower semicontinuous and proper. In Section 6 we consider the Moreau envelope and proximal mapping of the proximal average, in Section 7 its subdifferential operator as well as essential smoothness and essential strict convexity. In Section 8, it is shown that the arithmetic and epigraphical averages are pointwise limits of the proximal average. Epi-convergence properties are discussed in the final Section 9, where the arithmetic and epigraphical averages are shown to be limiting instances of the proximal average with respect to epi-convergence.

2 Standing Assumptions and Notation

Throughout this paper,

$$X \text{ is a real Hilbert space with inner product } \langle\cdot,\cdot\rangle \text{ and corresponding norm } \|\cdot\|. \tag{1}$$

Due to its repeated use, we abbreviate the quadratic energy function by

$$\mathsf{q} = \tfrac{1}{2}\|\cdot\|^2. \tag{2}$$

We set

$$\Gamma(X) = \big\{f\colon X \to \left]-\infty,+\infty\right] \mid f \text{ is convex, lower semicontinuous, and proper}\big\}. \tag{3}$$

We assume throughout that

$$n \in \{1, 2, 3, \dots\}, \tag{4}$$

that

$$f_1, \dots, f_n \text{ belong to } \Gamma(X), \tag{5}$$

that

$$\lambda_1, \dots, \lambda_n \text{ are nonnegative real numbers such that } \lambda_1 + \cdots + \lambda_n = 1, \tag{6}$$

and that

$$\mu \text{ is a strictly positive real number.} \tag{7}$$

The Fenchel conjugate of a function f is denoted by f^*. It will be convenient to set

$$\boldsymbol{f} = (f_1, \dots, f_n), \quad \boldsymbol{f}^* = (f_1^*, \dots, f_n^*), \quad \text{and} \quad \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n). \tag{8}$$

Other notation not explicitly defined here or later is standard in Convex Analysis and as in, e.g., [21, 22, 24]. Let f be a convex function and S be a set. Then we write $\operatorname{dom} f$, $\operatorname{epi} f$, ∂f, $\operatorname{cl} f$, $\inf f$, $\min f$, $\operatorname{argmin} f$, d_S, $\operatorname{conv} S$, $\operatorname{int} S$, ι_S, and N_S to denote the (effective) domain, epigraph, subdifferential operator, lower closure, infimum value, minimum value if the infimum value is attained, the set of minimizers, distance function, convex hull, interior, indicator function, and normal cone operator, respectively. The identity operator is represented by Id.

3 Auxiliary Results

We start by reviewing the key notions of epi-multiplication and epi-addition, following the viewpoint taken in [22, Section 1.H]. Let $\alpha \geq 0$, $f \in \Gamma(X)$, $g \in \Gamma(X)$, and $h \in \Gamma(X)$. Then

$$\alpha \star f = \begin{cases} \alpha f(\cdot/\alpha), & \text{if } \alpha > 0; \\ \iota_{\{0\}}, & \text{if } \alpha = 0. \end{cases} \tag{9}$$

The term "epi-multiplication" stems from the fact that $\operatorname{epi}(\alpha \star f) = \alpha \operatorname{epi}(f)$ when $\alpha > 0$. Epi-addition or infimal convolution is defined by

$$f \mathbin{\square} g \colon X \to [-\infty, +\infty] : x \mapsto \inf_{y+z=x} \big(f(y) + g(z)\big); \tag{10}$$

and the term "epi-addition" stems from the fact that the strict epigraph of $f \mathbin{\square} g$ is the Minkowski sum of the strict epigraphs of f and g, i.e., $\big\{(x, r) \in X \times \mathbb{R} \mid (f \mathbin{\square} g)(x) < r\big\} = \big\{(y, s) \in X \times \mathbb{R} \mid f(y) < s\big\} + \big\{(z, t) \in X \times \mathbb{R} \mid g(z) < t\big\}$. The epi-sum of finitely many functions is defined analogously.

To avoid excessive usage of parentheses, epi-multiplication and regular multiplication are given precedence over epi- and regular addition, i.e., $\alpha \star f + g = (\alpha \star f) + g$, $\alpha \star f \mathbin{\square} g = (\alpha \star f) \mathbin{\square} g$, $\alpha f + g = (\alpha f) + g$, and $\alpha f \mathbin{\square} g = (\alpha f) \mathbin{\square} g$. It will also be convenient to give epi-addition a higher precedence than regular addition or subtraction, i.e., $f \mathbin{\square} g + h = (f \mathbin{\square} g) + h$ and $f \mathbin{\square} g - h = (f \mathbin{\square} g) - h$.

The next three propositions are elementary. Proofs for the finite-dimensional case are in [22]; they extend without difficulty to the present Hilbert space setting.

Proposition 3.1 *Let $f \in \Gamma(X)$, let $\alpha \geq 0$, and let $\beta \geq 0$. Then the following hold.*

(i) $\alpha > 0 \Rightarrow \operatorname{epi}(\alpha \star f) = \alpha(\operatorname{epi} f)$.

(ii) $\operatorname{dom}(\alpha \star f) = \alpha(\operatorname{dom} f)$.

(iii) $f \oplus \iota_{\{0\}} = f$.

(iv) $\operatorname{dom}(f_1 \oplus \cdots \oplus f_n) = (\operatorname{dom} f_1) + \cdots + (\operatorname{dom} f_n)$.

(v) $\alpha \star (f_1 \oplus \cdots \oplus f_n) = \alpha \star f_1 \oplus \cdots \oplus \alpha \star f_n$.

(vi) $\alpha(f_1 \oplus \cdots \oplus f_n) = \alpha f_1 \oplus \cdots \oplus \alpha f_n$.

(vii) $\alpha \star (\beta \star f) = (\alpha\beta) \star f$.

(viii) $(\alpha + \beta) \star f = \alpha \star f \oplus \beta \star f$.

(ix) $\alpha > 0 \Rightarrow \alpha(\beta \star (\alpha^{-1} f)) = \beta \star f$.

Proof. The conclusions all follow readily from the definitions; see also [22, Exercise 1.28(a)] for (i), [22, page 25] for (v) and (vii), and [22, Exercise 2.24(c)] for (viii). ■

Proposition 3.2 *Let $\alpha \geq 0$. Then the following hold.*

(i) $(\alpha f)^* = \alpha \star f^*$.

(ii) $(\alpha \star f)^* = \alpha f^*$.

(iii) $(f_1 \oplus \cdots \oplus f_n)^* = f_1^* + \cdots + f_n^*$.

Proof. The statements are simple consequences of the definitions; see also [22, page 475] for (i) and (ii), and [22, Theorem 11.23(a)] for (iii). ■

Proposition 3.3 *Let $f \in \Gamma(X)$ and let $\alpha \geq 0$. Then the following hold.*

(i) $\mathsf{q}^* = \mathsf{q}$*; in fact, q is the only function equal to its Fenchel conjugate.*

(ii) $\alpha > 0 \Rightarrow \alpha^{-1} \star \mathsf{q} = \alpha\, \mathsf{q}$.

(iii) $(\alpha \star \mathsf{q})^* = \alpha\, \mathsf{q}$.

(iv) $(\alpha\, \mathsf{q})^* = \alpha \star \mathsf{q}$.

(v) $(f \oplus \mathsf{q}) + (f^* \oplus \mathsf{q}) = \mathsf{q}$.

Proof. (i): See, e.g., [22, Example 11.11]. (ii): An immediate consequence of the definition of $\mathfrak{q}$. (iii): Combine Proposition 3.2(ii) with (i). (iv): Combine Proposition 3.2(i) with (i). (v): See [19] or [22, Example 11.26]. ■

The next result is deep and stated as a fact.

Fact 3.4 *The following hold.*

(i) *If* $\operatorname{int}\operatorname{dom} f_1 \cap \cdots \cap \operatorname{int}\operatorname{dom} f_{n-1} \cap \operatorname{dom} f_n \neq \varnothing$, *then* $(f_1 + \cdots + f_n)^* = f_1^* \mathbin{\square} \cdots \mathbin{\square} f_n^*$ *and the epi-sum is exact, i.e., the infimum in the definition of the epi-sum is attained.*

(ii) *If* $\operatorname{int}\operatorname{dom} f_1^* \cap \cdots \cap \operatorname{int}\operatorname{dom} f_{n-1}^* \cap \operatorname{dom} f_n^* \neq \varnothing$, *then* $f_1 \mathbin{\square} \cdots \mathbin{\square} f_n$ *is exact and* $\operatorname{epi}(f_1 \mathbin{\square} \cdots \mathbin{\square} f_n) = (\operatorname{epi} f_1) + \cdots + (\operatorname{epi} f_n)$.

Proof. This is a consequence of [24, Theorem 2.8.7]. ■

The following result on the conjugate of the difference will be useful.

Fact 3.5 *Let* $g \in \Gamma(X)$ *and let* $h \in \Gamma(X)$ *such that both* h *and* h^* *have full domain. Then*

$$\big(\forall x^* \in X\big) \quad (g-h)^*(x^*) = \sup_{y^* \in X} \big(g^*(y^*) - h^*(y^* - x^*)\big). \tag{11}$$

Proof. This is a consequence of [11, Theorem 2.2]. ■

Corollary 3.6 *Let* $g \in \Gamma(X)$. *Then*

$$(g - \mu \star \mathfrak{q})^* = \mu(\mathfrak{q} - \mu^{-1} g^*)^* - \mu^{-1} \star \mathfrak{q}. \tag{12}$$

Proof. Set $h = \mu \star \mathfrak{q}$. Then $h^* = \mu\,\mathfrak{q}$ by Proposition 3.3(iii) and hence both h and h^* have full domain. Using Fact 3.5, we deduce that for every $x^* \in X$

$$\begin{aligned}
(g-h)^*(x^*) &= \sup_{y^* \in X} \big(g^*(y^*) - \mu\,\mathfrak{q}(y^* - x^*)\big) \\
&= \sup_{y^* \in X} \big(g^*(y^*) - \mu\,\mathfrak{q}(y^*) - \mu\,\mathfrak{q}(x^*) + \mu\langle y^*, x^*\rangle\big) \\
&= -\mu\,\mathfrak{q}(x^*) + \sup_{y^* \in X} \big(\langle y^*, \mu x^*\rangle - \big(\mu\,\mathfrak{q}(y^*) - g^*(y^*)\big)\big) \\
&= -\mu\,\mathfrak{q}(x^*) + \mu \sup_{y^* \in X} \big(\langle y^*, x^*\rangle - \big(\mathfrak{q}(y^*) - \mu^{-1} g^*(y^*)\big)\big) \\
&= -(\mu^{-1} \star \mathfrak{q})(x^*) + \mu(\mathfrak{q} - \mu^{-1} g^*)^*(x^*).
\end{aligned} \tag{13}$$

The proof is complete. ■

Lemma 3.7 $\big(\lambda_1 \star (f_1 + \mu \star \mathfrak{q}) \mathbin{\square} \cdots \mathbin{\square} \lambda_n \star (f_n + \mu \star \mathfrak{q})\big)^* = \lambda_1(f_1^* \mathbin{\square} \mu\,\mathfrak{q}) + \cdots + \lambda_n(f_n^* \mathbin{\square} \mu\,\mathfrak{q})$.

Proof. Using Proposition 3.2(iii), Proposition 3.2(ii), Fact 3.4(i), and Proposition 3.3(iii), we compute that

$$\begin{aligned}\big(\lambda_1\star(f_1+\mu\star\mathsf{q})\,\text{⨥}\cdots\text{⨥}\,\lambda_n\star(f_n+\mu\star\mathsf{q})\big)^* &= \big(\lambda_1\star(f_1+\mu\star\mathsf{q})\big)^*+\cdots+\big(\lambda_n\star(f_n+\mu\star\mathsf{q})\big)^*\\ &= \lambda_1(f_1+\mu\star\mathsf{q})^*+\cdots+\lambda_n(f_n+\mu\star\mathsf{q})^*\\ &= \lambda_1\big(f_1^*\,\text{⨥}\,(\mu\star\mathsf{q})^*\big)+\cdots+\lambda_n\big(f_n^*\,\text{⨥}\,(\mu\star\mathsf{q})^*\big)\\ &= \lambda_1(f_1^*\,\text{⨥}\,\mu\,\mathsf{q})+\cdots+\lambda_n(f_n^*\,\text{⨥}\,\mu\,\mathsf{q}).\end{aligned} \tag{14}$$

This completes the proof. ■

Fact 3.8 *Let* $(\forall i)$ $x_i \in \operatorname{dom} f_i$, *and set* $x = x_1+\cdots+x_n$. *Then the following implications hold.*

(i) $(f_1\,\text{⨥}\cdots\text{⨥}\,f_n)(x) = f_1(x_1)+\cdots+f_n(x_n) \Rightarrow \partial(f_1\,\text{⨥}\cdots\text{⨥}\,f_n)(x) = \partial f_1(x_1)\cap\cdots\cap\partial f_n(x_n)$.

(ii) $\partial f_1(x_1)\cap\cdots\cap\partial f_n(x_n)\neq\varnothing \Rightarrow (f_1\,\text{⨥}\cdots\text{⨥}\,f_n)(x) = f_1(x_1)+\cdots+f_n(x_n)$.

Proof. See [24, Corollary 2.4.7]. ■

Proposition 3.9 *Let* $f\in\Gamma(X)$ *and let* $\alpha>0$. *Then* $\partial(0\star f) = N_{\{0\}}$ *and* $\partial(\alpha\star f) = (\partial f)\circ(\alpha^{-1}\operatorname{Id})$.

Proof. Since $0\star f = \iota_{\{0\}}$, we deduce that $\partial(0\star f) = \partial\iota_{\{0\}} = N_{\{0\}}$. Also, $\partial(\alpha\star f) = \partial(\alpha f\circ(\alpha^{-1}\operatorname{Id}))$; the formula thus follows from Convex Calculus (see, e.g., [24, Theorem 2.8.3]). ■

4 Definition, Reformulations, Domain and Exactness

In Section 1, we have seen that the idea of computing the averaged Minkowski sum is doomed in general, due to the potential lack of coercivity properties of the terms. The proximal average can be interpreted as a three-step remedy of this idea: First, each function is "coercified" by epi-adding $\mu\star\mathsf{q}$. Second, the epi-average of the coercified terms is computed. The third step removes $\mu\star\mathsf{q}$ through subtraction. We are now ready to describe the proximal average.

Definition 4.1 (proximal average) *The* $\boldsymbol{\lambda}$-weighted proximal average of $\boldsymbol{f}$ with parameter μ *is*

$$p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \lambda_1\star(f_1+\mu\star\mathsf{q})\,\text{⨥}\cdots\text{⨥}\,\lambda_n\star(f_n+\mu\star\mathsf{q}) - \mu\star\mathsf{q}, \tag{15}$$

i.e., if $I = \big\{i\in\{1,\ldots,n\}\mid\lambda_i>0\big\}$, *then*

$$(\forall x\in X)\quad p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \frac{1}{\mu}\Big(-\tfrac{1}{2}\|x\|^2 + \inf_{\sum_{i\in I}x_i=x}\ \sum_{i\in I}\lambda_i\big(\mu f_i(x_i/\lambda_i)+\tfrac{1}{2}\|x_i/\lambda_i\|^2\big)\Big). \tag{16}$$

We also write $p(\boldsymbol{f},\boldsymbol{\lambda})$ *if* $\mu=1$, $p_\mu(\boldsymbol{f})$ *if all* λ_i *coincide, and* $p(\boldsymbol{f})$ *if* $\mu=1$ *and all* λ_i *coincide.*

Remark 4.2 Some immediate consequences of the definition are the following.

(i) $p_\mu(f_1, 1) = f_1$.

(ii) If $I = \big\{i \in \{1, \ldots, n\} \mid \lambda_i > 0\big\}$, $\widetilde{\boldsymbol{f}} = (f_i)_{i\in I}$ and $\widetilde{\boldsymbol{\lambda}} = (\lambda_i)_{i\in I}$, then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = p_\mu(\widetilde{\boldsymbol{f}}, \widetilde{\boldsymbol{\lambda}})$.

(iii) If π is a permutation of $I = \{1, \ldots, n\}$, $\widetilde{\boldsymbol{f}} = (f_{\pi(i)})_{i\in I}$ and $\widetilde{\boldsymbol{\lambda}} = (\lambda_{\pi(i)})_{i\in I}$, then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = p_\mu(\widetilde{\boldsymbol{f}}, \widetilde{\boldsymbol{\lambda}})$.

(iv) $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \mu^{-1} p_1(\mu \boldsymbol{f}, \boldsymbol{\lambda})$; equivalently, $p(\mu \boldsymbol{f}, \boldsymbol{\lambda}) = \mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

(v) If $\Lambda_{n-1} = \lambda_1 + \cdots + \lambda_{n-1} > 0$, then

$$\begin{aligned} p_1(\boldsymbol{f}, \boldsymbol{\lambda}) &= p_1\big((f_1, \ldots, f_n), (\lambda_1, \ldots, \lambda_n)\big) \\ &= p_1\Big(\Big(p_1\big((f_1, \ldots, f_{n-1}), \Lambda_{n-1}^{-1}(\lambda_1, \ldots, \lambda_{n-1})\big), f_n\Big), (\Lambda_{n-1}, \lambda_n)\Big). \end{aligned} \tag{17}$$

The identities in items (iv) and (v) may be useful if one wishes to develop the theory of results for a general $\mu > 0$ and a general $n \geq 2$ from the simpler case $\mu = 1$ and $n = 2$; however, the direct approach favoured in this paper is not only self-contained but it also yields proofs that we found much more readable. Nonetheless, (iv) and (v) may be convenient for the numerical computation of the proximal average — especially when the simpler case is already implemented [15].

Proposition 4.3 (reformulations)

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \big(\lambda_1(f_1^* \,\square\, \mu\,\mathsf{q}) + \cdots + \lambda_n(f_n^* \,\square\, \mu\,\mathsf{q})\big)^* - \mu^{-1}\,\mathsf{q} \tag{18}$$

$$= \big(\lambda_1(f_1 + \mu^{-1}\,\mathsf{q})^* + \cdots + \lambda_n(f_n + \mu^{-1}\,\mathsf{q})^*\big)^* - \mu^{-1}\,\mathsf{q} \tag{19}$$

and

$$\big(\forall x \in X\big) \quad p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \inf_{\sum \lambda_i y_i = x} \sum \lambda_i f_i(y_i) + \frac{1}{\mu}\Big(\Big(\sum \lambda_i\,\mathsf{q}(y_i)\Big) - \mathsf{q}(x)\Big). \tag{20}$$

Proof. By Proposition 3.1(iv), $(\forall i)$ $\operatorname{dom}(f_i^* \,\square\, \mu\,\mathsf{q}) = (\operatorname{dom} f_1^*) + (\operatorname{dom} \mu\,\mathsf{q}) = X$. Fact 3.4(i), Proposition 3.2(i), Proposition 3.2(iii), and Proposition 3.3(iv) imply that

$$\begin{aligned} \big(\lambda_1(f_1^* \,\square\, \mu\,\mathsf{q}) + \cdots + \lambda_n(f_n^* \,\square\, \mu\,\mathsf{q})\big)^* &= (\lambda_1(f_1^* \,\square\, \mu\,\mathsf{q}))^* \,\square\, \cdots \,\square\, (\lambda_n(f_n^* \,\square\, \mu\,\mathsf{q}))^* \\ &= \lambda_1 \star (f_1^* \,\square\, \mu\,\mathsf{q})^* \,\square\, \cdots \,\square\, \lambda_n \star (f_n^* \,\square\, \mu\,\mathsf{q})^* \\ &= \lambda_1 \star (f_1^{**} + (\mu\,\mathsf{q})^*) \,\square\, \cdots \,\square\, \lambda_n \star (f_n^{**} + (\mu\,\mathsf{q})^*) \\ &= \lambda_1 \star (f_1 + \mu \star \mathsf{q}) \,\square\, \cdots \,\square\, \lambda_n \star (f_n + \mu \star \mathsf{q}). \end{aligned} \tag{21}$$

This and Proposition 3.3(ii) yield (18). In turn, Fact 3.4(i) and Proposition 3.3(iv) imply (19). Changing variables, we see that (20) is equivalent to (16). ■

Remark 4.4 (some history) In [6], the proximal average was considered for $n = 2$ and $\mu = 1$, and written equivalently as

$$\big(\lambda_1(f_1^* \,\square\, \mathsf{q}) + \lambda_2(f_2^* \,\square\, \mathsf{q})\big)^* - \mathsf{q}; \tag{22}$$

see (18). The function (22) was utilized in [6] to explicitly illustrate Moreau's observation [19] that the set of proximal mappings is convex. More recently, the proximal average was considered in [4], again with $n = 2$ and $\mu = 1$, though it was written as (see (19))

$$\big(\lambda_1(f_1 + \mathsf{q})^* + \lambda_2(f_2 + \mathsf{q})^*\big)^* - \mathsf{q}. \tag{23}$$

Example 4.5 (connection to means of numbers) Let $\alpha_1, \ldots, \alpha_n$ be strictly positive numbers and suppose that $(\forall i)\ f_i = \alpha_i\,\mathsf{q}$. Using (19), we see that

$$p_{\mu^{-1}}(\boldsymbol{f}, \boldsymbol{\lambda}) = \left(\sum_{i=1}^n \lambda_i(\alpha_i\,\mathsf{q} + \mu\,\mathsf{q})^*\right)^* - \mu\,\mathsf{q} = \left(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i + \mu}\,\mathsf{q}\right)^* - \mu\,\mathsf{q} = \left(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i + \mu}\right)^{-1}\mathsf{q} - \mu\,\mathsf{q} \tag{24}$$

and thus

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \left(\left(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i + \mu^{-1}}\right)^{-1} - \mu^{-1}\right)\mathsf{q}. \tag{25}$$

Denote the coefficient of q in (25) by δ. Since δ is the difference of the weighted harmonic mean of $\alpha_1 + \mu^{-1}, \ldots, \alpha_n + \mu^{-1}$ and μ^{-1}, the Harmonic-Arithmetic Mean Inequality implies that δ does not exceed the weighted arithmetic mean

$$\sum_{i=1}^n \lambda_i\alpha_i. \tag{26}$$

As $\mu \to +\infty$, we note that δ converges to the weighted harmonic mean

$$\left(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i}\right)^{-1}, \tag{27}$$

while a calculus exercise shows that δ approaches, as $\mu \to 0^+$, the weighted arithmetic mean (26). In Remark 8.6, we revisit this example from a more general point of view.

The next result locates the domain of the proximal average exactly; moreover, it strengthens [4, Theorem 4.11], where equality was observed only for the closures and interiors.

Theorem 4.6 (domain) $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n$.

Proof. Using Proposition 3.1(iv) and Proposition 3.1(ii), we obtain $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \operatorname{dom}(\lambda_1 \star (f_1 + \mu \star \mathsf{q})) + \cdots + \operatorname{dom}(\lambda_1 \star (f_1 + \mu \star \mathsf{q})) = \lambda_1 \operatorname{dom}(f_1 + \mu \star \mathsf{q}) + \cdots + \lambda_n \operatorname{dom}(f_n + \mu \star \mathsf{q}) = \lambda_1 \operatorname{dom}(f_1) + \cdots + \lambda_n \operatorname{dom}(f_n)$. ■

Corollary 4.7 *Suppose that at least one function f_i has full domain and that $\lambda_i > 0$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ has full domain.*

Example 4.8 Assume each $\lambda_i > 0$ and each $f_i = \iota_{C_i}$, where C_i is a nonempty closed convex subset of X. In X^n, set $H = \big\{(z_i) \mid \sum \sqrt{\lambda_i} z_i = 0\big\}$ and $(\forall x \in X)$ D_x is the Cartesian product $\times \big(\sqrt{\lambda_i} C_i - \sqrt{\lambda_i} x\big)$. Then

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\colon X \to \left]-\infty, +\infty\right] : x \mapsto \tfrac{1}{2\mu} d^2_{H \cap D_x}(0). \tag{28}$$

Proof. Fix $x \in X$. Using (16), we obtain

$$\begin{aligned}
p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) &= \mu^{-1}\Big(-\tfrac{1}{2}\|x\|^2 + \inf_{\sum_i x_i = x} \sum \lambda_i \big(\mu\, \iota_{C_i}(x_i/\lambda_i) + \tfrac{1}{2}\|x_i/\lambda_i\|^2\big)\Big) \\
&= \mu^{-1} \inf_{\substack{\text{each } c_i \in C_i \\ \sum \lambda_i c_i = x}} \sum \lambda_i \big(\tfrac{1}{2}\|c_i\|^2 - \tfrac{1}{2}\|x\|^2\big) \\
&= \mu^{-1} \inf_{z = (z_i) \in H \cap D_x} \sum \lambda_i \big(\tfrac{1}{2}\big\|x + z_i/\sqrt{\lambda_i}\big\|^2 - \tfrac{1}{2}\|x\|^2\big) \\
&= \mu^{-1} \inf_{z = (z_i) \in H \cap D_x} \sum \tfrac{1}{2}\|z_i\|^2,
\end{aligned} \tag{29}$$

which completes the proof. ■

Remark 4.9 Consider Example 4.8 with $n = 2$, $\mu = 1$, $\lambda_1 > 0$, and $\lambda_2 > 0$. Then (28) simplifies to

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\colon X \to \left]-\infty, +\infty\right] : x \mapsto \frac{1}{2\lambda_1\lambda_2} d^2_{(\lambda_1(C_1 - x)) \cap (\lambda_2(x - C_2))}(0), \tag{30}$$

which is a formula first observed in [6, Theorem 6.1].

Theorem 4.10 (exactness) *For every $x \in \operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ there exist $y_i \in \lambda_i \operatorname{dom} f_i$ such that $x = y_1 + \cdots + y_n$ and $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = (\lambda_1 \star (f_1 + \mu \star \mathfrak{q}))(y_1) + \cdots + (\lambda_n \star (f_n + \mu \star \mathfrak{q}))(y_n) - (\mu \star \mathfrak{q})(x)$.*

Proof. Set $(\forall i)$ $g_i = \lambda_i \star (f_i + \mu \star \mathfrak{q})$. If $\lambda_i = 0$, then $g_i = \iota_{\{0\}}$ and hence $g_i^* = \iota_X$ has full domain. If $\lambda_i > 0$, then using Proposition 3.2(i), Fact 3.4(i), and Proposition 3.3(iv), we see that

$$g_i^* = \big(\lambda_i \star (f_i + \mu \star \mathfrak{q})\big)^* = \lambda_i (f_i + \mu \star \mathfrak{q})^* = \lambda_i (f_i^* \mathbin{\Box} (\mu \star \mathfrak{q})^*) = \lambda_i (f_i^* \mathbin{\Box} \mu\, \mathfrak{q}); \tag{31}$$

thus, g_i^* also has full domain. Therefore, by Fact 3.4(ii), the epi-sum

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) + \mu \star \mathfrak{q} = g_1 \mathbin{\Box} \cdots \mathbin{\Box} g_n \tag{32}$$

is *exact*. Since $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n$ by Theorem 4.6, the existence of the y_i is now clear. ■

5 Fenchel Conjugate

In this section, we compute the Fenchel conjugate of the proximal average. The explicit form obtained has several interesting consequences. We begin with a reformulation of Lemma 3.7:

$$(p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) + \mu \star \mathsf{q})^* = \lambda_1(f_1^* \Box \mu\,\mathsf{q}) + \cdots + \lambda_n(f_n^* \Box \mu\,\mathsf{q}). \tag{33}$$

We are now ready for a useful generalization of [6, Theorem 6.1] where $n = 2$ and $\mu = 1$.

Theorem 5.1 (Fenchel conjugate) $\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)^* = p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})$.

Proof. Set

$$g = p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) + \mu \star \mathsf{q}. \tag{34}$$

By (33), we have

$$g^* = \lambda_1(f_1^* \Box \mu\,\mathsf{q}) + \cdots + \lambda_n(f_n^* \Box \mu\,\mathsf{q}). \tag{35}$$

In view of (6), (35), Proposition 3.1(vi), Proposition 3.3(v), and Proposition 3.2(i), we obtain that

$$\begin{aligned}
\mathsf{q} - \mu^{-1} g^* &= \lambda_1\big(\,\mathsf{q} - \mu^{-1}(f_1^* \Box \mu\,\mathsf{q})\big) + \cdots + \lambda_n\big(\,\mathsf{q} - \mu^{-1}(f_1^* \Box \mu\,\mathsf{q})\big)\\
&= \lambda_1\big(\,\mathsf{q} - (\mu^{-1} f_1^* \Box \mathsf{q})\big) + \cdots + \lambda_n\big(\,\mathsf{q} - (\mu^{-1} f_1^* \Box \mathsf{q})\big)\\
&= \lambda_1\big((\mu^{-1} f_1^*)^* \Box \mathsf{q}\big) + \cdots + \lambda_n\big((\mu^{-1} f_n^*)^* \Box \mathsf{q}\big)\\
&= \lambda_1\big(\mu^{-1} \star f_1 \Box \mathsf{q}\big) + \cdots + \lambda_n\big(\mu^{-1} \star f_n \Box \mathsf{q}\big).
\end{aligned} \tag{36}$$

Consequently, using Fact 3.4(i), Proposition 3.2(i), Proposition 3.2(iii), Proposition 3.2(ii), Proposition 3.3(i), we see that

$$\begin{aligned}
\big(\,\mathsf{q} - \mu^{-1} g^*\big)^* &= \Big(\lambda_1\big(\mu^{-1} \star f_1 \Box \mathsf{q}\big) + \cdots + \lambda_n\big(\mu^{-1} \star f_n \Box \mathsf{q}\big)\Big)^*\\
&= \Big(\lambda_1\big(\mu^{-1} \star f_1 \Box \mathsf{q}\big)\Big)^* \Box \cdots \Box \Big(\lambda_n\big(\mu^{-1} \star f_n \Box \mathsf{q}\big)\Big)^*\\
&= \lambda_1 \star \big(\mu^{-1} \star f_1 \Box \mathsf{q}\big)^* \Box \cdots \Box \lambda_n \star \big(\mu^{-1} \star f_n \Box \mathsf{q}\big)^*\\
&= \lambda_1 \star \big((\mu^{-1} \star f_1)^* + \mathsf{q}^*\big) \Box \cdots \Box \lambda_n \star \big((\mu^{-1} \star f_n)^* + \mathsf{q}^*\big)\\
&= \lambda_1 \star \big(\mu^{-1} f_1^* + \mathsf{q}\big) \Box \cdots \Box \lambda_n \star \big(\mu^{-1} f_n^* + \mathsf{q}\big).
\end{aligned} \tag{37}$$

Now Proposition 3.1(vi), Proposition 3.1(ix), and Proposition 3.3(ii) imply that

$$\begin{aligned}
\mu\big(\,\mathsf{q} - \mu^{-1} g^*\big)^* &= \mu\Big(\lambda_1 \star \big(\mu^{-1}(f_1^* + \mu\,\mathsf{q})\big) \Box \cdots \Box \lambda_n \star \big(\mu^{-1}(f_n^* + \mu\,\mathsf{q})\big)\Big)\\
&= \mu\Big(\lambda_1 \star \big(\mu^{-1}(f_1^* + \mu\,\mathsf{q})\big)\Big) \Box \cdots \Box \mu\Big(\lambda_n \star \big(\mu^{-1}(f_n^* + \mu\,\mathsf{q})\big)\Big)\\
&= \lambda_1 \star (f_1^* + \mu\,\mathsf{q}) \Box \cdots \Box \lambda_n \star (f_n^* + \mu\,\mathsf{q})\\
&= \lambda_1 \star (f_1^* + \mu^{-1} \star \mathsf{q}) \Box \cdots \Box \lambda_n \star (f_n^* + \mu^{-1} \star \mathsf{q}).
\end{aligned} \tag{38}$$

Combining (34), Corollary 3.6, and (38), we conclude that

$$\begin{aligned}\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)^* &= (g-\mu\star\mathfrak{q})^* \\ &= \mu(\mathfrak{q}-\mu^{-1}g^*)^* - \mu^{-1}\star\mathfrak{q} \\ &= \lambda_1\star(f_1^*+\mu^{-1}\star\mathfrak{q}) \oplus\cdots\oplus \lambda_n\star(f_n^*+\mu^{-1}\star\mathfrak{q}) - \mu^{-1}\star\mathfrak{q} \\ &= p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda}),\end{aligned} \tag{39}$$

as claimed. ■

Corollary 5.2 (lower semicontinuity) $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ *is convex, lower semicontinuous, and proper.*

Proof. Applying Theorem 5.1 twice, we deduce that $(p_\mu(\boldsymbol{f},\boldsymbol{\lambda}))^{**} = (p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda}))^* = p_{(\mu^{-1})^{-1}}(\boldsymbol{f}^{**},\boldsymbol{\lambda}) = p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$. ■

The next result refines the corresponding two-function version [4, Proposition 4.8].

Example 5.3 $p(\boldsymbol{f},\boldsymbol{f}^*) = \mathfrak{q}$.

Proof. Theorem 5.1 readily implies that the $p(\boldsymbol{f},\boldsymbol{f}^*)$ is equal to its conjugate; consequently, it must be equal to $\mathfrak{q}$ by Proposition 3.3(i). ■

Theorem 5.4 (inequalities) $(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* \le p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \le \lambda_1 f_1 + \cdots + \lambda_n f_n$.

Proof. The right inequality follows from (20) (by setting $y_i = x$). Applying the right inequality to $\boldsymbol{f}^*$ and μ^{-1}, we learn that

$$p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda}) \le \lambda_1 f_1^* + \cdots + \lambda_n f_n^*. \tag{40}$$

Taking the Fenchel conjugate of (40) and utilizing Theorem 5.1, we deduce that $p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \big(p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})\big)^* \ge \big(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*\big)^*$. ■

Corollary 5.5 (infimum value)

$$\lambda_1 \inf f_1 + \cdots + \lambda_n \inf f_n \le \inf p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \le \inf(\lambda_1 f_1 + \cdots + \lambda_n f_n). \tag{41}$$

Corollary 5.6 (common minimizers) *Suppose that* $\bigcap_{i\colon \lambda_i>0} \operatorname{argmin}(f_i) \neq \varnothing$. *Then*

$$\min p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \sum_{i\colon \lambda_i>0} \lambda_i \min f_i \quad \textit{and} \quad \operatorname{argmin} p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \bigcap_{i\colon \lambda_i>0} \operatorname{argmin}(f_i). \tag{42}$$

Proof. Combine Theorem 5.4 and Corollary 5.5. ■

6 Moreau Envelope and Proximal Mapping

Definition 6.1 *Let* $f \in \Gamma(X)$. *The* Moreau envelope *of* f *with parameter* μ *is* $e_\mu f = f \,\square\, \mu \star \mathfrak{q}$.

Observe that

$$e_\mu f = (f^* + \mu\,\mathfrak{q})^*. \tag{43}$$

Theorem 6.2 (Moreau envelope and its Fenchel conjugate)

(i) $e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n$.

(ii) $\big(e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = \lambda_1 \star (e_\mu f_1)^* \,\square \cdots \square\, \lambda_n \star (e_\mu f_n)^*$.

Proof. Fix $y \in X$ and set $I = \big\{i \in \{1, \ldots, n\} \mid \lambda_i > 0\big\}$. Using (16), we obtain

$$\begin{aligned}
\big(e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)(y) &= \inf_x p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) + \frac{1}{2\mu}\|y - x\|^2 \\
&= \inf_x \inf_{\sum_{i\in I} x_i = x} \sum_{i\in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|x_i/\lambda_i\|^2\Big) + \frac{1}{2\mu}\|y\|^2 - \frac{1}{\mu}\langle x, y\rangle \\
&= \inf_x \inf_{\sum_{i\in I} x_i = x} \sum_{i\in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|x_i/\lambda_i\|^2 + \frac{1}{2\mu}\|y\|^2 - \frac{1}{\mu}\langle x_i/\lambda_i, y\rangle\Big) \\
&= \inf_x \inf_{\sum_{i\in I} x_i = x} \sum_{i\in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|y - x_i/\lambda_i\|^2\Big) \\
&= \inf_{x_i, i\in I} \sum_{i\in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|y - x_i/\lambda_i\|^2\Big) \\
&= \sum_{i\in I} \lambda_i \inf_{x_i} \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|y - x_i/\lambda_i\|^2\Big) \\
&= \sum_{i\in I} \lambda_i \big(e_\mu f_i\big)(y).
\end{aligned} \tag{44}$$

This implies (i), and (ii) follows by Fenchel conjugation. Alternatively, using Definition 6.1, Proposition 3.2(iii), Theorem 5.1, Proposition 3.3(iv), and Proposition 3.3(ii), one may prove (ii) via $\big(e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = \big(p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \,\square\, \mu \star \mathfrak{q}\big)^* = \big(p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* + \mu\,\mathfrak{q} = p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda}) + \mu^{-1} \star \mathfrak{q} = \lambda_1 \star (f_1^* + \mu^{-1} \star \mathfrak{q}) \,\square \cdots \square\, \lambda_n \star (f_n^* + \mu^{-1} \star \mathfrak{q}) = \lambda_1 \star (f_1^* + \mu\,\mathfrak{q}) \,\square \cdots \square\, \lambda_n \star (f_n^* + \mu\,\mathfrak{q}) = \lambda_1 \star (e_\mu f_1)^* \,\square \cdots \square\, \lambda_n \star (e_\mu f_n)^*$, and then deduces (i) by Fenchel conjugation. ■

The following result is well known.

Proposition 6.3 *Let* $f \in \Gamma(X)$. *Then* $\operatorname{argmin} e_\mu f = \operatorname{argmin} f$.

Proof. $\operatorname{argmin} e_\mu f = \partial(e_\mu f)^*(0) = \partial(f^* + \mu\,\mathfrak{q})(0) = (\partial f^* + \mu \operatorname{Id})(0) = \partial f^*(0) = \operatorname{argmin} f$. ■

Corollary 6.4 (minimizers) $\operatorname{argmin} p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \operatorname{argmin}\big(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n\big)$.

Proof. Combine Proposition 6.3 and Theorem 6.2(i). ■

Example 6.5 (least squares solutions) Let $C_1,\ldots,C_n$ be nonempty closed convex subsets of X and suppose that $(\forall i)$ $f_i = \iota_{C_i}$. Then $\operatorname{argmin} p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \operatorname{argmin}(\lambda_1 d_{C_1}^2 + \cdots + \lambda_n d_{C_n}^2)$.

Proof. This is a consequence of Corollary 6.4 since $(\forall i)$ $e_\mu f_i = e_\mu \iota_{C_i} = \iota_{C_i} \mathbin{\square} \mu \star \mathsf{q} = \mu^{-1}\iota_{C_i} \mathbin{\square} \mu^{-1}\,\mathsf{q} = \mu^{-1}(\iota_{C_i} \mathbin{\square} \mathsf{q}) = \mu^{-1}\frac{1}{2}d_{C_i}^2$. ■

Definition 6.6 *Let* $f \in \Gamma(X)$. *The* proximal mapping *of* f *with parameter* μ *is* $P_\mu f = (\operatorname{Id} + \mu\partial f)^{-1}$.

Observe that

$$\mu^{-1}(P_\mu f)^{-1} = \partial f + \mu^{-1}\operatorname{Id}, \tag{45}$$

that

$$P_\mu f = \big(\nabla(f + \mu^{-1}\,\mathsf{q})^*\big) \circ (\mu^{-1}\operatorname{Id}), \tag{46}$$

and that

$$(P_\mu f) \circ (\mu \operatorname{Id}) = \nabla(e_{\mu^{-1}}(f^*)). \tag{47}$$

We now show that the proximal mapping of the proximal average is simply the average of the individual proximal mappings. This result, which also explains how the proximal average got its name, was first proved in [6, Theorem 6.1] when $n = 2$ and $\mu = 1$.

Theorem 6.7 (proximal mapping) $P_\mu\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big) = \lambda_1 P_\mu f_1 + \cdots + \lambda_n P_\mu f_n$.

Proof. Theorem 5.1 and Theorem 6.2(i) (the latter applied to $\boldsymbol{f}^*$ and μ^{-1}) show that

$$e_{\mu^{-1}}\big((p_\mu(\boldsymbol{f},\boldsymbol{\lambda}))^*\big) = e_{\mu^{-1}}\big(p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})\big) = \lambda_1 e_{\mu^{-1}}(f_1^*) + \cdots + \lambda_n e_{\mu^{-1}}(f_n^*); \tag{48}$$

in turn, taking gradients yields

$$\nabla\big(e_{\mu^{-1}}\big((p_\mu(\boldsymbol{f},\boldsymbol{\lambda}))^*\big)\big) = \lambda_1 \nabla(e_{\mu^{-1}}(f_1^*)) + \cdots + \lambda_n \nabla\big(e_{\mu^{-1}}(f_n^*)\big). \tag{49}$$

Using (47), we see that this is equivalent to

$$\big(P_\mu\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)\big) \circ (\mu \operatorname{Id}) = \lambda_1(P_\mu f_1) \circ (\mu \operatorname{Id}) + \cdots + \lambda_n(P_\mu f_n) \circ (\mu \operatorname{Id}). \tag{50}$$

The result follows. ■

7 Subdifferential

Theorem 7.1 (subdifferential) *Let* $(\forall i)\ x_i \in \operatorname{dom} f_i$ *and set* $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$. *Then the following hold.*

(i) *If* $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = (\lambda_1 \star(f_1 + \mu \star \mathsf{q}))(\lambda_1 x_1) + \cdots + (\lambda_n \star(f_n + \mu \star \mathsf{q}))(\lambda_n x_n) - (\mu \star \mathsf{q})(x)$, *then*

$$\partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = -\mu^{-1}x + \bigcap_i \partial(\lambda_i \star(f_i + \mu \star \mathsf{q}))(\lambda_i x_i) \tag{51}$$

$$= -\mu^{-1}x + \bigcap_{i:\ \lambda_i>0} \big(\partial f_i(x_i) + \mu^{-1}x_i\big) \tag{52}$$

$$= -\mu^{-1}x + \bigcap_{i:\ \lambda_i>0} \big(\mu^{-1}(P_\mu f_i)^{-1}(x_i)\big). \tag{53}$$

(ii) *If* $\bigcap_{i:\ \lambda_i>0}(P_\mu f_i)^{-1}(x_i) \neq \varnothing$, *then*

$$p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = (\lambda_1 \star(f_1 + \mu \star \mathsf{q}))(\lambda_1 x_1) + \cdots + (\lambda_n \star(f_n + \mu \star j))(\lambda_n x_n) - (\mu \star \mathsf{q})(x). \tag{54}$$

Proof. Set $(\forall i)\ g_i = \lambda_i \star(f_i + \mu \star \mathsf{q})$. Theorem 4.6, Theorem 4.10, and Proposition 3.3(ii) imply that

$$g_1 \dot{+} \cdots \dot{+} g_n = p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) + \mu \star \mathsf{q} = p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) + \mu^{-1}\,\mathsf{q} \tag{55}$$

is exact on $\operatorname{dom}(g_1 \dot{+} \cdots \dot{+} g_n) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n = \operatorname{dom} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$. (i): (51), (52), and (53) follow from Fact 3.8(i), Proposition 3.9, and (45), respectively. (ii): Use Fact 3.8(ii). ■

Corollary 7.2 $(\forall x \in X)\ \bigcap_{i:\ \lambda_i>0} \partial f_i(x) \subseteq \partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$.

Proof. Take $x^* \in \bigcap_{i:\ \lambda_i>0} \partial f_i(x)$. Then $(\forall i)\ \lambda_i > 0 \Rightarrow \mu x^* + x \in \mu\partial f_i(x) + x = (P_\mu f_i)^{-1}(x)$. By Theorem 7.1(ii), $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = (\lambda_1 \star(f_1 + \mu \star \mathsf{q}))(\lambda_1 x) + \cdots + (\lambda_n \star(f_n + \mu \star j))(\lambda_n x) - (\mu \star \mathsf{q})(x)$. Using Theorem 7.1(i), we deduce that $x^* = -\mu^{-1}x + \mu^{-1}(\mu x^* + x) \in \partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$. ■

For the following results, it will be convenient to write $x = x_1 \oplus \cdots \oplus x_n$ if $x = x_1 + \cdots + x_n$ and $x_i \perp x_j$ for $i \neq j$. We also write $K_1 \oplus \cdots \oplus K_n = \big\{x_1 \oplus \cdots \oplus x_n \mid \text{each } x_i \in K_i \text{ and } x_i \perp x_j \text{ for } i \neq j\big\}$.

Corollary 7.3 *Let* $K_1, \ldots, K_n$ *be nonempty closed convex cones and set* $(\forall i)\ P_i = P_{K_i}$, *the orthogonal projector onto* K_i. *Suppose that*

$$(\forall x = x_1 \oplus \cdots \oplus x_n \in K_1 \oplus \cdots \oplus K_n)(\forall i) \quad P_i x = x_i, \tag{56}$$

that

$$(\forall x \in X) \quad x = P_1 x \oplus \cdots \oplus P_n x, \tag{57}$$

and that $(\forall i)\ f_i = \iota_{K_i}$ *and* $\lambda_i > 0$. *Then*

$$(\forall x \in X) \quad p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \frac{1}{2\mu}\sum_i \frac{(1-\lambda_i)}{\lambda_i}\|P_i x\|^2. \tag{58}$$

Proof. Observe that $(\forall i)\ P_\mu f_i = (\operatorname{Id} + \mu\partial\iota_{K_i})^{-1} = (\operatorname{Id} + \partial\iota_{K_i})^{-1} = P_i$. Take $x \in X$ and set

$$(\forall i) \quad x_i = \tfrac{1}{\lambda_i} P_i x = P_i\big(\tfrac{1}{\lambda_i} x\big). \tag{59}$$

Using (57), we obtain that

$$x = \lambda_1 x_1 \oplus \cdots \oplus \lambda_n x_n. \tag{60}$$

Now set

$$z = x_1 \oplus \cdots \oplus x_n. \tag{61}$$

By (56), we have $(\forall i)\ P_i z = x_i$. Thus $z \in \bigcap_i (P_\mu f_i)^{-1}(x_i)$. Therefore, by (60) and Theorem 7.1(ii),

$$\begin{aligned}
p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) &= (\lambda_1 \star (f_1 + \mu \star \mathfrak{q}))(\lambda_1 x_1) + \cdots + (\lambda_n \star (f_n + \mu \star \mathfrak{q}))(\lambda_n x_n) - (\mu \star \mathfrak{q})(x) \\
&= \mu^{-1}\lambda_1\, \mathfrak{q}(x_1) + \cdots + \mu^{-1}\lambda_n\, \mathfrak{q}(x_n) - \mu^{-1}\, \mathfrak{q}(x) \\
&= \frac{1}{2\mu}\big(\lambda_1\|x_1\|^2 + \cdots + \lambda_n\|x_n\|^2 - \|\lambda_1 x_1 + \cdots + \lambda_n x_n\|^2\big) \\
&= \frac{1}{2\mu}\sum_i \lambda_i(1-\lambda_i)\|x_i\|^2.
\end{aligned} \tag{62}$$

The conclusion thus follows from (59). ■

The following two examples are special cases of Corollary 7.3.

Example 7.4 Let $K_1, \ldots, K_n$ be closed subspaces that are pairwise orthogonal and such that $K_1 \oplus \cdots \oplus K_n = X$ and suppose that $f_i = \iota_{K_i}$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \mu^{-1}\sum_i (\lambda_i^{-1} - 1)(\,\mathfrak{q} \circ P_{K_i})$.

Example 7.5 (See also [4, Example 4.9].) Let K be a nonempty closed convex cone in X and let $\lambda \in\,]0,1[$. Then

$$(\forall x \in X) \quad p\big((\iota_K, \iota_{K^\ominus}), (1-\lambda, \lambda)\big)(x) = \frac{1}{2(1-\lambda)\lambda}\big(\lambda^2\|P_K x\|^2 + (1-\lambda)^2\|P_{K^\ominus} x\|^2\big), \tag{63}$$

where $K^\ominus$ is the polar cone of K.

Remark 7.6 We are now in a position to show that the inequalities in Theorem 5.4 can be strict. Suppose that $n = 2$, that $f_1 = \iota_K$ that $f_2 = \iota_{K^\ominus}$, where K is a nonempty closed convex cone in X, and that $\lambda_2 = \lambda \in\,]0,1[$. Using Example 7.5, we see that Theorem 5.4 becomes

$$(\forall x \in X) \quad \iota_X(x) \le \frac{1}{2(1-\lambda)\lambda}\big(\lambda^2\|P_K x\|^2 + (1-\lambda)^2\|P_{K^\ominus} x\|^2\big) \le \iota_{\{0\}}(x). \tag{64}$$

The inequalities are strict for every $x \in X \smallsetminus \{0\}$.

Let $f \in \Gamma(X)$. Following [3, Section 5], we say that f is *essentially smooth* if ∂f is at most single-valued and $\operatorname{int}\operatorname{dom} f$ is nonempty, that f is *essentially strictly convex* if f^* is essentially smooth, and that f is *Legendre* if f is both essentially smooth and essentially strictly convex. These notions coincide in our (reflexive) Hilbert space setting with the well known notions of the same name in Euclidean space (see [21, Section 26]).

The next three results extend corresponding results in [4, Section 6] considerably.

Corollary 7.7 (essential smoothness) *Suppose that at least one function f_i is essentially smooth and that $\lambda_i > 0$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is essentially smooth.*

Proof. Since f_i is essentially smooth, the set $\operatorname{dom} f_i$ has nonempty interior. Thus $\lambda_i \operatorname{dom} f_i$ and $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n$ (see Theorem 4.6) both have nonempty interiors as well. Now take $x \in \operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ and let $y_1, \ldots, y_n$ be as in Theorem 4.10, say $(\forall i)\ y_i = \lambda_i x_i$, where $x_i \in \operatorname{dom} f_i$. By Theorem 7.1(i), $\partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) \subseteq -\mu^{-1}x + \partial f_i(x_i) + \mu^{-1}x_i$. Because f_i is essentially smooth, the set $\partial f_i(x_i)$ is either empty or singleton. Thus $\partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x)$ is either empty or singleton. Altogether, $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is essentially smooth. ■

Corollary 7.8 (essential strict convexity) *Suppose that at least one function f_i is essentially strictly convex and that $\lambda_i > 0$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is essentially strictly convex.*

Proof. Since f_i is essentially strictly convex, its conjugate f_i^* is essentially smooth. By Corollary 7.7, $p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda})$ is essentially smooth. Hence $(p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda}))^*$ is essentially strictly convex. This last function is equal to $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ (by Theorem 5.1) and the proof is thus complete. ■

Corollary 7.9 (Legendre function) *Suppose that at least one function f_i is essentially smooth and that $\lambda_i > 0$. Furthermore, suppose that at least one function f_j is essentially strictly convex and that $\lambda_j > 0$. (It does not matter whether j and i are identical or distinct.) Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is both essentially smooth and essentially strictly convex, i.e., Legendre.*

Proof. Combine Corollary 7.7 and Corollary 7.8. ■

Before we formulate and prove the last result in this section, we briefly return to the Moreau envelope and the proximal mapping. Let $f \in \Gamma(X)$. Applying Proposition 3.3(v) to μf, we readily deduce that (see also [22, Example 11.26(b)])

$$\mu(e_\mu f) + \mu \star (e_{\mu^{-1}}(f^*)) = \mathfrak{q}. \tag{65}$$

Taking gradients and recalling (47) yields $\mathrm{Id} = P_\mu f + \mu(P_{\mu^{-1}}(f^*)) \circ (\mu^{-1}\,\mathrm{Id})$; equivalently, $\mu\,\mathrm{Id} = (P_\mu f) \circ (\mu\,\mathrm{Id}) + \mu P_{\mu^{-1}}(f^*)$ or

$$\mathrm{Id} = \mu^{-1}(P_\mu f) \circ (\mu\,\mathrm{Id}) + P_{\mu^{-1}}(f^*). \tag{66}$$

The following result generalizes [5, Theorem 4.22], where $n = 2$, $\lambda_1 = \lambda_2 = \frac{1}{2}$, and $\mu = 1$.

Theorem 7.10 *Suppose that $(a, a^*) \in X \times X$ satisfies $a^* \in \partial f_1(a) \cap \cdots \cap \partial f_n(a)$ and that $\{1, 2, \ldots, n\}$ is the disjoint union of two sets of indices I and J. Set $\lambda_J = \sum_{j \in J} \lambda_j$ and suppose that $\lambda_J > 0$. Then for every $z \in a + \big(\bigcap_{i \in I} N_{\operatorname{dom} f_i}(a) \cap \bigcap_{j \in J} N_{\operatorname{dom} f_j^*}(a^*)\big)$, we have*

$$a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a) \in \partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(z). \tag{67}$$

Consequently, $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is differentiable on $a + \operatorname{int}\big(\bigcap_{i \in I} N_{\operatorname{dom} f_i}(a) \cap \bigcap_{j \in J} N_{\operatorname{dom} f_j^*}(a^*)\big)$, with gradient $z \mapsto a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a)$.

Proof. Let z be as in the conclusion and set $y = z - a$. Fix $i \in I$. Now $a^* \in \partial f_i(a)$ and $\lambda_J^{-1} y \in N_{\operatorname{dom} f_i}(a) = \partial \iota_{\operatorname{dom} f_i}(a) = \partial \iota_{\operatorname{dom} \mu f_i}(a)$. Hence $\mu a^* \in \mu \partial f_i(a) = \partial(\mu f_i)(a)$. Thus $\mu a^* + \lambda_J^{-1} y \in \partial(\mu f_i)(a) + \partial(\iota_{\operatorname{dom} \mu f_i})(a) \subseteq \partial(\mu f_i + \iota_{\operatorname{dom} \mu f_i})(a) = \partial(\mu f_i)(a)$. It follows that

$$(\forall i \in I) \quad a = (P_\mu f_i)(\mu a^* + \lambda_J^{-1} y + a). \tag{68}$$

Next, fix $j \in J$. Then $a + \lambda_J^{-1} y \in \partial f_j^*(a^*)$ and $\mu^{-1} a + \mu^{-1}\lambda_J^{-1} y \in \partial(\mu^{-1} f_j^*)(a^*)$. Using (66), we thus have $a^* = (P_{\mu^{-1}} f_j^*)(\mu^{-1}a + \mu^{-1}\lambda_J^{-1} y + a^*) = \mu^{-1}a + \mu^{-1}\lambda_J^{-1} y + a^* - \mu^{-1}(P_\mu f_j)(a + \lambda_J^{-1} y + \mu a^*)$. Hence

$$(\forall j \in J) \quad a + \lambda_J^{-1} y = (P_\mu f_j)(a + \lambda_J^{-1} y + \mu a^*). \tag{69}$$

Now (68), (69), and Theorem 6.7 imply that

$$a + y = (P_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}))(a + \lambda_J^{-1} y + \mu a^*); \tag{70}$$

equivalently,

$$a^* + \mu^{-1}(\lambda_J^{-1} - 1) y \in \partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(a + y). \tag{71}$$

This verifies (67). Denote the intersection of the n normal cones by N. On $a + \operatorname{int} N$, the mapping $z \mapsto a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a)$ is thus a continuous selection of $\partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$; therefore, $\nabla p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(z) = a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a)$ by [20, Proposition 2.8]. ■

8 Pointwise Limits of the Proximal Average

Proposition 8.1 *Let $f \in \Gamma(X)$. Then $e_{\mu^{-1}}\big(f \circ (\mu \operatorname{Id})\big) = (e_\mu f) \circ (\mu \operatorname{Id})$.*

Proof. For every $x \in X$, we have $e_{\mu^{-1}}\big(f \circ (\mu \operatorname{Id})\big)(x) = \inf_y \big(f(\mu y) + \mu\, \mathfrak{q}(x - y)\big) = \inf_y \big(f(\mu y) + \mu^{-1}\, \mathfrak{q}(\mu x - \mu y)\big) = \inf_z \big(f(z) + \mu^{-1}\, \mathfrak{q}(\mu x - z)\big) = e_\mu f(\mu x)$. ■

Proposition 8.2 [22, Example 11.26(c)] *Let $f\colon X \to [-\infty, +\infty]$. Then*

$$(f + \mu\, \mathfrak{q})^* = (\mu\, \mathfrak{q} - e_{\mu^{-1}} f) \circ (\mu^{-1} \operatorname{Id}). \tag{72}$$

Proof. For every $x^* \in X$, we obtain that

$$\begin{aligned}
(f + \mu\, \mathfrak{q})^*(x^*) &= \sup_x \big(\langle x, x^*\rangle - f(x) - \mu\, \mathfrak{q}(x)\big) \\
&= \sup_x \big(\langle x, x^*\rangle - f(x) - \mu\, \mathfrak{q}(x - \mu^{-1}x^*) + \mu^{-1}\, \mathfrak{q}(x^*) - \langle x, x^*\rangle\big) \\
&= \mu^{-1}\, \mathfrak{q}(x^*) + \sup_x \big(-f(x) - \mu\, \mathfrak{q}(x - \mu^{-1}x^*)\big) \\
&= \mu^{-1}\, \mathfrak{q}(x^*) - \inf_x \big(f(x) + \mu\, \mathfrak{q}(\mu^{-1}x^* - x)\big) \\
&= \mu^{-1}\, \mathfrak{q}(x^*) - (f \mathbin{\Box} \mu\, \mathfrak{q})(\mu^{-1}x^*) \\
&= \mu\, \mathfrak{q}(\mu^{-1}x^*) - (f \mathbin{\Box} \mu^{-1} \star \mathfrak{q})(\mu^{-1}x^*) \\
&= \big(\mu\, \mathfrak{q} - e_{\mu^{-1}} f\big)(\mu^{-1}x^*).
\end{aligned} \tag{73}$$

The result follows. ■

The following alternative expression of the proximal average was discovered by Warren Hare for the case when $n = 2$ and $\mu = 1$.

Theorem 8.3 [9] $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = -e_\mu\big(-(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)\big)$.

Proof. Set $g = -(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)$. Taking the Fenchel conjugate on both sides of (33) leads to $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \big(\lambda_1(f_1^* \mathbin{⊕} \mu\,\mathsf{q}) + \cdots + \lambda_n(f_n^* \mathbin{⊕} \mu\,\mathsf{q})\big)^* - \mu \star \mathsf{q}$. On the other hand, $(\forall i)\ f_i^* \mathbin{⊕} \mu\,\mathsf{q} = (f_i + \mu \star \mathsf{q})^*$ by Fact 3.4(i) and Proposition 3.3(iii). Altogether,

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \big(\lambda_1(f_1 + \mu \star \mathsf{q})^* + \cdots + \lambda_n(f_n + \mu \star \mathsf{q})^*\big)^* - \mu \star \mathsf{q}. \tag{74}$$

Using (74), Proposition 3.3(ii), Proposition 8.2, and Proposition 8.1 we deduce that

$$\begin{aligned} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) &= \big(\lambda_1(f_1 + \mu \star \mathsf{q})^* + \cdots + \lambda_n(f_n + \mu \star \mathsf{q})^*\big)^* - \mu \star \mathsf{q} \\ &= \big(\lambda_1(f_1 + \mu^{-1}\,\mathsf{q})^* + \cdots + \lambda_n(f_n + \mu^{-1}\,\mathsf{q})^*\big)^* - \mu \star \mathsf{q}. \\ &= \big(\lambda_1(\mu^{-1}\,\mathsf{q} - e_\mu f_1) \circ (\mu\,\mathrm{Id}) + \cdots + \lambda_n(\mu^{-1}\,\mathsf{q} - e_\mu f_n) \circ (\mu\,\mathrm{Id})\big)^* - \mu \star \mathsf{q} \\ &= \big(\mu\,\mathsf{q} + g \circ (\mu\,\mathrm{Id})\big)^* - \mu \star \mathsf{q} \\ &= \big(\mu\,\mathsf{q} - e_{\mu^{-1}}(g \circ (\mu\,\mathrm{Id}))\big) \circ (\mu^{-1}\,\mathrm{Id}) - \mu \star \mathsf{q} \\ &= \mu^{-1}\,\mathsf{q} - \big(e_{\mu^{-1}}(g \circ (\mu\,\mathrm{Id}))\big) \circ (\mu^{-1}\,\mathrm{Id}) - \mu \star \mathsf{q} \\ &= -\big((e_\mu g) \circ (\mu\,\mathrm{Id})\big) \circ (\mu^{-1}\,\mathrm{Id}) \\ &= -e_\mu g. \end{aligned} \tag{75}$$

This verifies the result. ■

The *μ-proximal hull* of a function g is defined by $h_\mu g = -e_\mu(-e_\mu g)$; it satisfies $e_\mu g \le h_\mu g \le g$ and $e_\mu(h_\mu g) = e_\mu g$ (see [22, Example 1.44]). Theorem 8.3 shows that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ can be interpreted as some sort of weighted proximal hull of the functions $f_1, \ldots, f_n$. We now turn to the proximal hull of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

Corollary 8.4 (proximal hull) $h_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

Proof. By Theorem 6.2(i), $e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n$. Hence, using Theorem 8.3, $h_\mu(p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})) = -e_\mu(-e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})) = -e_\mu(-\lambda_1 e_\mu f_1 - \cdots - \lambda_n e_\mu f_n) = p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$. Since $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) + \mu \star \mathsf{q}$ is clearly convex and lower semicontinuous (by Corollary 5.2), the result follows alternatively from [22, Example 11.26(d)]. ■

Let us now determine the pointwise behaviour of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

Theorem 8.5 (pointwise limits) *Let $x \in X$. Then the function*

$$]0, +\infty[\ \to\ ]-\infty, +\infty] : \mu \mapsto p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) \quad \textit{is decreasing.} \tag{76}$$

Consequently, $\lim_{\mu\to 0^+} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ *and* $\lim_{\mu\to+\infty} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ *exist. In fact,*

$$\lim_{\mu\to 0^+} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \sup_{\mu>0} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \big(\lambda_1 f_1 + \cdots + \lambda_n f_n\big)(x) \tag{77}$$

and

$$\lim_{\mu\to+\infty} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \inf_{\mu>0} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \big(\lambda_1 \star f_1 \,⊕\, \cdots \,⊕\, \lambda_n \star f_n\big)(x). \tag{78}$$

Proof. The fact that $\mu \mapsto p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ is decreasing follows from (20); consequently, the two limits exist and the supremum/infimum descriptions are clear. Now $e_\mu\big(-(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)\big) \leq -(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)$. Thus, using Theorem 8.3, we deduce that $\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n \leq -e_\mu\big(-(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)\big) = p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$. On the other hand, Theorem 5.4 implies that $p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \leq \lambda_1 f_1 + \cdots + \lambda_n f_n$. Altogether,

$$\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n \leq p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \leq \lambda_1 f_1 + \cdots + \lambda_n f_n. \tag{79}$$

It is well known that Moreau envelopes converge pointwise to the underlying function as the parameter approaches 0; see, e.g., [1, Theorem 2.64] or [22, Theorem 1.25 and Theorem 2.26]. Thus $(\forall i)$ $\lim_{\mu\to 0^+} e_\mu f_i = f_i$ pointwise and (77) follows from taking the pointwise limit in (79) at x as $\mu \to 0^+$. Using (20), we deduce that

$$\begin{aligned}
\lim_{\mu\to+\infty} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) &= \inf_{\mu>0} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) \\
&= \inf_{\mu>0} \inf_{\sum \lambda_i y_i = x} \sum \lambda_i f_i(y_i) + \frac{1}{\mu}\Big(\Big(\sum \lambda_i\, \mathsf{q}(y_i)\Big) - \mathsf{q}(x)\Big) \\
&= \inf_{\sum \lambda_i y_i = x} \inf_{\mu>0} \sum \lambda_i f_i(y_i) + \frac{1}{\mu}\Big(\Big(\sum \lambda_i\, \mathsf{q}(y_i)\Big) - \mathsf{q}(x)\Big) \\
&= \inf_{\sum \lambda_i y_i = x} \sum \lambda_i f_i(y_i) \\
&= \inf_{\sum' x_i = x} {\sum}' \lambda_i f_i(x_i/\lambda_i) \\
&= \inf_{\sum' x_i = x} {\sum}' (\lambda_i \star f_i)(x_i) \\
&= \big(\lambda_1 \star f_1 \,⊕\, \cdots \,⊕\, \lambda_n \star f_n\big)(x),
\end{aligned} \tag{80}$$

where the indices in the $\sum'$ sums range over all i such that $\lambda_i > 0$. ■

The following nice observation, which is based on the comments of an anonymous referee, builds a bridge to [17].

Remark 8.6 (parallel sums) Suppose that $X = \mathbb{R}^N$, let $A_1, \ldots, A_n$ be positive definite $N \times N$ matrices, and suppose that $(\forall i)$ $f_i(x) = \frac{1}{2}\langle x, A_i x\rangle$, i.e., identify each A_i with its quadratic form. As $\mu \to 0^+$, $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ converges pointwise to $\lambda_1 f_1 + \cdots + \lambda_n f_n$ and, as $\mu \to +\infty$, $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ converges pointwise to $\lambda_1 \star f_1 \,⊕\, \cdots \,⊕\, \lambda_n \star f_n$. Using [17] (see also [12, Example IV.2.3.8], [14], and [16]), the matrices corresponding to the quadratic forms $\lambda_1 f_1 + \cdots + \lambda_n f_n$, $\lambda_1 \star f_1 \,⊕\, \cdots \,⊕\, \lambda_n \star f_n$, and $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$

are, respectively, the arithmetic average $\lambda_1 A_1 + \cdots + \lambda_n A_n$; the harmonic average $(\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1}$, i.e., the *parallel sum* of the matrices $\lambda_1^{-1} A_1, \ldots, \lambda_n^{-1} A_n$; and $(\lambda_1 (A_1 + \mu^{-1}\operatorname{Id})^{-1} + \cdots + \lambda_n (A_n + \mu^{-1}\operatorname{Id})^{-1})^{-1} - \mu^{-1}\operatorname{Id}$, i.e., a μ^{-1}-shifted version of the harmonic average (in accordance with the comment before Definition 4.1). Note that this provides another proof of Example 4.5 and that the theory for parallel sum extends to matrices that are only positive semidefinite.

9 Epi-Continuity and Epi-Limits of the Proximal Average

We now discuss the convergence behaviour of the proximal average with respect to the epi-topology. Analogously to [4, Section 5], we assume throughout this section that

$$X \text{ is finite-dimensional.} \tag{81}$$

Definition 9.1 (epi-convergence and epi-topology) (See [22, Chapter 6].) *Let g and $(g_k)_{k\in\mathbb{N}}$ be functions from X to $]-\infty, +\infty]$. Then $(g_k)_{k\in\mathbb{N}}$* epi-converges *to g, in symbols $g_k \xrightarrow{\text{e}} g$, if the following hold for every $x \in X$.*

(i) $\big(\forall\, (x_k)_{k\in\mathbb{N}}\big)\ x_k \to x \Rightarrow g(x) \leq \underline{\lim}\, g_k(x_k)$.

(ii) $\big(\exists\, (y_k)_{k\in\mathbb{N}}\big)\ y_k \to x$ *and* $\overline{\lim}\, g_k(y_k) \leq g(x)$.

The epi-topology *is the topology induced by epi-convergence.*

Fact 9.2 *Let g and $(g_k)_{k\in\mathbb{N}}$ be in $\Gamma(X)$ such that $g_k \xrightarrow{\text{e}} g$, and let h and $(h_k)_{k\in\mathbb{N}}$ be in $\Gamma(X)$ such that $h_k \xrightarrow{\text{e}} h$. Let ρ and $(\rho_k)_{k\in\mathbb{N}}$ be in $[0, +\infty[$ such that $\rho_k \to \rho$ and let $q\colon X \to \mathbb{R}$ be continuous. Then the following hold.*

(i) $g_k \pm q \xrightarrow{\text{e}} g \pm q$.

(ii) $\rho > 0 \Rightarrow \rho_k g_k \xrightarrow{\text{e}} \rho g$.

(iii) $\rho = 0$ *and* $\operatorname{dom} g = X \Rightarrow \rho_k g_k \xrightarrow{\text{e}} \rho g$.

(iv) $g_k^* \xrightarrow{\text{e}} g^*$.

(v) $0 \in \operatorname{int}(\operatorname{dom} g - \operatorname{dom} h) \Rightarrow g_k + h_k \xrightarrow{\text{e}} g + h$.

Proof. (i): See [22, Exercise 7.8(a)]. (ii): See [22, Exercise 7.8(d)]. (iii): See [4] or verify this directly. (iv): See [22, Theorem 11.34]. (v): See [22, Exercise 7.47(b)]. ■

Lemma 9.3 *Let $g_1, \ldots, g_n, h$ be in $\Gamma(X)$ and let $(g_{1,k})_{k\in\mathbb{N}}, \ldots, (g_{n,k})_{k\in\mathbb{N}}, (h_k)_{k\in\mathbb{N}}$ be sequences in $\Gamma(X)$ such that $(\forall i)$ $g_{i,k} \xrightarrow{\text{e}} g_i$ and $h_k \xrightarrow{\text{e}} h$. Let ρ and $(\rho_k)_{k\in\mathbb{N}}$ be in $[0, +\infty[$ such that $\rho_k \to \rho$. Suppose that* $\operatorname{dom} g_1^* = \cdots = \operatorname{dom} g_{n-1}^* = \operatorname{dom} h^* = X$ *and that* $(\forall i \in \{1, \ldots, n-1\})(\forall k)$ $\operatorname{dom} g_{i,k}^* = X$*. Then the following hold.*

(i) $g_{1,k} \mathbin{\text{⧺}} \cdots \mathbin{\text{⧺}} g_{n,k} \xrightarrow{\text{e}} g_1 \mathbin{\text{⧺}} \cdots \mathbin{\text{⧺}} g_n$.

(ii) $\rho_k \star h_k \xrightarrow{\text{e}} \rho \star h$.

Proof. (i): Fact 9.2(iv)&(v) imply that $g_{1,k}^* + \cdots + g_{n,k}^* \xrightarrow{\text{e}} g_1^* + \cdots + g_n^*$. Using Fact 9.2(iv), we see that $(g_{1,k}^* + \cdots + g_{n,k}^*)^* \xrightarrow{\text{e}} (g_1^* + \cdots + g_n^*)^*$, which is equivalent to $g_{1,k} \mathbin{\text{⧺}} \cdots \mathbin{\text{⧺}} g_{n,k} \xrightarrow{\text{e}} g_1 \mathbin{\text{⧺}} \cdots \mathbin{\text{⧺}} g_n$ by Fact 3.4(i). (ii): Fact 9.2(ii)–(iv) imply that $\rho_k h_k^* \xrightarrow{\text{e}} \rho h^*$. Using Fact 9.2(iv) once more, we deduce that $(\rho_k h_k^*)^* \xrightarrow{\text{e}} (\rho h^*)^*$, which is the same as the conclusion in view of Proposition 3.2(i). ■

Remark 9.4 Using the horizon functions associated with $g_1, \ldots, g_n$ and [22, Proposition 7.56], one may obtain a stronger version of Lemma 9.3 where the assumption on the functions $g_{i,k}^*$ is less restrictive; however, this is not needed in the sequel.

The next result extends [4, Theorem 5.4].

Theorem 9.5 (epi-continuity of the proximal average) *Let $(f_{i,k})_{k\in\mathbb{N}}$ be sequences in $\Gamma(X)$ such that $(\forall i)$ $f_{i,k} \xrightarrow{\text{e}} f_i$, let $(\lambda_{i,k})_{k\in\mathbb{N}}$ be sequences in $[0,1]$ such that $(\forall k)$ $\sum_i \lambda_{i,k} = 1$ and $(\forall i)$ $\lambda_{i,k} \to \lambda_i$, and let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\mu_k \to \mu$. Then*

$$p_{\mu_k}\big((f_{1,k}, \ldots, f_{n,k}), (\lambda_{1,k}, \ldots, \lambda_{n,k})\big) \xrightarrow{\text{e}} p_{\mu}\big((f_1, \ldots, f_n), (\lambda_1, \ldots, \lambda_n)\big) = p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}). \tag{82}$$

Proof. By Theorem 9.3(ii),

$$\mu_k \star \mathfrak{q} \xrightarrow{\text{e}} \mu \star \mathfrak{q}. \tag{83}$$

Furthermore,

$$(\forall i) \quad f_{i,k} + \mu_k \star \mathfrak{q} \xrightarrow{\text{e}} f_i + \mu \star \mathfrak{q} \tag{84}$$

by Fact 9.2(v) because $(\mu \star \mathfrak{q})^* = \mu\,\mathfrak{q}$ has full domain. Using (84), Lemma 9.3(ii), and the fact that $(\forall i)$ $(f_i + \mu \star \mathfrak{q})^* = (f_i^* \mathbin{\text{⧺}} (\mu \star \mathfrak{q})^*)^{**} = (f_i^* \mathbin{\text{⧺}} \mu\,\mathfrak{q})^{**}$ has full domain (and similarly for $(f_{i,k} + \mu_k \star \mathfrak{q})^*$), we deduce that

$$(\forall i) \quad \lambda_{i,k} \star (f_{i,k} + \mu_k \star \mathfrak{q}) \xrightarrow{\text{e}} \lambda_i \star (f_i + \mu \star \mathfrak{q}). \tag{85}$$

Since $(\forall i)$ $\big(\lambda_i \star (f_i + \mu \star \mathfrak{q})\big)^* = \lambda_i (f_i + \mu \star \mathfrak{q})^* = \lambda_i (f_i^* \mathbin{\text{⧺}} \mu \star \mathfrak{q})$ has full domain (and similarly for $(\lambda_{i,k} \star (f_{i,k} + \mu_k \star \mathfrak{q}))^*$), (85) and Lemma 9.3(i) yield

$$\lambda_{1,k} \star (f_{1,k} + \mu_k \star \mathfrak{q}) \mathbin{\text{⧺}} \cdots \mathbin{\text{⧺}} \lambda_{n,k} \star (f_{n,k} + \mu_k \star \mathfrak{q}) \xrightarrow{\text{e}} \lambda_1 \star (f_1 + \mu \star \mathfrak{q}) \mathbin{\text{⧺}} \cdots \mathbin{\text{⧺}} \lambda_n \star (f_n + \mu \star \mathfrak{q}). \tag{86}$$

In turn, (83), (86) and Fact 9.2(i) imply (82). ■

We now describe the behaviour of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ when μ approaches either 0 or $+\infty$ while $\boldsymbol{f}$ and $\boldsymbol{\lambda}$ are fixed.

Corollary 9.6 *$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{\text{e}} \lambda_1 f_1 + \cdots + \lambda_n f_n$ as $\mu \to 0^+$, and $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{\text{e}} \operatorname{cl}(\lambda_1 \star f_1 \mathbin{\text{⧺}} \cdots \mathbin{\text{⧺}} \lambda_n \star f_n)$ as $\mu \to +\infty$.*

Proof. Theorem 8.5 shows that $\mu \mapsto p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is pointwise increasing. In view of (77) and the lower semicontinuity of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ (see Corollary 5.2), an application of [22, Proposition 7.4(d)] yields that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{\text{e}} \lambda_1 f_1 + \cdots + \lambda_n f_n$ as $\mu \to 0^+$. Combining (78) with [22, Proposition 7.4(e)], we deduce similarly that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{\text{e}} \operatorname{cl}(\lambda_1 \star f_1 \mathbin{\Box\!\!\!+} \cdots \mathbin{\Box\!\!\!+} \lambda_n \star f_n)$ as $\mu \to +\infty$. ■

Corollary 9.6 and (77) show that as $\mu \to 0^+$, the pointwise and epigraphical limits of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ coincide. When $\mu \to +\infty$, the pointwise and epigraphical limits of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ may differ as we illustrate next.

Example 9.7 Suppose that $X = \mathbb{R}^2$, that $n = 2$, that $\lambda_1 > 0$, that $\lambda_2 > 0$, that $f_1 = \iota_{C_1}$, and that $f_2 = \iota_{C_2}$, where C_1 and C_2 are nonempty closed convex subsets of X such that $\lambda_1 C_1 + \lambda_2 C_2$ is not closed. Concretely, we may let C_1 and C_2 be the epigraphs of $x \mapsto \exp(x)$ and $x \mapsto \exp(-x)$, respectively. Then the pointwise limit (see (78))

$$\lim_{\mu \to +\infty} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \star f_1 \mathbin{\Box\!\!\!+} \lambda_2 \star f_2 = \iota_{\lambda_1 C_1 + \lambda_2 C_2} \tag{87}$$

is not lower semicontinuous, and hence different from the epigraphical limit (see Corollary 9.6) $\operatorname{cl}(\lambda_1 \star f_1 \mathbin{\Box\!\!\!+} \lambda_2 \star f_2)$, which is the indicator function of the closure of $\lambda_1 C_1 + \lambda_2 C_2$.

We now show that the limiting behaviour as $\mu \to +\infty$ cannot be obtained by conjugation.

Example 9.8 Suppose that $X = \mathbb{R}^2$, that $n = 2$, that $f_1 \colon (x, y) \mapsto -x + \iota_{\{0\}}(y)$, that $f_2 \colon (x, y) \mapsto x + \iota_{\{0\}}(y)$, that $\lambda_1 > 0$, and that $\lambda_2 > 0$. Now fix $(x, y) \in \mathbb{R}^2$. Using (16) and some calculus, we calculate

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x, y) = (\lambda_2 - \lambda_1)x + \iota_{\{0\}}(y) - 2\mu\lambda_1\lambda_2 = (\lambda_1 f_1 + \lambda_2 f_2)(x, y) - 2\mu\lambda_1\lambda_2. \tag{88}$$

Letting $\mu \to 0^+$ in (88) and in accordance with (77), we observe that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \to \lambda_1 f_1 + \lambda_2 f_2$ pointwise. Recalling (78) and letting $\mu \to +\infty$ in (88), we see that

$$(\lambda_1 \star f_1 \mathbin{\Box\!\!\!+} \lambda_2 \star f_2)(x, y) = \lim_{\mu \to +\infty} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x, y) = \begin{cases} -\infty, & \text{if } y = 0; \\ +\infty, & \text{if } y \neq 0. \end{cases} \tag{89}$$

Since $f_1^*(x, y) = \iota_{\{-1\}}(x)$ and $f_2^*(x, y) = \iota_{\{1\}}(x)$, we have $\operatorname{dom}(f_1^*) \cap \operatorname{dom}(f_2^*) = \varnothing$ and thus $\lambda_1 f_1^* + \lambda_2 f_2^* \equiv +\infty$. Altogether,

$$\lambda_1 \star f_1 \mathbin{\Box\!\!\!+} \lambda_2 \star f_2 \neq \big(\lambda_1 f_1^* + \lambda_2 f_2^*\big)^* \equiv -\infty. \tag{90}$$

Therefore, due to the absence of a constraint qualification on f_1^* and f_2^*, the epigraphical convergence of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ to the epigraphical average of f_1 and f_2 as $\mu \to +\infty$ could not have been obtained by conjugating the epigraphical convergence of $p_{\mu^{-1}}(f_1^*, f_2^*, \lambda_1, \lambda_2)$ to $\lambda_1 f_1^* + \lambda_2 f_2^*$ as $\mu \to +\infty$.

In the presence of a constraint qualification, we can use the proximal average to construct a homotopic curve with very nice properties.

Remark 9.9 (epigraphical and arithmetic averages are homotopic) Suppose that $\operatorname{int}\operatorname{dom} f_1^* \cap \cdots \cap \operatorname{int}\operatorname{dom} f_{n-1}^* \cap \operatorname{dom} f_n^* \neq \varnothing$. By Fact 3.4(i) and Proposition 3.2(i), we have $(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* = \lambda_1 \star f_1 \mathbin{\#} \cdots \mathbin{\#} \lambda_n \star f_n$. Therefore,

$$\operatorname{cl}(\lambda_1 \star f_1 \mathbin{\#} \cdots \mathbin{\#} \lambda_n \star f_n) = \lambda_1 \star f_1 \mathbin{\#} \cdots \mathbin{\#} \lambda_n \star f_n \tag{91}$$

and hence the pointwise and epigraphical limits of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ as either $\mu \to 0^+$ or $\mu \to +\infty$ coincide by Theorem 8.5 and Corollary 9.6. Now set

$$(\forall \rho \in [0,1]) \quad q_\rho \colon x \mapsto \begin{cases} (\lambda_1 f_1 + \cdots + \lambda_n f_n)(x), & \text{if } \rho = 0; \\ p_{\tan(\rho\pi/2)}(\boldsymbol{f}, \boldsymbol{\lambda})(x), & \text{if } 0 < \rho < 1; \\ (\lambda_1 \star f_1 \mathbin{\#} \cdots \mathbin{\#} \lambda_n \star f_n)(x), & \text{if } \rho = 1. \end{cases} \tag{92}$$

Then Theorem 8.5, Corollary 9.5, and Corollary 9.6 show that $(q_\rho)_{\rho\in[0,1]}$ is a decreasing, pointwise convergent, homotopic (with respect to the epi-topology) curve between the arithmetic average $\lambda_1 f_1 + \cdots + \lambda_n f_n$ and the epigraphical average $\lambda_1 \star f_1 \mathbin{\#} \cdots \mathbin{\#} \lambda_n \star f_n$.

Acknowledgment

We wish to thank Prof. J.-B. Hiriart-Urruty for lending us his copy of [16] and two anonymous referees for their comments. Heinz Bauschke was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. Yves Lucet was partially supported by the Natural Sciences and Engineering Research Council of Canada. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] H. Attouch, *Variational Convergence for Functions and Operators*, Pitman, 1984.

[2] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, “Fitzpatrick functions, cyclic monotonicity, and Rockafellar’s antiderivative”, *Nonlinear Analysis*, vol. 66, pp. 1198–1223, 2007.

[3] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, “Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces”, *Communications in Contemporary Mathematics*, vol. 3, pp. 615–647, 2001.

[4] H. H. Bauschke, Y. Lucet, and M. Trienis, “How to transform one convex function continuously into another”, to appear in *SIAM Review*.

[5] H. H. Bauschke, Y. Lucet, and X. Wang, “Primal-dual symmetric antiderivatives for cyclically monotone operators”, *SIAM Journal on Control and Optimization*, vol. 46, pp. 2031–2051, 2007.

[6] H. H. Bauschke, E. Matoušková, and S. Reich, “Projection and proximal point methods: Convergence results and counterexamples”, *Nonlinear Analysis*, vol. 56, pp. 715–738, 2004.

[7] H. H. Bauschke and X. Wang, "The kernel average for two convex functions and its applications to the extension and representation of monotone operators", to appear in *Transactions of the American Mathematical Society*.

[8] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland, 1973.

[9] W. L. Hare, personal communication 2006.

[10] W. L. Hare, "The proximal average of nonconvex functions: a proximal stability perspective", preprint, November 2007.

[11] J.-B. Hiriart-Urruty, "A general formula on the conjugate of the difference of functions", *Canadian Mathematical Bulletin*, vol. 29, pp. 482–485, 1986.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, Springer, New York, 1996.

[13] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II*, Springer, New York, 1996.

[14] J.-B. Hiriart-Urruty and M.-L. Mazure, "Formulations variationnelles de l'addition parallèle et de la soustraction parallèlle d'opérateurs semi-définis positifs", *Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique*, vol. 302, pp. 527–530, 1986.

[15] Y. Lucet, H. H. Bauschke, and M. Trienis, "The piecewise linear-quadratic model for computational convex analysis", to appear in *Computational Optimization and Applications*.

[16] M.-L. Mazure, *Analyse variationnelle des formes quadratriques convexes*, Thèse de doctorat de l'Université Paul Sabatier, Toulouse, France, 1986.

[17] M.-L. Mazure, "L'addition parallèle d'opérateurs interprétée comme inf-convolution de formes quadratiques convexes", *RAIRO Modélisation Mathématique et Analyse Numérique*, vol. 20, pp. 497–515, 1986.

[18] G. J. Minty, "Monotone (nonlinear) operators in Hilbert space", *Duke Mathematical Journal*, vol. 29, pp. 341–346, 1962.

[19] J. J. Moreau, "Proximité et dualité dans un espace hilbertien", *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[20] R. R. Phelps, *Convex Functions, Monotone Operators, and Differentiability* (second edition), Lecture Notes in Mathematics, vol. 1364, Springer-Verlag, 1993.

[21] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[22] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, 1998.

[23] S. Simons, *Minimax and Monotonicity*, Lecture Notes in Mathematics, vol. 1693, Springer-Verlag, 1998.

[24] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, 2002.

PREPRINT, submitted on July 7, 2006

HOW TO TRANSFORM ONE CONVEX FUNCTION CONTINUOUSLY INTO ANOTHER*

HEINZ H. BAUSCHKE†, YVES LUCET‡, AND MICHAEL TRIENIS§

Abstract. The proximal average operator provides a parametric family of convex functions that continuously transform one convex function into another even when the domains of the two functions do not intersect. We prove that the proximal average operator is a homotopy with respect to the epi-topology, study its properties, and present several explicit formulas for specific classes of functions. The parametric family inherits desirable properties such as differentiability and strict convexity from the given functions.

The results illustrate the powerful tools available in convex and variational analysis from both a theoretical and a computational point of view.

Key words. Attouch-Wets topology, continuous transformation, convex function, deformation, epi-convergence, epi-topology, Fenchel conjugate, homotopy, proximal average, proximal mapping, proximity operator

AMS subject classifications. 90C25, 26A51, 26B25, 26E60, 46B03, 90-01

PII. XXXXX

1. Introduction.

Throughout, we assume that

X is a finite-dimensional Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$.

The reader is invited to visualize X as $\mathbb{R}^n$ (equipped with the usual dot product); however, our slightly more general setting guarantees that all results hold true in matrix spaces such as $\mathbb{S}^n = \left\{M \in \mathbb{R}^{n\times n} \mid M = M^T\right\}$ (equipped with the trace inner product). In this note, we are concerned about the following question.

How do we continuously transform one convex function into another?

Given two functions f_0 and f_1, the most natural approach is to consider the (pointwise) *arithmetic average*

$$(\forall\lambda \in [0,1]) \qquad (1-\lambda)f_0 + \lambda f_1. \tag{1.1}$$

This approach works well if both functions are finite-valued (in which case the functions along with their arithmetic averages are continuous [19, Corollary 10.1.1]). Fig. 1.1(a) visualizes the arithmetic averages for a linear and a quadratic function.

However, what happens if the functions are not everywhere finite-valued? This is by no means an esoteric question. In modern optimization, *constraints are modeled*

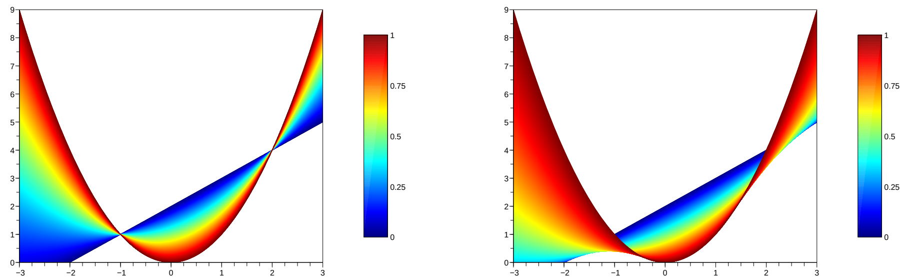
---

*Preprint, submitted on July 7, 2006.
http://www.siam.org/journals/

†Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada (heinz.bauschke@ubc.ca). Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

‡Computer Science, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada (yves.lucet@ubc.ca). Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

§Computer Science, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada (mjtrieni@interchange.ubc.ca). Research partially supported by the Pacific Institute for the Mathematical Sciences.



(a) From f_0 to f_1: Arithmetic Average (b) From f_0 to f_1: Proximal Average

FIG. 1.1. *Averages of the linear function* $f_0(x) = 2x + 2$ *and the quadratic function* $f_1(x) = x^2$.

via indicator functions: given an objective function f defined on X, and a constraint set C in X, the optimization problem

$$\text{minimize } f \text{ over } C$$

is written equivalently as

$$\text{minimize } f + \iota_C \text{ over } X, \tag{1.2}$$

where ι_C is the *indicator function* corresponding to C, i.e.,

$$\iota_C \colon X \to \left]-\infty, +\infty\right] = \mathbb{R} \cup \{+\infty\} \colon x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that (1.2) is an *unconstrained* optimization problem — the price we pay for such a nice problem is that we have to allow and work with $+\infty$. This is a common theme in modern convex analysis and optimization; see, e.g., [8, 9, 13, 19, 20]. These books also serve as reference material for notation and results not explicitly defined here. We shall work with the following standard class of functions.

$$\mathcal{F} := \big\{ f \colon X \to \left]-\infty, +\infty\right] \mid f \text{ is convex, lower semicontinuous, and proper} \big\}.$$

Let us review the concepts and fix $f \colon X \to \left]-\infty, +\infty\right]$. Then f is *convex*, if

$$\big(\forall x \in X\big)\big(\forall y \in X\big)\big(\forall \lambda \in \left]0,1\right[\big) \quad f\big((1-\lambda)x + \lambda y\big) \leq (1-\lambda)f(x) + \lambda f(y);$$

f is *lower semicontinuous*, if

$$x_n \to x \quad \Rightarrow \quad f(x) \leq \varliminf f(x_n);$$

(where $\varliminf$ denotes the limit inferior) and f is *proper*, if f is somewhere finite. A more geometric view on these three concept rests on the notion of the *epigraph* of f, which is defined by

$$\operatorname{epi} f := \big\{ (x,r) \in X \times \mathbb{R} \mid f(x) \leq r \big\}$$

and which consists of all points lying *on or above* the graph of f. Recall that a set C is convex if $(1-\lambda)x + \lambda y$ belongs to C whenever $x \in C$, $y \in C$, and $\lambda \in [0,1]$. Then the function f is convex, lower semicontinuous, and proper if and only if the set $\operatorname{epi} f$ is convex, closed, and nonempty, respectively. The (effective) *domain* of f is

$$\operatorname{dom} f := \big\{x \in X \mid f(x) \in \mathbb{R}\big\};$$

it coincides with the projection of $\operatorname{epi} f$ onto the X-component. One says that f *has full domain* if $\operatorname{dom} f = X$.

EXAMPLE 1.1. *Let C be a subset of X. Then the indicator function ι_C belongs to $\mathcal{F}$ if and only if the set C is convex, closed, and nonempty.*

EXAMPLE 1.2 (logarithmic barrier). *Suppose that $X = \mathbb{R}^n$, set $\mathbb{R}^n_{++} := \big\{x \in \mathbb{R}^n \mid$ each $x_i > 0\big\}$ and*

$$\operatorname{lb}\colon \mathbb{R}^n \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} -\sum_{i=1}^n \ln(x_i), & \text{if } x \in \mathbb{R}^n_{++}; \\ +\infty, & \text{otherwise.} \end{cases}$$

Then lb *belongs to $\mathcal{F}$.*

Note that the logarithmic barrier, which lies at the heart of interior point methods, has $\operatorname{dom} \operatorname{lb} = \mathbb{R}^n_{++}$ and it is thus not everywhere finite-valued. Moreover, the logarithmic barrier can also be defined at any matrix $M \in \mathbb{S}^n$, by mapping M to $-\ln\det(M)$, if M is positive definite; and to $+\infty$, if M is not positive definite.

Returning to our transformation problem, we now see that the arithmetic average is of no use if the domains of the functions are different. For instance, let us consider the logarithmic barrier and its reflection, i.e., let us set

$$f_0\colon \mathbb{R} \to \left]-\infty,+\infty\right] : x \mapsto \operatorname{lb}(-x) = -\ln(-x) \tag{1.3}$$

and

$$f_1\colon \mathbb{R} \to \left]-\infty,+\infty\right] : x \mapsto \operatorname{lb}(x) = -\ln(x). \tag{1.4}$$

Then $\operatorname{dom} f_0 \cap \operatorname{dom} f_1 = \left]-\infty,0\right[\cap \left]0,+\infty\right[= \varnothing$ and hence the arithmetic average $(1-\lambda)f_0 + \lambda f_1$ is identically equal to $+\infty$, for every $\lambda \in \left]0,1\right[$. This does not qualify as a continuous transformation between f_0 and f_1! At this point, it appears to be entirely unclear how a continuous transformation from f_0 to f_1 could be found. Quite surprisingly, convex analysis does offer a sensible way to transform any two functions from $\mathcal{F}$ into each other!

The goal of this note is to explain this transformation. The transformation requires basic "convex calculus" and it is thus accessible to students taking advanced undergraduate or beginning graduate courses in convex optimization and analysis. The material presented here provides a nice motivation for convex and variational analysis. It can also be used as a new topic in a corresponding course and — as we have not seen the main results elsewhere — possibly as a starting point for further research.

In contrast to the arithmetic average of (1.1), we call the average that provides the continuous transformation the *proximal average*:

$$\mathcal{P}(f_0,\lambda,f_1)\colon \mathcal{F} \times [0,1] \times \mathcal{F} \to \mathcal{F}.$$

Fig. 1.2 illustrates the proximal average between the two logarithmic barriers.

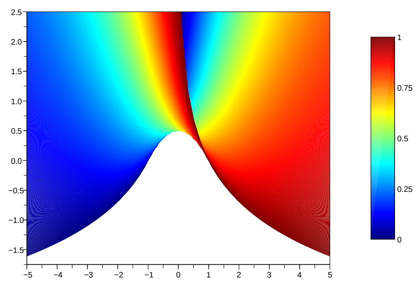


FIG. 1.2. *The proximal average of f_0 and f_1, as defined in* (1.3) *and* (1.4)*, respectively. Note that* $\operatorname{dom} f_0 \cap \operatorname{dom} f_1 = \varnothing$ *— which shows that the arithmetic average is of no use in such settings — and that for* $\lambda \in]0,1[$*, the proximal average* $\mathcal{P}(f_0, \lambda, f_1)$ *has full domain.*

However, before we are able to write down the formula for the proximal average, we require another notion from convex optimization. The *Fenchel conjugate* of a proper function $f\colon X \to]-\infty,+\infty]$ is defined by

$$f^*\colon X \to]-\infty,+\infty] : x^* \mapsto \sup_{x\in X} \langle x, x^*\rangle - f(x),$$

and the *Biconjugate Theorem* states that

$$\big(\forall f \in \mathcal{F}\big) \quad f^{**} = f. \tag{1.5}$$

The *energy* function

$$q := \tfrac{1}{2}\|\cdot\|^2$$

is the only self-conjugate function, i.e.,

$$\big(\forall f \in \mathcal{F}\big) \quad f = f^* \quad \Leftrightarrow \quad f = q. \tag{1.6}$$

Fenchel conjugates are crucial in convex optimization. The fundamental *Fenchel-Rockafellar Duality Theorem* states that given $f \in \mathcal{F}$, $g \in \mathcal{F}$, if $\operatorname{dom} f$ has a nonempty intersection with the interior of $\operatorname{dom} g$, then the value

$$\inf_{x\in X} f(x) + g(x)$$

of the primal optimization problem coincides with the value of the dual optimization problem

$$-\inf_{x^*\in X} f^*(-x^*) + g^*(x^*)$$

and the latter infimum is actually attained, i.e., a minimum.

Another basic operation is the so-called *infimal convolution* of f and g, defined by $f\Box g\colon X \to [-\infty,+\infty] : x \mapsto \inf_{y\in X} f(y) + g(x-y)$. Note that $\operatorname{dom}(f\Box g) = \operatorname{dom} f + \operatorname{dom} g$ and that $(f\Box g)^* = f^* + g^*$. If $\operatorname{dom} f$ meets the interior of $\operatorname{dom} g$, then one also has $(f+g)^* = f^*\Box g^*$. Thus, addition and infimal convolution are dual to each other.

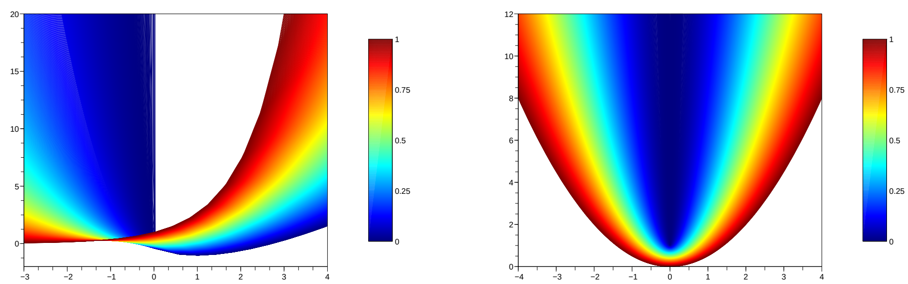
Another fundamental function in convex optimization is the so-called *negative entropy* which we define next.

EXAMPLE 1.3 (negative entropy). *Suppose that* $X = \mathbb{R}^n$ *and set* $\mathbb{R}^n_+ := \big\{x \in X \mid$ *each* $x_i \geq 0\big\}$. *With the standard convention that* $0\ln(0) = 0$, *define*

$$\operatorname{negent}\colon \mathbb{R}^n \to \left]-\infty, +\infty\right] : x \mapsto \begin{cases} \sum_{i=1}^n x_i \ln(x_i) - x_i, & \textit{if } x \in \mathbb{R}^n_+; \\ +\infty, & \textit{otherwise.} \end{cases}$$

Then negent *belongs to* $\mathcal{F}$ *and* $\operatorname{negent}^* = \exp$.

To whet the reader's appetite further, we visualize in Fig. 1.3 the proximal averages from the negative entropy to its conjugate, and from a quadratic function to an indicator function.



(a) From negent to exp

(b) From q to $\iota_{\{0\}}$

FIG. 1.3. *Proximal averages of two functions with different domains.*

The remainder of the paper is organized as follows. In Section 2, we introduce the formula for the proximal average and present basic properties and examples. Section 3 introduces epi-convergence of functions. The associated topology, the epi-topology, turns out to be exactly the right notion to establish continuity of the proximal average operator (Theorem 3.2). As demonstrated in Section 4, the proximal average inherits good properties — such as smoothness or strict convexity — from the functions it averages. As an application, we present a renorming result (Theorem 4.7). Additional examples of proximal averages are provided in the final Section 5, which can be viewed as a source of exercises.

The notation employed is standard. Let $S \subseteq X$ and $x \in X$. The *distance* from x to S is $d(x, S) := \inf_{s \in X} \|x - s\|$. The *closure* and *interior* of S is denoted by $\overline{S}$ and $\operatorname{int} S$, respectively.

2. Proximal average. The formula for the proximal average $\mathcal{P}$ that we now present appeared, in an equivalent form, first in [4], which dealt with aspects of convergence of fixed point iterations that lie beyond the scope of this paper. The proximal average was utilized there to provide an explicit constructive proof of Moreau's observation [17] that the set of all so-called proximal mappings is convex — this is the motivation for the term "proximal average".

DEFINITION 2.1 (proximal average). *The* proximal average operator $\mathcal{P}$ *is defined by*

$$\begin{aligned}(2.1)\quad \mathcal{P}\colon \mathcal{F}\times[0,1]\times\mathcal{F} &\to \big\{f \mid f\colon X\to[-\infty,+\infty]\big\}\\ (f_0,\lambda,f_1) &\mapsto \Big((1-\lambda)\big(f_0+\tfrac{1}{2}\|\cdot\|^2\big)^* + \lambda\big(f_1+\tfrac{1}{2}\|\cdot\|^2\big)^*\Big)^* - \tfrac{1}{2}\|\cdot\|^2.\end{aligned}$$

The next result is an immediate consequence of the definition and (1.5).

PROPOSITION 2.2. *Let* $f_0\in\mathcal{F}$*, let* $f_1\in\mathcal{F}$*, and let* $\lambda\in[0,1]$*. Then*

$$\mathcal{P}(f_0,0,f_1)=f_0,\quad \mathcal{P}(f_0,1,f_1)=f_1,\quad \textit{and}\ \ \mathcal{P}(f_0,\lambda,f_1)=\mathcal{P}(f_1,1-\lambda,f_0).$$

For $\lambda\in\,]0,1[$, it appears on first glance to be unlikely that the proximal average $\mathcal{P}(f_0,\lambda,f_1)$ is a convex function (since the difference between convex functions generally fails to be convex). However, $\mathcal{P}(f_0,\lambda,f_1)$ is not only convex, but *its conjugate is the corresponding proximal average of the conjugates!*

FACT 2.3. *Let* $f_0\in\mathcal{F}$*, let* $f_1\in\mathcal{F}$*, and let* $\lambda\in[0,1]$*. Then* $\mathcal{P}(f_0,\lambda,f_1)$ *belongs to* $\mathcal{F}$*, and*

$$\big(\mathcal{P}(f_0,\lambda,f_1)\big)^* = \mathcal{P}(f_0^*,\lambda,f_1^*).$$

Proof. This is a consequence of [4, Theorem 6.1]. ∎

REMARK 2.4. Let $f_0\in\mathcal{F}$, let $f_1\in\mathcal{F}$, and let $\lambda\in[0,1]$. Set $f_\lambda := \mathcal{P}(f_0,\lambda,f_1)$ and $(f^*)_\lambda := \mathcal{P}(f_0^*,\lambda,f_1^*)$. Then Fact 2.3 states that

$$(f_\lambda)^* = (f^*)_\lambda,$$

so we write these identical terms conveniently as f_λ^*.

We shall compute explicitly the proximal average of some indicator functions by specializing the following result.

EXAMPLE 2.5. *Let* C_0 *and* C_1 *be two nonempty closed convex subset of* X*, and let* $\lambda\in\,]0,1[$*. Set* $f_0:=\iota_{C_0}$*,* $f_1:=\iota_{C_1}$*, and* $f_\lambda:=\mathcal{P}(f_0,\lambda,f_1)$*. Then for every* $x\in X$*,*

$$f_\lambda(x) = (1-\lambda)\lambda\inf\big\{\tfrac{1}{2}\|c_0-c_1\|^2 \mid (1-\lambda)c_0+\lambda c_1 = x \ \textit{and each}\ c_i\ \textit{belongs to}\ C_i\big\}.$$

Moreover, $\operatorname{dom} f_\lambda = (1-\lambda)C_0+\lambda C_1$.

Proof. The "Moreover" part is clear from the formula for $f_\lambda(x)$, which follows from [4, Theorem 6.1]. ∎

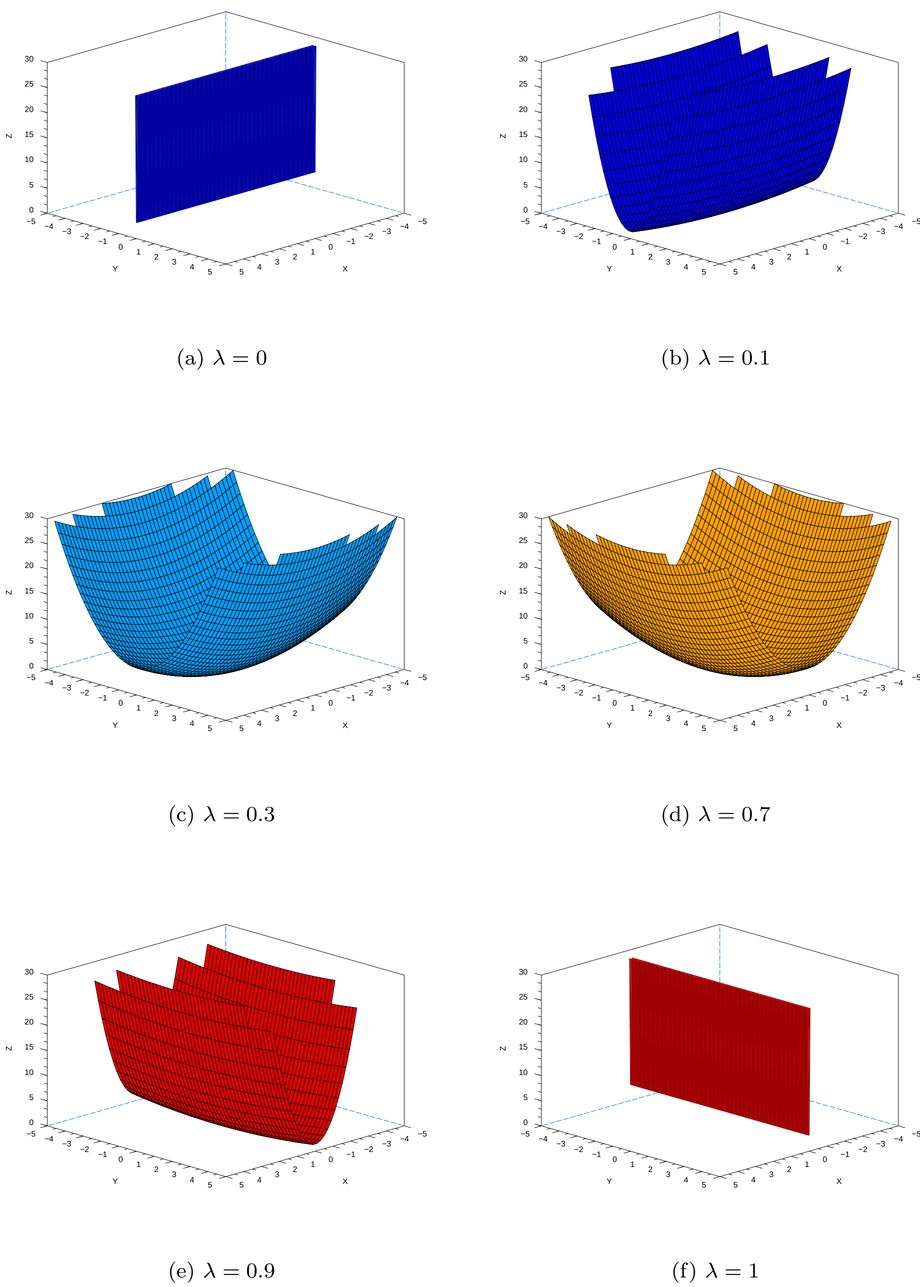
EXAMPLE 2.6 (subspace and its orthogonal complement). *Let* Y *be a subspace of* X *and let* $\lambda\in\,]0,1[$*. Set* $f_0:=\iota_Y$*,* $f_1:=\iota_{Y^\perp}$*, and* $f_\lambda:=\mathcal{P}(f_0,\lambda,f_1)$*. Then*

$$f_\lambda\colon X\to\,]-\infty,+\infty] : x\mapsto \frac{1}{2(1-\lambda)\lambda}\big(\lambda^2\|P_Y x\|^2 + (1-\lambda)^2\|P_{Y^\perp}x\|^2\big),$$

where P_Y *and* $P_{Y^\perp}$ *denote the projectors onto* Y *and its orthogonal complement* $Y^\perp$*, respectively.* Fig. 2.1 *illustrates the case* $Y=\mathbb{R}\times\{0\}$*.*

Proof. Fix $x\in X$. There is only one way to write x as a sum of two vectors from Y and $Y^\perp$, namely

$$x = P_Y x + P_{Y^\perp}x = (1-\lambda)P_Y\big(x/(1-\lambda)\big) + \lambda P_{Y^\perp}\big(x/\lambda\big).$$



(a) $\lambda = 0$ (b) $\lambda = 0.1$

(c) $\lambda = 0.3$ (d) $\lambda = 0.7$

(e) $\lambda = 0.9$ (f) $\lambda = 1$

FIG. 2.1. *The proximal average from the indicator of* $\mathbb{R} \times \{0\}$ *to the indicator of* $\{0\} \times \mathbb{R}$.

In view of Example 2.5, we conclude that

$$\begin{aligned} f_\lambda(x) &= \frac{(1-\lambda)\lambda}{2}\big\|P_Y\big(x/(1-\lambda)\big) - P_{Y^\perp}\big(x/\lambda\big)\big\|^2. \\ &= \frac{1}{2(1-\lambda)\lambda}\|\lambda P_Y x - (1-\lambda)P_{Y^\perp}x\|^2 \\ &= \frac{1}{2(1-\lambda)\lambda}\big(\lambda^2\|P_Y x\|^2 + (1-\lambda)^2\|P_{Y^\perp}x\|^2\big). \end{aligned}$$

This completes the proof. ☐

REMARK 2.7. Here are two special cases of the setting of Example 2.6.

(i) If $Y=\{0\}$, then $f_\lambda\colon x\mapsto \frac{1}{2}\big(\frac{1}{\lambda}-1\big)\|x\|^2$.

(ii) If $\lambda=\frac{1}{2}$, then $f_{\frac{1}{2}}=\frac{1}{2}\|\cdot\|^2$. We next show that this is not a coincidence: *the proximal "midpoint" between a function and its Fenchel conjugate is always the energy.* For example, no plot was taken for $\lambda=0.5$ in the sequence of images in Fig. 2.1, since it is the energy function.

PROPOSITION 2.8. *Let $f\in\mathcal{F}$. Then $\mathcal{P}(f,\frac{1}{2},f^*)=\frac{1}{2}\|\cdot\|^2$.*

Proof. Using Fact 2.3, (1.5), and Proposition 2.2, we see that

$$\big(\mathcal{P}(f,\tfrac{1}{2},f^*)\big)^* = \mathcal{P}(f^*,\tfrac{1}{2},f^{**}) = \mathcal{P}(f^*,1-\tfrac{1}{2},f) = \mathcal{P}(f,\tfrac{1}{2},f^*).$$

Thus $\mathcal{P}(f,\frac{1}{2},f^*)$ is the same as its Fenchel conjugate. On the other hand, only $q=\frac{1}{2}\|\cdot\|^2$ possesses this self-conjugacy property (see (1.6)). ☐

Another instructive case arises when we consider cones rather than subspaces. Recall that a subset K of X is a *convex cone* if K is convex and $(\forall\lambda>0)\ \lambda K=K$. Let K be a nonempty closed convex cone. Then its *polar cone* is

$$K^\ominus := \big\{x^*\in X \mid (\forall k\in K)\ \ \langle k,x^*\rangle\le 0\big\}.$$

Every subspace is a cone, but the converse is false in general (consider, e.g., $\mathbb{R}^n_+$). The polar cone of a subspace coincides with the orthogonal complement of the subspace; in general, these notions are different (e.g., $[0,+\infty[^\ominus =]-\infty,0]$ while $[0,+\infty[^\perp=\{0\}$). We also require the notion of the *projector* P_C of a nonempty closed convex set C in X. The operator P_C maps every $x\in X$ to its (unique) nearest point in C: $P_Cx\in C$, $d(x,C)=\|x-P_Cx\|$, and $(\forall c\in C\smallsetminus\{P_Cx\})\ d(x,C)<\|x-c\|$. We can now describe a classical result due to Moreau, which can be interpreted as the Pythagorean decomposition for cones.

FACT 2.9. *Let K be a nonempty closed convex cone in X. Then for any x in X, $x=P_Kx+P_{K^\ominus}x$ and $\langle P_Kx,P_{K^\ominus}x\rangle=0$.*

A second glance at the proof of Example 2.6 shows that the proof was easy because the decomposition of a vector into a sum of two vectors, one taken from the subspace and the other from the complement, is unique. This is no longer true for cones: e.g., if $K=[0,+\infty[$ in $\mathbb{R}$, then $K^\ominus=-K$ yet $2=2+0=3+(-1)$ can be written in more than one way as the sum of two elements in K and $K^\ominus$, respectively. Nonetheless, the Moreau decomposition is special in the following sense.

PROPOSITION 2.10. *Let K be a nonempty closed convex cone in X and let $x\in K$. Then the unique solution to the optimization problem*

$$\text{minimize}\ \ \|k\|\ \ \text{subject to}\ \ k\in K \text{ and } x-k\in K^\ominus \tag{2.2}$$

is the vector P_Kx.

Proof. Set $K^{\oplus} = -K^{\ominus}$. Then the optimization problem (2.2) is feasible (by Fact 2.9) and it is also equivalent to

$$\text{minimize} \quad \|k\| \quad \text{subject to} \quad k \in K \text{ and } k \in x + K^{\oplus}. \tag{2.3}$$

In turn, the unique solution of (2.3) is $P_{K\cap(x+K^{\oplus})}0$. On the other hand, [3, Remark 3.3] implies that,

$$P_{x+K^{\oplus}}0 = P_{0+K}x = P_K x.$$

Altogether, we conclude that $P_{K\cap(x+K^{\oplus})}0 = P_K x$. ☐

EXAMPLE 2.11 (cone and polar cone). *Let K be a nonempty closed convex cone in X and let $\lambda \in]0,1[$. Set $f_0 := \iota_K$, $f_1 := \iota_{K^{\ominus}}$, and $f_\lambda := \mathcal{P}(f_0,\lambda,f_1)$. Then*

$$f_\lambda \colon X \to]-\infty,+\infty] : x \mapsto \frac{1}{2(1-\lambda)\lambda}\big(\lambda^2\|P_K x\|^2 + (1-\lambda)^2\|P_{K^{\ominus}}x\|^2\big).$$

Proof. Fix $x \in X$. Utilizing Example 2.5, Fact 2.9, and Proposition 2.10, we estimate that

$$\begin{aligned}
f_\lambda(x) &\leq \frac{(1-\lambda)\lambda}{2}\|P_K(\frac{x}{1-\lambda}) - P_{K^{\ominus}}(\frac{x}{\lambda})\|^2\\
&= \frac{(1-\lambda)\lambda}{2}\left(\frac{1}{(1-\lambda)^2}\|P_K x\|^2 + \frac{1}{\lambda^2}\|P_{K^{\ominus}}x\|^2\right)\\
&= \frac{(1-\lambda)\lambda}{2}\inf_{k+l=x,k\in K,l\in K^{\ominus}}\left(\frac{1}{(1-\lambda)^2}\|k\|^2 + \frac{1}{\lambda^2}\|l\|^2\right)\\
&= \frac{(1-\lambda)\lambda}{2}\inf_{(1-\lambda)u+\lambda v=x,u\in K,v\in K^{\ominus}}\big(\|u\|^2+\|v\|^2\big)\\
&\leq \frac{(1-\lambda)\lambda}{2}\inf_{(1-\lambda)u+\lambda v=x,u\in K,v\in K^{\ominus}}\big(\|u\|^2 - 2\langle u,v\rangle + \|v\|^2\big)\\
&= \frac{(1-\lambda)\lambda}{2}\inf_{(1-\lambda)u+\lambda v=x,u\in K,v\in K^{\ominus}}\|u-v\|^2\\
&= f_\lambda(x).
\end{aligned}$$

Therefore,

$$f_\lambda(x) = \frac{(1-\lambda)\lambda}{2}\left(\frac{1}{(1-\lambda)^2}\|P_K x\|^2 + \frac{1}{\lambda^2}\|P_{K^{\ominus}}x\|^2\right)$$

and the result follows. ☐

REMARK 2.12. Consider Example 2.11. Even though the given functions f_0 and f_1 are indicator functions, the proximal average f_λ is *not* an indicator function.

We conclude this section with an observation that is complementary to the "Moreover" part of Example 2.5. However, since it requires a more technical proof, it may be skipped on first reading.

THEOREM 2.13. *Let $f_0 \in \mathcal{F}$, let $f_1 \in \mathcal{F}$, and let $\lambda \in]0,1[$. Set $f_\lambda := \mathcal{P}(f_0,\lambda,f_1)$. Then*

$$\overline{\operatorname{dom} f_\lambda} = \overline{(1-\lambda)\operatorname{dom} f_0 + \lambda \operatorname{dom} f_1}$$

and

$$\operatorname{int}\operatorname{dom} f_\lambda = \operatorname{int}\big((1-\lambda)\operatorname{dom} f_0 + \lambda \operatorname{dom} f_1\big).$$

Proof. Let us recall the required notions from convex analysis. Fix $f \in \mathcal{F}$. The *subdifferential operator* ∂f is the set-valued mapping

$$x \mapsto \big\{x^* \in X \mid (\forall h \in X)\ \ f(x) + \langle h, x^*\rangle \le f(x+h)\big\},$$

with domain $\operatorname{dom}\partial f := \big\{x \in X \mid \partial f(x) \neq \varnothing\big\}$. Recall that $q = \frac{1}{2}\|\cdot\|^2$. Then the *proximity operator* of $f \in \mathcal{F}$ is $\partial(q+f)^*$. Note that $\operatorname{ran}\partial(q+f)^* = \operatorname{ran}((\partial(q+f))^{-1}) = \operatorname{dom}\partial(q+f) = \operatorname{dom}\partial f \subseteq \operatorname{dom} f$ and thus $\overline{\operatorname{ran}\partial(q+f)^*} = \overline{\operatorname{dom} f}$; see, e.g., [19, Chapters 23 and 24]. Abbreviate the proximity operator of f_μ by P_μ, for $\mu \in \{0, \lambda, 1\}$. Then [4, Theorem 2.6] states that

$$P_\lambda = (1-\lambda)P_0 + \lambda P_1.$$

Set $A := (1-\lambda)P_0$ and $B := \lambda P_1$. Observe that A and B are subdifferential operators of functions in $\mathcal{F}$, namely of $(1-\lambda)(q+f_0)^*$ and $\lambda(q+f_1)^*$, respectively. Thus A and B are *maximal monotone* (see [19, Corollary 31.5.2] and also [10, 21] for further information), and both satisfy condition $(*)$ of Brézis and Haraux [11]. Since $A + B = P_\lambda$ is also maximal monotone, [11, Théorème 3] (see also [21, Section 19]) implies that $\overline{\operatorname{ran}(A+B)} = \overline{\operatorname{ran} A + \operatorname{ran} B}$ and $\operatorname{int}(\operatorname{ran}(A+B)) = \operatorname{int}(\operatorname{ran} A + \operatorname{ran} B)$. Hence $\overline{\operatorname{dom} f_\lambda} = \overline{\operatorname{ran}\partial(q+f_\lambda)^*} = \overline{\operatorname{ran} P_\lambda} = \overline{\operatorname{ran}(A+B)} = \overline{\operatorname{ran} A + \operatorname{ran} B} = \overline{(1-\lambda)\operatorname{ran} P_0 + \lambda \operatorname{ran} P_1} = \overline{(1-\lambda)\operatorname{dom}\partial f_0 + \lambda \operatorname{dom}\partial f_1} = \overline{(1-\lambda)\operatorname{dom} f_0 + \lambda \operatorname{dom} f_1}$. Consequently, $\operatorname{int}\operatorname{dom} f_\lambda = \operatorname{int}\overline{\operatorname{dom} f_\lambda} = \operatorname{int}\overline{(1-\lambda)\operatorname{dom} f_0 + \lambda \operatorname{dom} f_1} = \operatorname{int}((1-\lambda)\operatorname{dom} f_0 + \lambda \operatorname{dom} f_1)$. ☐

COROLLARY 2.14. *Let $f_0 \in \mathcal{F}$, let $f_1 \in \mathcal{F}$, and let $\lambda \in\ ]0,1[$. Suppose that f_0 or f_1 has full domain. Then $\mathcal{P}(f_0,\lambda,f_1)$ has full domain as well.*

Proof. Assume that f_0 has full domain. Then $\operatorname{int}\operatorname{dom} f_\lambda = \operatorname{int}((1-\lambda)\operatorname{dom} f_0 + \lambda \operatorname{dom} f_1) = \operatorname{int}((1-\lambda)X + \lambda \operatorname{dom} f_1) = X$. ☐

REMARK 2.15. Theorem 2.13 makes is clear why we observed in Fig. 1.2 that $\mathcal{P}(f_0,\lambda,f_1)$ has full domain for $\lambda \in\ ]0,1[$. The results in this section admit extensions to a general (possibly infinite-dimensional) Hilbert space setting; the required results from convex analysis can be found in [22].

3. Homotopy. As the above examples show, the proximal average is not pointwise continuous, i.e., for $f_0 \in \mathcal{F}$, $f_1 \in \mathcal{F}$, and $x \in X$, the mapping $[0,1] \to\ ]-\infty,+\infty] : \lambda \mapsto \mathcal{P}(f_0,\lambda,f_1)(x)$ is in general *not* continuous. Nonetheless, the above examples and figures seem to suggest that *some* form of continuity is present. This is indeed the case when we utilize the *epi-topology*, a basic tool of modern variational analysis. Especially in our finite-dimensional setting, the corresponding theory is well developed, beautiful, and ready to use. We will utilize results from the seminal book by Rockafellar and Wets [20].

DEFINITION 3.1 (epi-convergence and epi-topology). *Let f and $(f_n)_{n\in\mathbb{N}}$ be functions from X to $]-\infty,+\infty]$. Then $(f_n)_{n\in\mathbb{N}}$* epi-converges *to f, if the following hold for every $x \in X$.*

(i) *For every sequence $(x_n)_{n\in\mathbb{N}}$ in X converging to x, one has $f(x) \le \underline{\lim}\, f_n(x_n)$.*
(ii) *There exists a sequence $(x_n)_{n\in\mathbb{N}}$ in X converging to x such that $\overline{\lim}\, f_n(x_n) \le f(x)$.*

The epi-topology *is the topology induced by epi-convergence.*

We are now able to provide a simple proof of the continuity of the proximal average operator.

THEOREM 3.2 (continuity). *Suppose that $\mathcal{F}$ is equipped with the epi-topology. Then the proximal average operator $\mathcal{P}\colon \mathcal{F} \times [0,1] \times \mathcal{F} \to \mathcal{F}$ is continuous.*

Proof. The following operations are all continuous.

(i) $\mathcal{F} \to \mathcal{F}\colon f \mapsto f + q$ (see [20, Exercise 7.8(a)]).
(ii) $\mathcal{F} \to \mathcal{F}\colon f \mapsto f^*$ (see [20, Theorem 11.34]).
(iii) $[0,1] \times \big\{f \in \mathcal{F} \mid \operatorname{dom} f = X\big\} \to \big\{h \in \mathcal{F} \mid \operatorname{dom} h = X\big\}\colon (\mu, f) \mapsto \mu f$ (for $]0,1[\times \big\{f \in \mathcal{F} \mid \operatorname{dom} f = X\big\}$ see [20, Exercise 7.8(d)]; if the scalar is 0, use the working definition [20, Proposition 7.2] directly).
(iv) $\mathcal{F} \times \mathcal{F} \to \mathcal{F}\colon (f, g) \mapsto f + g$ under a constraint qualification (see [20, Exercise 7.47]).
(v) $\mathcal{F} \to \big\{h \mid h\colon X \to \,]-\infty, +\infty]\big\}\colon f \mapsto f - q$ (see [20, Exercise 7.8(a)]).

Since the proximal average operator is built on these operations (see (2.1)), the result follows. ∎

COROLLARY 3.3 (homotopy). *Let $f_0 \in \mathcal{F}$ and let $f_1 \in \mathcal{F}$. Then*

$$\mathcal{P}(f_0, f_1, \cdot)\colon [0,1] \to \mathcal{F}$$

provides a homotopy between f_0 and f_1. Consequently, all functions in $\mathcal{F}$ are homotopic.

REMARK 3.4. In infinite-dimensional spaces, the theory is more complicated and there are a number of topologies one might consider (most of which coincide in finite dimensions). We believe the most useful topology for carrying the results of this section over is the *Attouch-Wets topology*, for which summation and Fenchel conjugation are continuous operations. We refer the interested reader to the books [1, 6, 15] as well as to the articles [2, 18].

4. Inheritance.

Recall that $f \in \mathcal{F}$ is *strictly convex*, if

$$\left.\begin{array}{c} x \in \operatorname{dom} f \\ y \in \operatorname{dom} f \\ \lambda \in \,]0,1[\\ x \neq y \end{array}\right\} \quad \Rightarrow \quad f\big((1-\lambda)x + \lambda y\big) < (1-\lambda) f(x) + \lambda f(y).$$

Strict convexity is important in optimization because it guarantees that there is at most one minimizer of any objective function with this property. Fenchel conjugacy shows that strict convexity and differentiability are dual to each other in the following immensely useful and beautiful sense.

FACT 4.1. *Let $f \in \mathcal{F}$ such that both f and f^* have full domain. Then*

$$f \text{ is differentiable} \Leftrightarrow f^* \text{ is strictly convex.}$$

We now show that the proximal average inherits good properties such as differentiability or strict convexity from the given functions.

THEOREM 4.2 (inheritance). *Let $f_0 \in \mathcal{F}$, let $f_1 \in \mathcal{F}$, and let $\lambda \in \,]0,1[$. Set $f_\lambda := \mathcal{P}(f_0, \lambda, f_1)$. Suppose that f_0 or f_1 has full domain, and that f_0^* or f_1^* has full domain. Then the following hold.*

(i) *Both f_λ and f_λ^* have full domain.*
(ii) *If f_0 or f_1 is differentiable everywhere, then so is f_λ.*

(iii) *If f_0 or f_1 is strictly convex and its Fenchel conjugate has full domain, then f_λ is strictly convex.*

Proof. (i): f_λ has full domain by Corollary 2.14. A dual argument (recall Remark 2.4) shows that f_λ^* has full domain as well. (ii): Assume that f_0 is differentiable with full domain. Then so is $f_0 + q$ and hence (by Fact 4.1) $(f_0 + q)^* = f_0^* \Box q$ is strictly convex with full domain. Thus $(1-\lambda)(f_0+q)^* + \lambda(f_1+q)^*$ is strictly convex with full domain. Item (i) and Fact 4.1 imply that $f_\lambda + q$ is differentiable with full domain. Therefore, f_λ is differentiable with full domain. (iii): Assume that f_0 is strictly convex and that $\operatorname{dom} f_0^* = X$. Then f_0^* is differentiable everywhere and hence so is f_λ^* (by (ii) applied to f_0^* and f_1^*). We conclude that f_λ is strictly convex. □

Corollary 4.3. *Let $f_0 \in \mathcal{F}$, let $f_1 \in \mathcal{F}$, and let $\lambda \in \,]0,1[$. Suppose that f_0, f_1, and their Fenchel conjugates have full domain, that f_0 is strictly convex, and that f_1 is differentiable. Set $f_\lambda := \mathcal{P}(f_0, \lambda, f_1)$. Then both f_λ and f_λ^* are differentiable and strictly convex.*

Remark 4.4. Assume the hypothesis of Corollary 4.3 holds, and that f_1 is *not* differentiable. The conclusion of Corollary 4.3 guarantees that the proximal average f_λ *is differentiable.* In contrast, the arithmetic average $(1-\lambda)f_0 + \lambda f_1$ *is not differentiable.* This illustrates another advantage of the proximal average over the arithmetic average. Note that both averages produce a strictly convex function.

The next result will be used later.

Proposition 4.5 (proximal average vs arithmetic average). *Let $f_0 \in \mathcal{F}$, let $f_1 \in \mathcal{F}$, and let $\lambda \in \,]0,1[$. Suppose that $0 \le f_0$, $0 \le f_1$, and set $f_\lambda := \mathcal{P}(f_0, \lambda, f_1)$. Then*

$$0 \le f_\lambda \le (1-\lambda) f_0 + \lambda f_1.$$

Proof. Clearly, $q \le q + f_0$ and $q \le q + f_1$. Conjugate these inequalities, scale the results (by $1-\lambda$ and λ, respectively), and then add to obtain $q \ge (1-\lambda)(q+f_0)^* + \lambda(q+f_1)^*$. Conjugate one more time to obtain $q \le ((1-\lambda)(q+f_0)^* + \lambda(q+f_1)^*)^*$. It follows that

$$\begin{aligned} q &\le \big((1-\lambda)(q+f_0)^* + \lambda(q+f_1)^*\big)^* \\ &\le (1-\lambda)(q+f_0)^{**} + \lambda(q+f_1)^{**} \\ &= q + (1-\lambda) f_0 + \lambda f_1. \end{aligned}$$

The result follows by subtracting q. □

Remark 4.6. Suppose that f_{lower}, f_0, f_1, and f_{upper} belong to $\mathcal{F}$ and that $f_{\text{lower}} \le \min\{f_0, f_1\} \le \max\{f_0, f_1\} \le f_{\text{upper}}$. We leave it to the reader to show that

$$(\forall \lambda \in \,]0,1[) \quad f_{\text{lower}} \le f_\lambda \le f_{\text{upper}}.$$

Recall that a *norm* $[\![\cdot]\!]$ (on X) is a function $X \to \mathbb{R}$ satisfying the following three properties: $(\forall x \in X)$ $[\![x]\!] \ge 0$; $(\forall x \in X)(\forall r \in \mathbb{R})$ $[\![rx]\!] = |r|[\![x]\!]$; and $(\forall x \in X)(\forall y \in X)$ $[\![x+y]\!] \le [\![x]\!] + [\![y]\!]$. Every norm $[\![\cdot]\!]$ can be written as $[\![\cdot]\!] = 2\sqrt{f}$, where the unique function $f \in \mathcal{F}$ is nonnegative, it vanishes only at 0, and it is *quadratic-homogeneous*, i.e., $(\forall x \in X)(\forall r \in \mathbb{R})$ $f(rx) = r^2 f(x)$; we shall say that f *is the function associated with* $[\![\cdot]\!]$, and that $[\![\cdot]\!]$ *is the norm associated with* f. Let $[\![\cdot]\!]$ be a norm with associated function f. Then f^* is associated with the corresponding *dual norm* $[\![\cdot]\!]^\circ \colon X \to \mathbb{R} \colon x^* \mapsto \sup\big\{\langle x, x^*\rangle \mid [\![x]\!] \le 1\big\}$. (See [19, Chapter 15].) The norm $[\![\cdot]\!]$ is *strictly convex*, if f is; and $[\![\cdot]\!]$ is *smooth*, if f is differentiable everywhere. Note that Fact 4.1 implies that $[\![\cdot]\!]$ is strictly convex $\Leftrightarrow$ $[\![\cdot]\!]^\circ$ is smooth, and that $[\![\cdot]\!]$ is smooth $\Leftrightarrow$ $[\![\cdot]\!]^\circ$ is strictly convex.

We now show that the proximal average can be employed to create a new norm that inherits good properties from two given norms.

THEOREM 4.7 (renorming). *Let $[\![\cdot]\!]_0$ and $[\![\cdot]\!]_1$ be two norms on X, and denote their associated functions by f_0 and f_1. Let $\lambda \in\,]0,1[$ and set $f_\lambda := \mathcal{P}(f_0,\lambda,f_1)$. Then f_λ is associated with some norm $[\![\cdot]\!]_\lambda$, f_λ^* is associated with $[\![\cdot]\!]_\lambda^\circ$, and*

$$0 \leq \Big(\frac{1}{1-\lambda}f_0\Big)\Box\Big(\frac{1}{\lambda}f_1\Big) \leq f_\lambda \leq (1-\lambda)f_0 + \lambda f_1. \tag{4.1}$$

If $[\![\cdot]\!]_0$ is strictly convex and $[\![\cdot]\!]_1$ is smooth, then $[\![\cdot]\!]_\lambda$ and $[\![\cdot]\!]_\lambda^\circ$ are strictly convex and smooth.

Proof. Since $0 \leq f_0$ and $0 \leq f_1$, the first inequality in (4.1) follows immediately. The third inequality in (4.1) was already recorded in Proposition 4.5. Applying this proposition to f_0^* and f_1^*, which are the functions associated with the dual norms $[\![\cdot]\!]_0^\circ$ and $[\![\cdot]\!]_1^\circ$, we see that $f_\lambda^* \leq (1-\lambda)f_0^* + \lambda f_1^*$. Conjugate this inequality to obtain

$$f_\lambda \geq \Big(\big((1-\lambda)f_0^*\big)^*\Big)\Box\Big(\big(\lambda f_1^*\big)^*\Big) = \Big((1-\lambda)f_0\big(\,\cdot\,/(1-\lambda)\big)\Big)\Box\Big(\lambda f_0\big(\,\cdot\,/\lambda\big)\Big).$$

The second inequality in (4.1) thus follows from the quadratic homogeneity of f_0 and f_1. Let us show that f_λ is associated with some norm $[\![\cdot]\!]_\lambda$. It is clear that f_λ belongs to $\mathcal{F}$, and (4.1) yields $0 \leq f_\lambda$. Suppose that $x \in X$ satisfies $f_\lambda(x) = 0$. Since f_0 and f_1 are both coercive, the infimal convolution of (4.1) is exact at x. Hence there exist vectors x_0 and x_1 in X such that $x = x_0+x_1$ and $0 \leq (1/(1-\lambda))f_0(x_0)+(1/\lambda)f_1(x_1) \leq f_\lambda(x) = 0$; Thus $f_0(x_0) = f_1(x_1) = 0$, which implies $x_0 = 0 = x_1$ and further $x = 0$. It follows that f_λ vanishes only at 0. The definition of f_λ now leads readily to the quadratic homogeneity of f_λ. Therefore, f_λ is associated with some norm $[\![\cdot]\!]_\lambda$ and f_λ^* is associated with $[\![\cdot]\!]_\lambda^\circ$. Finally, we assume that f_0 is strictly convex and f_1 is differentiable everywhere. In view of Corollary 4.3, f_λ and f_λ^* are strictly convex and differentiable; equivalently, $[\![\cdot]\!]_\lambda$ and $[\![\cdot]\!]_\lambda^\circ$ are strictly convex and smooth. ∎

Define two norms at any point $(x_1,x_2) \in \mathbb{R}^2$ via

$$[\![(x_1,x_2)]\!]_0 = \begin{cases} \frac{1}{2}|x_1|, & \text{if } |x_1| > 2|x_2|; \\ \frac{1}{2}|x_2|, & \text{if } |x_2| > 2|x_1|; \\ |x_1| + |x_2| - \sqrt{2|x_1||x_2|}, & \text{otherwise,} \end{cases} \tag{4.2}$$

and

$$[\![(x_1,x_2)]\!]_1 = \frac{|x_1| + \sqrt{4|x_1|^2 + 3|x_2|^2}}{3}. \tag{4.3}$$

Note that $[\![\cdot]\!]_0$ is smooth but not strictly convex, and that $[\![\cdot]\!]_1$ is strictly convex but not smooth. The unit spheres of these norms — as well as those of all norms $[\![\cdot]\!]_\lambda$ obtained by Theorem 4.7 — are illustrated in Fig. 4.1. As predicted by the theory, the intermediate norms $[\![\cdot]\!]_\lambda$ are both smooth and strictly convex.

5. Exercises. In this final section, we present a collection of examples some of which were previously considered by the first-named author and A. Jarvis [14]. We do leave the details of these examples to the interested reader as exercises and we

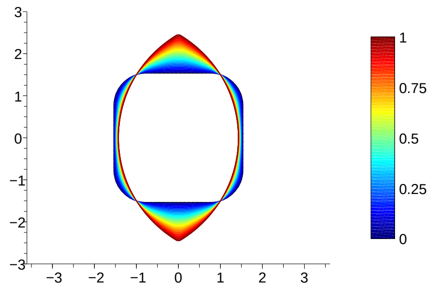


FIG. 4.1. *An illustration of Theorem 4.7 when* $[\![\cdot]\!]_0$ *and* $[\![\cdot]\!]_1$ *are defined by* (4.2) *and* (4.3)*, respectively. The unit spheres of the norms* $[\![\cdot]\!]_\lambda$*, where* $\lambda \in [0,1]$*, are visualized.*

encourage him/her to utilize available software such as [5, 7, 12, 16], which was very beneficial for the creation of the examples presented in this paper.

In all these examples, we shall specify two functions f_0 and f_1 in $\mathcal{F}$, and we shall assume that

$$\lambda \in \,]0,1[\quad \text{and} \quad f_\lambda := \mathcal{P}(f_0,\lambda,f_1).$$

The first example shows that the proximal average coincides with the arithmetic average if one function is simply a (vertical) translate of the other.

EXAMPLE 5.1. *Let* $f_0 \in \mathcal{F}$*, let* $\gamma \in \mathbb{R}$*, and set* $f_1 := f_0 + \gamma$*. Then* $f_\lambda = f_0 + \lambda\gamma = (1-\lambda)f_0 + \lambda f_1$*.*

On the other hand, if we consider the proximal average between 0 and a nonzero linear functional, then we will not obtain the arithmetic average.

EXAMPLE 5.2. *Let* $a \in X \smallsetminus \{0\}$*, set* $f_0 := 0$ *and* $f_1 := \langle \cdot, a\rangle$*. Then* $f_\lambda = \lambda\langle\cdot,a\rangle - (1-\lambda)\lambda q(a)$*.*

Next let us demonstrate that simple scaling may lead to relatively complicated proximal averages.

EXAMPLE 5.3. *Let* $\alpha > 0$*, set* $f_0 := q$ *and* $f_1 := \alpha q$*. Then* $f_\lambda = \dfrac{(1+\lambda)\alpha + 1 - \lambda}{(1-\lambda)\alpha + 1 + \lambda} q$*.*

By discussing the case of true quadratics and linear functions separately, it is possible to obtain the following explicit formula for transforming between two quadratics. Note that the proximal average is a quadratic as well unless both functions are linear.

EXAMPLE 5.4 (two quadratics). *Let* $\alpha_0 \ge 0$*, let* $\alpha_1 \ge 0$*, let* $b_0 \in X$*, let* $b_1 \in X$*, let* $\gamma_0 \in \mathbb{R}$*, and let* $\gamma_1 \in \mathbb{R}$*. Define* $f_0 \colon x \mapsto \alpha_0\|x\|^2 + \langle x, b_0\rangle + \gamma_0$ *and* $f_1 \colon x \mapsto \alpha_1\|x\|^2 + \langle x, b_1\rangle + \gamma_1$*, and set*

(i) $\delta_\lambda := 1 + 2(\lambda\alpha_0 + (1-\lambda)\alpha_1)$,

(ii) $\alpha_\lambda := \dfrac{2\alpha_0\alpha_1 + (1-\lambda)\alpha_0 + \lambda\alpha_1}{\delta_\lambda}$,

(iii) $b_\lambda := \dfrac{(1-\lambda)(2\alpha_1+1)b_0 + \lambda(2\alpha_0+1)b_1}{\delta_\lambda}$*, and*

(iv) $\gamma_\lambda := (1-\lambda)\gamma_0 + \lambda\gamma_1 - \dfrac{(1-\lambda)\lambda\|b_0 - b_1\|^2}{2\delta_\lambda}$.

Then

$$f_\lambda \colon x \mapsto \alpha_\lambda\|x\|^2 + \langle x, b_\lambda\rangle + \gamma_\lambda.$$

Fig. 1.1(b) *illustrates the case* $\alpha_0 = 0, b_0 = 1, \gamma_0 = 2$*, and* $\alpha_1 = 1, b_1 = 0, \gamma_1 = 0$.

The transformation from a quadratic to a singleton indicator follows along a quadratic path.

Example 5.5. *Let* $\alpha_0 \geq 0$*, let* $b_0 \in X$*, let* $b_1 \in X$*, let* $\gamma_0 \in \mathbb{R}$*, and let* $\gamma_1 \in \mathbb{R}$*. Define* $f_0\colon x \mapsto \alpha_0\|x\|^2 + \langle x, b_0\rangle + \gamma_0$ *and* $f_1 := \iota_{\{b_1\}} + \gamma_1$*, and set*

(i) $\alpha_\lambda := \dfrac{2\alpha_0 + \lambda}{2(1-\lambda)}$,

(ii) $b_\lambda := \dfrac{(1-\lambda)b_0 - \lambda(1+2\alpha_0)b_1}{1-\lambda}$, *and*

(iii) $\gamma_\lambda := \dfrac{2(1-\lambda)^2\gamma_0 + 2\lambda(1-\lambda)\gamma_1 + \lambda\langle b_1, (1+2\lambda\alpha_0)b_1 - 2(1-\lambda)b_0\rangle}{2(1-\lambda)}$.

Then

$$f_\lambda\colon x \mapsto \alpha_\lambda\|x\|^2 + \langle x, b_\lambda\rangle + \gamma_\lambda.$$

Corollary 4.3 implies that for each $\lambda \in \,]0,1[$*, the proximal average* f_λ *is both* strictly convex and differentiable. Fig. 1.3(b) *illustrates the case when* $\alpha_0 = \frac{1}{2}, b_0 = 0, \gamma_0 = 0$*, and* $b_1 = 0, \gamma_1 = 1$*. If* $\alpha_0 = 0$*, then* f_0 *is differentiable* but not strictly convex*, and* f_1 *is strictly convex* but not differentiable.

We conclude the paper with two singleton indicator functions.

Example 5.6. *Let* $b_0 \in X$*, let* $b_1 \in X$*, let* $\gamma_0 \in \mathbb{R}$*, and let* $\gamma_1 \in \mathbb{R}$*. Set* $f_0 := \iota_{\{b_0\}} + \gamma_0$ *and* $f_1 := \iota_{\{b_1\}} + \gamma_1$*. Then*

$$f_\lambda = \iota_{\{(1-\lambda)b_0 + \lambda b_1\}} + (1-\lambda)\gamma_0 + \lambda\gamma_1 + \frac{(1-\lambda)\lambda}{2}\|b_0 - b_1\|^2.$$

Acknowledgment. We wish to thank Roger Wets and Jon Vanderwerff for helpful discussions concerning the material in Section 3 and Section 4, respectively.

REFERENCES

[1] Hédy Attouch, *Variational Convergence for Functions and Operators*, Applicable Mathematics Series, Pitman (Advanced Publishing Program), Boston, MA, 1984.

[2] Hédy Attouch and Roger J.-B. Wets, *Isometries for the Legendre-Fenchel transform*, Trans. Amer. Math. Soc., 296 (1986), pp. 33–60.

[3] Heinz H. Bauschke, *Duality for Bregman projections onto translated cones and affine subspaces*, J. Approx. Theory, 121 (2003), pp. 1–12.

[4] Heinz H. Bauschke, Eva Matoušková, and Simeon Reich, *Projection and proximal point methods: Convergence results and counterexamples*, Nonlinear Anal., 56 (2004), pp. 715–738.

[5] Heinz H. Bauschke and Martin von Mohrenschildt, *Symbolic computation of Fenchel conjugates*, SIGSAM Bull., (2006).

[6] Gerald Beer, *Topologies on Closed and Closed Convex Sets*, vol. 268 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1993.

[7] Jonathan M. Borwein and Chris H. Hamilton, *Symbolic computation of multidimensional Fenchel conjugate*. Submitted, 2006.

[8] Jonathan M. Borwein and Adrian S. Lewis, *Convex Analysis and Nonlinear Optimization*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3, Springer-Verlag, New York, 2000. Theory and examples.

[9] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[10] H. Brézis, *Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert*, North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

[11] H. Brézis and Alain Haraux, *Image d'une somme d'opérateurs monotones et applications*, Israel J. Math., 23 (1976), pp. 165–186.
[12] Chris H. Hamilton, *Symbolic convex analysis*, master's thesis, Simon Fraser University, 2005.
[13] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal, *Convex Analysis and Minimization Algorithms*, vol. 305–306 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993. Vol I: Fundamentals, Vol II: Advanced theory and bundle methods.
[14] Amanda J. Jarvis, *A brief manual to using Lucet's code*, tech. report, University of Guelph, June 2004. Manuscript prepared for an Undergraduate Summer Research Project.
[15] Roberto Lucchetti, *Convexity and Well-posed Problems*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 22, Springer, New York, 2006.
[16] Yves Lucet, *Fast Moreau envelope computation I: Numerical algorithms*, tech. report, University of British Columbia Okanagan, 2005.
[17] Jean-Jacques Moreau, *Proximité et dualité dans un espace Hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[18] Jean-Paul Penot and Constantin Zălinescu, *Continuity of usual operations and variational convergences*, Set-Valued Anal., 11 (2003), pp. 225–256.
[19] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New york, 1970.
[20] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[21] Stephen Simons, *Minimax and Monotonicity*, vol. 1693 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1998.
[22] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.

SIAM J. CONTROL OPTIM.
Vol. 46, No. 6, pp. 2031–2051

PRIMAL-DUAL SYMMETRIC INTRINSIC METHODS FOR FINDING ANTIDERIVATIVES OF CYCLICALLY MONOTONE OPERATORS*

HEINZ H. BAUSCHKE†, YVES LUCET‡, AND XIANFU WANG†

Abstract. A fundamental result due to Rockafellar states that every cyclically monotone operator A admits an antiderivative f in the sense that the graph of A is contained in the graph of the subdifferential operator ∂f. Given a method $\mathfrak{m}$ that assigns every finite cyclically monotone operator A some antiderivative $\mathfrak{m}_A$, we say that the method is *primal-dual symmetric* if $\mathfrak{m}$ applied to the inverse of A produces the Fenchel conjugate of $\mathfrak{m}_A$. Rockafellar's antiderivatives do not possess this property. Utilizing Fitzpatrick functions and the proximal average, we present novel primal-dual symmetric intrinsic methods. The antiderivatives produced by these methods provide a solution to a problem posed by Rockafellar in 2005. The results leading to this solution are illustrated by various examples.

Key words. antiderivative, convex function, cyclically monotone operator, Fenchel conjugate, Fitzpatrick function, maximal monotone operator, n-cyclically monotone operator, proximal average, Rockafellar's antiderivative, Rockafellar function, subdifferential operator

AMS subject classifications. Primary 47H05; Secondary 26B25, 52A41, 90C25

DOI. 10.1137/060675794

1. Introduction.

Suppose that X is a real Banach space with continuous dual X^*, dual pairing $\langle \cdot, \cdot \rangle$, and norm $\|\cdot\|$. We start by recalling some known notions and results concerning (cyclically) monotone operators. These operators play a fundamental role in modern optimization as well as convex and variational analysis; see, e.g., [11, 12, 18, 20, 21, 22, 24] for further information and notation not explicitly defined here. Let A be a *set-valued operator* from X to X^*, i.e., $(\forall x \in X)\ Ax \subseteq X^*$; thus, A is a mapping from X to the power set of X^*. We use the notation $A\colon X \rightrightarrows X^*$ and remark that A can be identified with its *graph* $\operatorname{gra} A := \{(x, x^*) \in X \times X^* \mid x^* \in Ax\}$. The *domain* of A is $\operatorname{dom} A := \{x \in X \mid Ax \neq \varnothing\}$ and the *range* of A is $\operatorname{ran} A := A(X) = \bigcup_{x \in X} Ax$. The *inverse* of A is the operator $A^{-1}\colon X^* \rightrightarrows X$, defined by $x \in A^{-1}x^* \Leftrightarrow x^* \in Ax$. Furthermore, let $n \in \{2, 3, \dots\}$. Then A is *n-cyclically monotone* [1, 2, 3, 7, 10, 23] if the implication

$$\text{(1)} \qquad \left.\begin{array}{c} (a_1, a_1^*) \in \operatorname{gra} A, \\ \vdots \\ (a_n, a_n^*) \in \operatorname{gra} A \\ a_{n+1} := a_1 \end{array}\right\} \quad \Rightarrow \quad \sum_{i=1}^{n} \langle a_{i+1} - a_i, a_i^* \rangle \leq 0$$

*Received by the editors November 23, 2006; accepted for publication (in revised form) August 16, 2007; published electronically November 30, 2007.
http://www.siam.org/journals/sicon/46-6/67579.html

†Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, BC V1V 1V7, Canada (heinz.bauschke@ubc.ca, shawn.wang@ubc.ca). The first author was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. The third author was partially supported by the Natural Sciences and Engineering Research Council of Canada.

‡Computer Science, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, BC V1V 1V7, Canada (yves.lucet@ubc.ca). This author was partially supported by the Natural Sciences and Engineering Research Council of Canada.

holds. Note that 2-monotonicity simplifies to

$$\big(\forall (x,x^*)\in \operatorname{gra} A\big)\big(\forall (y,y^*)\in \operatorname{gra} A\big) \quad \langle x-y, x^*-y^*\rangle \geq 0, \tag{2}$$

i.e., to ordinary *monotonicity*. *Cyclic monotonicity* describes the situation when A is m-cyclically monotone for every $m \in \{2,3,\dots\}$. The operator A is *maximal n-cyclically monotone* if A is n-cyclically monotone and no proper extension (in the sense of inclusion of graphs) of A is n-cyclically monotone. Zorn's lemma guarantees that every n-cyclically monotone operator admits a maximal n-cyclically monotone extension. At one end of the spectrum of maximal n-cyclically monotone operators are the maximal 2-monotone, i.e., the *maximal monotone* operators. At the other end are the *maximal cyclically monotone operators*, which Rockafellar in a groundbreaking paper [19] (see Fact 3.4 below) revealed to be precisely the subdifferential operators of functions that are convex, lower semicontinuous, and proper.

This paper is motivated by the following question posed by Rockafellar in 2005 during open-problem sessions at conferences in Borovets (Bulgaria) and Banff (Canada).

> *Given a cyclically monotone operator A with a finite graph, find a method that produces an antiderivative of A that preserves the natural symmetry induced by convex duality.*

One motivation for the above question that we feel will become particularly relevant for applications as numerical convex analysis matures is the efficient storage and representation of convex functions. This is a surprisingly difficult problem. The perhaps most natural approach of storing grid points (x_i, y_i) causes significant problems because Lagrangian interpolation can fail to recover a convex function [13]. We now describe three other possible approaches. First, one could solve for subgradients x_i^* at each point x_i, or store such data in the first place, and then recover the function via $f(x) = \max_i \big(\langle x - x_i, x_i^*\rangle + y_i\big)$. The resulting function is piecewise linear with a full domain; thus, its conjugate has a bounded domain. Second, one could restrict the model of the function to the convex hull of the points x_i and set the function equal to $+\infty$ outside. Third, one could store the points and subgradients (x_i, x_i^*) along with a scalar y_0 and then recover a function f that satisfies $x_i^* \in \partial f(x_i)$ and $f(x_0) = y_0$. However, the existing representations in the literature [9, 19] are based on piecewise linear functions; so, in the finite graph case, one has to unavoidably privilege either the primal or the dual space in the very model used to recover the function.

In this paper, we provide constructive answers to Rockafellar's question. In fact, we shall exhibit methods for constructing antiderivatives that we call *primal-dual symmetric*. These methods have the property that, when they are applied to A^{-1} instead of A, the Fenchel conjugate of the antiderivative of A is obtained. The mere existence of such methods struck us initially as quite remarkable since antiderivatives are at best unique up to additive constants. These methods also allow for the design of models of convex functions that inherit the symmetry induced by convex duality in the given discrete data. Our constructions are based on Rockafellar's classical construction of an antiderivative as well as on recent work on Fitzpatrick functions and the proximal average operator, which has a close connection to fundamental objects of optimization such as Moreau envelopes and proximal mappings [4, 6]. Another pleasant consequence of primal-dual symmetric methods is their "slope 1" property—we believe that this will aid in efforts to represent convex functions in a numerically stable way (the "slope 1" property guarantees that the derivatives outside the domain of interest have slopes that are neither too small nor too large in magnitude). This is an area of active research that lies beyond the scope of this paper; see [15] for a one-dimensional framework that is capable to express such antiderivatives and that serves as a starting point for further research.

The remainder of this paper can be summarized as follows. In section 2, we introduce the common ancestor and Fitzpatrick functions [2]. These functions have turned out to be immensely useful in the study of—and they are intimately tied to—n-cyclic monotonicity. We provide a recursion formula for the Fitzpatrick functions (Proposition 2.13) and show that they stabilize when applied to cyclically monotone operators with a finite graph (Theorem 2.16). In section 3, we revisit Rockafellar's classical antiderivative result (Fact 3.4) in the context of Fitzpatrick functions. In fact, his antiderivative satisfies a certain minimality property (Theorem 3.5), it is related to the common ancestor function (Corollary 3.11), and a closed form can be found for some finite-graph operators on $\mathbb{R}$ (Theorem 3.14)—parts of these results, of which we were unaware during the preparation of the originally submitted version of this paper, were previously obtained by Lambert et al. in their interesting work [14] in which they focus on finding upper and lower bounds for antiderivatives using a linear programming formulation. The supremum of all Rockafellar antiderivatives is expressible in terms of a Fitzpatrick function (Theorem 3.15 and Corollary 3.16). Section 4 introduces the notion of a *primal-dual symmetric* method for antiderivatives (Definition 4.6). Such methods provide antiderivatives that depend only on the graph—which makes them *intrinsic*—and that return the Fenchel conjugate of the antiderivative when applied to the inverse operator. Neither Rockafellar's classical antiderivatives nor simple symmetrizations of them have this property (Proposition 4.7). Based on recent work on the proximal average operator, we proceed to present our main result which provides a general construction of primal-dual symmetric methods (Theorem 4.13). Concrete instances are proximal-average-based symmetrizations of the maximum and of the average of Rockafellar's antiderivatives (Examples 4.19 and 4.20). We then present a result (Corollary 4.23) that leads to a resolution of Rockafellar's problem (Corollary 4.26 and Remark 4.27). We conclude the paper with a numerical example (Example 4.28).

Our notation is standard. The subdifferential operator of a convex function f is denoted by ∂f, its Fenchel conjugate by f^*, and its domain by $\operatorname{dom} f$. For a set S, we use $\operatorname{conv} S$, $\overline{\operatorname{conv}}\, S$, $\operatorname{int} S$, and $\overline{S}$ to denote its convex hull, its closed convex hull, its interior, and its closure, respectively. For a nonempty convex subset C of X and a point $x \in C$, the tangent and the normal cone of C at x are denoted by $T_C(x)$ and by $N_C(x)$, respectively. Finally, the set of all functions from X to $]-\infty, +\infty]$ that are convex, lower semicontinuous, and proper is denoted by $\Gamma(X)$ or simply by Γ.

2. The common ancestor and Fitzpatrick functions.

DEFINITION 2.1 (see [2, Definition 2.1]). *Let $A\colon X \rightrightarrows X^*$, and let $(a_1, a_1^*) \in \operatorname{gra} A$. The* common ancestor functions *are defined by*

$$C_{A,2,(a_1,a_1^*)}\colon X \times X^* \to\ ]-\infty,+\infty] : (x,x^*) \mapsto \langle x, a_1^*\rangle + \langle a_1, x^*\rangle - \langle a_1, a_1^*\rangle, \tag{3}$$

and, for every $n \in \{3, 4, \dots\}$, by

$$\begin{aligned} C_{A,n,(a_1,a_1^*)}\colon X\times X^* &\to\ ]-\infty,+\infty] \\ (x,x^*) &\mapsto \sup_{\substack{(a_2,a_2^*)\in \operatorname{gra} A,\\ \vdots \\ (a_{n-1},a_{n-1}^*)\in \operatorname{gra} A}} \left(\sum_{i=1}^{n-2}\langle a_{i+1}-a_i, a_i^*\rangle\right) + \langle x-a_{n-1}, a_{n-1}^*\rangle + \langle a_1, x^*\rangle. \end{aligned} \tag{4}$$

We also set

$$C_{A,\infty,(a_1,a_1^*)} = \sup_{n\in\{2,3,\dots\}} C_{A,n,(a_1,a_1^*)}. \tag{5}$$

It is clear that

$$\big(\forall n \in \{2,3,\dots\}\big) \quad C_{A,n,(a_1,a_1^*)} \text{ is convex and lower semicontinuous,} \tag{6}$$

that the sequence

$$\big(C_{A,n,(a_1,a_1^*)}\big)_{n\in\{2,3,\dots\}} \text{ is increasing and pointwise convergent to } C_{A,\infty,(a_1,a_1^*)}, \tag{7}$$

and that $C_{A,\infty,(a_1,a_1^*)}$ is convex and lower semicontinuous. Moreover,

$$\operatorname{gra} A \text{ finite} \quad\Rightarrow\quad \big(\forall n \in \{2,3,\dots\}\big)\ C_{A,n,(a_1,a_1^*)} \text{ is polyhedral and continuous.} \tag{8}$$

The next result shows that common ancestor functions are closely related to n-cyclic monotonicity. The proof is straightforward and thus omitted.

PROPOSITION 2.2. *Let $A\colon X \rightrightarrows X^*$, and let $n \in \{2,3,\dots\}$. Then A is n-cyclically monotone if and only if*

$$\big(\forall (a,a^*) \in \operatorname{gra} A\big)\big(\forall (b,b^*) \in \operatorname{gra} A\big) \quad C_{A,n,(a,a^*)}(b,b^*) \le \langle b,b^*\rangle. \tag{9}$$

Computationally convenient is the following recursive formula.

PROPOSITION 2.3 (recursion). *Let $A\colon X \rightrightarrows X^*$, let $(a_1,a_1^*) \in \operatorname{gra} A$, let $n \in \{2,3,\dots\}$, and let $(x,x^*) \in X \times X^*$. Then*

$$C_{A,n+1,(a_1,a_1^*)}(x,x^*) = \sup_{(a,a^*)\in\operatorname{gra} A} C_{A,n,(a_1,a_1^*)}(a,x^*) + \langle x-a,a^*\rangle. \tag{10}$$

Proof. By definition, $C_{A,n+1,(a_1,a_1^*)}(x,x^*)$ is the supremum of the terms

$$\begin{aligned}&\left(\sum_{i=1}^{n-1}\langle a_{i+1}-a_i,a_i^*\rangle\right) + \langle x-a_n,a_n^*\rangle + \langle a_1,x^*\rangle \\ &\quad= \left(\sum_{i=1}^{n-2}\langle a_{i+1}-a_i,a_i^*\rangle\right) + \langle a_n-a_{n-1},a_{n-1}^*\rangle + \langle a_1,x^*\rangle + \langle x-a_n,a_n^*\rangle,\end{aligned} \tag{11}$$

where $(a_2,a_2^*),\dots,(a_n,a_n^*)$ in $\operatorname{gra} A$. Supremizing first over $(a_2,a_2^*),\dots,(a_{n-1},a_{n-1}^*)$, followed by supremizing over (a_n,a_n^*), we obtain the conclusion. ∎

Due to their implementability, operators with finite graphs are of particular interest. The next result demonstrates that, if a sufficiently high order of cyclic monotonicity is achieved, the common ancestor functions stabilize.

THEOREM 2.4. *Let $A\colon X \rightrightarrows X^*$, let $(a_1,a_1^*) \in \operatorname{gra} A$, and let $n \in \{2,3,\dots\}$. Suppose that A is n-cyclically monotone and that* $\operatorname{gra} A$ *has at most n points. Then* $C_{A,\infty,(a_1,a_1^*)} = C_{A,n+1,(a_1,a_1^*)}$.

Proof. Take $(x,x^*) \in X \times X^*$, and take $m \in \{n+2,n+3,\dots\}$. It suffices to show that

$$C_{A,m,(a_1,a_1^*)}(x,x^*) \le C_{A,m-1,(a_1,a_1^*)}(x,x^*), \tag{12}$$

since this and (7) then imply that $C_{A,n+1,(a_1,a_1^*)} = C_{A,n+2,(a_1,a_1^*)} = \cdots = C_{A,\infty,(a_1,a_1^*)}$. Take $(a_2,a_2^*),\dots,(a_{m-1},a_{m-1}^*)$ in $\operatorname{gra} A$. Since $\operatorname{gra} A$ contains at most n points and since $m-1 \geq n+1$, there exist integers k and l such that $1 \leq k < l \leq m-1$ and $a_k = a_l$. Hence

$$\begin{aligned}(13)\quad &\sum_{i=1}^{m-2}\langle a_{i+1}-a_i, a_i^*\rangle + \langle x - a_{m-1}, a_{m-1}^*\rangle + \langle a_1, x^*\rangle \\ &= \sum_{i=1}^{k-1}\langle a_{i+1}-a_i, a_i^*\rangle + \sigma + \sum_{i=l}^{m-2}\langle a_{i+1}-a_i, a_i^*\rangle + \langle x - a_{m-1}, a_{m-1}^*\rangle + \langle a_1, x^*\rangle,\end{aligned}$$

where

$$\sigma = \sum_{i=k}^{l-1}\langle a_{i+1}-a_i, a_i^*\rangle. \tag{14}$$

We claim that

$$\sigma \leq 0. \tag{15}$$

Note that σ contains $l-k$ terms. If $l-k \leq n$, then the n-cyclic monotonicity of A implies that $\sigma \leq 0$. Otherwise, $l-k > n$, and we may analogously and recursively split up σ until it is a finite sum of negative terms. This verifies (15). Now (13) implies (12). ∎

Example 2.5. Suppose that X is a Hilbert space. Let $e \in X$ be such that $\|e\| = 1$, and define A via $\operatorname{gra} A := \{(-e,-e),(e,e)\}$. Then A is (2-cyclically) monotone, and for every $(x,x^*) \in X \times X$ we have

$$C_{A,2,(-e,-e)}(x,x^*) = -\langle x + x^*, e\rangle - 1, \tag{16}$$
$$C_{A,2,(e,e)}(x,x^*) = \langle x + x^*, e\rangle - 1, \tag{17}$$
$$C_{A,3,(-e,-e)}(x,x^*) = \max\big\{-\langle x + x^*, e\rangle - 1, \langle x - x^*, e\rangle - 3\big\}, \tag{18}$$
$$C_{A,3,(e,e)}(x,x^*) = \max\big\{\langle x^* - x, e\rangle - 3, \langle x + x^*, e\rangle - 1\big\}. \tag{19}$$

THEOREM 2.6. *Let $A\colon X \rightrightarrows X^*$, and let $(a_1,a_1^*) \in \operatorname{gra} A$. Suppose that A is not cyclically monotone. Then $C_{A,\infty,(a_1,a_1^*)} \equiv +\infty$.*

Proof. There exist n points $(a_2,a_2^*),\dots,(a_{n+1},a_{n+1}^*)$ in $\operatorname{gra} A$, where $n \in \{2,3,\dots\}$, such that

$$\sigma := \sum_{i=2}^{n+1}\langle a_{i+1}-a_i, a_i^*\rangle > 0, \quad \text{where } a_{n+2} := a_2. \tag{20}$$

Take $(x,x^*) \in X \times X^*$. Take $k \in \{2,3,\dots\}$, and define

$$a_{kn+2} := a_{(k-1)n+2} := \cdots := a_2, \tag{21}$$
$$a_{kn+1} := a_{(k-1)n+1} := \cdots := a_{n+1}, \tag{22}$$
$$a_{kn} := a_{(k-1)n} := \cdots := a_n, \tag{23}$$
$$\vdots \tag{24}$$
$$a_{(k-1)n+3} := a_{(k-2)n+3} := \cdots := a_3, \tag{25}$$

and analogously for $a^*_{n+2}, \dots, a^*_{kn+2}$. Then

$$C_{A,kn+3,(a_1,a_1^*)}(x,x^*) \geq \sum_{i=1}^{kn+1} \langle a_{i+1} - a_i, a_i^* \rangle + \langle x - a_{kn+2}, a^*_{kn+2} \rangle + \langle a_1, x^* \rangle \tag{26}$$

$$= \langle a_2 - a_1, a_1^* \rangle + k\sigma + \langle x - a_2, a_2^* \rangle + \langle a_1, x^* \rangle \tag{27}$$

$$\to +\infty \quad \text{as } k \to +\infty. \tag{28}$$

Therefore, $\lim_{k\to+\infty} C_{A,kn+3,(a_1,a_1^*)}(x,x^*) = +\infty$, and the result now follows from (7). ☐

Example 2.7. Suppose that X is a Hilbert space. Let $e \in X$ be such that $\|e\| = 1$, and define A via $\operatorname{gra} A := \{(-e,e),(e,-e)\}$. Then A is not monotone, and for every $k \in \{2,3,\dots\}$ and $(x,x^*) \in X \times X$ we have

$$C_{A,2,(-e,e)}(x,x^*) = \langle x - x^*, e \rangle + 1, \tag{29}$$

$$C_{A,2,(e,-e)}(x,x^*) = \langle x^* - x, e \rangle + 1, \tag{30}$$

$$C_{A,2k-1,(-e,e)}(x,x^*) = 4(k-1) - 2 + \max\big\{\langle x - x^*, e \rangle - 1, -\langle x + x^*, e \rangle + 1\big\}, \tag{31}$$

$$C_{A,2k-1,(e,-e)}(x,x^*) = 4(k-1) - 2 + \max\big\{\langle x + x^*, e \rangle + 1, \langle x^* - x, e \rangle - 1\big\}, \tag{32}$$

$$C_{A,2k,(-e,e)}(x,x^*) = 4(k-1) + \max\big\{\langle x - x^*, e \rangle + 1, -\langle x + x^*, e \rangle - 1\big\}, \tag{33}$$

$$C_{A,2k,(e,-e)}(x,x^*) = 4(k-1) + \max\big\{\langle x + x^*, e \rangle - 1, \langle x^* - x, e \rangle + 1\big\}, \tag{34}$$

$$C_{A,\infty,(-e,e)}(x,x^*) = +\infty, \tag{35}$$

$$C_{A,\infty,(e,-e)}(x,x^*) = +\infty. \tag{36}$$

We now turn to Fitzpatrick functions.

DEFINITION 2.8 (Fitzpatrick functions [2, Definition 2.2]). *Let $A\colon X \rightrightarrows X^*$. For every $n \in \{2,3,\dots\}$, the* Fitzpatrick function of A of order n *is*

$$F_{A,n} := \sup_{(a,a^*)\in \operatorname{gra} A} C_{A,n,(a,a^*)}. \tag{37}$$

The Fitzpatrick function of A of infinite order *is*

$$F_{A,\infty} := \sup_{n\in\{2,3,\dots\}} F_{A,n} = \sup_{(a,a^*)\in \operatorname{gra} A} C_{A,\infty,(a,a^*)}. \tag{38}$$

It is clear that each $F_{A,n}$ is convex and lower semicontinuous; moreover, if $\operatorname{gra} A$ is finite, then each $F_{A,n}$ is polyhedral and continuous. The sequence $(F_{A,n})_{n\in\{2,3,\dots\}}$ is increasing and pointwise convergent to $F_{A,\infty}$, which is convex and lower semicontinuous. An immediate consequence of Definition 2.8 is the following result.

PROPOSITION 2.9 ([2, Proposition 2.3]). *Let $A\colon X \rightrightarrows X^*$, and let $n \in \{2,3,\dots\}$. Then $F_{A,n}\colon X \times X^* \to [-\infty,+\infty]$ is convex and lower semicontinuous. At $(x,x^*) \in X \times X^*$, the value of $F_{A,n}$ is given by*

$$\sup_{\substack{(a_1,a_1^*)\in \operatorname{gra} A,\\ \vdots \\ (a_{n-1},a^*_{n-1})\in \operatorname{gra} A}} \left(\sum_{i=1}^{n-2} \langle a_{i+1} - a_i, a_i^* \rangle \right) + \langle x - a_{n-1}, a^*_{n-1} \rangle + \langle a_1, x^* \rangle. \tag{39}$$

Moreover,

$$F_{A,n} \geq \langle \cdot,\cdot \rangle \ \textit{on } \operatorname{gra} A. \tag{40}$$

PROPOSITION 2.10. *Let* $A\colon X \rightrightarrows X^*$, *let* $n \in \{2,3,\dots\}$, *and let* $(x,x^*) \in X \times X^*$. *Then* $F_{A^{-1},n}(x^*,x) = F_{A,n}(x,x^*)$ *and* $F_{A^{-1},\infty}(x^*,x) = F_{A,\infty}(x,x^*)$.

Proof. Take $(b_1^*,b_1),\dots,(b_{n-1}^*,b_{n-1})$ in $\operatorname{gra} A^{-1}$ and set

$$(41)\qquad \big(\forall i \in \{1,\dots,n-1\}\big) \quad (a_i,a_i^*) := (b_{n-i},b_{n-i}^*) \in \operatorname{gra} A.$$

Then

$$(42)\qquad \begin{aligned}&\sum_{i=1}^{n-2}\langle b_i,b_{i+1}^*-b_i^*\rangle + \langle b_{n-1},x^*-b_{n-1}^*\rangle + \langle x,b_1^*\rangle\\ &= \sum_{i=1}^{n-2}\langle b_i,b_{i+1}^*\rangle - \sum_{i=1}^{n-1}\langle b_i,b_i^*\rangle + \langle b_{n-1},x^*\rangle + \langle x,b_1^*\rangle\\ &= \sum_{i=1}^{n-2}\langle a_{i+1},a_i^*\rangle - \sum_{i=1}^{n-1}\langle a_i,a_i^*\rangle + \langle a_1,x^*\rangle + \langle x,a_{n-1}^*\rangle\\ &= \sum_{i=1}^{n-2}\langle a_{i+1}-a_i,a_i^*\rangle + \langle x-a_{n-1},a_{n-1}^*\rangle + \langle a_1,x^*\rangle.\end{aligned}$$

The result follows by supremizing. ∎

FACT 2.11 ([2, Proposition 2.4 and Corollary 2.5]). *Let* $A\colon X \rightrightarrows X^*$, *and let* $n \in \{2,3,\dots\}$. *Then*

$$(43)\qquad A \textit{ is } n\textit{-cyclically monotone} \Leftrightarrow F_{A,n} \le \langle\cdot,\cdot\rangle \textit{ on } \operatorname{gra} A \Leftrightarrow F_{A,n} = \langle\cdot,\cdot\rangle \textit{ on } \operatorname{gra} A,$$

and

$$(44)\qquad A \textit{ is cyclically monotone} \Leftrightarrow F_{A,\infty} \le \langle\cdot,\cdot\rangle \textit{ on } \operatorname{gra} A \Leftrightarrow F_{A,\infty} = \langle\cdot,\cdot\rangle \textit{ on } \operatorname{gra} A.$$

COROLLARY 2.12. *Let* $A\colon X \rightrightarrows X^*$, *and let* $n \in \{2,3,\dots\}$. *Then* A *is* n*-cyclically monotone if and only if* A^{-1} *is.*

The recursion formula for Fitzpatrick functions that we present next is an immediate consequence of Proposition 2.3. (A special case of it was utilized in [3].)

PROPOSITION 2.13 (recursion). *Let* $A\colon X \rightrightarrows X^*$, *let* $n \in \{2,3,\dots\}$, *and let* $(x,x^*) \in X \times X^*$. *Then*

$$(45)\qquad F_{A,n+1}(x,x^*) = \sup_{(a,a^*)\in\operatorname{gra} A} F_{A,n}(a,x^*) + \langle x,a^*\rangle - \langle a,a^*\rangle.$$

Combining Fact 2.11 and Proposition 2.13, we obtain the following result which underlines the importance of the values of the Fitzpatrick function on $\operatorname{dom} A \times \operatorname{ran} A$.

COROLLARY 2.14. *Let* $A\colon X \rightrightarrows X^*$, *and let* $n \in \{3,4,\dots\}$. *Then* A *is* n*-cyclically monotone if and only if*

$$(46)\qquad \big(\forall (a,a^*) \in \operatorname{gra} A\big)\big(\forall (b,b^*) \in \operatorname{gra} A\big) \quad F_{A,n-1}(a,b^*) - \langle a,b^*\rangle \le \langle a-b,a^*-b^*\rangle.$$

Example 2.15. Let $A\colon X \rightrightarrows X^*$ be monotone such that its graph contains two points, and let $n \in \{2,3,\dots\}$. Then $F_{A,n} = \langle\cdot,\cdot\rangle$ on $\operatorname{dom} A \times \operatorname{ran} A$; consequently, A is cyclically monotone.

Proof. The fact that $F_{A,n} = \langle\cdot,\cdot\rangle$ is proved readily by induction. The cyclic monotonicity of A now follows from Corollary 2.14 and from the monotonicity of A. (Alternatively, use Corollary 2.18 below.) ∎

THEOREM 2.16. *Let $A\colon X \rightrightarrows X^*$ be such that* $\operatorname{gra} A$ *contains at most n points, where $n \in \{2, 3, \dots\}$. Suppose that A is n-cyclically monotone. Then A is $(n+1)$-cyclically monotone and*

$$F_{A,n+1} = F_{A,n+2} = \cdots = F_{A,\infty}. \tag{47}$$

Proof. Take

$$\big\{(b_1, b_1^*), \dots, (b_{n+1}, b_{n+1}^*)\big\} \subseteq \operatorname{gra} A. \tag{48}$$

We must show that

$$\sigma := \sum_{i=1}^{n+1} \langle b_{i+1} - b_i, b_i^* \rangle \le 0, \quad \text{where } b_{n+2} := b_1. \tag{49}$$

Since $\operatorname{gra} A$ contains no more than n points, there exist integers k and l such that

$$b_k = b_l \quad \text{and} \quad 1 \le k < l \le n+1. \tag{50}$$

Then

$$\sigma = \sigma_1 + \sigma_2, \tag{51}$$

where

$$\sigma_1 := \sum_{i=k}^{l-1} \langle b_{i+1} - b_i, b_i^* \rangle \quad \text{and} \quad \sigma_2 := \sum_{i=l}^{n+1} \langle b_{i+1} - b_i, b_i^* \rangle + \sum_{i=1}^{k-1} \langle b_{i+1} - b_i, b_i^* \rangle \tag{52}$$

are two cyclic sums, each of which contains at least one term and hence at most n terms. Since A is n-cyclically monotone, we see that $\sigma_1 \le 0$ and that $\sigma_2 \le 0$. Therefore, $\sigma = \sigma_1 + \sigma_2 \le 0$. The statement concerning the Fitzpatrick functions follows from Theorem 2.4 and Definition 2.8. ∎

Example 2.17. Let $A\colon X \rightrightarrows X^*$ be such that $\operatorname{gra} A = \{(a, a^*)\}$ for some $(a, a^*) \in X \times X^*$. Then A is cyclically monotone, and for every $(x, x^*) \in X \times X^*$ we have

$$F_{A,2}(x, x^*) = F_{A,3}(x, x^*) = \cdots = F_{A,\infty}(x, x^*) = \langle a, x^* \rangle + \langle x, a^* \rangle - \langle a, a^* \rangle. \tag{53}$$

COROLLARY 2.18. *Let $A\colon X \rightrightarrows X^*$ be such that* $\operatorname{gra} A$ *contains at most n points, where $n \in \{2, 3, \dots\}$. Suppose that A is n-cyclically monotone. Then A is cyclically monotone.*

Proof. By Theorem 2.16, A is $(n+1)$-cyclically monotone and

$$F_{A,n+1} = F_{A,\infty}. \tag{54}$$

On the other hand, Fact 2.11 yields

$$F_{A,n+1} = \langle \cdot, \cdot \rangle \text{ on } \operatorname{gra} A. \tag{55}$$

The result follows by combining (54), (55), and Fact 2.11. ∎

COROLLARY 2.19. *Let $A\colon X \rightrightarrows X^*$ be such that* $\operatorname{gra} A$ *contains at most n points, where $n \in \{2, 3, \dots\}$. Then A is cyclically monotone if and only if*

$$\big(\forall (a, a^*) \in \operatorname{gra} A\big) \quad F_{A,n}(a, a^*) = \langle a, a^* \rangle. \tag{56}$$

Proof. "$\Rightarrow$": On $\operatorname{gra} A$, we always have $\langle\cdot,\cdot\rangle \le F_{A,n} \le F_{A,\infty}$. Since A is cyclically monotone, $F_{A,\infty} = \langle\cdot,\cdot\rangle$ on $\operatorname{gra} A$ and hence $F_{A,n} = \langle\cdot,\cdot\rangle$ on $\operatorname{gra} A$. "$\Leftarrow$": By Fact 2.11, A is n-cyclically monotone. The result now follows from Corollary 2.18. ∎

Example 2.20. Suppose that X is a Hilbert space. Let $e \in X$ such that $\|e\| = 1$, and define A via $\operatorname{gra} A = \{(-e,-e),(e,e)\}$. Then A is cyclically monotone but $F_{A,2} \neq F_{A,3}$; in fact, for every $(x,x^*) \in X \times X$, we have

$$F_{A,2}(x,x^*) = \max\big\{-1 \pm \langle x+x^*, e\rangle\big\} \tag{57}$$

and

$$F_{A,3}(x,x^*) = \cdots = F_{A,\infty}(x,x^*) = \max\big\{-1 \pm \langle x+x^*, e\rangle, -3 \pm \langle x-x^*, e\rangle\big\}. \tag{58}$$

Proof. The operator A is cyclically monotone since $\operatorname{gra} A \subset \operatorname{gra}\operatorname{Id} = \operatorname{gra}\partial\frac{1}{2}\|\cdot\|^2$. The formulas for $F_{A,2}$ and $F_{A,3}$ follow from Example 2.5. Theorem 2.16 shows that $F_{A,3} = \cdots = F_{A,\infty}$. Finally, we note that $F_{A,2}(2e,-2e) = -1$, whereas $F_{A,3}(2e,-2e) = 1$. ∎

The next example, which is an immediate consequence of Example 2.7 and Definition 2.8, illustrates the nonmonotone case.

Example 2.21. Suppose that X is a Hilbert space. Let $e \in X$ be such that $\|e\| = 1$, and define A via $\operatorname{gra} A := \{(-e,e),(e,-e)\}$. Then A is not monotone, and for every $k \in \{2,3,\dots\}$ and $(x,x^*) \in X \times X$ we have

$$F_{A,2}(x,x^*) = 1 + \max\big\{\pm\langle x-x^*, e\rangle\big\}, \tag{59}$$

$$F_{A,2k-1}(x,x^*) = 4(k-1) - 2 + \max\big\{-1 \pm \langle x-x^*, e\rangle, 1 \pm \langle x+x^*, e\rangle\big\}, \tag{60}$$

$$F_{A,2k}(x,x^*) = 4(k-1) + \max\big\{1 \pm \langle x-x^*, e\rangle, -1 \pm \langle x+x^*, e\rangle\big\}, \tag{61}$$

$$F_{A,\infty}(x,x^*) = +\infty. \tag{62}$$

3. Rockafellar functions.

DEFINITION 3.1. *Let $A\colon X \rightrightarrows X^*$, and let $f \in \Gamma$. Then f is an* antiderivative *of A if*

$$\operatorname{gra} A \subseteq \operatorname{gra}\partial f. \tag{63}$$

The following result will turn out to be useful.

PROPOSITION 3.2. *Suppose that X is reflexive. Let $A\colon X \rightrightarrows X^*$, let $f \in \Gamma$, and suppose that f is an antiderivative of A such that $\overline{\operatorname{ran}\partial f} \subseteq \overline{\operatorname{conv}}\operatorname{ran} A$. Then $\overline{\operatorname{dom} f^*} = \overline{\operatorname{conv}}\operatorname{ran} A$.*

Proof. On the one hand, since f is an antiderivative of A, we deduce that $\operatorname{gra} A \subseteq \operatorname{gra}\partial f \Leftrightarrow \operatorname{gra} A^{-1} \subseteq \operatorname{gra}(\partial f)^{-1} = \operatorname{gra}\partial f^* \Rightarrow \operatorname{ran} A = \operatorname{dom} A^{-1} \subseteq \operatorname{dom}\partial f^* \subseteq \operatorname{dom} f^* \Rightarrow \overline{\operatorname{conv}}\operatorname{ran} A \subseteq \overline{\operatorname{conv}}\operatorname{dom} f^* = \overline{\operatorname{dom} f^*}$. Because $\overline{\operatorname{ran}\partial f} \subseteq \overline{\operatorname{conv}}\operatorname{ran} A$, we see that $\overline{\operatorname{dom} f^*} \subseteq \overline{\operatorname{dom} f^*} = \overline{\operatorname{dom}\partial f^*} = \overline{\operatorname{dom}(\partial f)^{-1}} = \overline{\operatorname{ran}\partial f} \subseteq \overline{\operatorname{conv}}\operatorname{ran} A$. Altogether, $\overline{\operatorname{dom} f^*} = \overline{\operatorname{conv}}\operatorname{ran} A$. ∎

DEFINITION 3.3 (Rockafellar function). *Let $A\colon X \rightrightarrows X^*$. Then the* Rockafellar functions *are defined by*

$$\big(\forall (a,a^*) \in \operatorname{gra} A\big) \quad R_{A,(a,a^*)}\colon X \to \left]-\infty,+\infty\right] : x \mapsto \sup_{n\in\{2,3,\dots\}} C_{A,n,(a,a^*)}(x,0). \tag{64}$$

The importance of the Rockafellar functions stems from a fundamental result due to Rockafellar (see [19] or [24, Proposition 2.4.3, Theorem 3.2.8, and Corollary 3.2.11]), which states that maximal cyclically monotone operators are precisely the subdifferential operators of convex, lower semicontinuous, and proper functions. The following part of Rockafellar's result will be utilized later.

FACT 3.4 (Rockafellar [19] or [24, Proposition 2.4.3 and Corollary 3.2.11]). *Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, and let $(a, a^*) \in \operatorname{gra} A$. Then the following hold:*

(i) *$R_{A,(a,a^*)}$ is convex, lower semicontinuous, and proper, $R_{A,(a,a^*)}(a) = 0$, and $R_{A,(a,a^*)}$ is an antiderivative of A.*

(ii) *If A is maximal cyclically monotone, then any two antiderivatives of A differ only by a constant.*

Among all antiderivatives, Rockafellar functions have a special status due to the following minimality property, which was first observed in [14, Theorem 3.4] for cyclically monotone operators with a finite graph.

THEOREM 3.5. *Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, and let $a \in \operatorname{dom} A$. Then*

$$\big(\forall a^* \in Aa\big) \quad R_{A,(a,a^*)} = \min\big\{f \in \Gamma(X) \mid f \textit{ is an antiderivative of } A \textit{ with } f(a) \geq 0\big\} \tag{65}$$

$$= \min\big\{f \in \Gamma(X) \mid f \textit{ is an antiderivative of } A \textit{ with } f(a) = 0\big\}. \tag{66}$$

Proof. Suppose that $f \in \Gamma$ is an antiderivative of A with $f(a) \geq 0$, and take $a^* \in Aa$ and $x \in X$. Then, for every $x \in X$, $n \in \{1, 2, \dots\}$, and $(a_1, a_1^*), \dots, (a_n, a_n^*)$ belonging to $\operatorname{gra} A$, we have

$$f(x) \geq f(x) - f(a_n) + \left(\sum_{i=1}^{n-1} f(a_{i+1}) - f(a_i)\right) + f(a_1) - f(a)$$

$$\geq \langle x - a_n, a_n^*\rangle + \left(\sum_{i=1}^{n-1} \langle a_{i+1} - a_i, a_i^*\rangle\right) + \langle a_1 - a, a^*\rangle. \tag{67}$$

This implies

$$f \geq R_{A,(a,a^*)}. \tag{68}$$

In view of Fact 3.4(i), the proof is complete. ◻

COROLLARY 3.6. *Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, let $a \in \operatorname{dom} A$, let $a_1^* \in Aa$, and let $a_2^* \in Aa$. Then $R_{A,(a,a_1^*)} = R_{A,(a,a_2^*)}$.*

Corollary 3.6 and Theorem 2.6 make the following definition well-defined.

DEFINITION 3.7. *Let $A\colon X \rightrightarrows X^*$, and let $a \in \operatorname{dom} A$. Then we set*

$$R_{A,a} := R_{A,(a,a^*)}, \tag{69}$$

where a^* is an arbitrary point in Aa.

COROLLARY 3.8. *Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, and let $a \in \operatorname{dom} A$. Set $B\colon X \rightrightarrows X^*\colon x \mapsto \operatorname{conv}(Ax)$. Then B is cyclically monotone, and $R_{B,a} = R_{A,a}$.*

Proof. It is readily verified that B is cyclically monotone. Hence $R_{B,a}$ is an antiderivative of B and of A such that $R_{B,a}(a) = 0$. By Theorem 3.5, $R_{B,a} \geq R_{A,a}$. On the other hand, $R_{A,a}$ is also an antiderivative of B; thus, again by Theorem 3.5, $R_{A,a} \geq R_{B,a}$. Altogether, $R_{B,a} = R_{A,a}$. ◻

COROLLARY 3.9. *Suppose that X is reflexive. Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, and let $(a, a^*) \in \operatorname{gra} A$. Then*

$$R_{A,a}^* = \max\big\{g \in \Gamma(X^*) \mid g \textit{ is an antiderivative of } A^{-1} \textit{ and } g(a^*) = \langle a, a^*\rangle\big\}. \tag{70}$$

Proof. Take $g \in \Gamma(X^*)$ such that g is an antiderivative of A^{-1} and $g(a^*) = \langle a, a^* \rangle$. Then $g^*(a) = 0$, and g^* is an antiderivative of A. By Theorem 3.5, $g^* \geq R_{A,a}$, and therefore $g^{**} = g \leq R^*_{A,a}$. ☐

Corollary 3.9 results in the following interesting counterpart to Theorem 3.5; see also [14, Proposition 4.2].

COROLLARY 3.10. *Suppose that X is reflexive. Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, and let $(a, a^*) \in \operatorname{gra} A$. Then*

$$R^*_{A^{-1},a^*} - \langle a, a^* \rangle = \max\big\{ f \in \Gamma(X) \mid f \text{ is an antiderivative of } A \text{ and } f(a) = 0 \big\}. \tag{71}$$

The next result will be used later.

COROLLARY 3.11. *Let $A\colon X \rightrightarrows X^*$, let $(a, a^*) \in \operatorname{gra} A$, and let $n \in \{2, 3, \dots\}$. Suppose that $\operatorname{gra} A$ contains at most n points and that A is n-cyclically monotone. Then A is cyclically monotone, and for every $x \in X$ we have*

$$R_{A,a}(x) = C_{A,n+1,(a,a^*)}(x, 0) \tag{72}$$
$$= \max_{\substack{(a_2,a_2^*)\in \operatorname{gra} A,\\ \vdots \\ (a_n,a_n^*)\in \operatorname{gra} A}} \langle x - a_n, a_n^* \rangle + \langle a_n - a_{n-1}, a^*_{n-1} \rangle + \cdots + \langle a_2 - a, a^* \rangle. \tag{73}$$

Consequently, $R_{A,a}$ is a polyhedral and continuous antiderivative of A with $\operatorname{ran} \partial R_{A,a} \subset \operatorname{conv} \operatorname{ran} A$.

Proof. This follows from Corollary 2.18, Theorem 2.4, (8), Fact 3.4(i), and the Ioffe–Tikhomirov theorem (see, e.g., [24, Theorem 2.4.18]). ☐

Fact 3.4(ii) implies that, if A is maximal cyclically monotone, the Rockafellar functions $\{R_{A,a}\}_{a \in \operatorname{dom} A}$ differ only by constants. For finite-graph operators, this is no longer true as the following consequence of Example 2.5 and Definition 3.3 shows.

Example 3.12. Suppose that X is a Hilbert space. Let $e \in X$ be such that $\|e\| = 1$, and define A via $\operatorname{gra} A := \{(-e, -e), (e, e)\}$. Then for every $x \in X$ we have

$$R_{A,-e}(x) = \max\big\{ -\langle x, e \rangle - 1, \langle x, e \rangle - 3 \big\} = -2 + |\langle x, e \rangle - 1| \tag{74}$$

and

$$R_{A,e}(x) = \max\big\{ \langle x, e \rangle - 1, -\langle x, e \rangle - 3 \big\} = -2 + |\langle x, e \rangle + 1|. \tag{75}$$

Consequently, $R_{A,e} \not\geq R_{A,-e}$ and $R_{A,e} \not\leq R_{A,-e}$.

Remark 3.13. Let $A\colon X \rightrightarrows X^*$, and suppose that A is not cyclically monotone. Then Theorem 2.6 and Definition 3.3 imply that the Rockafellar functions $\{R_{A,a}\}_{a \in \operatorname{dom} A}$ are all identically equal to $+\infty$. For a concrete example, see Example 2.7.

Turning momentarily to the case when $X = \mathbb{R}$, we now present not only a considerable generalization of Example 3.12 but also an explicit formula for any Rockafellar function and its subdifferential operator of a cyclically monotone operator with a finite graph. See also [14, section 7].

THEOREM 3.14. *Let $A\colon \mathbb{R} \rightrightarrows \mathbb{R}$ have a finite graph, and suppose that the graph of $B\colon \mathbb{R} \rightrightarrows \mathbb{R}\colon x \mapsto \operatorname{conv}(Ax)$ is*

$$\bigcup_{i=1}^{n} \big(\{a_i\} \times [b_i^-, b_i^+] \big), \tag{76}$$

where $n \in \{1, 2, \ldots\}$*,* $a_1 < a_2 < \cdots < a_n$*, and* $b_1^- \le b_1^+ \le b_2^- \le \cdots \le b_n^- \le b_n^+$*. Set* $a_0 := -\infty$ *and* $a_{n+1} := +\infty$*. Suppose that* $k \in \{1, \ldots, n\}$*. Then* R_{A,a_k} *is given by*

$$\mathbb{R} \to \mathbb{R}\colon x \mapsto \begin{cases} (x - a_i)b_i^- + \displaystyle\sum_{j=i+1}^{k} (a_{j-1} - a_j)b_j^- & \text{if } a_{i-1} < x \le a_i \le a_k, \\ (x - a_i)b_i^+ + \displaystyle\sum_{j=k}^{i-1} (a_{j+1} - a_j)b_j^+ & \text{if } a_k \le a_i \le x < a_{i+1}, \end{cases} \tag{77}$$

and $\partial R_{A,a_k}$ *is given by*

$$\mathbb{R} \rightrightarrows \mathbb{R}\colon x \mapsto \begin{cases} \{b_i^-\} & \text{if } a_{i-1} < x < a_i \le a_k, \\ [b_i^-, b_{i+1}^-] & \text{if } x = a_i < a_k, \\ [b_k^-, b_k^+], & \text{if } x = a_k, \\ [b_{i-1}^+, b_i^+] & \text{if } a_k < x = a_i, \\ \{b_i^+\} & \text{if } a_k \le a_i < x < a_{i+1}. \end{cases} \tag{78}$$

Proof. Clearly, A and B are cyclically monotone. Denote the function described in (77) by R, and observe that R is piecewise linear, continuous everywhere, and well-defined at a_k, with

$$R(a_k) = 0. \tag{79}$$

Moreover, (77) implies that ∂R is given by (78), which is clearly monotone. Thus

$$R \in \Gamma(\mathbb{R}). \tag{80}$$

Take $i \in \{1, 2, \ldots, n\}$. If $i < k$, then $Aa_i \subseteq Ba_i = [b_i^-, b_i^+] \subseteq [b_i^-, b_{i+1}^-] = \partial R(a_i)$. If $i = k$, then $Aa_i = Aa_k \subseteq Ba_k = [b_k^-, b_k^+] = \partial R(a_k)$. If $k < i$, then $Aa_i \subseteq Ba_i = [b_i^-, b_i^+] \subseteq [b_{i-1}^+, b_i^+] = \partial R(a_i)$. Thus,

$$R \text{ is an antiderivative of } A. \tag{81}$$

Since $\bigcup_{i=1}^n \{(a_i, b_i^-), (a_i, b_i^+)\} \subseteq \operatorname{gra} A$, we deduce from (73) and (77) that

$$R_{A,a_k} \ge R. \tag{82}$$

Hence (79), (80), (81), (82), and Theorem 3.5 imply that $R = R_{A,a_k}$. ∎

The next result links Rockafellar functions to Fitzpatrick functions.

THEOREM 3.15. *Let* $A\colon X \rightrightarrows X^*$*. Then*

$$\big(\forall (x, x^*) \in X \times X^*\big) \quad F_{A,\infty}(x, x^*) = \sup_{a \in \operatorname{dom} A} \langle a, x^* \rangle + R_{A,a}(x). \tag{83}$$

Proof. (See also the proof of [2, Theorem 3.5] for a variant.) Take $(x, x^*) \in X \times X^*$. Using Definitions 2.8, 2.1, 3.3, and 3.7, we see that

$$F_{A,\infty}(x, x^*) = \sup_{n \in \{2,3,\ldots\}} F_{A,n}(x, x^*) = \sup_{n \in \{2,3,\ldots\}} \sup_{(a,a^*) \in \operatorname{gra} A} C_{A,n,(a,a^*)}(x, x^*) \tag{84}$$

$$= \sup_{(a,a^*) \in \operatorname{gra} A} \sup_{n \in \{2,3,\ldots\}} C_{A,n,(a,a^*)}(x, 0) + \langle a, x^* \rangle \tag{85}$$

$$= \sup_{(a,a^*) \in \operatorname{gra} A} \langle a, x^* \rangle + R_{A,(a,a^*)}(x) \tag{86}$$

$$= \sup_{a \in \operatorname{dom} A} \langle a, x^* \rangle + R_{A,a}(x), \tag{87}$$

as required. ∎

We deduce that the Fitzpatrick function of infinite order with the second variable set to zero is exactly the supremum of all Rockafellar functions.

COROLLARY 3.16. *Let* $A\colon X \rightrightarrows X^*$*. Then*

$$F_{A,\infty}(\cdot,0) = \sup_{a\in\operatorname{dom} A} R_{A,a}. \tag{88}$$

Remark 3.17. Let $A\colon X \rightrightarrows X^*$ be maximal cyclically monotone. Rockafellar [19] (see Fact 3.4) proved that $A = \partial f$, where $f \in \Gamma$ is uniquely determined up to additive constants. By [2, Theorem 3.5], $F_{A,\infty} = f \oplus f^*$. Thus Corollary 3.16 implies that

$$\sup_{a\in\operatorname{dom} A} R_{A,a} \equiv +\infty \quad\Leftrightarrow\quad 0 \notin \operatorname{dom} f^* \quad\Leftrightarrow\quad \inf f(X) = -\infty. \tag{89}$$

COROLLARY 3.18. *Let* $A\colon X \rightrightarrows X^*$ *be cyclically monotone with a finite graph, let* $x^* \in X^*$*, and set* $f := F_{A,\infty}(\,\cdot\,, x^*)$*. Then* f *is polyhedral, continuous, with a full domain,* $\operatorname{gra} A \subset \operatorname{gra}\partial f$*, and* $\overline{\operatorname{ran}\partial f} \subseteq \operatorname{conv}\operatorname{ran} A$.

Proof. Since $\operatorname{gra} A$ is finite, Theorem 3.15 yields

$$f = \max_{a\in\operatorname{dom} A} \langle a, x^*\rangle + R_{A,a}. \tag{90}$$

The function f is continuous, polyhedral, with a full domain, as it is the finite maximum of such functions (see Corollary 3.11). Fix $x \in X$, and set $D_x := \big\{a \in \operatorname{dom} A \mid f(x) = \langle a, x^*\rangle + R_{A,a}(x)\big\}$. On the one hand, using the Ioffe–Tikhomirov theorem (see, e.g., [24, Theorem 2.4.18]) and Corollary 3.11, we have

$$\partial f(x) = \overline{\operatorname{conv}}^{\,*} \bigcup_{a\in D_x} \partial R_{A,a}(x) \subseteq \overline{\operatorname{conv}}^{\,*} \bigcup_{a\in D_x} \operatorname{conv}\operatorname{ran} A = \operatorname{conv}\operatorname{ran} A, \tag{91}$$

where $\overline{\operatorname{conv}}^{\,*}$ denotes the weak* closed convex hull operator. Hence $\operatorname{ran}\partial f \subseteq \operatorname{conv}\operatorname{ran} A$, and thus $\overline{\operatorname{ran}\partial f} \subseteq \operatorname{conv}\operatorname{ran} A$, since $\operatorname{conv}\operatorname{ran} A$ is compact as a convex hull of finitely many points. On the other hand, Fact 3.4(i) implies that

$$\big(\forall a \in \operatorname{dom} A\big) \quad \operatorname{gra} A \subset \operatorname{gra}\partial R_{A,a}. \tag{92}$$

Combining (91) and (92), we conclude altogether that $\operatorname{gra} A \subset \operatorname{gra}\partial f$. ∎

Example 3.19. Suppose that X is a Hilbert space. Let $e \in X$ such that $\|e\| = 1$, and define A via $\operatorname{gra} A = \{(-e,-e),(e,e)\}$. Then A is cyclically monotone, and for every $x \in X$ we have

$$F_{A,\infty}(x,0) = \max\big\{R_{A,-e}(x), R_{A,e}(x)\big\} = \max\big\{-1 \pm \langle x,e\rangle\big\} = -1 + |\langle x,e\rangle|. \tag{93}$$

Proof. Combine Example 3.12 and Corollary 3.16. ∎

We conclude this section with a result which illustrates how Fitzpatrick functions give rise to the smallest nonnegative antiderivative.

COROLLARY 3.20. *Let* $A : X \rightrightarrows X^*$ *be cyclically monotone with a finite graph. Then*

$$F_{A,\infty}(\cdot,0) = \min\big\{f \in \Gamma(X) \mid f \text{ is an antiderivative of } A \text{ such that } f \geq 0 \text{ on } \operatorname{dom} A\big\}. \tag{94}$$

Proof. Take $f \in \Gamma(X)$ such that f is an antiderivative of A and $f \geq 0$ on $\operatorname{dom} A$. Then Theorem 3.5 implies that $(\forall a \in \operatorname{dom} A)$ $f \geq R_{A,a}$; hence, by Corollary 3.16,

$$f \geq \max_{a\in\operatorname{dom} A} R_{A,a} = F_{A,\infty}(\cdot,0). \tag{95}$$

In view of Corollary 3.18 and Fact 3.4(i), $F_{A,\infty}(\cdot,0)$ is an antiderivative of A that is nonnegative on $\operatorname{dom} A$. ∎

4. Intrinsic and primal-dual symmetric methods. From now on,

(96)
$\mathcal{A}$ is the set of all cyclically monotone operators on X with finite nonempty graphs.

DEFINITION 4.1. *An* intrinsic method for finding antiderivatives*—or simply an* intrinsic method*—is a mapping* $m\colon \mathcal{A} \to \Gamma\colon A \mapsto m_A$ *such that, for every* $A \in \mathcal{A}$*,* m_A *is an antiderivative of* A*.*

Example 4.2. Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, and let $(a, a^*) \in \operatorname{gra} A$. Then the Rockafellar function $R_{A,(a,a^*)} = R_{A,a}$ is an antiderivative (see Fact 3.4) but—due to the dependency on a and the resulting nonuniqueness of Rockafellar functions—there is no corresponding intrinsic method m that produces Rockafellar functions. See Example 3.12 for a concrete example.

Remark 4.3. Given $A \in \mathcal{A}$, an intrinsic method m provides an antiderivatives m_A as a mapping depending only on A or, equivalently, only on the (unordered) graph of A. This key property of intrinsic methods explains why the process of providing Rockafellar functions considered in Example 4.2 is not intrinsic. Similarly, if a method computes antiderivatives by using an enumeration of the graph of A and if a different enumeration may result in a different antiderivative, then such a method cannot be intrinsic.

We now provide two intrinsic methods.

Example 4.4. Let $m\colon \mathcal{A} \to \Gamma\colon A \mapsto F_{A,\infty}(\cdot, 0) = \max_{(a,a^*)\in \operatorname{gra} A} R_{A,(a,a^*)}$. Corollaries 3.16 and 3.18 imply that m is an intrinsic method. Moreover, for every $A \in \mathcal{A}$, the antiderivative m_A has a full domain and $\overline{\operatorname{ran} \partial m_A} \subseteq \operatorname{conv} \operatorname{ran} A$.

Example 4.5. Let $A\colon X \rightrightarrows X^*$ be cyclically monotone such that $\operatorname{gra} A$ contains exactly n points, where $n \in \{1, 2, \ldots\}$, and set

$$m_A := \sum_{(a,a^*)\in \operatorname{gra} A} \frac{1}{n} R_{A,(a,a^*)}. \tag{97}$$

Then m_A is an antiderivative of A that is polyhedral and continuous with a full domain and $\overline{\operatorname{ran} \partial m_A} \subseteq \operatorname{conv} \operatorname{ran} A$. Furthermore, the corresponding method $m\colon \mathcal{A} \to \Gamma\colon A \mapsto m_A$ is intrinsic.

Proof. Note that m_A is continuous and polyhedral with a full domain, as a finite sum of such functions. The sum rule (see, e.g., [24, Theorem 2.8.7(iii)]) and Fact 3.4 imply that, for every $(x, x^*) \in \operatorname{gra} A$, we have $x^* \in Ax \subseteq \sum_{(a,a^*)\in \operatorname{gra} A} \frac{1}{n} Ax \subseteq \sum_{(a,a^*)\in \operatorname{gra} A} \frac{1}{n} \partial R_{A,(a,a^*)}(x) = \partial m_A(x)$. Corollary 3.11 shows that, for every $x \in X$, we have $\partial m_A(x) = \sum_{(a,a^*)\in \operatorname{gra} A} \frac{1}{n} \partial R_{A,(a,a^*)}(x) \subseteq \sum_{(a,a^*)\in \operatorname{gra} A} \frac{1}{n} \operatorname{conv} \operatorname{ran} A = \operatorname{conv} \operatorname{ran} A$. Consequently, $\operatorname{ran} \partial m_A \subseteq \operatorname{conv} \operatorname{ran} A$, and hence $\overline{\operatorname{ran} \partial m_A} \subseteq \overline{\operatorname{conv} \operatorname{ran} A} = \operatorname{conv} \operatorname{ran} A$. It is clear that m is intrinsic. ∎

We assume from now on that

$$X \text{ is a Hilbert space.} \tag{98}$$

DEFINITION 4.6. *An intrinsic method* $\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto \mathfrak{m}_A$ *is* primal-dual symmetric *if*

$$(\forall A \in \mathcal{A}) \quad \mathfrak{m}_{A^{-1}} = \mathfrak{m}_A^*. \tag{99}$$

PROPOSITION 4.7. *While intrinsic, neither*

$$\mathcal{A} \to \Gamma\colon A \mapsto F_{A,\infty}(\cdot, 0) = \max_{(a,a^*)\in \operatorname{gra} A} R_{A,(a,a^*)} \tag{100}$$

nor

(101)
$$\mathcal{A} \to \Gamma \colon A \mapsto \sum_{(a,a^*)\in \operatorname{gra} A} \frac{1}{n_A} R_{A,(a,a^*)}, \quad \textit{where } n_A \textit{ is the number of points in } \operatorname{gra} A,$$

is primal-dual symmetric.

Proof. On the one hand, both methods produce polyhedral continuous functions with a *full domain.* On the other hand, the Fenchel conjugates of such functions have a *bounded domain.* ∎

Since antiderivatives are only (and at best; see Example 3.12) unique up to a constant, it is perhaps surprising that primal-dual symmetric methods even exist. The remainder of this section is devoted to the derivation of such methods. We shall require several known notions, which we review now.

Let $A\colon X \rightrightarrows X$ be a monotone operator. The *resolvent* of A is (the single-valued, firmly nonexpansive operator) $J_A := (\operatorname{Id} + A)^{-1}$, where Id denotes the identity operator. A classical result due to Minty [16] asserts that J_A has a full domain if and only if A is maximal monotone. The proof of the following result is straightforward and hence omitted.

PROPOSITION 4.8. *Let $A\colon X \rightrightarrows X$ and $B\colon X \rightrightarrows X$ be monotone operators. Then the following are equivalent:*

(i) $\operatorname{gra} A \subseteq \operatorname{gra} B$.
(ii) $\operatorname{gra} A^{-1} \subseteq \operatorname{gra} B^{-1}$.
(iii) *J_B is an extension of J_A; i.e., $J_B = J_A$ on $\operatorname{dom} J_A = \operatorname{ran}(\operatorname{Id} + A)$.*

We further recall that, given $f \in \Gamma$, the *proximal mapping* [17] of f is $\operatorname{Prox}(f) := J_{\partial f}$. It is clear from the definition that, for two points x and x^* in X, one has

$$x^* \in \partial f(x) \quad \Leftrightarrow \quad x = \operatorname{Prox}(f)(x + x^*). \tag{102}$$

PROPOSITION 4.9. *Let $f \in \Gamma$, let $(a, a^*) \in \operatorname{gra} \partial f$, and suppose that $y \in N_{\operatorname{dom} f}(a)$. Then $a = \operatorname{Prox}(f)(2y + a + a^*)$.*

Proof. Since $N_{\operatorname{dom} f}(a)$ is a cone, we have $2y \in N_{\operatorname{dom} f}(a)$. Hence $2y + a + a^* \in a + a^* + \partial\iota_{\operatorname{dom} f}(a) \subseteq a + \partial f(a) + \partial\iota_{\operatorname{dom} f}(a) \subseteq a + \partial(f + \iota_{\operatorname{dom} f})(a) = a + \partial f(a) = (\operatorname{Id} + \partial f)(a)$, and thus $a = \operatorname{Prox}(f)(2y + a + a^*)$. ∎

We need one more notion.

DEFINITION 4.10. *Let $f_0 \in \Gamma$, and let $f_1 \in \Gamma$. The* proximal midpoint average *of f_0 and f_1 is the function*

$$\mathcal{P}(f_0, f_1) := \Big(\tfrac{1}{2}\big(f_0 + \tfrac{1}{2}\|\cdot\|^2\big)^* + \tfrac{1}{2}\big(f_1 + \tfrac{1}{2}\|\cdot\|^2\big)^*\Big)^* - \tfrac{1}{2}\|\cdot\|^2. \tag{103}$$

The proximal average, which is a generalization of the proximal midpoint average with a parameter $\lambda \in [0, 1]$ (the choice $\lambda = \frac{1}{2}$ yields the proximal midpoint average), was introduced in [6] and further studied in [4, 5, 8].

We require the following properties.

FACT 4.11. *Let $f_0 \in \Gamma$, and let $f_1 \in \Gamma$. Then the following hold:*

(i) $\mathcal{P}(f_0, f_1) = \mathcal{P}(f_1, f_0)$.
(ii) $(\mathcal{P}(f_0, f_1))^* = \mathcal{P}(f_0^*, f_1^*)$.
(iii) $\mathcal{P}(f_0, f_1) \in \Gamma$.
(iv) $\operatorname{Prox}(\mathcal{P}(f_0, f_1)) = \frac{1}{2}\operatorname{Prox}(f_0) + \frac{1}{2}\operatorname{Prox}(f_1)$.
(v) *If $f_0 \le f_1$, then $f_0 \le \mathcal{P}(f_0, f_1) \le f_1$.*
(vi) $(\forall \gamma \in \mathbb{R})\ \mathcal{P}(f_0, f_0 + \gamma) = f_0 + \frac{1}{2}\gamma$.

Proof. (i): This is clear from the definition. (ii): See [6, Theorem 6.1]. (iii): This follows from (ii). (iv): See [6, Theorem 6.1]. (v) and (vi) follow readily from the definition. (See also [5, Remark 4.15 and Example 7.1].) □

COROLLARY 4.12. *Let $A\colon X \rightrightarrows X$ be cyclically monotone, and let f_0 and f_1 be antiderivatives of A. Then $\mathcal{P}(f_0, f_1)$ is also an antiderivative of A.*

Proof. By assumption, $\operatorname{gra} A \subseteq \operatorname{gra} \partial f_0$ and $\operatorname{gra} A \subseteq \operatorname{gra} \partial f_1$. Using Proposition 4.8, we see that both $\operatorname{Prox}(f_0)$ and $\operatorname{Prox}(f_1)$ extend J_A. Thus, by Fact 4.11(iv), $\operatorname{Prox}(\mathcal{P}(f_0, f_1))$ also extends J_A. Utilizing Proposition 4.8 once more, we deduce that $\mathcal{P}(f_0, f_1)$ is an antiderivative of A. □

We are now ready for our main result.

THEOREM 4.13 (symmetrization). *Let $m\colon \mathcal{A} \to \Gamma\colon A \mapsto m_A$ be an intrinsic method. Set*

$$\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto \mathcal{P}\big(m_A, m^*_{A^{-1}}\big). \tag{104}$$

Then $\mathfrak{m}$ is a primal-dual symmetric intrinsic method.

Proof. Fix $A \in \mathcal{A}$. Observe that Corollaries 2.12 and Corollary 4.12 imply that $\mathfrak{m}_A$ is an antiderivative of A; thus, $\mathfrak{m}$ is an intrinsic method. On the one hand, the definitions and Fact 4.11(i) yield

$$\mathfrak{m}_{A^{-1}} = \mathcal{P}\big(m_{A^{-1}}, m^*_{(A^{-1})^{-1}}\big) = \mathcal{P}\big(m_{A^{-1}}, m^*_A\big) = \mathcal{P}\big(m^*_A, m_{A^{-1}}\big). \tag{105}$$

On the other hand, Fact 4.11(ii) implies

$$\mathfrak{m}^*_A = \Big(\mathcal{P}\big(m_A, m^*_{A^{-1}}\big)\Big)^* = \mathcal{P}\big(m^*_A, m^{**}_{A^{-1}}\big) = \mathcal{P}\big(m^*_A, m_{A^{-1}}\big). \tag{106}$$

Altogether, we obtain that $\mathfrak{m}_{A^{-1}} = \mathfrak{m}^*_A$. Therefore, $\mathfrak{m}$ is primal-dual symmetric. □

Example 4.14. Let $m\colon \mathcal{A} \to \Gamma\colon A \mapsto m_A$ be intrinsic, and set $\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto \mathcal{P}(m_A, m^*_{A^{-1}})$. Let $A \in \mathcal{A}$ be such that $\operatorname{gra} A \subset \operatorname{gra} \operatorname{Id}$. Then $\mathfrak{m}_A = \frac{1}{2}\|\cdot\|^2$.

Proof. Since $A = A^{-1}$, we have $\mathfrak{m}^*_A = (\mathcal{P}(m_A, m^*_{A^{-1}}))^* = (\mathcal{P}(m_A, m^*_A))^* = \mathcal{P}(m^*_A, m^{**}_A) = \mathcal{P}(m^*_A, m_A) = \mathcal{P}(m_A, m^*_A) = \mathcal{P}(m_A, m^*_{A^{-1}}) = \mathfrak{m}_A$, and the result follows. □

Before we present further applications of Theorem 4.13, let us discuss a nonintrinsic variant based on the original Rockafellar function.

THEOREM 4.15. *Let $A\colon X \rightrightarrows X^*$ be cyclically monotone, and let $(a, a^*) \in \operatorname{gra} A$. Set*

$$f_{A,(a,a^*)} := \mathcal{P}\big(R_{A,(a,a^*)}, R^*_{A^{-1},(a^*,a)}\big). \tag{107}$$

Then $f^*_{A,(a,a^*)} = f_{A^{-1},(a^*,a)} := \mathcal{P}\big(R_{A^{-1},(a^*,a)}, R^*_{A,(a,a^*)}\big)$. Moreover,

$$A \text{ is maximal cyclically monotone} \;\Rightarrow\; f_{A,(a,a^*)} = R_{A,(a,a^*)} + \tfrac{1}{2}\langle a, a^*\rangle. \tag{108}$$

Proof. The proof of $f^*_{A,(a,a^*)} = f_{A^{-1},(a^*,a)}$ is analogous to the one of Theorem 4.13. Now assume that A is maximal cyclically monotone. Since $R_{A,(a,a^*)}$ is an antiderivative of A and $R_{A^{-1},(a^*,a)}$ is an antiderivative of A^{-1}, there exists $\gamma \in \mathbb{R}$ such that

$$R^*_{A^{-1},(a^*,a)} = R_{A,(a,a^*)} + \gamma. \tag{109}$$

Conjugating (109) followed by evaluating at a^* yields $0 = R^*_{A,(a,a^*)}(a^*) - \gamma = \langle a, a^*\rangle - R_{A,(a,a^*)}(a) - \gamma = \langle a, a^*\rangle - \gamma$. Hence

$$R^*_{A^{-1},(a^*,a)} = R_{A,(a,a^*)} + \langle a, a^*\rangle, \tag{110}$$

and this readily implies that $f_{A,(a,a^*)} = \mathcal{P}(R_{A,(a,a^*)}, R_{A,(a,a^*)} + \langle a, a^*\rangle) = R_{A,(a,a^*)} + \frac{1}{2}\langle a, a^*\rangle$. ◻

If the intrinsic method m in Theorem 4.13 produces "nice" antiderivatives, then so does sometimes the symmetrized method $\mathfrak{m}$. Before we state the corresponding result more precisely, we recall the required properties of the proximal midpoint average. Since these properties were stated in finite-dimensional Hilbert spaces, we assume from now on that

$$X \text{ is a finite-dimensional Hilbert space.} \tag{111}$$

Recall that $f \in \Gamma$ is *piecewise linear-quadratic* if $\operatorname{dom} f$ can be written as a finite union of polyhedral sets on which f is of the form $\langle x, Ax\rangle + \langle x, b\rangle + \gamma$, where $A\colon X \to X$ is linear, $b \in X$, and $\gamma \in \mathbb{R}$. The piecewise linear-quadratic functions on X have many nice properties; see [21, sections 10.E and 11.D].

FACT 4.16. *Let $f_0 \in \Gamma$, and let $f_1 \in \Gamma$ be such that f_0 and f_1^* have a full domain. Then the following hold:*

(i) *$\mathcal{P}(f_0, f_1)$ and $\mathcal{P}(f_0^*, f_1^*)$ have a full domain.*

(ii) *If f_0 and f_1 are piecewise linear-quadratic, then so is $\mathcal{P}(f_0, f_1)$.*

(iii) *If f_0 is differentiable and f_1 is strictly convex, then both $\mathcal{P}(f_0, f_1)$ and its conjugate are differentiable and strictly convex.*

Proof. (i): See [5, Theorem 6.2.(i)]. (ii): The functions f_0, f_1, and $\frac{1}{2}\|\cdot\|^2$ are piecewise linear-quadratic. The operations employed to create $\mathcal{P}(f_0, f_1)$ do not lead outside the class of piecewise linear-quadratic functions; consequently, $\mathcal{P}(f_0, f_1)$ is piecewise linear-quadratic as well. (See also [15, Corollary 5.3].) (iii): This follows from (i) and [5, Theorems 6.2.(ii) and 6.2.(iii)]. ◻

COROLLARY 4.17. *Let $m\colon \mathcal{A} \to \Gamma\colon A \mapsto m_A$ be an intrinsic method that produces antiderivatives with a full domain. Set*

$$\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto \mathcal{P}\big(m_A, m^*_{A^{-1}}\big). \tag{112}$$

Then $\mathfrak{m}$ is a primal-dual symmetric intrinsic method, and the following hold:

(i) *$(\forall A \in \mathcal{A})$ $\mathfrak{m}_A$ and $\mathfrak{m}_A^*$ have a full domain.*

(ii) *If $(\forall A \in \mathcal{A})$ m_A is piecewise linear-quadratic, then $(\forall A \in \mathcal{A})$ $\mathfrak{m}_A$ and $\mathfrak{m}_A^*$ are both piecewise linear-quadratic.*

Proof. Theorem 4.13 states that $\mathfrak{m}$ is primal-dual symmetric. Now fix $A \in \mathcal{A}$, set $f_0 := m_A$, and set $f_1 := m^*_{A^{-1}}$. Then $\mathfrak{m}_A = \mathcal{P}(f_0, f_1)$ and, by Fact 4.11(i) and (ii), $\mathfrak{m}_A^* = \mathcal{P}(f_0^*, f_1^*) = \mathcal{P}(f_1^*, f_0^*)$. (i): Since f_0 and f_1^* have a full domain, Fact 4.16(i) (applied to f_0 and f_1 and to f_1^* and f_0^*) implies that $\mathfrak{m}_A$ and $\mathfrak{m}_A^*$ have a full domain. (ii): This is clear from Fact 4.16(ii). ◻

Remark 4.18. Consider Corollary 4.17. In general, antiderivatives are neither differentiable nor strictly convex. However, if for a particularly nice instance $A \in \mathcal{A}$ both m_A and $m_{A^{-1}}$ are differentiable, then we deduce from Fact 4.16(iii) that $\mathfrak{m}_A$ and $\mathfrak{m}_A^*$ are both differentiable and strictly convex. Analogous comments can be made for other symmetrizations of (not necessarily intrinsic) methods based on the proximal midpoint average.

We are now able to provide two examples of primal-dual symmetric intrinsic methods with very nice properties. These are in striking contrast to Proposition 4.7.

Example 4.19. Set $m\colon \mathcal{A} \to \Gamma\colon A \mapsto \max_{(a,a^*)\in \operatorname{gra} A} R_{A,(a,a^*)}$ and $\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto \mathcal{P}(m_A, m^*_{A^{-1}})$. Then $\mathfrak{m}$ is a primal-dual symmetric intrinsic method, and for every $A \in \mathcal{A}$ both $\mathfrak{m}_A$ and $\mathfrak{m}_A^*$ have a full domain and are piecewise linear-quadratic antiderivatives of A and A^{-1}, respectively.

Proof. For every $A \in \mathcal{A}$, m_A is a convex polyhedral (hence piecewise linear-quadratic) antiderivative of A with full domain (Example 4.4). The conclusion is now a consequence of Corollary 4.17. ∎

Example 4.20. Set $m\colon \mathcal{A} \to \Gamma\colon A \mapsto \frac{1}{n_A}\sum_{(a,a^*)\in \operatorname{gra} A} R_{A,(a,a^*)}$, where n_A is the number of points in $\operatorname{gra} A$, and $\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto \mathcal{P}(m_A, m^*_{A^{-1}})$. Then $\mathfrak{m}$ is a primal-dual symmetric intrinsic method, and for every $A \in \mathfrak{m}$ both $\mathfrak{m}_A$ and $\mathfrak{m}^*_A$ have a full domain and are piecewise linear-quadratic antiderivatives of A and A^{-1}, respectively.

Proof. For every $A \in \mathcal{A}$, m_A is a convex polyhedral (hence piecewise linear-quadratic) antiderivative of A with a full domain (Example 4.5). The conclusion follows from Corollary 4.17. ∎

In the remainder of this section, we aim to extract further nice properties enjoyed by these two methods. We require the following results on the proximal midpoint average.

PROPOSITION 4.21. *Let $f_0 \in \Gamma$, let $f_1 \in \Gamma$, and set $f := \mathcal{P}(f_0, f_1)$. Suppose that $a^* \in \partial f_0(a) \cap \partial f_1(a)$. Then*

$$\big(\forall x \in N_{\operatorname{dom} f_0}(a) \cap N_{\operatorname{dom} f_1^*}(a^*)\big) \quad x + a^* \in \partial f(x + a). \tag{113}$$

Proof. Take $x \in N_{\operatorname{dom} f_0}(a) \cap N_{\operatorname{dom} f_1^*}(a^*)$. Proposition 4.9 yields

$$a = \operatorname{Prox}(f_0)(2x + a + a^*). \tag{114}$$

The same result (applied to f_1^*) shows that $a^* = \operatorname{Prox}(f_1^*)(2x + a^* + a)$, which is equivalent to $a^* = (\operatorname{Id} - \operatorname{Prox}(f_1))(2x + a + a^*)$, i.e., to

$$2x + a = \operatorname{Prox}(f_1)(2x + a + a^*). \tag{115}$$

Add (114) and (115), divide the result by 2, and recall Fact 4.11(iv) to deduce that

$$x + a = \operatorname{Prox}(f)\big((x + a) + (x + a^*)\big). \tag{116}$$

The conclusion now follows from (102). ∎

THEOREM 4.22. *Let $f_0 \in \Gamma$, let $f_1 \in \Gamma$, and set $f := \mathcal{P}(f_0, f_1)$. Suppose that $a^* \in \partial f_0(a) \cap \partial f_1(a)$, and set $N := (N_{\operatorname{dom} f_0}(a) \cap N_{\operatorname{dom} f_1^*}(a^*)) \cup (N_{\operatorname{dom} f_1}(a) \cap N_{\operatorname{dom} f_0^*}(a^*))$. Then the following hold:*

(i) $(\forall y \in a + N)$ $y + a^* - a \in \partial f(y)$.

(ii) *f is differentiable at every point $y \in a + \operatorname{int} N$, with $\nabla f(y) = y + a^* - a$.*

Proof. Proposition 4.21 implies (i). On $a + \operatorname{int} N$, we note that $y \mapsto y + a^* - a$ is a continuous selection of ∂f; therefore, $\nabla f(y) = y + a^* - a$ by [18, Proposition 2.8], and (ii) holds. ∎

COROLLARY 4.23. *Let $m\colon \mathcal{A} \to \Gamma\colon A \mapsto m_A$ be an instrinsic method such that, for every $A \in \mathcal{A}$, $\overline{\operatorname{ran} \partial m_A} \subseteq \operatorname{conv} \operatorname{ran} A$. Set*

$$\mathfrak{m}\colon \mathcal{A} \to \Gamma\colon A \mapsto \mathcal{P}\big(m_A, m^*_{A^{-1}}\big), \tag{117}$$

take $A \in \mathcal{A}$, take $(a, a^*) \in \operatorname{gra} A$, and set $N := N_{\operatorname{conv} \operatorname{dom} A}(a) \cap N_{\operatorname{conv} \operatorname{ran} A}(a^*)$. Then the following hold:

(i) $(\forall y \in a + N)\ y + a^* - a \in \partial \mathfrak{m}_A(y)$.

(ii) $(\forall y \in a + \operatorname{int} N)$ $\mathfrak{m}_A$ *is differentiable at* y, *with* $\nabla \mathfrak{m}_A(y) = y + a^* - a$.

Proof. Set $f_0 := m_A$, and set $f_1 := m^*_{A^{-1}}$ so that $\mathfrak{m}_A = \mathcal{P}(f_0, f_1)$. Proposition 3.2 implies that $\operatorname{dom} f_0^* = \operatorname{dom} m_A^* = \operatorname{conv} \operatorname{ran} A$ and that $\operatorname{dom} f_1 = \operatorname{dom} m^*_{A^{-1}} = \operatorname{conv} \operatorname{ran} A^{-1} = \operatorname{conv} \operatorname{dom} A$. The conclusion is therefore a consequence of Theorem 4.22. ☐

Remark 4.24. In view of Examples 4.4 and 4.5, we observe that Corollary 4.23 is applicable when m is either as in Example 4.19 or as in Example 4.20.

Example 4.25. Let m and $\mathfrak{m}$ be as in Corollary 4.23, let $(a, a^*) \in X \times X$, and suppose that $A\colon X \rightrightarrows X$ is given by $\operatorname{gra} A = \{(a, a^*)\}$. Then there exists $\gamma \in \mathbb{R}$ such that $\mathfrak{m}_A = \frac{1}{2}\|\cdot\|^2 + \langle \cdot, a^* - a\rangle + \gamma$.

Proof. Indeed, the set N in Corollary 4.23 is the entire space X, and hence item (ii) of that result implies that $\nabla \mathfrak{m}_A\colon X \to X\colon x \mapsto x + a^* - a$. ☐

We observe next that, on the real line, the subdifferential extending A is actually single-valued—i.e., it corresponds to a gradient—with slope one outside the box $(\operatorname{conv} \operatorname{dom} A) \times (\operatorname{conv} \operatorname{ran} A)$.

COROLLARY 4.26. *Suppose that* $X = \mathbb{R}$, *let* m *and* $\mathfrak{m}$ *be as in Corollary* 4.23, *and let* $A\colon \mathbb{R} \rightrightarrows \mathbb{R}$ *have finite graph* $\{(a_1, a_1^*), \ldots, (a_n, a_n^*)\}$, *where* $a_1 \leq a_2 \leq \cdots \leq a_n$ *and* $a_1^* \leq a_2^* \leq \cdots \leq a_n^*$. *Then*

$$(\forall x < a_1) \quad \mathfrak{m}'_A(x) = x - a_1 + a_1^* \qquad \textit{and} \qquad (\forall x > a_n) \quad \mathfrak{m}'_A(x) = x - a_n + a_n^*. \tag{118}$$

Proof. Since $\operatorname{conv} \operatorname{dom} A = [a_1, a_n]$ and $\operatorname{conv} \operatorname{ran} A = [a_1^*, a_n^*]$, the result follows from Corollary 4.23(ii) (applied at (a_1, a_1^*) and at (a_n, a_n^*)). ☐

Remark 4.27. Primal-dual symmetry and the "slope one" property of the extension of the cyclically monotone operator A in Corollary 4.26 were properties deemed desirable by Rockafellar. In view of Remark 4.24, there exist two explicit methods that generate antiderivatives with these desirable properties. Although not the product of an intrinsic method, the function $f_{A,(a,a^*)}$ of Theorem 4.15 has the same properties.

We conclude this paper by numerically illustrating an antiderivative produced by the primal-dual symmetric intrinsic method of Example 4.19.

Example 4.28. Define $A\colon \mathbb{R} \rightrightarrows \mathbb{R}$ via $\operatorname{gra} A := \{(a, \exp(a)) \mid a \in \{0, \pm\frac{1}{2}, \pm 1\}\}$. Because $\operatorname{gra} A \subset \operatorname{gra}(\exp)$, the operator A is cyclically monotone. We interpret A as a 5-point sample of the gradient of the exponential function. Let m and $\mathfrak{m}$ be as in Example 4.19. Figure 1 visualizes the exponential function, the antiderivative m_A, the antiderivative $m^*_{A^{-1}}$, and the antiderivative $\mathfrak{m}_A$ produced by the primal-dual symmetric intrinsic method $\mathfrak{m}$. As predicted by Corollary 3.20, the function m_A is nonnegative on $\operatorname{dom} A$. In Figure 2, we visualized the derivative of the exponential function, its 5-point sample corresponding to A, and the maximal cyclically monotone extension $\partial \mathfrak{m}_A$. Note that by Theorem 3.14 the Rockafellar functions are piecewise linear, and hence their subdifferential operators have a "staircase" graph. On the other hand, $\mathfrak{m}_A$ is piecewise linear-quadratric, and its subdifferential operator displays the "slope one" property guaranteed by Corollary 4.26 outside the rectangle $\operatorname{dom} A \times \operatorname{ran} A$. Both plots were generated in Scilab utilizing software packages discussed in [15]; further details on the numerical implementation will appear elsewhere.

Acknowledgments. The authors thank Jean-Baptiste Hiriart-Urruty for making them aware of [14] and two referees for their pertinent comments.

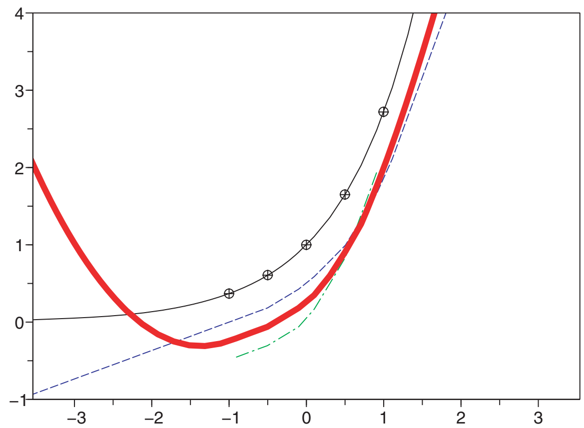


Fig. 1. *The graph of the exponential function (thin black curve) and the* 5 *points (circled) on its graph that led to the operator* A*, the antiderivative* m_A *(dashed blue curve), the antiderivative* $m^*_{A^{-1}}$ *(dashed-dotted green curve), and the proximal-average-based antiderivative* $\mathfrak{m}_A$ *(thick red curve) are shown. Note that* $m_A \geq 0$ *on* $\operatorname{dom} A$*, in accordance with Corollary* 3.20*.*

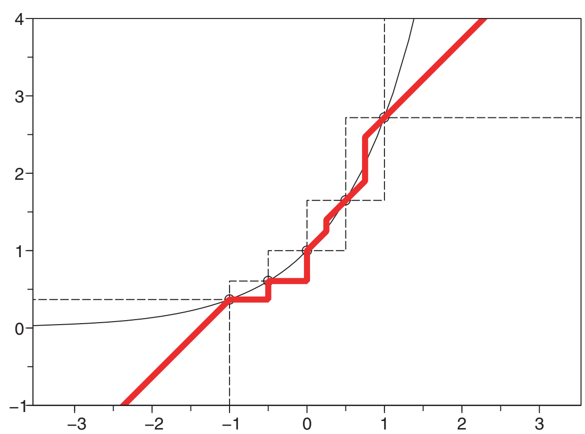


Fig. 2. *The finite graph operator* A *is shown as points (circled) on the graph of the exponential function (thin black curve), which is the same as its derivative. The reconstructed subdifferential operator* $\partial\mathfrak{m}_A$ *(thick red curve) stays inside the rectangles (dashed line) imposed on any monotone extension of* A*. Note the "slope one" property of* $\partial\mathfrak{m}_A$ *outside the rectangle* $\operatorname{conv}\operatorname{dom} A \times \operatorname{conv}\operatorname{ran} A$*, as guaranteed by Corollary* 4.26*.*

REFERENCES

[1] E. Asplund, *A Monotone Convergence Theorem for Sequences of Nonlinear Mappings*, Nonlinear Functional Analysis, in Proceedings of Symposia in Pure Mathematics XVIII Part 1, American Mathematical Society, Chicago, 1970, pp. 1–9.

[2] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, *Fitzpatrick functions, cyclic monotonicity, and Rockafellar's antiderivative*, Nonlinear Anal., 66 (2007), pp. 1198–1223.

[3] H. H. Bauschke, J. M. Borwein, and X. Wang, *Fitzpatrick functions and continuous linear monotone operators*, SIAM J. Optim., 18 (2007), pp. 789–809.

[4] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, *The proximal average: Basic theory*, SIOPT, submitted.

[5] H. H. Bauschke, Y. Lucet, and M. Trienis, *How to transform one convex function continuously into another*, SIAM Rev., to appear.
[6] H. H. Bauschke, E. Matoušková, and S. Reich, *Projection and proximal point methods: Convergence results and counterexamples*, Nonlinear Anal., 56 (2004), pp. 715–738.
[7] H. H. Bauschke and X. Wang, *A convex-analytical approach to extension results for n-cyclically monotone operators*, Set-Valued Anal., 15 (2007), pp. 297–306.
[8] H. H. Bauschke and X. Wang, *The Kernel Average for Two Convex Functions and Its Application to the Extension and Representation of Monotone Operators*, preprint available at http://www.optimization-online.org/DB_HTML/2007/05/1658.html
[9] J. Benoist and A. Daniilidis, *Subdifferential representation of convex functions: Refinements and applications*, J. Convex Anal., 12 (2005), pp. 255–265.
[10] J. M. Borwein, *Maximal monotonicity via convex analysis*, J. Convex Anal., 13 (2006), pp. 561–586.
[11] J. M. Borwein and Q. J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, Berlin, 2005.
[12] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland, Amsterdam, 1973.
[13] P. Choné and H. V. J. Le Meur, *Non-convergence result for conformal approximation of variational problems subject to a convexity constraint*, Numer. Funct. Anal. Optim., 22 (2001), pp. 529–547.
[14] D. Lambert, J.-P. Crouzeix, V. H. Nguyen, and J.-J. Strodiot, *Finite convex integration*, J. Convex Anal., 11 (2004), pp. 131–146.
[15] Y. Lucet, H. H. Bauschke, and M. Trienis, *The piecewise linear-quadratic model for computational convex analysis*, Comput. Optim. Appl., to appear.
[16] G. J. Minty, *On the maximal domain of a monotone function*, Michigan Math. J., 8 (1961), pp. 135–137.
[17] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[18] R. R. Phelps, *Convex Functions, Monotone Operators, and Differentiability*, second ed., Lecture Notes in Math. 1364, Springer-Verlag, Berlin, 1993.
[19] R. T. Rockafellar, *On the maximal monotonicity of subdifferential mappings*, Pacific J. Math., 33 (1970), pp. 209–216.
[20] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[21] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[22] S. Simons, *Minimax and Monotonicity*, Lecture Notes in Math. 1693, Springer-Verlag, Berlin, 1998.
[23] M. D. Voisei, *Extension theorems for k-monotone operators*, Stud. Cercet. Stiinţ. Ser. Mat. Univ. Bacău, 9 (1999), pp. 235–242.
[24] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.

Journal of Convex Analysis
Volume 7 (2000), No. 2, 375–393

Compactly epi-Lipschitzian Convex Sets and Functions in Normed Spaces*

Jonathan Borwein
CECM, Department of Mathematics and Statistics, Simon Fraser University, Burnaby BC, V5A 1S6, Canada.
e-mail: jborwein@cecm.sfu.ca

Yves Lucet
Department of Mathematics and Statistics, University of Victoria, PO BOX 3045 STN CSC, Victoria BC, V8W 3P4, Canada.
e-mail: lucet@na-net.ornl.gov

Boris Mordukhovich
Department of Mathematics, Wayne State University, Detroit, Michigan 48202, USA.
e-mail: boris@math.wayne.edu

Received November 9, 1998
Revised manuscript received January 11, 2000

We provide several characterizations of compact epi-Lipschitzness for closed convex sets in normed vector spaces. In particular, we show that a closed convex set is compactly epi-Lipschitzian if and only if it has nonempty relative interior, finite codimension, and spans a closed subspace.

Next, we establish that all boundary points of compactly epi-Lipschitzian sets are proper support points. We provide the corresponding results for functions by using inf-convolutions and the Legendre–Fenchel transform. We also give an application to constrained optimization with compactly epi-Lipschitzian data via a generalized Slater condition involving relative interiors.

Keywords: Compactly epi-Lipschitzian set, convex set

1. Introduction

The concept of compactly epi-Lipschitzian (CEL) sets in locally convex topological spaces was introduced by Borwein and Strojwas [6]. It is an extension of Rockafellar's concept of epi-Lipschitzian sets [36]. An advantage of the CEL property is that it always holds in finite dimensional spaces and, in contrast to its epi-Lipschitzian predecessor, makes it possible to recapture much of the detailed information available in finite dimensions. The original motivation for introducing the CEL concept was to select class of closed sets in infinite dimensions (primarily in Banach spaces) for which the Clarke tangent and normal cones [11] adequately measure boundary behavior. A number of strong results were obtained in this direction; see [3], [6], [7], [8], and references therein. At the same time it was clarified that the CEL property is not sufficient for the (weak-star) locally compactness of the Clarke normal cone at boundary points [3, Example 4.1]. To get

*Research supported by NSERC and by the Shrum endowment of Simon Fraser University for the first author, research partly supported by the Pacific Institute for the Mathematical Sciences for the second author, and research partly supported by the National Science Foundation under grant DMS-9704751 and by an NSERC Foreign Researcher Award for the third author.

ISSN 0944-6532 / $ 2.50

the latter result, Borwein introduced [3] the notion of epi-Lipschitz-like (ELL) sets that takes an intermediate place between the epi-Lipschitzian and CEL properties. He also proved in [3] that, for any closed subset $C \subset X$ of a normed space and its boundary point $c \in \operatorname{bd} C$, the Clarke normal cone is locally compact if C is CEL and tangentially regular at this point. The mentioned regularity property seems to be restrictive even in the finite dimensional setting. In particular, it never holds for the so-called Lipschitzian manifolds [38] which are locally homeomorphic to graphs of nonsmooth Lipschitz continuous vector functions; see [26, Section 3] for more details.

Further research showed that the regularity assumption can be avoided in infinite dimensions and the CEL property alone ensures the required local compactness if the Clarke normal cone is replaced with a different concept of generalized normals. The first result in this direction was obtained by Loewen [23] who proved, for any closed sets in reflexive spaces, that the CEL property implies the local compactness of the so-called limiting Fréchet normal cone [22] which is an infinite dimensional generalization of the normal cone introduced by Mordukhovich [24], [25]. An extension of this result to the more general case of weakly compactly generated Asplund spaces was obtained by Mordukhovich and Shao [30].

In the case of general Banach spaces a similar result was established by Jourani and Thibault [19] for the so-called G-normal cone of Ioffe [15] that is another infinite dimensional extension of Mordukhovich's construction. In [17], Jourani proved that the CEL property implies the local compactness of Ioffe's A-normal cone (which may be bigger than the G-cone) if the space is "weakly trustworthy." Recently Ioffe [16] established several characterizations of the CEL property in terms of normal cones satisfying certain requirements in corresponding Banach spaces. We refer the reader to [4], [10], [16], [17], [18], [19], [20], [21], [29], [28], [30] and their bibliographies for various applications of the CEL property to subdifferential calculus, metric regularity, Lipschitzian stability, necessary optimality conditions, and related aspects of nonlinear analysis and optimization in Banach spaces.

The primary goal of this paper is to provide intrinsic characterizations of the CEL property of closed convex sets in normed spaces. We are not familiar with any of such characterizations for either CEL or ELL convex sets even in particular infinite dimensional spaces. On the other hand, it has been known for a long time that a convex set is epi-Lipschitzian if and only if its interior is nonempty; see Rockafellar [37]. In this paper we prove that the CEL and ELL properties of closed convex sets agree in any normed spaces. Our main Theorem 2.5 in Section 2 contains also eight other characterizations of the CEL property one of which requires the additional Baire structure of the normed space in question. In particular, we show that a closed convex subset of a normed space is CEL if and only if its relative interior is nonempty and its span is a closed subspace of finite codimension.

In Section 3 we study supporting properties of CEL sets. The main characterization theorem of Section 2 allows us to establish that any boundary point of a CEL closed convex set in an arbitrary normed space is a proper support point of the set. In the case of Banach spaces we give a variational proof of this result and discuss its nonconvex generalizations. Section 4 concerns with characterizations of the CEL property for closed convex functions that are derived from the corresponding set characterizations applied to epigraphs. The final Section 5 contains some applications of the obtained results to constrained optimization via a generalized Slater interiority condition. Throughout the paper we use standard notation and terminology.

2. Characterizations of compactly epi-Lipschitzian Convex Sets

We give our main characterization theorem and proceed to prove it.

2.1. The main results: formulations and discussions

In the following, X denotes a normed linear space, and $\|.\|$ its norm. First we recall the definition of CEL and ELL sets [3, Definition 3.1]. We use the terminology compactly epi-Lipschitzian while in the literature it is sometimes referred as compactly epi-Lipschitz.

Definition 2.1. (i) A set C in X is *compactly epi-Lipschitzian (CEL)* if for all x in C, there are N_x a neighborhood of x, U a neighborhood of the origin, a positive ϵ and K a convex compact set such that

$$0 < \lambda < \epsilon \Longrightarrow C \cap N_x + \lambda U \subset C + \lambda K.$$

(ii) A subset C of X is *epi-Lipschitz-like (ELL)* if for all x in C there are N_x a neighborhood of x, Ω a convex set with polar set Ω^0 weakly$*$ locally compact, and a positive ϵ such that

$$0 < \lambda < \epsilon \Longrightarrow C \cap N_x + \lambda\Omega \subset C.$$

To characterize CEL convex sets, we recall what is a precompact set and a Baire space.

Definition 2.2. (i) A set P is *precompact* or *totally bounded* if for any open set U, there is a finite set F with $P \subset F + U$.

(ii) A set Σ is a *polytope* if it is the convex hull of a finite number of points.

(iii) A normed linear space X is a *Baire space* or *of the second category* if any countable covering of X with closed sets A_n, contains a set A_{n_0} with nonempty interior.

Remark 2.3. The Baire category theorem tells us that if X is a complete metric space, every nonempty open subset of X is of the second category [39, 41]. In particular, any complete metric space is a Baire space (see [41] for a reference on Baire spaces). Hence Baire spaces include Banach spaces, but there are examples of Baire spaces which are not complete as the following example shows [41, Exercise 3-1-4] (see also [1]) .

Example 2.4. Take X an infinite dimensional Banach space. It contains Y a nonclosed subspace with countable non-finite codimension. Name $(e_i)_i$ the sequence of linearly independent vectors such that $\operatorname{span}(Y \cup \cup_{i=1}^{\infty}\{e_i\}) = X$. Define $X_N := \operatorname{span}\{Y, e_1, \ldots, e_N\}$. Then $X = \cup X_N$ and there is $\bar{N}$ with $X_{\bar{N}}$ of second category. Since $X_{\bar{N}}$ is not closed, it is a non-Banach Baire space.

Now we can state our main theorem.

Theorem 2.5. *Let C denote a closed convex set in a normed linear space X. The following are equivalent:*

(i) *The set C is CEL.*

(ii) *There is a convex compact set K with $0 \in \operatorname{int}(C + K)$.*

(iii) *There is a precompact set P with $0 \in \operatorname{int}(C + P)$.*

(iv) *There is a convex polytope* Σ *with* $0 \in \operatorname{int}(C + \Sigma)$.
(v) *There is a finite dimensional space* E *with* $\operatorname{int}(C + E) \neq \emptyset$.
(vi) *There is a point* x *in* C *such that the polar set of* $C - x$, $(C-x)^0$ *is weak-star locally compact.*
(vii) *The subspace spanned by* C, $\operatorname{span} C$, *is a finite-codimensional closed subspace and the relative interior* $\operatorname{ri} C := \operatorname{int}_{\operatorname{span} C} C$ *is nonempty.*
(viii) *The set* C *is ELL.*
(ix) *The projection* $Q : X \to M$ *on the subspace* $M := \operatorname{span} C$ *is linear continuous with its null space* $\mathcal{N}(Q)$ *finite dimensional and* $\operatorname{int}(Q(C))$ *nonempty.*

When X *is a Baire normed space, these are also equivalent to:*

(x) *There is a continuous linear open map* $Q : X \to Y$ *having a finite dimensional nullspace such that* $\operatorname{int}(Q(C))$ *is nonempty.*

In fact, our proof shows that the theorem still holds when Baire normed space is replaced by convex Baire spaces (see [40] for a reference on convex Baire spaces). That notion is slightly more general since there are examples of convex Baire spaces which are not Baire normed space.

This theorem calls for several kind of remarks.

Remark 2.6. First are some straightforward remarks:

- Clearly compact epi-Lipschitzness is invariant under translation.
  Hence we may always assume $0 \in C$ to simplify some of our arguments.
- The intersection of CEL convex sets need not be CEL. Indeed, take C_1 a pointed cone with interior in X an infinite dimensional subspace. Let C_2 be $-C_1$. Then C_1 and C_2 are CEL (since they have nonempty interior) but $C_1 \cap C_2 = \{0\}$ is not CEL (otherwise 0 would have a compact neighborhood, so X would be finite dimensional). For example, take the positive cone in c or in l_∞.
- Compact epi-Lipschitzness is clearly an infinite dimensional notion. Indeed, any set in a finite dimensional normed linear space is CEL (take K equals to the unit ball).

Remark 2.7. Next all three parts of Properties (vii) are needed:

- We do need $\operatorname{span} C$ to be closed. Take Φ a discontinuous linear function on an infinite dimensional Banach space and consider $C := \Phi^{-1}(0)$ an hyperplane. Then there is e in X such that $C + \mathbb{R}e = X$, so $\operatorname{span} C$ is finite codimensional. However, we have $\operatorname{int}(C + [-1,1]e)$ is empty.
  Indeed, suppose there is an open set U contained in $C + [-1,1]e$. Then $\Phi(U) \subset \Phi(C) + [-1,1]\Phi(e) = [-1,1]\Phi(e)$. So Φ is bounded which contradicts the fact Φ is not continuous. Consequently, there is no compact convex set K such that $\operatorname{int}(C + K) \neq \emptyset$, which means C is not CEL.
- Assuming $\operatorname{ri} C$ is nonempty and $\operatorname{span} C$ closed does not imply C is CEL. For example take $C := \mathbb{R}e$ in an infinite dimensional Banach space. Then $\operatorname{ri} C \neq \emptyset$ but C is not finite codimensional hence not CEL.
- We do need $\operatorname{ri} C \neq \emptyset$. Consider $X := l_2(\mathbb{N})$ the space of sequences with the norm $\|\cdot\|_2$, and $C := l_2^+$ the positive cone. Then $\operatorname{span} C = X$ is finite codimensional closed but $\operatorname{ri} C = \emptyset$.

- Property (vii) may be rewritten as: "there is a finite codimensional closed subspace M such that $C \subset M$ and $\operatorname{int}_M(C)$ is nonempty". This seems a weaker statement since Property (vii) merely states that it holds for $M = \operatorname{span} C$. However note that for any set S in a normed space X, $\operatorname{int}(S)$ nonempty yields $\operatorname{span} S = X$; consequently, the set M in the rephrased statement must be $\operatorname{span} C$. That fact follows from arguing by contradiction: take any s in $\operatorname{int}(S)$ and assume there is x not in $\operatorname{span} S$. Then build $x_n := x/n + s$. For n large enough, x_n is in $\operatorname{int}(S)$, so $x = n(x_n - s)$ is in $\operatorname{span} S$.

Remark 2.8. The convexity is very important in our result. Indeed, take $X := l_\infty(\mathbb{N})$ the bounded sequences in the supremum norm, define $f(x) := \liminf_{n\to\infty}|x_n|$, and consider $C := \{x \in X \;:\; f(x) \leq 0\}$. Then C is CEL but neither ELL nor convex [3, Example 4.1].

Similarly, $\{(x, r) : \liminf x_k \leq r \text{ in } l_2(\mathbb{R})\}$ is nonconvex CEL but not ELL.

Remark 2.9. In Property (iv) not only can we find a polytope Σ satisfying $\operatorname{int}(C + \Sigma) \neq \emptyset$, but also can we take $\Sigma = \sum_{k=1}^{N}[-1, 1]e_k$ where $e_1, \dots, e_N$ are linearly independent vectors. This fact will be used several time to simplify our proofs.

2.2. Proof of the main theorem: general relations

The proof of Theorem 2.5 is split into several lemmas. We first show how all parts fit together.

Proof. First the relations (i)$\Leftrightarrow$(ii)$\Rightarrow$(iii)$\Rightarrow$(iv)$\Leftrightarrow$(v) are implied by Proposition 2.10 and Lemma 2.11. Next Lemma 2.14 gives (iv)$\Rightarrow$(x)$\Rightarrow$(ii) and Lemma 2.12, 2.15, and 2.16 give (iv)$\Leftrightarrow$(vii)$\Leftrightarrow$(ix); so all properties (i)–(v), (vii), (ix), and (x) are equivalent.

Now use Lemma 2.17 to get (vii) and (iv) imply (viii) and [3, Proposition 3.1 (a)] to obtain (viii)$\Rightarrow$(i).

Finally, [3, Lemma 2.1] gives (v)$\Rightarrow$(vi)$\Rightarrow$(ii). □

Let us start with the proof of our first characterization: (i)$\Leftrightarrow$ (ii). Note that it holds without any closedness assumption, so any convex set containing a CEL set is CEL (properties (i)–(vi) are clearly preserved under inclusion). In particular, if C is convex CEL, its closure $\bar{C}$ is CEL.

Proposition 2.10. *Let C be a convex subset of X. Then C is CEL if and only if there is K a convex compact set such that $0 \in \operatorname{int}(C + K)$*

Proof. Apply the notation of Definition 2.1. If C is CEL, take x in C and $\lambda := \epsilon/2$. Since

$$x + \lambda U \subset C \cap N_x + \lambda U \subset C + \lambda K,$$

we obtain $\lambda U \subset C + (\lambda K - x)$. Hence $0 \in \operatorname{int}(C + (\lambda K - x))$, which proves the necessary condition.

Fix $\bar{x} \in C$. Let V be a neighborhood of zero with $V \subset C + K$ and let U be a convex neighborhood of zero with $U+U\subset V$. Then for any $\lambda\in(0,1)$ one has

$$\begin{aligned} C\cap(\bar{x}+U)+\lambda U &\subset (1-\lambda)C+\lambda(\bar{x}+U)+\lambda U\\ &\subset (1-\lambda)C+\lambda(U+U)+\lambda\bar{x}\\ &\subset (1-\lambda)C+\lambda C+\lambda K+\lambda\bar{x}\\ &\subset C+\lambda(K+xb). \end{aligned}$$

This completes the proof because $K+\bar{x}$ is compact. □

Lemma 2.11. *In Theorem* 2.5, *the following relations hold:*
(ii)⇒(iii)⇒(iv) ⇔(v).

Proof. Clearly, (ii)⇒(iii) holds (a compact set is precompact) as well as (iv)⇒(v) (span Σ is finite dimensional) and (v)⇒(iv)(because for any basis $e_1,\dots,e_N$ of E, $\Sigma=\sum_{i=1}^{N}[-1,1]e_i$ is a convex polytope spanning E and so satisfying $\operatorname{int}(C+\Sigma)\neq\emptyset$).

So the only remaining implication to prove is (iii)⇒(iv). Let U be a nonempty bounded open set with $U\subset C+P$. Using precompactness, there is a finite set F such that $P\subset F+U/2$. Define $\Sigma:=\operatorname{co}F$. We have

$$U\subset C+F+\frac{U}{2}.$$

By induction, we deduce

$$\begin{aligned} \frac{U}{2} &\subset \frac{C}{2}+\frac{\Sigma}{2}+\frac{U}{4}\subset\frac{C}{2}+\frac{C}{4}+\frac{\Sigma}{2}+\frac{\Sigma}{4}+\frac{U}{8},\\ &\subset \cdots\subset(\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^n})C+(\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^n})\Sigma+\frac{U}{2^{n+1}}. \end{aligned}$$

Consequently,

$$\frac{U}{2}\subset(1-\frac{1}{2^n})(C+\Sigma)+\frac{U}{2^{n+1}}.$$

Taking the limit when n goes to infinity, we find $U/2\subset\overline{C+\Sigma}$. Since C is closed and Σ compact, we obtain $\operatorname{int}(C+\Sigma)\neq\emptyset$. □

Lemma 2.12. *In Theorem* 2.5, *the following relation holds:* (vii)⇔(ix).

Proof. The set $M:=\operatorname{span}C$ is a finite codimensional closed subspace. We can write $M+N=X$ with $M\cap N=\{0\}$ and the null space $N:=\mathcal{N}(Q)$ finite dimensional. The result follows since X and $M\times N$ are isomorphic. □

Remark 2.13. Note that we only need the projection to be continuous with finite dimensional null space and $\operatorname{int}(Q(C))$ nonempty. Indeed, Q is open is always true for a projection (at least in a normed space) and all projections are linear.

Next we show the particular part involving Baire spaces.

Lemma 2.14. *In Theorem* 2.5, *the following relations hold:*
(iv)⇒(x)⇒(ii).

Proof. Assume Property (iv) holds. Call $N := \operatorname{span}\Sigma$ the finite dimensional space spanned by Σ. Let $Y = X/N$ be the quotient space, and define $Qx := x + N$. Then N, the nullspace of Q, is finite dimensional. In addition, $Q(C) = C + N$ contains $C + \Sigma$. Hence it has nonempty interior.

Note that [39, p. 60, Proposition 3] shows that Q is open. So to end the proof, we show that Property (x) implies Property (ii).

Assume (x) holds and X is a Baire space. Name $N := \mathcal{N}(Q)$ and $B_r := \{x \in X : \|x\| \leq r\}$ the closed ball of radius r in X. The set $C_r := C + N \cap B_r$ is closed (since it is the sum of a closed set and a compact set). The set $Q(C+N) = Q(C)$ has nonempty interior in Y, i.e. , there is an open set V such that $V \subset Q(C+N)$. We deduce that $Q^{-1}(V) \subset Q^{-1}(Q(C+N))$. The continuity of Q implies that $Q^{-1}(V)$ is open and its linearity implies $Q^{-1}(Q(C+N)) = C+N$. Indeed take $x \in Q^{-1}(Q(C+N))$, there is $c \in C$ such that $Q(x) = Q(c)$. Hence $x = (x-c)+c \in N + C$; the reverse inclusion is obvious. Consequently $C+N$ has nonempty interior. So there is an open set U contained in $C + N = \cup_{r=1}^{\infty} C_r$. We deduce that there is r_0 with $\operatorname{int}(C_{r_0})$ nonempty. Therefore Property (ii) holds with $K = N \cap B_{r_0}$. □

Next we prove the part involving finite codimensional spaces.

2.3. Proof of the main theorem: finite codimension property

Lemma 2.15. *The following relation holds in Theorem* 2.5*:* (iv)⇒(vii)*.*

Proof. Without loss of generality, we can assume $0 \in C$. There are $\Sigma_N := \sum_{k=1}^{N}[-1,1]e_k$ with $e_1, \ldots, e_N$ linearly independent and U an open neighborhood of 0 with $U \subset C + \Sigma_N$.

Step 1: The subspace $\operatorname{span} C$ is finite codimensional and closed.

Take $e \notin \operatorname{span} C$ and assume $C + [-1,1]e$ has interior. We are going to prove $\operatorname{span} C$ is closed. For contradiction we suppose the open unit ball $B[0;1]$ is contained in $C+[-1,1]e$. Take $r > \|e\|$. If $\operatorname{span} C$ is not closed, it is dense. Since $e \notin \operatorname{span} C$ there is a sequence $x_n \in \operatorname{span} C$ with $x_n \to e \in B[0;r]$. Eventually $x_n \in B[0;r] \subset rC + [-r,r]e$ and so $x_n \in rC$ for n large enough. Using the closedness of C, we obtain the contradiction: $e \in rC \subset \operatorname{span} C$.

Now build $C_0 := C$, $C_k := C_{k-1} + [-1,1]e_k$. All C_k have Property (vi) (since $C \subset C_k$ and that property is preserved under inclusion). Hence the previous argument shows C_{k-1} is closed in $\operatorname{span} C_k$ for $k = 1, \ldots, N$. Since $\operatorname{span} C_N = X$ (see Remark 2.7), all $\operatorname{span} C_k$ are closed. In particular, $\operatorname{span} C_0 = \operatorname{span} C$ is closed.

Since $C_N = C + \Sigma_N$ has nonempty interior, $\operatorname{span} C_N = X$ is finite codimensional. Moreover $\operatorname{span}(C_{k-1} + [-1,1]e_k) = \operatorname{span} C_k$ yields $\operatorname{codim} C_{k-1} \leq \operatorname{codim} C_k < \infty$. In particular, $\operatorname{span} C_0 = \operatorname{span} C$ is finite codimensional which ends Step 1.

Step 2: The relative interior of C, $\operatorname{ri} C := \operatorname{int}_{\operatorname{span} C} C$ is nonempty.

To prove the claim we use the two properties: $0 \leq \alpha \leq \beta \Rightarrow \alpha C \subset \beta C$ (since $\alpha/\beta C + (1-\alpha/\beta)0 \subset C$), and $\alpha, \beta \geq 0 \Rightarrow (\alpha+\beta)C = \alpha C + \beta C$ (another use of convexity).

Note $M := \operatorname{span} C = \operatorname{cone} C - \operatorname{cone} C$, $E_1 := \sum_{e_i \in M}[-1,1]e_i$, and $E_2 := \sum_{e_i \notin M}[-1,1]e_i$.

There is an open set $U \subset C + E_1 + E_2$ which satisfies

$$0 \in U \cap M \subset (C + E_1 + E_2) \cap M \subset C + E_1.$$

For $e_i \in M = \operatorname{cone} C - \operatorname{cone} C$, there are $\alpha_i^1, \alpha_i^2 \geq 0$ and $c_i^1, c_i^2 \in C$ such that $e_i = \alpha_i^1 c_i^1 - \alpha_i^2 c_i^2$. So any $u \in U \cap M \subset C + E_1$ can be written with $\lambda_i \in [-1, 1]$ as

$$\begin{aligned} u &= c + \sum_{i \in I} \lambda_i(\alpha_i^1 c_i^1 - \alpha_i^2 c_i^2), \\ &= c + \sum_{i \in I^+} (\lambda_i \alpha_i^1 c_i^1 + (1 - \lambda_i)\alpha_i^2 c_i^2 + \alpha_i^1 c_i^1) + \\ &\quad \sum_{i \in I^-} (-\lambda_i \alpha_i^2 c_i^2 + (1 + \lambda_i)\alpha_i^1 c_i^1 + \alpha_i^2 c_i^2) - \bar{x}, \end{aligned}$$

where $I^+ := \{i \in I \ : \ \lambda_i \in [0, 1]\}$, $I^- := \{i \in I \ : \ \lambda_i \in [-1, 0]\}$, $I := \{i \ : \ e_i \in M\}$, $c \in C$ and $\bar{x} := \sum_{i \in I} (\alpha_i^1 c_i^1 + \alpha_i^2 c_i^2)$. So

$$U \cap \operatorname{span} C \subset [1 + 2 \sum_i (\alpha_i^1 + \alpha_i^2)] C - \bar{x}$$

which implies $\operatorname{ri} C$ is not empty. □

Lemma 2.16. *The following relation holds in Theorem* 2.5*:* (vii)⇒(iv).

Proof. Without loss of generality, we can assume $0 \in \operatorname{int}_{\operatorname{span} C} C$. Call $X_0 := \operatorname{span} C$ and $X_1 := X_0 + \mathbb{R} e_1$ with $X = X_0 + \operatorname{span}\{e_1, \dots, e_N\}$ with $\|e_i\| = 1$. There is a ball B_0 with radius r_0 centered at 0 with $B_0 \cap X_0 \subset C$. We claim that there is another ball $B_1 \subset B_0$ with $B_1 \cap X_1 \subset C + [-1, 1]e_1$. Indeed, take y in $B_0 \cap X_1$ and not in X_0, then $y = \hat{c} + \alpha e_1$ for some $\hat{c}$ in C. Consider the linear functional $u : X_1 \to \mathbb{R}$ defined for all $x = c + \beta e_1$ by $u(x) = \beta$. Since $X_0 = \mathcal{N}(u)$ is closed, u is continuous [39, p. 382, Proposition 4]. Therefore $u(B_0)$ is bounded and hence there is r_1, $0 < r_1 < r_0$ such that for the ball B_1 centered at 0 and with radius r_1 one has $u(B_1) \subset [-1, 1]$, and for $y \in B_1 \cap X_1$ not in X_0, $y \in C + [-1, 1]e_1$. Moreover $B_1 \cap X_0 \subset B_0 \cap X_0 \subset C$. All in all, $B_1 \cap X_1 \subset C + [-1, 1]e_1$.

Now consider $X_2 := X_1 + \mathbb{R} e_2$. Applying the same argument gives the existence of a ball B_2 with $B_2 \cap X_2 \subset C + [-1, 1]e_1 + [-1, 1]e_2$. Consequently there is B_N in $X_N = X$ with $B_N \subset C + \sum_{i=1}^N [-1, 1]e_i$ which means (iv). □

Lemma 2.17. *The following relations hold in Theorem* 2.5*:*
(vii) *and* (iv)⇒(viii).

Proof. We can always assume $0 \in \operatorname{ri} C$.

Take $x = 0 \in C$, N_x a convex neighborhood of x such that $N_x \cap C - N_x \cap C \subset C$ ($0 \in \operatorname{ri} C$), $\epsilon := 1$, $0 < \lambda < \epsilon$, and $\Omega := x - C \cap N_x$. Then Ω is convex and

$$C \cap N_x + \lambda\Omega \ = C \cap N_x - \lambda C \cap N_x \ \subset C \cap N_x - C \cap N_x \subset C$$

since $\lambda(C \cap N_x) \subset C \cap N_x$ because $C \cap N_x$ is a convex set containing zero.

Taking (i)⇒(vi) and the definition of ELL into account, it remains to be proved that Ω is CEL. But there is $K := \sum_{i=1}^{N}[-1,1]e_i$ compact convex and an open ball B centered at 0 with $B \subset C + K$ which yields $B \subset B \cap C + K \subset C + K$ (same argument as in the previous lemma, we heavily use the closedness of $\operatorname{span} C$). Hence $B \cap C$ is CEL, which ends the proof. □

2.4. Proof of the main theorem: quasi-relative interior

Using quasi-relative interiors [5], we give an alternative proof of (iv)⇒(vii) under the additional assumption that C has nonempty quasi-relative interior. Even if the assumption is stronger, it clarifies the relation with quasi-relative interiors. Note that the proof uses Theorem 3.2 whose second proof does not depend on Theorem 2.5.

Definition 2.18. A point x is in the *quasi-relative interior* of a convex set C if for all nonzero continuous linear functional λ:

$$\lambda(C - x) \geq 0 \Rightarrow \lambda(C - x) = 0.$$

The set of quasi-relative interior points is denoted by $\operatorname{qri} C$.

In other words, x is a quasi-relative interior point of C if and only if $T_C(x)$, the tangent cone to C at x, is a subspace. In $\mathbb{R}^n$, the quasi-relative interior of a convex set is its relative interior.

If X is a separable Banach space and C a closed convex set, Borwein and Lewis [5] proved that C has a nonempty quasi-relative interior. So the next proposition applies in separable Banach spaces.

Proposition 2.19. *Assume C is a CEL closed convex set with nonempty quasi-relative interior. Then there is a finite codimensional closed subspace M such that $C \subset M$ and $\operatorname{int}_M(C)$ is nonempty.*

Proof. If 0 does not belong to C, it belongs to $C - \bar{c}$ where $\bar{c} \in C$. Since $\operatorname{qri}(C - \bar{c}) = \operatorname{qri}(C) - \bar{c}$ and $C - \bar{c}$ is closed convex CEL, we apply the proposition to $C - \bar{c}$. So we can always assume $0 \in C$.

Proposition 2.10 and Lemma 2.11 give (v): there is a finite dimensional subspace Σ_0 and a nonempty open set U with $U \subset C + \Sigma_0$. We name $\hat{c}$ a point in $\operatorname{qri} C$.

If $\operatorname{int}(C)$ is nonempty, the whole space $M := X$ satisfies the theorem: $U \cap X \subset C \subset X$ and $\operatorname{codim} X = 0$. Otherwise $\operatorname{int}(C) = \emptyset$, so $\hat{c}$ belongs to $\operatorname{bd} C$, the boundary of C. Applying Theorem 3.2, we deduce that $\hat{c}$ is a support point of C. So there is a nonzero continuous linear functional $\lambda_0 \in X^*$ such that $\lambda_0(C - \hat{c}) \geq 0$. Since $\hat{c}$ is in $\operatorname{qri} C$ and $0 \in C$, we obtain $\lambda_0(C) = 0$, i.e. , C is in the nullspace $M_1 := \mathcal{N}(()\lambda_0)$ of λ_0. Naming $\Sigma_1 := \Sigma_0 \cap M_1$ gives

$$U \cap M_1 \subset C + \Sigma_1 \subset M_1.$$

Now either the proposition holds with $M = M_1$ or we can build inductively a sequence of finite codimensional closed subspaces M_k and a sequence of finite dimensional subspaces Σ_k with $\operatorname{codim} M_k = k$ and $\dim \Sigma_k < \dim \Sigma_{k-1}$ (in fact $\dim \Sigma_k + 1 = \dim \Sigma_{k-1}$) such that

$$U \cap M_k \subset C + \Sigma_k \subset M_k.$$

Indeed, that property holds for $k = 1$. Suppose it holds for k, then either $\mathrm{int}_{M_k}(C)$ is nonempty so the theorem holds, or $\mathrm{int}_{M_k}(C)$ is empty and $\hat{c}$ belongs to $\mathrm{bd}\, C \subset \mathrm{supp}\, C \subset M_k$. Hence there is a nonzero continuous linear functional $\lambda_k \in (M_k)^*$ such that $\lambda_k(C - \hat{c}) \geq 0$. We apply the Hahn-Banach theorem [39, p. 77, Theorem 1] to extend λ_k to all X. The same argument as above gives $\lambda_k(C) = 0$. We define $M_{k+1} := N(\lambda_k)$ and $\Sigma_{k+1} := \Sigma_k \cap M_{k+1}$ to obtain

$$U \cap M_{k+1} \subset C + \Sigma_{k+1} \subset M_{k+1}$$

with $\dim \Sigma_{k+1} < \dim \Sigma_k$. Consequently both sequences exist.

To conclude, note that $n := \dim \Sigma_0$ is finite implies $\dim \Sigma_n = 0$. So $\Sigma_n = \{0\}$ and $\mathrm{codim}\, M_n = n$. Consequently the proposition holds with $M = M_n$. □

Remark 2.20. The assumption "C has nonempty quasi-relative interior" (which is true if the space is Banach and separable) is needed in the proof of the proposition.

Because of this assumption, the proof above does not cover all cases since the quasi-relative interior is empty for sets like $S := l_p^+(\mathbb{R})$ with $1 \leq p < \infty$ (the positive cone in $l_p(\mathbb{R}) := \{s : \mathbb{R} \to \mathbb{R} \ : \ \sum_{r \in \mathbb{R}} |s(r)|^p < \infty\}$). Indeed recall that $\sum_{r \in \mathbb{R}} |s(r)|^p = \sup_{F \text{ finite}} \sum_{r \in \mathbb{R} \cap F} |s(r)|$ and that the sum being finite implies the support of s is countable in S. Take any $\bar{s}$ in $l_p^+(\mathbb{R})$. Since the support of $\bar{s}$ is countable, there is $\bar{r}$ in $\mathbb{R}$ with $\bar{s}(\bar{r}) = 0$. Define the linear functional $f(s) := s(\bar{r})$. Using $f(\bar{s}) = \bar{s}(\bar{r}) = 0$ and $f(S) = \{s(\bar{r}) \ : \ s \in S\} \geq 0$ yields $\langle f, S - \bar{s} \rangle \geq 0$. Since there is $\hat{s}$ with $f(\hat{s}) = 1$, f is not always 0 on S. Hence $\bar{s}$ is a support point of S and $T_S(\bar{s})$ is not a subspace. All in all, $\mathrm{qri}\, S$ is empty.

3. Supporting properties of compactly epi-Lipschitzian sets

Starting from our main characterization theorem, we study supporting properties of CEL sets. An alternate proof of the main theorem in that section is presented for CEL sets in Banach spaces. The extension to nonconvex sets is then discussed.

3.1. Convex sets in normed spaces

In addition to characterizing CEL closed convex sets, we show that they have "nice" boundaries. Indeed, our next theorem will show that every boundary point of such a set is also a support point. First we need a technical lemma.

Lemma 3.1. *Let X be a normed linear space, K be a convex compact set, and C be a closed convex set. If $\bar{c}$ is in $\mathrm{bd}\, C$, the boundary of C, then there is $\bar{k}$ in $\mathrm{bd}\, K$ such that $\bar{c} + \bar{k} \in \mathrm{bd}(C + K)$.*

Proof. Take $\bar{c}$ in $\mathrm{bd}\, C$. Suppose, to the contrary, that for all $\bar{k}$ in $\mathrm{bd}\, K$, $\bar{c} + \bar{k}$ does not belong to $\mathrm{bd}(C + K)$. Then it must belong to $\mathrm{int}\,(C + K)$. Since K has empty interior (X has infinite dimension) the inclusion holds for all $k \in K$. Since K is compact, there is a positive ϵ such that

$$\bar{c} + K + \epsilon B \subset C + K \tag{3.1}$$

Now applying the Rådström cancellation principle [33] gives $\bar{c} + \epsilon B \subset C$ which contradicts the fact that $\bar{c}$ belongs to $\mathrm{bd}\, C$. □

Here is the announced theorem. Recall that a proper support point means that C is supported by an hyperplane and C is not included in the hyperplane. We will need such a proper hyperplane in the proof of Lemma 5.1.

Theorem 3.2. *Suppose X is a normed space and C a CEL closed convex set. Then every boundary point of C is also a proper support point of C.*

Our proof relies on the two classical separation theorems: with nonempty interior and in finite dimensions. See the next remark for an alternate proof and links to the Bishop-Phelps theorem.

Proof. Take $\bar{c}$ in the boundary of C. First suppose $\operatorname{int}(C)$ is nonempty. Then we can separate C from $\{\bar{c}\}$ (with the Hahn–Banach theorem for example): there is a nonzero continuous linear functional l such that $l(\bar{c}) = \sup_{c \in C} l(c)$ which means that $\bar{c}$ is a support point of C.

Now assume $\operatorname{int}(C)$ is empty. Theorem 2.5 implies C has nonempty relative interior. So we can apply the above argument to obtain a nonzero continuous linear functional l, defined on $\operatorname{span} C$, with $l(\bar{c}) = \sup_{c \in C} l(c)$. Now $\operatorname{span} C$ being finite codimensional allows us to extend l to a continuous linear functional $\hat{l}$ defined on all X. For example define $\hat{l}(x) := l(\operatorname{Proj}_{\operatorname{span} C}(x))$, where $\operatorname{Proj}_{\operatorname{span} C}$ is the projection onto $\operatorname{span} C$. So $\bar{c}$ is a proper support point of C. $\square$

Note that every point in C is a support point of C. Indeed if $\bar{c} \in C \setminus \operatorname{ri} C$ the above proof shows $\bar{c}$ is a proper support point. Otherwise $\bar{c} \in \operatorname{ri} C$. Write $X = \operatorname{span} C \oplus \mathbb{R}e \oplus Y$. Fix $x = c + re + y$ and take $l(x) = r$. Then $l(C) = 0$ and hence $\bar{c}$ is a support point of C.

Remark 3.3. Theorem 3.2 can be obtained from Lemma 4 in [2] in the case of Banach spaces based on our characterizations of CEL sets.

Remark 3.4. Theorem 3.2 does not always hold for non-CEL closed convex sets. Indeed, Fonf [13] proved that in every incomplete normed space there is a closed bounded convex set C with no support point. Theorem 3.2 tells us that such a set C cannot be CEL.

A more striking case when $\operatorname{bd} C$ equals $\operatorname{supp} C$ (the set of support points of C) is provided by the following example in which $C(S)$ could be ℓ_∞.

Example 3.5. Let S be a compact Hausdorff space, and $F \subset S$ be a closed non G_δ set. Then $C := \{f \in C(S) \,:\, f \geq 0 \text{ and } f(x) = 0 \text{ for } x \in F\}$ verifies

$$\operatorname{bd} C = C = \operatorname{supp} C.$$

We end that subsection with a way to generate CEL closed convex sets. Indeed, a consequence of the next proposition is that in any normed linear space, projections of CEL closed convex sets are CEL: note that any closed convex set containing a CEL closed convex set is CEL. Note also that the closed convex hull of a CEL set is necessarily CEL by Proposition 2.10.

Proposition 3.6. *Assume C is CEL, $Q : X \to Y$ is an open linear continuous map, and X, Y are normed linear spaces. Then $Q(C)$ is CEL.*

Proof. There is a convex compact set K and an open set U with $U \subset C + K$. Thus, $Q(U) \subset Q(C) + Q(K)$. Now since Q is open, $Q(U)$ is open and since Q is continuous, $Q(K)$ is compact. Hence, we found a compact set $K' = Q(K)$ with $\operatorname{int}(Q(C) + K')$ nonempty, which means that $Q(C)$ is CEL. □

3.2. Variational arguments in Banach spaces

Let us present another proof of Theorem 3.2 in the case of Banach spaces X, i.e. , under an additional completeness assumption in the theorem. This proof is fully independent of Theorem 2.5: it is actually based on variational arguments and admits generalizations to nonconvex sets; see the next subsection. For the case of convex closed sets C under consideration we use the classical Bishop-Phelps theorem on the density of support points in the boundary of C; see, e.g., [32, Theorem 3.18].

Take any $\bar{c}$ in the boundary of C. According to the Bishop-Phelps theorem, we find sequences $\{c_n\} \subset X$ and $\{\xi_n\} \in X^*$ satisfying $c_n \to \bar{c}$ as $n \to \infty$, $\|\xi_n\| = 1$, and

$$\xi_n \in N(c_n; C) := \{\xi \in X^* |\ \langle \xi, c - c_n \rangle \leq 0\} \tag{3.2}$$

for all $n = 1, 2, \ldots$, where $N(\cdot; C)$ signifies the normal cone of convex analysis. Since C is CEL, we can apply Loewen's result in [23, Proposition 3.7] to conclude that there exist a compact set $S \subset X$, a neighborhood U of $\bar{c}$, and a number $\gamma > 0$ such that

$$N(c; C) \subset K_\gamma(S) := \{\xi \in X^* |\ \gamma \|\xi\| \leq \max_{s \in S} |\langle \xi, s \rangle|\} \tag{3.3}$$

for all $c \in C \cap U$. Note that the mentioned result of [23] covers the general case of nonconvex sets $C \subset X$ where $N(c; C)$ in (3.3) is replaced with the so-called *Fréchet normal cone* to C at $c \in C$ defined by

$$\widehat{N}(c; C) := \{\xi \in X^* |\ \limsup_{x \to c,\, x \in C} \frac{\langle \xi, x - c \rangle}{\|x - c\|} \leq 0\}. \tag{3.4}$$

It is well known that constructions (3.2) and (3.4) agree for convex sets.

Due to the weak* compactness of the unit ball in the dual space X^* we select a subnet $\xi_\nu \in N(c_\nu; C)$ in (3.2) which weakly* converges to some $\bar{\xi} \in X^*$. Passing to the limit in (3.2), we easily get that $\langle \bar{\xi}, c - \bar{c} \rangle \leq 0$.

It remains to prove that $\|\bar{\xi}\| \neq 0$. Assume on the contrary that $\bar{\xi} = 0$. Using the compactness of the set $S \subset X$ in (3.3), we conclude that $\langle \xi_\nu, s \rangle \to 0$ uniformly in S. Thus (3.3) implies that $\xi_\nu \to 0$ in the *norm* topology of X^*. But this is impossible due to $\|\xi_\nu\| = 1$ for all ν. The obtained contradiction completes the proof of Theorem 3.2 in Banach spaces.

3.3. Nonconvex generalizations

The above arguments can be extended to the nonconvex case using variational principles and appropriate concepts of normal cones in nonsmooth analysis. If X is an *Asplund space* (that is, a Banach space where every convex continuous function is generically Fréchet differentiable, in particular, any reflexive space; see [32]), then a proper analogue of the

Bishop-Phelps theorem is obtained by Mordukhovich and Shao [27] via the density of the set

$$c \in \text{bd } C \ \text{ with } \ \widehat{N}(c; C) \neq \{0\} \tag{3.5}$$

involving the Fréchet normal cone (3.4). Moreover, the density of (3.5) for every closed set $C \subset X$ is shown to be a *characterization* of Asplund space; see [12]. Now using Loewen's result mentioned above, we conclude similarly to Subsection 3.2 that for any closed set C, CEL at $\bar{c}$, one has

$$N(\bar{c}; C) \neq \{0\} \ \text{ at every } \ \bar{c} \in \text{bd } C \tag{3.6}$$

in terms of the *limiting normal cone*

$$N(\bar{c}; C) := \limsup_{c \to \bar{c},\, c \in C} \widehat{N}(c; C) \tag{3.7}$$

introduced in [22] as an extension of the finite dimensional construction of Mordukhovich [24]. In (3.7) "limsup" connotes the *sequential* Painlevé-Kuratowski upper limit of multifunctions with respect to the norm topology in X and the weak* topology in X^*. Note that we can use the sequential vs. topological (net) upper limit in (3.6) and (3.7) since a bounded set in X^* is weakly* sequentially compact for any Asplund space X; see [32].

When C is convex, the normal cone (3.7) reduces to the normal cone of convex analysis. Thus (3.6) can be viewed as an extention of our support point theorem (Theorem 3.2) to nonconvex CEL sets in Asplund spaces. Note that in this form it does not hold outside of Asplund spaces. In fact it was shown by Fabian and Mordukhovich [12] that in any non-Asplund space X there is an *epi-Lipschitzian* set $C \subset X$ for which (3.6) is violated at *every* boundary point. To cover the case of arbitrary Banach spaces, one needs to use a different normal cone for nonconvex sets.

An appropriate construction was introduced by Ioffe under the name of the (approximate) *G-normal cone* denoted by $N_G(\cdot; C)$, see [15]. This construction is another infinite dimensional extension of [24] being generally more complicated than (3.6). It agrees with (3.6) in certain most important situations but may be bigger (never smaller) than (3.6) even for epi-Lipschitzian sets in spaces with Fréchet smooth renorms; see [29, Section 9] for more details and further references.

Similarly to the arguments in [27] one can show that the density result (3.5) holds in terms of $N_G(\cdot; C)$ for any closed set C in a Banach space X. Then we need to pass to the limit and get an analogue of (3.6) in terms of the G-normal cone for any closed CEL set. To make it possible, we can use the result of Jourani and Thibault [19, Lemma 3] that ensures a local compactness property of type (3.3) for the G-normal cone under the CEL assumption on C at $\bar{c}$. This justifies the N_G-analogue of Theorem 3.2 for CEL sets in arbitrary Banach spaces.

Note that density (properness) results of type (3.5) where obtained by Borwein and Strojwas [7, 8] for various normal cones dual to some tangent cones in Banach spaces. However, for arbitrary CEL sets those normal cones may not possess a local compactness property of type (3.3), called "normal compactness" in [29], that is crucial for the limiting procedure. In particular, the Clarke normal cone is not locally compact unless C is ELL; see [3]. The normal compactness property with the Clarke normal cone is shown to be satisfied for ELL sets in [20]. Ioffe recently proved [16] that the CEL property of $C \subset X$ at $\bar{c}$ is actually *equivalent* to property (3.3), if $N(\cdot; C)$ is either the Fréchet normal cone on an Asplund space X or the G-normal cone on an arbitrary Banach space. It follows from Borwein [3, Example 4.1] that this result does not hold for the Clarke normal cone in $X = l_\infty$.

4. Characterizations of compactly epi-Lipschitzian Convex Functions

The analogous of our main theorem for CEL sets is presented for CEL closed convex functions. As usual the link between functions and sets is provided by the epigraph

$$\operatorname{epi} f := \{(x, r) \in X \times \mathbb{R} \ : \ f(x) \leq r\}.$$

We say that a closed convex function $f : X \to \mathbb{R} \cup \{+\infty\}$ is CEL if its epigraph is CEL. As for sets, we can always assume $f(0) = 0$.

The next theorem characterizes CEL convex functions. The notation δ_K denotes the indicator function of the set K: $\delta_K(x) = 0$ if $x \in K$, $+\infty$ otherwise. We denote by f^* the Legendre–Fenchel conjugate of f (see [14, 35])

$$f^*(s) := \sup_{x \in X} [\langle s, x \rangle - f(x)],$$

and $f \Box g$ the inf-convolution of f and g

$$f \Box g(x) := \inf_{y \in X} [f(y) + g(x - y)].$$

The core denotes the algebraic interior of a set

$$\operatorname{core} C := \{x \in C \ : \ \forall d \in X, \exists T > 0 \ : \ |t| \leq T \Rightarrow x + td \in C\}.$$

Theorem 4.1. *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper closed convex function. The following are equivalent:*

(i) *The function f is CEL.*
(ii) *There is a convex compact set K such that $f \Box \delta_K$ is continuous at 0.*
(iii) *There is a convex compact set K such that $f^* + \delta_K^*$ has bounded level sets.*
(iv) *There is a convex compact set K such that* $0 \in \operatorname{core}(\operatorname{dom} f \Box \delta_K)$.

To prove the theorem, we need several basic steps. First we recall the following well-known result. Its simple proof is included for the sake of self-containedness.

Lemma 4.2. *Take $\sigma : X \to \mathbb{R} \cup \{+\infty\}$.*

- *If* $0 \in \operatorname{int}(\operatorname{epi} \sigma)$, *then σ is bounded on some neighborhood of* 0.
- *If σ is finite and continuous at* 0, *then* $\operatorname{int}(\operatorname{epi} \sigma)$ *is nonempty.*

Proof. There is an open ball B of center 0 and radius r with $0 \in B \subset \operatorname{epi} \sigma$. If there is a sequence x_n converging to 0 with $\sigma(x_n) = \infty$, then for n large enough, $(x_n, \sigma(x_n)) \in B$, i.e. , $\|x_n\| + |\sigma(x_n)| < r$ which contradicts $\sigma(x_n) = \infty$. This proves the first part of the lemma.

Next, assume σ is (finite and) continuous at 0. Set $I := (\sigma(0)-1, \sigma(0)+1)$. Then

$$\begin{aligned}(0,2)+\sigma^{-1}(I)\times I &= \sigma^{-1}(I)\times(\sigma(0)+1,\sigma(0)+3)\\ &\subset \{(x,\alpha)\ :\ \sigma(x)<\alpha\}\subset \operatorname{epi}\sigma.\end{aligned}$$

Since $\sigma^{-1}(I)$ is open, $\operatorname{int}(\operatorname{epi}\sigma)$ is nonempty. □

Remark 4.3. Note that even if a function is continuous CEL convex it may not send bounded sets to bounded sets.

Indeed take $X = l_2$ and $f(x) = \sum_{n=1}^{\infty}|x_n|^{2n}$. Then f is convex and continuous (since it is lower semi-continuous and finite in a Banach space). Moreover f is CEL since its epigraph spans $l_2\times\mathbb{R}$ (which is clearly closed and of finite codimension), and has nonempty interior (by the previous lemma since f is continuous at 0). However, $f(2e_n) = 2^{2n}$ implies that if we denote B the unit ball, $f(2B)$ is unbounded.

So f is a CEL convex continuous function that does not send bounded sets to bounded sets.

Since the strict epigraph of the inf-convolution is the sum of the strict epigraphs (see Remark [14, IV.2.3.3]) we need to relate compact epi-Lipschitzness to strict epigraphs. Recall that

$$\operatorname{epi}_s f := \{(x,r)\in X\times\mathbb{R}\ :\ f(x)<r\}.$$

Lemma 4.4. *If f is a proper convex function,* $\operatorname{span}(\operatorname{epi} f) = \operatorname{span}(\operatorname{epi}_s f)$, *and* $\operatorname{ri}(\operatorname{epi} f) = \operatorname{ri}(\operatorname{epi}_s f)$. *In particular,* $\operatorname{epi} f$ *is CEL if and only if* $\operatorname{epi}_s f$ *is CEL.*

Proof. The second assertion clearly follows by using Theorem 2.5(vii).

Step 1: $\operatorname{span}(\operatorname{epi} f) = \operatorname{span}(\operatorname{epi}_s f)$.

Take y in $\operatorname{span}(\operatorname{epi} f)$. There are $\alpha_i \geq 0$, $x_i \in X$, and $r_i \in \mathbb{R}$ with $f(x_i) \leq r_i$ for $i = 1, 2$ and $y = \alpha_1(x_1, r_1) - \alpha_2(x_2, r_2)$. Since without loss of generality we can assume $(0, 0) \in \operatorname{epi} f$, the points $(0, a)$ are in $\operatorname{span}(\operatorname{epi} f)$ for any $a \in \mathbb{R}$. (The convexity allows to write $\operatorname{span}(\operatorname{epi} f) = \operatorname{cone}(\operatorname{epi} f) - \operatorname{cone}(\operatorname{epi} f)$).

Either $\alpha_2 = 0$ and we can write

$$y = \alpha_1(x_1, r_1) = \alpha_1(x_1, r_1 + 1) - (0, \alpha_1)$$

so y belongs to $\operatorname{span}(\operatorname{epi}_s f)$. Or $\alpha_2 \neq 0$ and we write

$$y = \alpha_1(x_1, r_1) - \alpha_2(x_2, r_2) = \alpha_1(x_1, r_1 + 1) - \alpha_2(x_2, r_2 + \frac{\alpha_1}{\alpha_2}).$$

So y is again in $\operatorname{span}(\operatorname{epi}_s f)$. The reverse inclusion is obvious.

Step 2: $\operatorname{ri}(\operatorname{epi} f) = \operatorname{ri}(\operatorname{epi}_s f)$.

Take $y \in \operatorname{ri}(\operatorname{epi} f)$, and note $M := \operatorname{span}(\operatorname{epi} f) = \operatorname{span}(\operatorname{epi}_s f)$, and $B(y, \delta)$ the open ball of center y and radius δ. There is $\delta > 0$ such that $B(y, \delta) \cap M \subset \operatorname{epi} f$. Take $u := (x_u, r_u) \in B(y, \delta/2)\cap M$ and consider $u' := u-(0, \delta/3)$. Using the triangle inequality we deduce $u' \in B(y, \delta)$. Since u' is also in M, it is in $\operatorname{epi} f$. Hence u is in $\operatorname{epi}_s f$. We conclude that $B(y, \delta/2) \cap M \subset \operatorname{epi}_s f$. The reverse inclusion is clear. □

Now we prove Theorem 4.1.

Proof. Step 1: (ii) implies (i).

Assume (ii). Apply Lemma 4.2 and Lemma 4.4 to obtain

$$\emptyset \neq \operatorname{int}(\operatorname{epi}(f \square \delta_K)) = \operatorname{int}(\operatorname{epi}_s(f \square \delta_K)) = \operatorname{int}(\operatorname{epi}_s f + \operatorname{epi}_s \delta_K).$$

Note that $\operatorname{epi}_s \delta_K \subset K \times (0,\infty)$ and $\operatorname{epi}_s f + K \times (0,\infty) \subset \operatorname{epi}_s f + K \times \{0\}$ to get $\operatorname{int}(\operatorname{epi}_s f + K') \neq \emptyset$ with $K' := K \times \{0\}$ a compact convex set. Consequently $\operatorname{epi}_s f$, and so $\operatorname{epi} f$, is CEL.

Step 2: (i) implies (ii).

Conversely, assume (i). Then there is K' a compact convex set with $0 \in \operatorname{int}(\operatorname{epi}_s f + K')$. Define $K := \operatorname{Proj}_X K'$ and $\bar{t} := \min_{(x,t)\in K'} t$. Since both projections Proj_X and $\operatorname{Proj}_{\mathbb{R}}$ are continuous and K' is compact, K is compact convex, $\bar{t}$ is well-defined and we have

$$\operatorname{epi}_s f + K' \subset \operatorname{epi}_s f + K \times \{\bar{t}\}.$$

We obtain $0 \in \operatorname{int}(\operatorname{epi}_s \tilde{f} + K \times (0,\infty))$ where $\tilde{f} := f(.) - \bar{t}$. Applying again Remark [14, IV.2.3.3] we deduce $0 \in \operatorname{int}(\operatorname{epi}_s(\tilde{f} \square \delta_K))$. Applying Lemma 4.2, the function $\tilde{f} \square \delta_K$ is bounded on a neighborhood of 0.

To conclude we note that $\tilde{f} \square \delta_K$ is a convex function, so it is continuous at 0 (for example see [14, Lemma IV.3.1.1] whose proof still holds in a normed linear space). Since $\tilde{f} \square \delta_K = f \square \delta_K - \bar{t}$, (ii) holds.

Step 3: (ii) $\Leftrightarrow$ (iii) $\Leftrightarrow$ (iv).

The key result we use is a theorem proved by Moreau [31] and by Rockafellar [34] which implies that a proper convex function on a normed linear space is strongly continuous at 0 if and only if its conjugate has bounded level sets.

More precisely, [34, Theorem 7A(a)] and [34, Corollary 4D] give (ii) $\Leftrightarrow$ (iii), while [34, Theorem 4C] gives (iii)$\Leftrightarrow$(iv). □

5. Applications to constrained optimization

In order to apply the Fenchel duality theorem of [5], we first prove that the relative interior is equal to the quasi-relative interior for CEL closed convex sets.

Lemma 5.1. *Assume C is CEL, closed, and convex. Then* $\operatorname{ri} C = \operatorname{qri} C$.

Proof. Without loss of generality we can take $x = 0$ to prove both inclusions.

Step 1: We show that $\operatorname{ri} C \subset \operatorname{qri} C$.

Take $x = 0 \in \operatorname{ri} C$ and let us prove $\operatorname{cl}(\mathbb{P}C) = \operatorname{span} C$. Fix a neighborhood V of zero with $V \cap \operatorname{span} C \subset C$. Then for any $\lambda > 0$ one has

$$(\lambda V) \cap \operatorname{span} C = \lambda V \cap \lambda \operatorname{span} C \subset \lambda C.$$

So $(\mathbb{P}V) \cap \operatorname{span} C \subset \mathbb{P}C$ and hence, since $\mathbb{P}V = X$ one gets $(\mathbb{P}V) \cap \operatorname{span} C = \operatorname{span} C$. So $\operatorname{span} C \subset \mathbb{P}C$. As $\mathbb{P}C \subset \operatorname{span} C$, one has $\operatorname{span} C = \mathbb{P}C$. In particular $\operatorname{span} C = \operatorname{cl}(\mathbb{P}C)$ since $\operatorname{span} C$ is closed.

Step 2: We show that $x \notin \operatorname{ri} C \Rightarrow x \notin \operatorname{qri} C$.

If $x \notin$ ri and $x \in C$ then $x \in \operatorname{bd} C$. Applying Theorem 3.2, there is a nonzero continuous linear functional λ such that $\lambda(C \setminus \{x\} - x) > 0$. Since X is infinite dimensional and C is CEL, C is not reduced to $\{x\}$. So there is $x' \in C$ with $\lambda(x' - x) > 0$. Consequently we found a nonzero continuous linear functional λ such that $\lambda(C - x) \geq 0$ but $\lambda(C - x) \neq 0$. In other words, $x \notin \operatorname{qri} C$. □

To write our next theorem, we use the same notations as in [5]: X is a normed space, $g : X \to (-\infty, \infty]$ and $h : \mathbb{R}^n \to (-\infty, \infty]$ are convex proper; and $A : X \to \mathbb{R}^n$ is a continuous linear function.

Theorem 5.2. *If g is closed convex CEL and*

$$\text{either} \quad A(\operatorname{ri}(\operatorname{dom} g)) \cap \operatorname{ri}(\operatorname{dom} h) \neq \emptyset$$
$$\text{or} \quad A(\operatorname{ri}(\operatorname{dom} g)) \cap \operatorname{dom} h \neq \emptyset \qquad \text{and } h \text{ is polyhedral,}$$

then

$$\inf\{g(x) + h(Ax) \ : \ x \in X\} = \max\{-g^*(A^T\lambda) - h^*(-\lambda) \ : \ \lambda \in \mathbb{R}^n\}.$$

Proof. If the function g is CEL convex, then its domain is also CEL by Proposition 3.6. So Lemma 5.1 gives $\operatorname{qri}(\operatorname{dom} g) = \operatorname{ri}(\operatorname{dom} g)$. Applying Fenchel duality theorem [5, Corollary 4.3] gives the same generalized Slater condition as in finite dimension: we need to find a point in $\operatorname{ri}(\operatorname{dom} g)$ with image by A in $\operatorname{ri}(\operatorname{dom} h)$. □

So the CEL property ensures sufficient amount of compactness to recover the finite dimensional results (see our discussions in Section 1 and in Subsections 3.2 and 3.3).

The subsequent sum rule and minimax theorem have been stated in [5]. In fact, Ye used our main characterization theorem (Theorem 2.5) to apply a sum rule in her proof of necessary conditions for optimal control of strongly monotone variational inequalities [42].

Acknowledgements. We are grateful to Lionel Thibault for a very careful reading of the manuscript and for suggesting several simplifications in our proofs. Many thanks to Simon Fitzpatrick for the proof of the closedness of the span of a CEL convex set and to Jafar Zafarani for pointing out a missing link in a first version of the proof of Lemma 2.15.

References

[1] J. Arias de Reyna: Dense hyperplanes of first category, Mathematische Annalen 249 (1980) 111–114.

[2] E. Bishop, R. R. Phelps: The support functionals of a convex set, Proc. of Symposia in Pure Math., Amer. Math. Soc., Providence, R.I., Convexity VII (1963) 27–35.

[3] J. M. Borwein: Epi-Lipschitz-like sets in Banach space: theorems and examples, Nonlinear Anal. 11 (1987) 1207–1217.

[4] J. M. Borwein, A. Jofré: A nonconvex separation property in Banach spaces, Math. Methods Oper. Res. 48 (1998) 169–179.

[5] J. M. Borwein, A. S. Lewis: Partially finite convex programming, part 1: Quasi relative interiors and duality theory, Math. Programming 57 (1992) 15–48.

[6] J. M. Borwein, H. M. Strójwas: Tangential approximations, Nonlinear Anal. 9 (1985) 1347–1366.

[7] J. M. Borwein, H. M. Strójwas: Proximal analysis and boundaries of closed sets in Banach space, I. Theory, Canad. J. Math. 38 (1986) 431–452.

[8] J. M. Borwein, H. M. Strójwas: Proximal analysis and boundaries of closed sets in Banach space, II. Applications, Canad. J. Math. 39 (1987) 428–472.

[9] J. M. Borwein, J. D. Vanderwerff: Banach spaces that admit support sets, Proc. Amer. Math. Soc. 124 (1996) 751–755.

[10] J. M. Borwein, Q. J. Zhu: A survey of subdifferential calculus with applications, J. Nonlinear Analysis: Th. Meth. Appl. 38 (1999) 687–773.

[11] F. H. Clarke: Optimization and Nonsmooth Analysis, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 1990.

[12] M. Fabian, B. S. Mordukhovich: Nonsmooth characterizations of Asplund spaces and smooth variational principles, Set-Valued Anal. 6 (1998) 381–406.

[13] V. P. Fonf: On supportless convex sets in incomplete normed spaces, Proc. Amer. Math. Soc. 120 (1994) 1173–1176.

[14] J.-B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms, A Series of Comprehensive Studies in Mathematics, Springer-Verlag, 1993.

[15] A. D. Ioffe: Approximate subdifferentials and applications, III. The metric theory, Mathematika 36 (1989) 1–38.

[16] A. D. Ioffe: Codirectional compactness, metric regularity and subdifferential calculus, Tech. Rep., In Press, 1997.

[17] A. Jourani: Compactly epi-Lipschitzian sets and A-subdifferentials in WT-spaces, Optimization 34 (1995) 1–17.

[18] A. Jourani: The role of locally compact cones in nonsmooth analysis, Comm. Appl. Nonlinear Anal. 5 (1998) 1–35.

[19] A. Jourani, L. Thibault: Metric regularity for strongly compactly Lipschitzian mappings, Nonlinear Anal. 24 (1995) 229–240.

[20] A. Jourani, L. Thibault: Extensions of subdifferential calculus rules in Banach spaces, Canad. J. Math. 48 (1996) 834–848.

[21] A. Jourani, L. Thibault: Coderivatives of multivalued mappings, locally compact cones and metric regularity, Nonlinear Anal. 35 (1999) 925–945.

[22] A. Y. Kruger, B. S. Mordukhovich: Extremal points and the Euler equation in nonsmooth optimization problems, Dokl. Akad. Nauk BSSR 24 (1980) 684–687, 763.

[23] P. D. Loewen: Limits of Fréchet normals in nonsmooth analysis, in: Optimization and Nonlinear Analysis (Haifa, 1990), vol. 244 of Pitman Res. Notes Math. Ser., Longman Sci. Tech., Harlow (1992) 178–188.

[24] B. S. Mordukhovich: Maximum principle in the problem of time optimal response with nonsmooth constraints, J. Appl. Math. Mech. 40 (1976) 960–969.

[25] B. S. Mordukhovich: Approximation Methods in Problems of Optimization and Control, "Nauka", Moscow, 1988. (Russian). Second English ed. to appear.

[26] B. S. Mordukhovich: Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis, Trans. Amer. Math. Soc. 343 (1994) 609–657.

[27] B. S. Mordukhovich, Y. Shao: Extremal characterizations of Asplund spaces, Proc. Amer. Math. Soc. 124 (1996) 197–205.

[28] B. S. Mordukhovich, Y. Shao: Nonconvex differential calculus for infinite-dimensional multifunctions, Set-Valued Anal. 4 (1996) 205–236.

[29] B. S. Mordukhovich, Y. Shao: Nonsmooth sequential analysis in Asplund spaces, Trans. Amer. Math. Soc. 348 (1996) 1235–1280.

[30] B. S. Mordukhovich, Y. Shao: Stability of set-valued mappings in infinite dimensions: point criteria and applications, SIAM J. Control Optim. 35 (1997) 285–314.

[31] J.-J. Moreau: Sur la fonction polaire d'une fonction semi-continue supérieurement, C. R. Acad. Sci. Paris, 258 (1964) 1128–1130.

[32] R. R. Phelps: Convex Functions, Monotone Operators and Differentiability, vol. 1364 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, second ed. 1993.

[33] H. Rådström: An embedding theorem for spaces of convex sets, Proc. Amer. Math. Soc. 3 (1952) 165–169.

[34] R. T. Rockafellar: Level sets and continuity of conjugate convex functions, Trans. Amer. Math. Soc. 123 (1966) 46–63.

[35] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton, New York, 1970.

[36] R. T. Rockafellar: Directionally Lipschitzian functions and subdifferential calculus, Proc. London Math. Soc. (3) 39 (1979) 331–355.

[37] R. T. Rockafellar: Generalized directional derivatives and subgradients of nonconvex functions, Canad. J. Math. 32 (1980) 257–280.

[38] R. T. Rockafellar: Maximal monotone relations and the second derivatives of nonsmooth functions, Ann. Inst. H. Poincaré. Anal. Non Linéaire 2 (1985) 167–184.

[39] C. Swartz: An Introduction to Functional Analysis, Pure and Applied Mathematics, A Series of Monographs and Textbooks, Marcel Dekker Inc., New York, 1992.

[40] M. Valdivia: Topics in Locally Convex Spaces, vol. 67 of North-Holland Mathematics Studies, North-Holland Publishing Co., Amsterdam, 1982. Notas de Matemática [Mathematical Notes] 85.

[41] A. Wilansky: Modern Methods in Topological Vector Spaces, McGraw-Hill International Book Co., London, New York, 1978.

[42] J. Ye: Necessary optimality conditions for control of strongly monotone variational inequalities, in: Control of Distributed Parameter and Stochastic Systems, S. Chen et al. (ed.), Kluwer, Boston, (1999) 153–160.

Journal of Convex Analysis
Volume 14 (2007), No. 3, 657–666

Parametric Computation of the Legendre-Fenchel Conjugate with Application to the Computation of the Moreau Envelope*

Jean-Baptiste Hiriart-Urruty

Laboratoire de Mathématiques pour l'Industrie et la Physique,
Institut de Mathématiques, Université Paul Sabatier,
118 route de Narbonne, 31 062 Toulouse Cedex, France
jbhu@cict.fr

Yves Lucet

Computer Science, I. K. Barber School of Arts and Sciences,
University of British Columbia Okanagan, 3333 University Way,
Kelowna BC V1V 1V7, Canada
yves.lucet@ubc.ca

Received: April 28, 2006

A new algorithm, named the Parametric Legendre Transform (PLT) algorithm, to compute the Legendre-Fenchel conjugate of a convex function of one variable is presented. It returns a parameterization of the graph of the conjugate except for some affine parts corresponding to nondifferentiable points of the function. The approach is extended to the computation of the Moreau envelope, resulting in a simple yet efficient algorithm.

Theoretical results, the description (and extension) of the algorithm, its approximation error and the convergence, as well as the comparison with known algorithms are included.

Keywords: Legendre-Fenchel transform, Fenchel conjugate, Moreau envelope, Moreau-Yosida regularization, Fast algorithm, Computational convex analysis

2000 Mathematics Subject Classification: 65Y20, 52A41

1. Introduction

In convex duality the Legendre-Fenchel transform, which associates with a convex extended-valued function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ its Legendre-Fenchel conjugate

$$f^*(s) = \sup_{x \in \mathbb{R}^d} \langle s, x \rangle - f(x)$$

allows to relate primal optimization problems of the type

$$p = \inf_{x \in \mathbb{R}^d} f(x) + g(Ax)$$

where $A \in \mathbb{R}^{n \times d}$ and f, g are extended-valued proper lower semi-continuous (lsc) convex functions, with the dual problem

$$d = \sup_{x \in \mathbb{R}^m} -f^*(-A^T z) - g^*(-z),$$

*This work was partly supported by the second author NSERC Discovery grant.

ISSN 0944-6532 / \$ 2.50

(where A^T denotes the transpose of A) by Fenchel's duality theorem [5, Theorem 3.3.5]. Convex duality encompasses linear programming duality, the Min-Max Theorem from game theory, and also allows to state duality results for semi-definite optimization problems. Earlier work on duality used the notion of Legendre transform as detailed in [21, Section 26].

Recent work has focused on the numerical computation of the Legendre-Fenchel conjugate. Symbolic computation was explored in [3, 2, 10] where a specific class of functions was identified. It relies on inverting the gradient mapping with recent advances focusing on extension to multidimensional functions. When symbolic computation fails, or when the function is only known through a black box or at specific points, numerical algorithms still allow the computation of an approximation of the conjugate. Work in that area started with the Fast Legendre Algorithm (FLT) [6, 16], and was later improved with the Linear-time Legendre Transform (LLT) algorithm [17]. Recent work [18] noticed the close relationship between the Fenchel conjugate and the Moreau envelope

$$M_\lambda(s) = \inf_{x\in\mathbb{R}^d} f(x) + \frac{\|s-x\|^2}{2\lambda},$$

resulting in several linear-time algorithms for both transforms.

While the FLT and LLT algorithms initially focused on solving numerically partial differential equations, namely the Hamilton-Jacobi equation, more specialized algorithms have been developed in image processing to compute the so-called square Euclidean distance transform

$$DT^2(p) = \|p\|^2 - g^*(p)$$

of a binary image $p \mapsto f(p) \in \{0,1\}$, where $g(q) = \|q\|^2/2 + I(q)$ and I is the indicator function of the background: $I(q) = 0$ if $f(q) = 1$, $+\infty$ otherwise. Algorithms from image processing take additional advantage of the simplified grid $\{1,\dots,n_1\} \times \{1,\dots n_2\}$ on which the function is defined, and of the integer value of f to speed up the computation by using integer arithmetic instead of floating point operations. See [14, 15, 8, 9, 7, 1, 11, 20, 23, 18] for the latest Euclidean distance transform algorithms.

The framework of the present work is slightly different. We assume $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a one-variable function with f and f' available through a black box. In that context, we investigate a parameterization of the Fenchel conjugate in Section 2, which allows us to define the PLT algorithms in Section 3. Section 4 focuses on the approximation error and the convergence, while Section 5 extends the approach to the computation of the Moreau envelope. Numerical comparison is performed in Section 6, and Section 7 concludes the paper.

2. Theoretical preliminaries

Note $\Gamma(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\} |\ f$ is convex, lower-semicontinuous, proper $\}$, where f proper means that $\operatorname{dom} f = \{x \in \mathbb{R} | f(x) < +\infty\}$ is nonempty. Additionally, we note $\tilde{\Gamma}(\mathbb{R})$ the set of functions $f \in \Gamma(\mathbb{R})$ whose domain is not a singleton. Among the set of all lsc convex functions, Γ disregards the special case $f \equiv +\infty$, and $\tilde{\Gamma}$ the case f is a needle function *i.e.* $f = I_{\{x_0\}}$ for some $x_0 \in \mathbb{R}$. Define

$$G = \{(s,z) \in \mathbb{R}^2 \mid \exists x \in \mathbb{R},\ f \text{ differentiable at } x,\ s = f'(x), \text{ and } z = sx - f(x)\}.$$

We first prove that the parameterization that describes G allows us to recover the graph of the conjugate.

Lemma 2.1. *If $f \in \tilde{\Gamma}(\mathbb{R})$, then $G \neq \emptyset$.*

Proof. By definition of $\tilde{\Gamma}$, there are $x_1, x_2 \in \operatorname{dom} f$ with $x_1 \neq x_2$. Since f is convex, $\operatorname{dom} f$ is also convex, so the segment $[x_1, x_2]$ is contained in $\operatorname{dom} f$. Applying [21, Theorem 25.3], f is differentiable almost everywhere in the open interval (x_1, x_2). So there is $x \in (x_1, x_2)$ with f differentiable at x, and G is nonempty since it contains the point $(f'(x), f'(x)x - f(x))$. □

Remark 2.2. The assumption that there are two distinct points in $\operatorname{dom} f$ is needed since $G = \emptyset$ when $f = I_{\{x_0\}}$ is a needle function. It is equivalent in our context to the function f^* being lower-bounded by two affine functions with different slopes. See the so-called class of epi-pointed functions introduced in [4] for a generalization to higher dimension.

Proposition 2.3. *For any $f \in \tilde{\Gamma}(\mathbb{R})$, $\overline{co}G = \operatorname{epi} f^*$, where $\overline{co}G$ denotes the closed convex hull of G.*

Proof. We invoke [21, Corollary 25.12] which states that G is the set of exposed points of $\operatorname{epi} f^*$. Since the lemma above ensures $G \neq \emptyset$, taking the closed convex hull gives the result. □

Hence the graph of the conjugate is described by the set G up to some affine parts.

To better understand the parameterization, we now recall the relationship between the conjugate and the closed perspective function $\tilde{f}$ (also called epi-multiplication, see [22, Section 1.H] or dilation) defined by [13, Section 2.2 and Example IV.3.2.4]:

$$\tilde{f}(x,t) = \begin{cases} tf(\frac{x}{t}) & \text{for } t > 0, \\ f'_\infty(x) & \text{for } t = 0, \\ +\infty & \text{otherwise;} \end{cases} \tag{1}$$

where f'_∞ denotes the asymptotic function, or recession function [13, Section IV.3.2] (also called the horizon function [22, Section 3.C]) of f, and is defined by:

$$f'_\infty(d) = \sup_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} = \lim_{t\to+\infty} \frac{f(x_0 + td) - f(x_0)}{t},$$

where x_0 is any point in $\operatorname{dom} f$.

Since $\tilde{f}$ is a positively homogeneous lower-semicontinuous (lsc) convex function, it is the support function of a closed convex set in $\mathbb{R}^2$. We note $\partial f(x)$ the (convex) subdifferential of a function f at a point $x \in \mathbb{R}$:

$$\partial f(x) = \{s \in \mathbb{R} \mid \forall y \in \mathbb{R},\ f(y) \geq f(x) + \langle s, y - x\rangle\}.$$

For simplicity we state the next proposition for functions defined on $\mathbb{R}$, although it is valid on $\mathbb{R}^d$.

Proposition 2.4. *Assume f is a proper lsc convex function on $\mathbb{R}$. Then* (*i*) *the closed perspective function* $\tilde{f}$ *is the support function of the set*

$$C_f = \{(x, -r) | (x, r) \in \operatorname{epi} f^*\}$$

the symmetric of epi f^* *with respect to the* x*-axis.*

(*ii*) *the set* C_f *is the subdifferential of* $\tilde{f}$ *at the origin and*

$$C_f = \partial \tilde{f}(0,0) = \overline{co} \overrightarrow{\nabla} \tilde{f}(0,0) + N_{\operatorname{dom} \tilde{f}}(0,0)$$

where $N_{\operatorname{dom} \tilde{f}}(0,0)$ *is the normal cone at the origin to* dom $\tilde{f}$*, the domain of the function* $\tilde{f}$*, and* $\overline{co} \overrightarrow{\nabla} \tilde{f}(0,0)$ *is the closed convex hull of the upper limit set, i.e. it is the set of all the limits of sequences* $\nabla \tilde{f}(x_n, t_n)$*, where* $\tilde{f}$ *is differentiable at* (x_n, t_n) *and* $(x_n, t_n) \to (0,0)$*.*

Proof. (*i*) See [21, Corollary 13.5.1].

(*ii*) Since $\tilde{f}$ is the support function of C_f, we have $C_f = \partial \tilde{f}(0,0)$. The remaining part is [21, Theorem 25.6]. □

While $\overline{\text{co}} \overrightarrow{\nabla} \tilde{f}(0,0)$ is difficult to compute as a general rule, in our case it is constant along the half-lines (x, t) with $t > 0$ (since $\tilde{f}$ is positively homogeneous). We deduce

$$\nabla \tilde{f}(x,t) = (f'(\frac{x}{t}), f(\frac{x}{t}) - \frac{x}{t} f'(\frac{x}{t}))$$

as soon as $\tilde{f}$ is differentiable at (x, t) with $t > 0$. Consequently, a parameterization of the border of C_f, except for some affine parts, is given by

$$\begin{cases} x = f'(s), \\ y = f(s) - s f'(s). \end{cases} \tag{2}$$

For functions of one variable, the subdifferential of f, a lsc proper convex function, is either empty, a single point, or a segment where the endpoints are the left- and right-derivative (see Section 4 below). So handling nondifferentiable points is not too difficult and the parameterization can be fully captured (at differentiable and nondifferentiable points) by the formula:

$$\begin{cases} s \in \partial f(x), \\ f^*(s) = sx - f(x). \end{cases}$$

3. The Parametric Legendre Transform algorithm

Assume $f \in \tilde{\Gamma}(\mathbb{R})$. Given $a \in \operatorname{dom} f$, a simple binary search allows the computation of dom f, so we can always assume our input consists of $a, b \in \operatorname{dom} f$, $a < b$, and a black box returning f, and f' (wherever it is defined).

The PLT algorithm returns a set S of slopes in dom f^* and the values of f^* on S. It can be stated as follows.

(i) Build an uniform grid $x_i = a + i(b-a)/(n-1)$ between a and b.

(ii) Compute $s_i = f'(x_i)$ and $f^*(s_i) = x_i s_i - f(x_i)$. Remove duplicate values to obtain a (non-uniform) grid s_i. Note that since f is convex, $(s_i)_i$ is a nondecreasing sequence.
(iii) Interpolate as needed between $[s_i, s_{i+1}]$ (See Section 4).
(iv) Extrapolate outside $[s_1, s_m]$.

The extrapolation can be performed either by

- assuming that $\operatorname{dom} f$ is bounded; in other words, we only consider our input to be valid on a compact set, which amounts to replacing the function f with $f + I_{[a,b]}$ *i.e.* f is $+\infty$ outside $[a, b]$. The resulting computation returns $f^* \Box \sigma_{[a,b]}$: the inf-convolution of f^* with the support function of $[a, b]$. Then we use [12] to obtain a very strong convergence result when $a \to +\infty$ and $b \to +\infty$. This is the approach used by fast algorithms [16, 17].
- computing $f'_\infty(-1)$ (resp. $f'_\infty(1)$). If it is finite, extrapolate by $+\infty$ at slopes $s \in (-\infty, f'_\infty(-1)] \cup [f'_\infty(1), +\infty)$, otherwise perform a linear extrapolation. We can use the estimates

$$f'_\infty(-1) \simeq \frac{f(a-t) - f(a)}{t} \quad \text{and} \quad f'_\infty(1) \simeq \frac{f(b+t) - f(b)}{t},$$

  for a large value of $t > 0$. Then we approximate by assuming that the supremum is attained at the boundary of $\operatorname{dom} f$ *i.e.* at b when $s > f'_\infty(1)$, and at a when $s < f'_\infty(-1)$.

4. Approximation error and convergence results

We will say that a numerical approximation of f^* at a slope s is *exact* if it equals $f^*(s)$ up to the usual errors occurring when using floating point operations. These errors are much smaller than the errors due to sampling and extrapolation, so we will ignore them.

The PLT algorithm returns a set of slopes $S = \{s_i | i = 1, \dots n\}$ and (an approximation of) the value of f^* at these slopes $f_n^*(s_i)$. Assume $s \in \mathbb{R}$, and consider the following four possible cases.

(i) The slope s is outside $f'([a, b])$ *i.e.* $s > f'_\infty(1)$ or $s < f'_\infty(-1)$. Then we approximate $f^*(s)$ by performing an extrapolation using one of the two possibilities explained previously.
(ii) The slope s equals s_i for some index i. Since $(s_i, f^*(s_i)) \in G$, the computation is exact.
(iii) The slope s is between two given slopes: $s_i \leq s < s_{i+1}$ for some index i. We consider two subcases:
  (a) $(s, f^*(s)) \in G$. We need to interpolate between the point $(s_i, f^*(s_i))$ with derivative x_i, and $(s_{i+1}, f^*(s_{i+1}))$ with derivative x_{i+1}. There are several possibilities: Using the line segment and disregarding the first-order information, building a polyhedral function using the first-order information, using Hermite interpolation polynomials (but we must ensure the resulting polynomial is convex between s_i and s_{i+1}), or using other shape-preserving interpolation scheme. We choose to build a polyhedral function using the first-order information, *i.e.*

$$f_n^*(s) = \max(x_i(s - s_i) + f^*(s_i), x_{i+1}(s - s_{i+1}) + f^*(s_{i+1})).$$

(b) $(s, f^*(s)) \notin G$. Proposition 2.3 ensures that f^* is affine between s_i and s_{i+1}. So we interpolate linearly to obtain an exact computation.

In practice, we may have a black box to compute f at any point x, but the computation of f' may not be available. However, using finite differences and [21, Theorem 24.1], we can compute

$$f'_+(x) = \lim_{t \to 0+} \frac{f(x+t) - f(x)}{t} \quad \text{and} \quad f'_-(x) = \lim_{t \to 0+} \frac{f(x-t) - f(x)}{t}$$

and detect points $x \in \operatorname{int} \operatorname{dom} f$ where f is not differentiable. They are the points were $\partial f(x) = [f'_-(x), f'_+(x)]$ is not a singleton. So we can easily distinguish between case (iii)(a) and case (iii)(b) above.

The convergence of the PLT algorithm results directly from convergence results in [17] for both the extrapolation (when $a \to -\infty$ and $b \to +\infty$) and the interpolation parts (when $n \to +\infty$ *i.e.* a thinner grid is used).

5. Parametric Moreau Envelope

By expanding the squared norm in the definition of the Moreau envelope, we can relate it to the conjugacy operation

$$M_\lambda(s) = \frac{\|s\|^2}{2\lambda} - \frac{1}{\lambda} g^*_\lambda(s), \text{ with } g_\lambda(y) = \lambda f(y) + \frac{\|y\|^2}{2}. \tag{3}$$

Hence we obtain immediately a parameterization of the Moreau envelope.

Proposition 5.1. *Assume f is a proper lsc convex function on $\mathbb{R}^d$. Then the Moreau envelope can be parameterized by*

$$\begin{cases} z = x + \lambda \nabla f(x), \\ M_\lambda(z) = f(x) + \frac{\lambda}{2}\|\nabla f(x)\|^2, \end{cases}$$

where $x \in \mathbb{R}^d$ is a point where f is differentiable. Missing parts of the graph of M_λ are recovered by piecewise quadratic interpolation (knowing that M_λ is C^1).

Proof. Apply Formula (3) with the conjugate parameterization of Formula (2).

Alternatively, apply [22, Example 10.2] to evaluate the proximal mapping

$$x = P_\lambda(z) = (I + \lambda \partial f)^{-1}(z)$$

(the proximal mapping is the set of points where the infimum in the definition of the Moreau envelope is attained; when f is lsc proper convex, the infimum is always attained). So $z = x + \lambda \partial f(x)$, and when f is differentiable at x we obtain the first part of the parameterization. The second part follows from substituting $x = P_\lambda(z)$ into the definition of $M_\lambda(z)$.

Finally, to recover the full graph of M_λ let us see how to operate at any point x. If f is differentiable at x, the corresponding point on the graph of M_λ is recovered by the parameterization process. Assume f is not differentiable at x. Then g_λ is not differentiable at x either. Hence the point $(s, g_\lambda^*(s))$ with $s = \nabla g_\lambda(x)$ belongs to the boundary of $\operatorname{epi} g_\lambda^*$ but is not an exposed point (by Proposition 2.3 and its proof). So it is a convex combination of exposed points, which means that g_λ^* is affine around $s = \nabla g_\lambda(x)$. Then by Formula (3) M_λ is quadratic around x. So we can recover the graph of M_λ by performing a piecewise quadratic interpolation. □

Example 5.2. Consider the function

$$f(x) = \begin{cases} -x-1 & \text{when } x \leq -1, \\ 0 & \text{when } -1 \leq x \leq 1, \\ x-1 & \text{when } 1 < x. \end{cases}$$

The function f is the convex envelope of $x \mapsto ||x|-1|$. Its Moreau envelope is

$$M_\lambda(x) = \begin{cases} -x-1-\frac{\lambda}{2} & \text{when } x \leq -\lambda-1, \\ \frac{(x+1)^2}{2\lambda} & \text{when } -\lambda-1 \leq x \leq -1, \\ 0 & \text{when } -1 \leq x \leq 1, \\ \frac{(x-1)^2}{2\lambda} & \text{when } 1 \leq x \leq \lambda+1, \\ x-1-\frac{\lambda}{2} & \text{when } 1+\lambda < x. \end{cases}$$

The parameterization formula gives

$$\begin{cases} z = x-\lambda \text{ and } M_\lambda(z) = -x-1+\frac{\lambda}{2} & \text{when } x < -1, \\ z = x \text{ and } M_\lambda(z) = 0 & \text{when } -1 < x < 1, \\ z = x+\lambda \text{ and } M_\lambda(z) = x-1+\frac{\lambda}{2} & \text{when } 1 < x. \end{cases}$$

So we deduce

$$M_\lambda(z) = \begin{cases} -z-\frac{\lambda}{2}-1 & \text{when } z < -\lambda-1, \\ 0 & \text{when } -1 < z < 1, \\ z-\frac{\lambda}{2}-1 & \text{when } \lambda+1 < z. \end{cases}$$

The two remaining segments, $[-\lambda-1, -1]$ and $[1, \lambda+1]$, are recovered by quadratic interpolation.

Like for the conjugate, if one wants to deal with nondifferentiable points of f, there is no need to recover quadratic parts. The complete parameterization formula is then

$$\begin{cases} z \in x + \lambda\partial f(x), \\ M_\lambda(z) = f(x) + \frac{\|z-x\|^2}{2\lambda}. \end{cases}$$

6. Numerical comparisons

While the PLT algorithm has clearly a linear-time complexity, we compare it with similar algorithms [19, 18, 8] for computing the Moreau envelope. All the algorithms have a linear worst-case time complexity. The Parabolic Envelope (PE) algorithm computes the lower envelope of quadratic functions. It relies on the fact that computing the intersection of two quadratic functions can be done in constant time $O(1)$. The Linear-time

Legendre Transform (LLT) algorithm reduces the computation of the Moreau envelope to the computation of the Legendre-Fenchel transform, and then uses the LLT algorithm, which takes advantage of convexity to run in linear time. Finally, the Non-Expansive Prox (NEP) algorithm uses the non-expansiveness of the proximal mapping to reduce its complexity. Note that the proximal mapping is nonexpansive for convex functions (and is Lipschitz for slightly more general functions like prox-regular functions [22]).

n	PE	LLT	NEP	PLT
1,000	0.3	0.43	0.16	0.01
3,000	0.99	1.37	0.54	0.
5,000	1.75	2.41	0.98	0.
7,000	2.59	3.57	1.5	0.01
9,000	3.51	4.84	2.11	0.
11,000	4.51	6.21	2.8	0.
13,000	5.57	7.69	3.55	0.01
15,000	6.73	9.28	4.42	0.01
17,000	7.97	10.99	5.35	0.01
19,000	9.3	12.82	6.33	0.02
21,000	10.67	14.72	7.44	0.02
23,000	12.16	16.72	8.58	0.01
25,000	13.82	18.92	9.87	0.02
27,000	15.35	21.33	11.25	0.01
29,000	17.39	23.72	12.77	0.01

Table 6.1: Numerical comparison of the PLT, NEP, LLT, and PE algorithms for the function $f(x) = x^2/2$ on the interval $[-n/2, n/2]$.

Table 6.1 shows the result. The computation was run on a Pentium IV 2.8 GHz processor using Scilab v3.1.1 under Linux Mandriva 2005 LE.

The PLT algorithm is the simplest to code and takes naturally advantage of the optimized vectorized operations of Scilab to achieve a speed improvement by a factor of more than 400. The typical slopes of the least-squares approximation are shown in Table 6.2. They correspond to the multiplicative constant factor associated with each algorithm (each algorithm runs in $Cn + o(n)$ and Table 6.2 shows an approximation of the constant C).

While the NEP and PLT algorithms are restricted to convex functions, the PLT algorithm additionally only computes a parameterization of the graph of the Moreau envelope (instead of an evaluation of the Moreau envelope on a given grid). As such, it relies on more restrictive assumptions to achieve a much faster computation.

If the function is convex but a parameterization is not suitable, the NEP algorithm is

Algorithm	Least-squares slope
PLT	0.000000
NEP	0.000447
LLT	0.000603
PE	0.000835

Table 6.2: Slopes of the least-squares approximation for the data on Table 6.1.

the fastest. When the function is not convex, the PE and LLT algorithms are applicable. While they both share a linear-time complexity, the asymptotic multiplicative constant in the $O(n)$ notation depends mostly on the convexity of the input data.

7. Conclusion

We have presented and proven the correctness of a new algorithm to compute the Legendre-Fenchel conjugate and the Moreau envelope. The algorithm is the simplest to implement and naturally takes advantage of optimized vector operations occurring in Scilab to achieve a more than 400 fold speed improvement. However, the algorithm does not output the value of the transform on a grid but only a parameterization. It is also restricted to convex data.

The PLT algorithm could be extended to nonconvex functions by first computing the convex envelope as the LLT algorithm does, and then computing the parameterization of the conjugate of the piecewise linear function going through the vertices of the convex envelope. However, the computation would not be significantly faster than the LLT algorithm, since most of the computation cost occurs in computing the vertices of the convex hull.

Acknowledgements. The algorithms were implemented on Scilab v3.1.1. Some examples were investigated using the Computational Convex Analysis Toolkit (SCAT) [10].

References

[1] D. G. Bailey: An efficient Euclidean distance transform, in: Combinatorial Image Analysis (Auckland, 2004), R. Klette, J. Žunić (eds.), Lecture Notes in Computer Science 3322, Springer, Berlin (2004) 394–408.

[2] H. Bauschke, M. von Mohrenschildt: Symbolic computation of Fenchel conjugates, ACM SIGSAM Bulletin, to appear.

[3] H. H. Bauschke, M. von Mohrenschildt: Fenchel conjugates and subdifferentials in Maple, Tech. Rep. CORR 97-23, Department of Combinatorics and Optimization, University of Waterloo, Waterloo (1997).

[4] J. Benoist, J.-B. Hiriart-Urruty: What is the subdifferential of the closed convex hull of a function?, SIAM J. Math. Anal. 27 (1996) 1661–1679.

[5] J. Borwein, A. S. Lewis: Convex Analysis and Nonlinear Optimization. Theory and Examples, CMS Books in Mathematics 3, Springer, New York (2000).

[6] L. Corrias: Fast Legendre-Fenchel transform and applications to Hamilton Jacobi equations and conservation laws, SIAM J. Numer. Anal. 33 (1996) 1534–1558.

[7] O. Cuisenaire, B. Macq: Fast and exact signed Euclidean distance transformation with linear complexity, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (Phoenix, 1999), IEEE Computer Society Press (1999) 3293–3296.

[8] P. F. Felzenszwalb, D. P. Huttenlocher: Distance transforms of sampled functions, Tech. Rep. TR2004-1963, Cornell Computing and Information Science, Cornell University, Cornell (2004).

[9] M. Gavrilova, M. H. Alsuwaiyel: Two algorithms for computing the euclidean distance transform, Tech. Rep. 2000-661-13, Computer Science Technical Reports, University of Calgary, Calgary (2000).

[10] C. H. Hamilton: Symbolic Convex Analysis, M.Sc. Thesis, Department of Mathematics, Simon Fraser University, Burnaby (2005).

[11] W. Hesselink: A linear-time algorithm for euclidean feature transform sets, available at http://www.cs.rug.nl/∼wim/pub/mans.html (2005).

[12] J.-B. Hiriart-Urruty: Lipschitz r-continuity of the approximate subdifferential of a convex function, Math. Scand. 47 (1980) 123–134.

[13] J.-B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms, I: Fundamentals; II: Advanced Theory and Bundle Methods, Comprehensive Studies in Mathematics 305, 306, Springer, Berlin (1993).

[14] J. Toriwaki, K. Mori: Distance transformation and skeletonization of 3d pictures and their applications to medical images, in: Digital and Image Geometry: Advanced Lectures, G. Bertrand et al. (ed.), Lecture Notes in Computer Science 2243, Springer, Berlin (2001) 412–429.

[15] C. R. Maurer Jr., R. Qi, V. V. Raghavan: A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 265–270.

[16] Y. Lucet: A fast computational algorithm for the Legendre-Fenchel transform, Comput. Optim. Appl. 6 (1996) 27–57.

[17] Y. Lucet: Faster than the fast Legendre transform, the linear-time Legendre transform, Numer. Algorithms 16 (1997) 171–185.

[18] Y. Lucet: Fast Moreau envelope computation I: Numerical algorithms, Tech. Rep., University of British Columbia Okanagan, Kelowna (2005).

[19] Y. Lucet: A linear euclidean distance transform algorithm based on the linear-time Legendre transform, in: Proc. Second Canadian Conference on Computer and Robot Vision (Victoria, 2005), IEEE Computer Society Press (2005) 262–267.

[20] S. Mauch: A fast algorithm for computing the closest point and distance transform, Tech. Rep. 077, ASC Center for Simulation of Dynamic Response of Materials, California Institute of Technology, Pasadena (2000).

[21] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton, New York (1970).

[22] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Springer, Berlin (1998).

[23] C. Sigg, R. Peikert, M. Gross: Signed distance transform using graphics hardware, in: Proceedings of IEEE Visualization (Zürich, 2003), IEEE Computer Society Press (2003) 83–90.

A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform

Yves Lucet,
Computer Science Department,
University of British Columbia - Okanagan Campus
Okanagan University College,
3333 University Way,
Kelowna, BC V1V 1V7, Canada
ylucet@ouc.bc.ca

Abstract

We introduce a new exact Euclidean distance transform algorithm for binary images based on the Linear-time Legendre Transform algorithm. The three-step algorithm uses dimension reduction and convex analysis results on the Legendre–Fenchel transform to achieve linear-time complexity. First, computation on a grid (the image) is reduced to computation on a line, then the convex envelope is computed, and finally the squared Euclidean distance transform is obtained. Examples and an extension to non-binary images are provided.

1. Introduction

Distance transforms have been investigated for decades [25] due to their diverse applications (see for example [5, 15] and references therein). Various algorithms were introduced to compute the Euclidean Distance Transform (EDT), and recent research focuses on simplifying the algorithms while still achieving linear-time complexity. Various properties were used to achieve linear-time. For example, the fact that the Euclidean distance transform computation is equivalent to computing the lower envelope of quadratic functions was exploited in [6, 8] to achieve a simple linear-time algorithm. Other algorithms based on monotonicity or neighborhood properties also managed to achieve linear complexity [11, 27].

Independently of the distance transform computation, the numerical computation of the Legendre–Fenchel Transform (LFT) was studied as the solution to some Hamilton-Jacobi differential equations. First a $\log$-linear algorithm named the Fast Legendre Transform (FLT for short, by analogy with the Fast Fourier Transform) was introduced [3, 4, 17, 22, 26] to be subsequently improved by a linear-time algorithm: The Linear-time Legendre transform (LLT) [18]. The FLT and the faster LLT algorithms have been used in efficient numerical simulation of the Burger's equation [1, 2, 9, 10, 12, 13, 21, 28]. The LLT has also found applications in robotics [16], and network communication [14].

Computing distance transforms reduces to computing the LFT. Henceforth a strong connection exists between convex analysis, image processing, differential calculus and dynamical systems. The resulting area was named differential morphology [19, 20].

In the present paper, we apply the Linear-time Legendre Transform (LLT) algorithm [18] to the computation of the Euclidean distance transform for binary images, and obtain a new linear-time algorithm to compute the exact Euclidean distance transform for binary images. We then show how the algorithm readily extends to non-binary images in any dimension.

The paper is organized as follows. Section 2 explains the relation between the Euclidean distance transform and the (Legendre) conjugate. Section 3 details the dimension reduction step, and Section 4 explains how computing the (lower) convex envelope can be done while still achieving linear complexity. The core idea of the LLT is explained in Section 5: Using convex analysis results we show that computing the conjugate is equivalent to merging two sorted lists. We state the algorithm in Section 6. Then Section 7 provides examples. Finally, Section 8 explains how to compute distance transforms for non-binary images as suggested in [8] and we conclude the paper with Section 9.

2. From EDT to LFT

We note as p (resp. q) a pixel with coordinate (i, j) (resp. (k, l)). An $n \times n$ binary image is modeled as an applica-

Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV'05)
0-7695-2319-6/05 $ 20.00 **IEEE**

tion $p \mapsto f(p) \in \{0,1\}$ with $1 \le i,j \le n$. A pixel p with $f(p)=1$ is called a feature. The *distance transform* DT associates to each pixel the distance to its nearest feature. The *feature transform* associates to each pixel its nearest feature. For the Euclidean distance, the squared distance transform is

$$DT^2(p) = \min_q\{\|p-q\|^2; f(q)=1\}.$$

Defining the indicator function I as $I(p)=0$ if $f(p)=1$, otherwise $I(p)=+\infty$, and expanding the quadratic form gives

$$DT^2(p) = \|p\|^2 - 2\max_q\{<p,q> - (\frac{1}{2}\|q\|^2 + I(q))\}.$$

In convex analysis the conjugate of a function h is defined as $h^*(s) := \sup_x <s,x> - h(x)$, and the mapping $h \mapsto h^*$ is called the Legendre–Fenchel transform or LFT for short (we refer to [24] for convex analysis results). So by defining $g(q) = \frac{1}{2}\|q\|^2 + I(q)$ we obtain

$$DT^2(p) = \|p\|^2 - g^*(p). \qquad (1)$$

Various authors [12, 22, 26] noted that the computation of the Moreau envelope $\min_y\{f(y) + \frac{1}{2}||x-y||^2\}$ of a function f is equivalent to the computation of a Legendre conjugate by expanding the quadratic term as above. Hence computing the distance transform DT is equivalent to computing the conjugate g^* for all pixels p of the image.

While computing the conjugate g^* at one point p has a linear worst-case time complexity, computing g^* at many points can still be achieved in linear-time by using convexity. The next sections detail the three steps of the Linear-time Legendre Transform algorithm (LLT).

3. From grid to line

First the computation of the conjugate g^* for all pixels is reduced to the computation of a conjugate on a line. Indeed,

$$\begin{aligned} g^*(i,j) &= \max_k[ik + \max_l[jl - g(k,l)]] \\ &= \max_k[ik - h_j(k)] \end{aligned}$$

with $h_j(k) = -\max_l[jl - g(k,l)]$. So computing g^* for all pixels (i,j) amounts to

1. For all k, compute $h_j(k)$ at all j. In other words, we compute n conjugates at points j on a line.
2. For all j, compute $g^*(i,j) = h_j^*(i)$ at all i.

Both steps only involve the repeated computation of the conjugate for functions of one variable. The process can be repeated so that computing the conjugate for a function in dimension d amounts to computing several conjugates for functions of one variable.

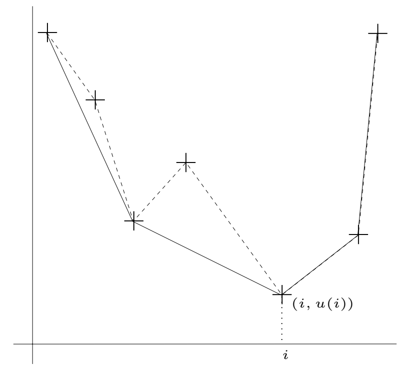


Figure 1. The function u and its lower convex envelope. The dotted line goes through all the points $(i, u(i))$ while the full line goes only through the vertices of the lower convex envelope.

4. Convexity

Based on the previous section, we focus on computing the conjugate $u^*(k) = \max_i[ki - u(i)]$ for a function u defined at each index i. It is a well-known fact in convex analysis [24, Theorem 11.1] that the conjugate only depends on the convex envelope *i.e.* the conjugate of u is the same as the conjugate of its lower convex envelope. The result still holds true in our discrete setting [4]. So provided we can do it in linear time, we can compute the lower convex envelope of the set of points $(i, u(i))$ *i.e.* we can ignore any point $(i, u(i))$ which is not on the convex envelope of u. Figure 1 shows a function u and its lower convex envelope. The next section will show that the resulting convexity greatly simplifies the computation of the conjugate u^*.

In fact, computing the convex envelope of a set of points in the plane with sorted x-coordinates can be done in linear-time using either the Beneath-Beyond algorithm or the Divide-and-Conquer algorithm [7, 23].

Moreover, considering the special form of the function g, there is no need to compute its convex hull. Indeed, name co1 the $n \times n$ matrix with values 1 at index (i,j) when (i,j) is a vertex of the lower convex hull (computed row by row) of g, and 0 otherwise. Then co1 $= 1-$IMG[1]. Note, however, that the computation of the convex hull is necessary for computing the full conjugate of g from its partial conjugate (see Section 6 for the full algorithm).

The present step reduces the number of points from n to the number of vertices of the lower convex envelope of u. While it does not reduce the overall worst-case complex-

1 if $(f(q)=1$, $g(q)=\infty$ so obviously q is not a vertex and co1$(q)=0$; when $f(q)=0$ prove by contradiction using the strict convexity of the square Euclidean norm.

ity, it can speed up computation significantly in practise for highly nonconvex functions. However, its main advantage is to simplify the computation in the following step.

5. The Merge step

According to the above, computing the conjugate reduces to computing the conjugate of a convex set $(i, u(i))$ for $1 \le i \le n'$ with n' the number of vertices on the lower convex envelope of $(i, u(i))$ for $1 \le i \le n$. We define the slopes

$$c_i := \frac{u(i+1) - u(i)}{(i+1) - i}$$

for $1 \le i \le n' - 1$, and recall [18, Lemma 2]:

- If $c_{i-1} < i < c_i$, the maximum in computing the conjugate $u^*(k) = \max_j [kj - u(j)]$ is attained uniquely for $j = i$.
- If $c_i = i$, the maximum is attained for both $j = i$ and $j = i + 1$.

Since the set $(i, u(i))_i$ is convex, the sequence $(c_i)_i$ is increasing. So computing the conjugate amounts to merging the sequences c_i with the sequence i. Both sequences are increasing so the merge can be done in linear-time.

Figure 2 gives a graphical interpretation of the merge.

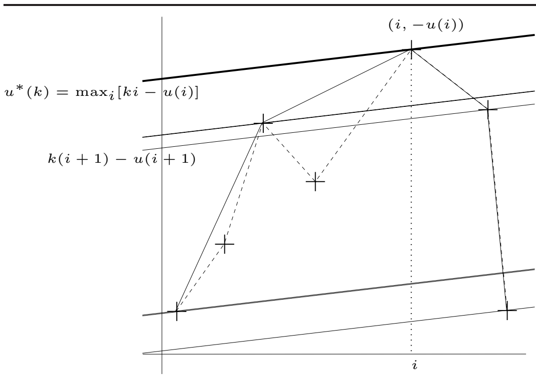


Figure 2. Graphical interpretation of the conjugate. The maximum is attained at the point $(i, -u(i))$ where the line with slope k touches the envelope.

6. The algorithm

Putting all steps together gives the following pseudocode for the algorithm (the notation $g(i_1, :)$ means the i_1th row of matrix g, and similarly Conj$(:, j_2)$ refers to the j_2th column of matrix Conj).

1. Compute the matrix g associated with the function g defined from the binary image IMG:

$$g(i,j) = \begin{cases} \frac{i^2+j^2}{2} & \text{if IMG}(i,j) = 0, \\ \infty & \text{otherwise.} \end{cases}$$

2. for each row i_1, compute the one dimensional conjugate Conjpartial$(i_1, :) = -$dlt$(g(i_1, :))$
3. for each column j_2, compute the one dimensional conjugate Conj$(:, j_2) =$ dlt(Conjpartial$(:, j_2)$)
4. Compute the square distance transform DT$^2(i, j) = i^2 + j^2 - 2$ Conj(i, j).

The function dlt is the one-dimensional transform that takes a vector Y and returns its conjugate Conj. Here is its code:

1. Compute the lower convex hull of $(i, Y(i))$
2. Compute $C(i) = Y(i+1) - Y(i)$
3. Merge C and S to obtain Conj as follows[2]:
   for $j = 1$ to m {
     while ($j > C(i)$) { $i = i + 1$ }
     Conj$(i) = ji - Y(i)$
   }

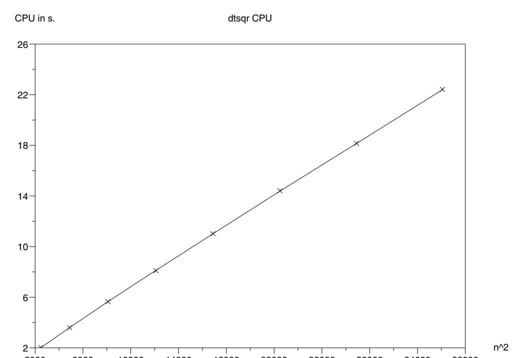


Figure 3. Linear computation time of the LLT algorithm

Figure 3 verifies experimentally that the complexity is linear with respect to the size of the image.

Note that with obvious modifications, the algorithm extends to general rectangular grids.

2 We assume $m \le C(n)$, otherwise adjust accordingly.

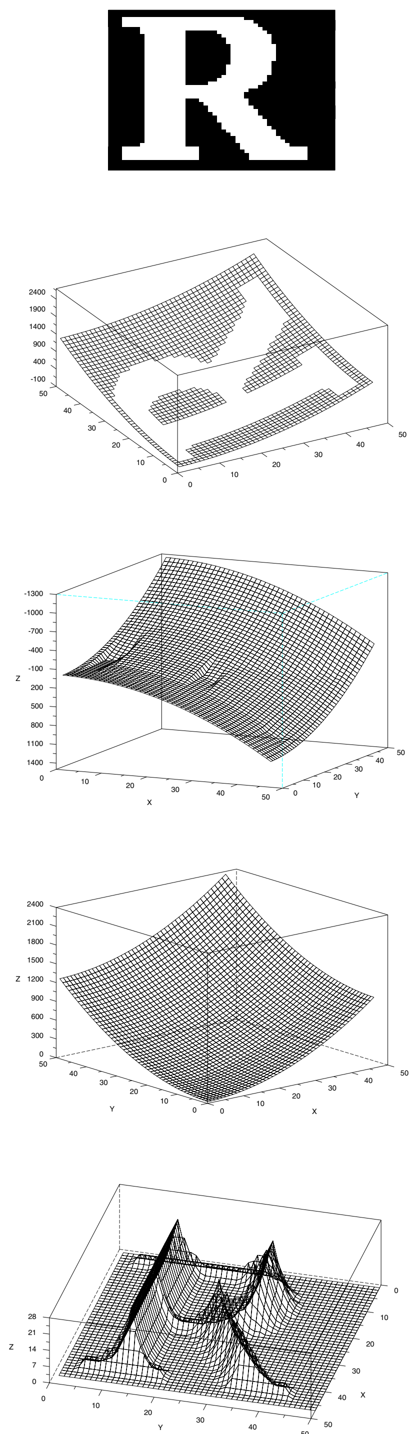


Figure 4. 47×50 **image of the letter R, its associated function** g**, its partial conjugate, its conjugate, and its squared Euclidean distance transform**

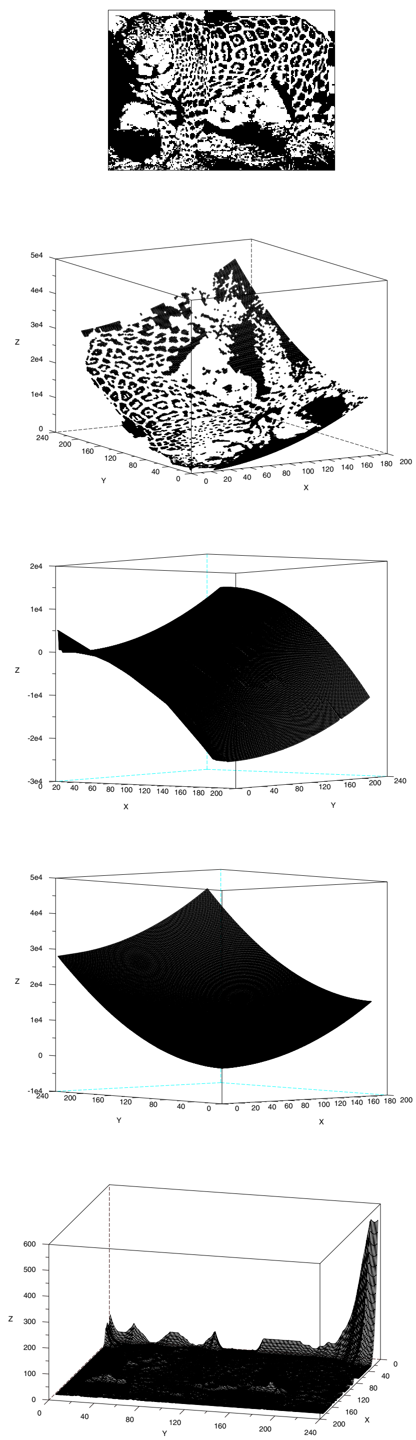


Figure 5. 182×238 **binary image, its associated function** g**, its partial conjugate, its conjugate, and its squared Euclidean distance transform**

7. Examples

To illustrate the algorithm, consider the simple binary image

$$\text{IMG} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

its associated function g is

$$g = \begin{bmatrix} 1. & 2.5 & 5. & 8.5 & 13. \\ 2.5 & \infty & \infty & \infty & 14.5 \\ 5. & \infty & \infty & \infty & 17. \\ 8.5 & \infty & \infty & \infty & 20.5 \\ 13. & 14.5 & 17. & 20.5 & 25. \end{bmatrix}.$$

Computing the conjugate for each row gives the partial conjugate Conjpartial:

$$\begin{bmatrix} 0. & -1.5 & -4. & -7.5 & -12. \\ 1.5 & 0.5 & -0.5 & -5.5 & -10.5 \\ 4. & 3. & 2. & -3. & -8. \\ 7.5 & 6.5 & 5.5 & 0.5 & -4.5 \\ 12. & 10.5 & 8. & 4.5 & 0. \end{bmatrix}.$$

Then computing the conjugate for each column, we obtain the conjugate of g:

$$\text{Conj} = \begin{bmatrix} 1. & 2.5 & 5. & 8.5 & 13. \\ 2.5 & 3.5 & 6. & 9.5 & 14.5 \\ 5. & 6. & 7. & 12. & 17. \\ 8.5 & 9.5 & 12. & 15.5 & 20.5 \\ 13. & 14.5 & 17. & 20.5 & 25. \end{bmatrix},$$

from which we deduce the square Euclidean distance transform

$$\text{DT}^2 = \begin{bmatrix} 0. & 0. & 0. & 0. & 0. \\ 0. & 1. & 1. & 1. & 0. \\ 0. & 1. & 4. & 1. & 0. \\ 0. & 1. & 1. & 1. & 0. \\ 0. & 0. & 0. & 0. & 0. \end{bmatrix}.$$

Figure 4 illustrates the algorithm on a simple binary image representing the letter R. A more complex example is illustrated with Figure 5.

8. Extension to non-binary images

The LLT algorithm can be applied to more general images. As in [8], we consider the more general case of computing the squared Euclidean distance transform of a function f, also known as the Moreau envelope of f:

$$D^2(p) = \min_q [\|p - q\|^2 + f(q)],$$

where f is not necessarily equal to g, and q moves on a $n \times m$ grid not necessarily on $\{1, \ldots, n\}^2$. The LLT algorithm still fully applies with obvious adjustments. Figure 6 illustrates such a computation on a grayscale image. Note that contrary to binary images, for a general function f we can no longer skip the computation of the convex hull during the iteration on rows.

The LLT algorithm also applies in any dimension with the same linear-time complexity with respect to the size of the grid.

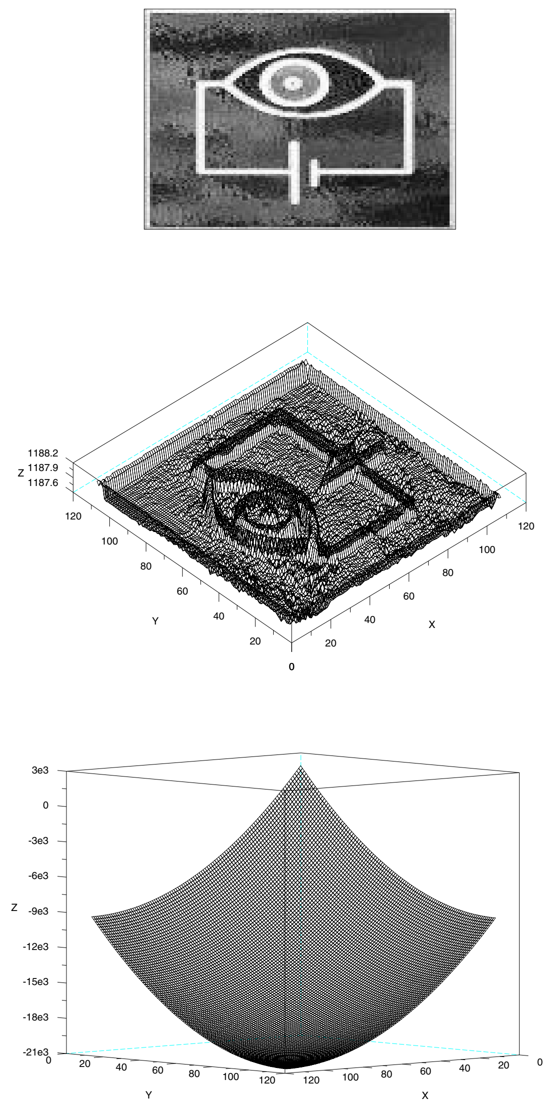


Figure 6. 108×108 **grayscale image, its associated function** f**, and its squared Euclidean distance transform**

9. Conclusions

We have presented a new exact Euclidean distance transform algorithm for binary images with a linear-time complexity. The algorithm reduces to simple calculations on a line: Convex hull and merging two sorted sequences. It also scales to higher dimensions and extends to more general distance transforms. As such, it is an interesting alternative to [8].

Acknowledgements: Part of this work was funded through the Okanagan University College Grant-In-Aid program.

All computations were done on the mathematical software Scilab using the SIP (Scilab Image Processing) package on a Pentium IV 1.8 GHz computer. The implementation of the LLT algorithm presented in the present paper is available through the GPL license.

References

[1] E. Aurell, S.-g. N. Gurbatov, and S.-g. I. Simdyankin. Numerical proof of self-similarity in burgers' turbulence. Technical Report http://arxiv.org/abs/patt-sol/9602005, arXiv.org, 1996.

[2] J. Bec, Frisch, U., and K. Khanin. Kicked burger turbulence. *J. Fluid Mech.*, 416:239–267, 2000.

[3] Y. Brenier. Un algorithme rapide pour le calcul de transformée de Legendre–Fenchel discrètes. *C. R. Acad. Sci. Paris Sér. I Math.*, 308:587–589, 1989.

[4] L. Corrias. Fast Legendre–Fenchel transform and applications to Hamilton–Jacobi equations and conservation laws. *SIAM J. Numer. Anal.*, 33(4):1534–1558, Aug. 1996.

[5] O. Cuisenaire. *Distance Transformations: Fast Algorithms and Applications to Medical Image Processing*. PhD thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, Oct. 1999.

[6] L. Deniau and J. Blanc-Talon. Fractal analysis with hausdorff distance under affine transformations. Technical report, ETCA-CREA-SP, 1995.

[7] H. Edelsbrunner. *Algorithms in Combinatorial Geometry*. EATC Monographs on Theoretical Computer Science. Springer-Verlag, 1987.

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Distance transforms of sampled functions. Technical Report TR2004-1963, Cornell Computing and Information Science, Sept. 2004.

[9] U. Frisch and J. Bec. Burgulence. In A. M. Lesieur and e. F. David, editors, *Les Houches 2000: New Trends in Turbulence*, pages 341–383. Springer EDP-Sciences, 2001.

[10] U. Frisch, J. Bec, and B. Villone. Singularities and the distribution of density in the burgers/adhesion model. *Physica D*, 152-153:620–635, 2001.

[11] M. Gavrilova and M. H. Alsuwaiyel. Two algorithms for computing the euclidean distance transform. Technical Report 2000-661-13, Computer Science Technical Reports, University of Calgary, 2000.

[12] S. N. Gurbatov, S. I. Simdyankin, E. Aurell, U. Frisch, and G. Tth. On the decay of burgers turbulence. *J. Fluid Mech.*, 344:339–374, 1997.

[13] S. N. Gurbatov and A. V. Troussov. The decay of multiscale signals - deterministic model of the burgers turbulence. Technical report, arxiv.org, 2000.

[14] T. Hisakado, K. Okumura, V. Vukadinovic, and L. Trajkovic. Characterization of a simple communication network using legendre transform. In *Proc. IEEE Int. Symp. Circuits and Systems*, volume 3, pages 738–741, May 2003.

[15] J. ichiro Toriwaki and K. Mori. Distance transformation and skeletonization of 3d pictures and their applications to medical images. In R. K. E. G. Bertrand, A. Imiya, editor, *Digital and Image Geometry: Advanced Lectures*, volume 2243 / 2001 of *Lecture Notes in Computer Science*, pages 412–429. Springer-Verlag Heidelberg, 2001.

[16] B. Koopen. Contact of bodies in 2D-space: Implementing the discrete legendre transform. AI master's thesis, Intelligent Autonomous Systems Group, University of Amsterdam, Feb. 2002.

[17] Y. Lucet. A fast computational algorithm for the Legendre–Fenchel transform. *Computational Optimization and Applications*, 6(1):27–57, July 1996.

[18] Y. Lucet. Faster than the fast Legendre transform, the linear-time Legendre transform. *Numer. Algorithms*, 16(2):171–185, Jan. 1997.

[19] P. Maragos. Differential morphology and image processing. *IEEE Transactions on Image Processing*, 5(6):922–927, June 1996.

[20] P. Maragos. Differential morphology. In S. K. Mitra and G. L. Sicuranza, editors, *Nonlinear Image Processing*, Academic Press Series in Communications, Networking, and Multimedia, chapter 10, pages 289–329. Academic Press, 2001.

[21] A. Noullez, S. N. Gurbatov, E. Aurell, and S. I. Simdyankin. The global picture of self-similar and not self-similar decay in burgers turbulence. Technical Report nlin.CD/0409022, arXiv.org eprint archive, Sept. 2004.

[22] A. Noullez and M. Vergassola. A fast Legendre transform algorithm and applications to the adhesion model. *Journal of Scientific Computing*, 9(3):259–281, 1994.

[23] F. P. Preparata and M. I. Shamos. *Computational Geometry*. Texts and monographs in computer science. Springer-Verlag, third edition, 1990.

[24] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Springer-Verlag, Berlin, 1998.

[25] A. Rosenfeld and J. L. Pfaltz. Sequential operations in digital picture processing. *J. ACM*, 13(4):471–494, 1966.

[26] Z.-S. She, E. Aurell, and U. Frisch. The inviscid burgers equation with initial data of brownian type. *Comm. Math. Phys.*, 148(3):623–641, 1992.

[27] F. Y. Shih and Y.-T. Wu. Fast euclidean distance transformation in two scans using a 3 x 3 neighborhood. *Comput. Vis. Image Underst.*, 93(2):195–205, 2004.

[28] M. Vergassola, B. Dubrulle, U. Frisch, and N. A. Burger's equation, devil's staircases and the mass distribution function for large-scale structures. *Astronomy and Astrophysics*, 289:325–356, 1994.

Numer Algor (2006) 43:235–249
DOI 10.1007/s11075-006-9056-0

ORIGINAL PAPER

Fast Moreau envelope computation I: numerical algorithms

Yves Lucet

Received: 29 September 2005 / Revised: 29 November 2006 /
Accepted: 29 November 2006 / Published online: 5 January 2007

Abstract The present article summarizes the state of the art algorithms to compute the discrete Moreau envelope, and presents a new linear-time algorithm, named NEP for NonExpansive Proximal mapping. Numerical comparisons between the NEP and two existing algorithms: The Linear-time Legendre Transform (LLT) and the Parabolic Envelope (PE) algorithms are performed. Worst-case time complexity, convergence results, and examples are included. The fast Moreau envelope algorithms first factor the Moreau envelope as several one-dimensional transforms and then reduce the brute force quadratic worst-case time complexity to linear time by using either the equivalence with Fast Legendre Transform algorithms, the computation of a lower envelope of parabolas, or, in the convex case, the non expansiveness of the proximal mapping.

Keywords Moreau–Yosida approximate · Legendre–Fenchel transform · conjugate · discrete Legendre transform · computational convex analysis

Mathematics Subject Classifications (2000) 52B55 · 65D99

1 Introduction

The Moreau envelope [71] of an extended real-valued function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$, also called the Moreau–Yosida approximate, Yosida Approximate [1] or Moreau–Yosida regularization

$$M_\lambda(s) := \inf_{x\in\mathbb{R}^d} \left[\frac{\|s-x\|^2}{2\lambda} + f(x) \right] \tag{1}$$

Y. Lucet (✉)
Computer Science, I. K. Barber School of Arts and Sciences-Unit 4, The University of British Columbia Okanagan, 3333 University Way, Kelowna, BC V1V 1V7, Canada
e-mail: yves.lucet@ubc.ca

has been studied extensively both theoretically and algorithmically for its regularization properties. Its origin goes back to the work of Yosida [77] on maximal monotone operators, and its behaviour is well known in the field of convex analysis [58–60, 69] and variational analysis [71, Chapter 12]. Under general conditions, M_λ is C^1 with Lipschitz continuous gradient, and critical points of f are fixed points of the proximal mapping

$$P_\lambda(s) := \operatorname*{Argmin}_{x \in \mathbb{R}^d} \left[\frac{\|s-x\|^2}{2\lambda} + f(x) \right]. \tag{2}$$

When f is convex lower semi-continuous and proper, the proximal mapping is a maximal monotone operator and its fixed points are the minima of f. More precise smoothness results are known under various hypothesis on f [12, 24, 45, 46, 48, 67] and recent developments focus on the notion of prox-regularity [4–6, 44, 65, 66].

The Moreau envelope is an attractive regularization transform considering that M_λ converges pointwise to $f(x)$ when $\lambda \searrow 0$, and that it shares the same critical points of f. On the practical side, the proximal point algorithm exploits the fixed point property of the proximal mapping to converge to a minimum of f [70]. Its convergence properties are well known [21, 34], and variants have been introduced to speed it up [7, 9–11, 13, 33, 47, 57, 61, 62, 68, 73, 76]. Extensions to non-quadratic kernels like entropy methods and Bregman distances have also been studied [17, 30, 74].

In addition to proximal point algorithm variants and extensions, bundle methods are another family of numerical optimization algorithms that is intrinsically linked to the Moreau envelope (see [47], and [28, Chapter XV]). Recent developments in that direction focus on $\mathcal{VU}$-decomposition [35–39, 49–56].

While the present article is concerned with the numerical computation of the Moreau envelope, contrary to [25] we do not consider computing its value at one point but instead we tackle the problem of computing the Moreau envelope on a grid.

Similar algorithms have been developed to compute the Legendre conjugate (also named Legendre–Fenchel conjugate, and Legendre–Fenchel Transform)

$$f^*(s) = \sup_{x \in \mathbb{R}^n} [\langle s, x\rangle - f(x)] \tag{3}$$

motivated by the study of some Hamilton–Jacobi differential equations. A log-linear algorithm named the Fast Legendre Transform (FLT for short, by analogy with the Fast Fourier Transform) was first introduced [8, 14, 40, 64, 72] to be subsequently improved by a linear-time algorithm: The Linear-time Legendre transform (LLT) [41]. The FLT and the faster LLT algorithms have been used in efficient numerical simulations of the Burger's equation [2, 3, 19, 20, 22, 23, 63, 75]. The LLT has also found applications in robotics [31], network communication [29], pattern recognition [43], numerical simulation of multiphasic flows [26], and analysis of the distribution of chemical compounds in the atmosphere [32].

Applications spanning a wide range of fields (Image processing, robot navigation, partial differential equations, etc.) are forthcoming in a companion paper [42].

The paper is organized as follow. Section 2 introduces our framework and convergence results, Section 3 presents the LLT, and Section 4 the PE algorithms. We introduce our new algorithm in Section 5, validate our complexity results in Section 6, and conclude the paper in Section 7.

2 The discrete Moreau envelope

Let us fix our notations. Our objective is the numerical computation of the discrete Moreau envelope

$$M_{\lambda,X}(s) := \min_{x\in X}\left[\frac{\|s-x\|^2}{2\lambda}+f(x)\right], \tag{4}$$

at many points $s \in S \subset \mathbb{R}^d$, knowing the values of f on the set X. Both sets S and X are discrete sets in $\mathbb{R}^d$ containing m and n points respectively. A brute force computation of the Moreau envelope involves computing a minimum over n points x for m values of the parameters s. Hence, it has a quadratic worst-case time complexity. Our objective is to investigate algorithms that produce the same results with a linear worst-case time complexity.

We name the set of minima in (4) the discrete proximal mapping

$$P_{\lambda,X}(s) := \operatorname*{Argmin}_{x\in X}\left[\frac{\|s-x\|^2}{2\lambda}+f(x)\right]. \tag{5}$$

When there is no ambiguity, we will drop the index λ and X.

To obtain more efficient algorithms, we are going to restrict the set X we consider. The Moreau envelope (1) can always be factored as d one-dimensional envelopes. However, we need to restrict X to be a grid $X = X_1 \times \cdots \times X_d \subset \mathbb{R}^d$ to take advantage of that property. In that case, we have for $x = (x_1, \ldots, x_d)$ and $s = (s_1, \ldots, s_d)$

$$M_{\lambda,X}(s) = \inf_{x_1\in X_1}\left[\frac{|s_1-x_1|^2}{2\lambda}+\cdots+\inf_{x_d\in X_d}\left[\frac{|s_d-x_d|^2}{2\lambda}+f(x)\right]\ldots\right], \tag{6}$$

and a fast one-dimensional algorithm will give a fast d-dimensional algorithm by applying it repeatedly. So for computation on a grid, we can restrict ourselves to computing the Moreau envelope for functions of one variable. Consequently, in all the remainder of the paper and unless otherwise specified, X will be a discrete subset in $\mathbb{R}$.

For numerical computation purposes, we can ignore the parameter λ since

$$M_{\lambda,X}(s) = \frac{1}{2\lambda}\min_{x\in X}\left[\|s-x\|^2+2\lambda f(x)\right].$$

So an algorithm computing $M_{1/2}$ for a function f also computes M_λ when applied to the function $2\lambda f$ *i.e.* $M_\lambda f = \frac{1}{2\lambda}M_{1/2}2\lambda f$.

Remark 1 If we restrict ourselves to computing the discrete Moreau envelope on the same grid the function f is sampled (*i.e.* take $S = X$) and of using a regularly sampled grid, then it is not restrictive to assume the grid X is $\{1, \ldots, n\}$. Under these assumptions we have

$$M_j := M_\lambda(x_j) = \frac{h^2}{2\lambda}\min_i\left[\|j-i\|^2+\frac{2\lambda}{h^2}f(x_0+ih)\right]$$

where $x_i = x_0 + ih$ for $1 \leq i \leq n$ and $h > 0$ is the stepsize of the grid. While we will not restrict ourselves to regular grids, the above setting is similar to the distance transform framework in image processing and allows for a simpler implementation (if not faster due to the restriction of some operations to integer arithmetic), see [43].

The discrete Moreau envelope converges to the Moreau envelope, and the smoother the function f, the faster the convergence. To state convergence results, we consider an interval $[a, b] \subset \mathbb{R}$ with $a < b$, and name $M_{[a,b]}$ the Moreau envelope of the function $f + I_{[a,b]}$ *i.e.*

$$M_{[a,b]} = M_\lambda(f + I_{[a,b]}),$$

where $I_{[a,b]}$ is the indicator function of the interval $[a, b]$: $I_{[a,b]}(x) = 0$ when $x \in [a, b]$ and $+\infty$ otherwise. We consider a grid $X \subset [a, b]$ with n regularly spaced points and note M_n the discrete Moreau envelope: $M_n = M_{\lambda,X}$.

We will say a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is proper if there is $x \in \mathbb{R}^d$ for which $f(x) < +\infty$.

Proposition 1 *Assume* $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ *is proper.*

(1) *If f is upper semi-continuous on $[a, b]$, M_n converges pointwise to $M_{[a,b]}$.*
(2) *If f is continuous on $[a.b]$, M_n converges uniformly on $[a.b]$ to $M_{[a,b]}$.*
(3) *If f is Lipschitz continuous on $[a, b]$,*

$$\|M_{[a,b]} - M_n\|_{L^\infty([a,b])} \leq \frac{c_1}{n}$$

where c_1 is a constant depending only on the Lipschitz constant of f, on a, and on b.
(4) *If f is C^2 on a neighborhood of $[a, b]$,*

$$\|M_{[a,b]} - M_n\|_{L^\infty([a,b])} \leq \frac{c_2}{n^2}$$

where c_2 is a constant depending only on $\max_{[a,b]} f''$, and on $b - a$.
(5) *The convergence results above also hold when f is defined in $\mathbb{R}^d$.*

Proof Use Formula (7) below with the convergence results for the discrete Legendre transform contained in [14, 40, 41]. □

Since the previous proposition only guarantees the convergence of $M_\lambda f$ to $M_\lambda(f + I_{[a,b]})$, to obtain the convergence on an unbounded domain involves taking a set $[a, b]$ large enough to contain $P_\lambda(s)$ for all s we want to approximate the Moreau envelope at. The next result states that for a grid large enough, the Moreau envelope of the function f is equal to the Moreau envelope of the function $f + I_{[a,b]}$ for $b - a$ large enough.

Proposition 2 *The following equivalence holds for any $s \in \mathbb{R}$, and any $a > 0$*

$$\partial f^*(s) \cap [-a, a] \neq \emptyset \Leftrightarrow M_{[-a,a]} = M.$$

Proof The full proof is contained in [27]. See also [41, Proposition 2] for the same statement in the context of discrete Legendre transforms. □

3 The Linear-time Legendre Transform algorithm (LLT)

The computation of the Moreau envelope is equivalent to the computation of the Legendre–Fenchel conjugate as the following proposition shows.

Proposition 3 (See [71, Example 11.26]) *Let f be a function defined on $\mathbb{R}^n$, and assume $\lambda > 0$. Then*

$$M_\lambda(s) = \frac{\|s\|^2}{2\lambda} - \frac{1}{\lambda} g_\lambda^*(s). \tag{7}$$

where $g_\lambda(x) := \|x\|^2/2 + \lambda f(x)$.

Proof Expand the quadratic form to obtain

$$\begin{aligned} M_\lambda(s) &:= \inf_{x\in\mathbb{R}} \left[\frac{\|s-x\|^2}{2\lambda} + f(x) \right], \\ &= \frac{\|s\|^2}{2\lambda} + \inf_{x\in\mathbb{R}} \left[-\frac{1}{\lambda} \left(\langle s, x\rangle - \left(\frac{\|x\|^2}{2} + \lambda f(x) \right) \right) \right], \\ &= \frac{\|s\|^2}{2\lambda} - \frac{1}{\lambda} \sup_{x\in\mathbb{R}} [\langle s, x\rangle - g_\lambda(x)], \end{aligned}$$

and use the definition of g_λ to obtain (7). □

Remark 2 The Moreau–Yosida approximate of a nonconvex function may not be convex. In fact, Formula (7) shows it is a difference of convex functions. Nevertheless we can use (7) with discrete Legendre–Fenchel transform algorithms such as [41] to compute the Moreau envelope of even nonconvex functions f in linear time.

Remark 3 The equivalence between the computation of the Legendre–Fenchel conjugate and the Moreau envelope has been noted by several authors [22, 64, 71, 72].

Remark 4 As Formula (7) only involves computing one Legendre–Fenchel conjugate, it is more efficient than the suggestions in [40, 41] to use the fact that for lower semi-continuous convex functions $F_\lambda = (f^* + \frac{\lambda}{2}\|.\|^2)^*$.

For the sake of completion we recall the Linear-time Legendre transform algorithm (full details are presented in [41]). A similar factorization as Formula (6) holds for the Legendre–Fenchel conjugate, so without loss of generality we can assume the sets X_n, and S_m are subsets of $\mathbb{R}$. (Figure 1) shows the algorithm.

The key step is the explicit computation of the convex hull. Using the fact that the set X_n is already sorted, the convex hull is computed in linear time. The resulting vertices give rise to an increasing sequence C, since the derivative of a convex function is always non-decreasing. Then computing the argmax amounts to merging the two increasing sequences C and S.

In higher dimension, the algorithm computes several one-dimensional conjugate. For example, for a function f of two variables, the LLT algorithm first computes the conjugate for each row, then for each column. The resulting complexity in dimension d is $O(dN)$ where N is the number of points on the grid.

Fig. 1 LLT algorithm

Input: The sets X_n, Y_n, and S_m where $Y_n[i]$ is an approximation of $f(X_n[i])$.
Output: The set Z_m where $Z_m[j]$ is an approximation of $f^*(S_m[j])$.
Step 1: Compute the (lower) convex hull of the points $(X[i], Y[i])$ using the Beneath-Beyond convex hull algorithm. Rename the sequence X and Y to be the resulting vertices.
Step 2: Compute the set of slopes $C[i] := (Y[i+1] - Y[i])/(X[i+1] - X[i])$ and for each $S[j]$ find the index i such that $C[i] \leq S[j] < C[i+1]$. Then $X[i]$ is a point where the maximum is attained and $Z[j] = S[j] * X[i] - Y[i]$.

4 The Parabolic Envelope algorithm (PE)

Felzenszwalb and Huttenlocher [18] proposed a linear-time algorithm to compute the Moreau envelope. Their algorithm uses Formula (6) to reduce the computation to one dimension. Then they compute the lower envelope of parabolas by noting that the intersection of two parabolas can be computed in constant time. In this paper, their algorithm will be named PE for parabolic envelope. It is summarized in (Fig. 2).

In Step 1, each parabola is added once and may be deleted at most once. Since adding or deleting a parabola is done in constant time and we consider n parabolas, Step 1 executes in $O(n)$. Step 2 takes $O(m)$ to give a $O(n+m)$ worst-case time complexity for a function of one variable.

Felzenszwalb and Huttenlocher further note that in dimension d the complexity becomes $O(dN)$ where N is the number of points in the grid ($N = n + m$ in the algorithm description).

Remark 5 The idea to compute a parabolic envelope is also contained in [15, 16].

5 The NonExpansive Proximal mapping algorithm (NEP)

The regularity of the proximal mapping allows the development of faster algorithms. Let us first recall the following properties of the proximal mapping.

Proposition 4 *Assume f is a proper lower semi-continuous function and $\lambda > 0$. Then the proximal mapping P_λ is monotone.*

If f is also convex, then P_λ is maximal monotone and non-expansive (hence single-valued).

Proof See [71, Proposition 12.19].

We give here a simple proof for the convex case.

Name ∂f the subdifferential in the sense of convex analysis

$$\partial f(x) := \{s \in \mathbb{R}^d : f(y) \geq f(x) + \langle s, y - x \rangle \text{ for all } y\}$$

Fig. 2 PE algorithm

Input: The sets X_n, Y_n, and S_m where $Y_n[i]$ is an approximation of $f(X_n[i])$.
Output: The set Z_m where $Z_m[j]$ is an approximation of $f^*(S_m[j])$.
Step 1: Compute the (lower) envelope of the n parabolas.
Step 2: Fill in Z_m.

and apply [71, Example 10.2] to obtain

$$\frac{x - p(x)}{\lambda} \in \partial f(p(x))$$

where $p(x) \in P_\lambda(x)$ is a selection in P_λ.

Note that although [71, Example 10.2] is stated with the general subgradient (which is the limit of regular subgradients), it still holds with the subdifferential in the sense of convex analysis since both subgradients are equal in the convex case.

For any two points x and x', we have

$$\frac{x - p(x)}{\lambda} \in \partial f(p(x)), \text{ and } \frac{x' - p(x')}{\lambda} \in \partial f(p(x')).$$

Now use the monotonicity of the subdifferential to obtain

$$\left\langle \frac{x - p(x)}{\lambda} - \frac{x' - p(x')}{\lambda}, p(x) - p(x') \right\rangle \geq 0,$$

in other words

$$\langle x - x', p(x) - p(x') \rangle \geq \|p(x) - p(x')\|^2. \tag{8}$$

So p is *strongly monotone* for any proper function f. In particular, for univariate functions $(x - x')(p(x) - p(x') \geq 0$ *i.e.* p is increasing. □

Now, using the same scheme as in [14, 40, 64] we can build a log-linear $O((n+m)\ln(n+m))$ worst-case time algorithm to compute the Moreau envelope at m points, where n is the number of points at which we sample the function f. Of course, it is outperformed by the PE and the LLT linear-time algorithms.

However, in the convex case, we can build a very simple linear-time algorithm using the smoothness of the proximal mapping by carefully selecting grids as follow. Apply the Cauchy-Schwarz inequality to (8) to obtain

$$\|p(x) - p(x')\| \leq \|x - x'\|,$$

in other words, P_λ is non-expansive. So any selection p of the proximal mapping P_λ is 1-Lipschitz. Take two partitions, $x_1 < \cdots < x_n$ and $s_1 < \cdots < s_m$, and assume

$$x_{i+1} - x_i = s_{j+1} - s_j =: h$$

for any integer i, j with $1 \leq i \leq n-1$ and $1 \leq j \leq m-1$. The algorithm NEP described in (Fig. 3) computes the Moreau-Yosida approximate at all the point on the grid $(s_j)_j$ by approximating the infimum with the computation of the minimum on the grid $(x_i)_i$.

The complexity of the NEP algorithm is linear since Step 1 costs $O(n)$ and Step 2 runs in $O(m)$ (each $p(s_j)$ is computed in constant time for $j > 1$).

Fig. 3 NEP algorithm

Step 1: Compute $p(s_1)$ by a linear search.
Step 2: Compute a selection of the proximal mapping using

$$0 \leq p(s_{j+1}) - p(s_j) \leq h$$

so for each j, either $p(s_{j+1}) = p(s_j)$ or $p(s_{j+1}) = p(s_j) + h$.

Table 1 LLT and brute force (direct) algorithms for computing the discrete Legendre transform as included in lft.sci

Algorithms	Discrete Legendre transform functions
lft_llt	Compute the discrete Legendre transform using the LLT algorithm (*main function*).
lft_llt_d	Same as lft_llt but computation is performed on $\{1, \ldots, n\}$ for a function defined on $\{1, \ldots, n\}$.
bb	Compute the lower convex envelope of a set of points in the plane using the Beneath-Beyond algorithm. Assume the points are sorted along the x-axis.
lft_direct	Compute the discrete Legendre transform with a quadratic (brute force) algorithm (for comparison only).
lft_direct_d	Same as lft_direct on the grid $\{1, \ldots, n\}$ for a function defined on $\{1, \ldots, n\}$.
fusion	Merge two increasing sequences using a linear algorithm (internal function).
fusionsci	Merge two increasing sequences using Scilab syntax resulting in a fast but nonlinear algorithm (internal function).

Lemma 1 *The NEP algorithm computes the Moreau envelope in linear-time when the function f is convex.*

Proof We only need to prove that the algorithm computes the Moreau envelope. Since

$$0 \leq p(s_j) - p(s_{j-1}) \leq s_j - s_{j-1} \leq h$$

the only possibilities for $p(s_j) - p(s_{j-1})$ are either 0 or h. In the first case, $p(s_j) = p(s_{j-1})$ and in the second $p(s_j)$ is the successor of $p(s_{j-1})$ in the grid $x_1 < \cdots < x_n$. So the result is indeed a selection of the proximal mapping. □

Table 2 LLT, PE, NEP, and brute force algorithms for computing the discrete Moreau envelope as included in me.sci

Algorithms	Discrete Moreau envelope functions
	Main functions (one dimension)
me_llt	Compute the discrete Moreau envelope (LLT algorithm).
me_pe	Compute the discrete Moreau envelope (PE algorithm).
me_nep	Compute the discrete Moreau envelope (NEP algorithm).
	Main functions (two dimensions)
me_llt2d	Same as me_llt for a function of two variables.
me_pe2d	Same as me_pe for a function of two variables.
me_nep2d	Same as me_nep for a function of two variables.
	Functions provided for comparison only
me_direct	Compute the discrete Moreau envelope with a quadratic (brute force) algorithm.
me_direct2d	Same as me_direct for a function of two variables but uses separability to achieve a $O(N^{3/2})$ complexity.
me_brute2d	Compute the discrete Moreau envelope with a quadratic (brute force) algorithm for a function of two variables.

Table 3 Examples and unit test files

Script name	Function demonstrated and tested
test_bb.sci	bb
test_llt.sci	lft_llt, lft_llt_d, lft_direct, lft_direct_d
test_me_llt.sci	me_llt, me_llt2d, me_direct, me_direct2d
test_me_nep.sci	me_nep, me_nep2d, me_direct
test_me_pe.sci	me_pe, me_pe2d
test_fusion.sci	Fusion and fusionsci (internal functions, unit tests only).

6 Numerical validation

The CCA (Computational Convex Analysis) toolbox for Scilab 4.0 associated with the paper runs under both Ms Windows and Linux, and is freely available from www.netlib.org/numeralgo/na24. (Table 1) lists the functions available in the package for computing the discrete Legendre transform, (Table 2) lists the functions for computing the discrete Moreau envelope, and (Table 3) shows the unit tests and examples available. The package consists of two main files, lft.sci and me.sci, with several unit test files named test_∗.sci called from test.sci. The unit test files serve a double purpose: To check the functions and to provide examples. Finally (Table 4) lists the demonstration files that also illustrate the use of the functions. (Due to space constraints, the result of some of the demonstrations were not included in the paper but are available in the package).

We now present numerical results that validate our complexity results. Computations were performed on a 1.8GHz Pentium 4 using Scilab v4.0.

(Figure 4) compares the three Moreau envelope algorithms with the direct algorithm to compute the function $f(x) = x^2$ on the grid $x_i = x_0 + ih$, with $x_0 = 1$,

Table 4 Demonstration files

Demo	Script name	Description
1	time_me.sci	Comparison of Moreau Envelope algorithms (LLT, PE, NEP, direct). Produces (Fig. 4).
2	abs.sci	Moreau Envelope and proximal mapping of the absolute value function. Produces (Figs. 6a–6b).
3	sqr.sci	Moreau Envelope and proximal mapping of the square function. Produces (Figs. 5a–5b).
4	nonconvexabs.sci	Moreau Envelope and proximal mapping of $f(x) = \|\|x\| - 1\|$. Produces (Figs. 7a–7b).
5	nonconvexsqr.sci	Moreau Envelope and proximal mapping of $f(x) = (x^2 - 1)^2$. Produces (Figs. 8a–8b).
6	bb.sci	Computes the lower convex envelope of the smooth nonconvex function $f(x) = (x^2 - 1)^2$ on the grid $X = \{-2, -1.5, \dots, 2\}$.
7	lft_direct.sci	Illustrates the convergence of the discrete Legendre-Fenchel transform when the domain is enlarged.
8	time_lft.sci	Comparison of Legendre-Fenchel transform algorithms (lft_direct, lft_llt using fusion, and lft_llt using fusionsci).

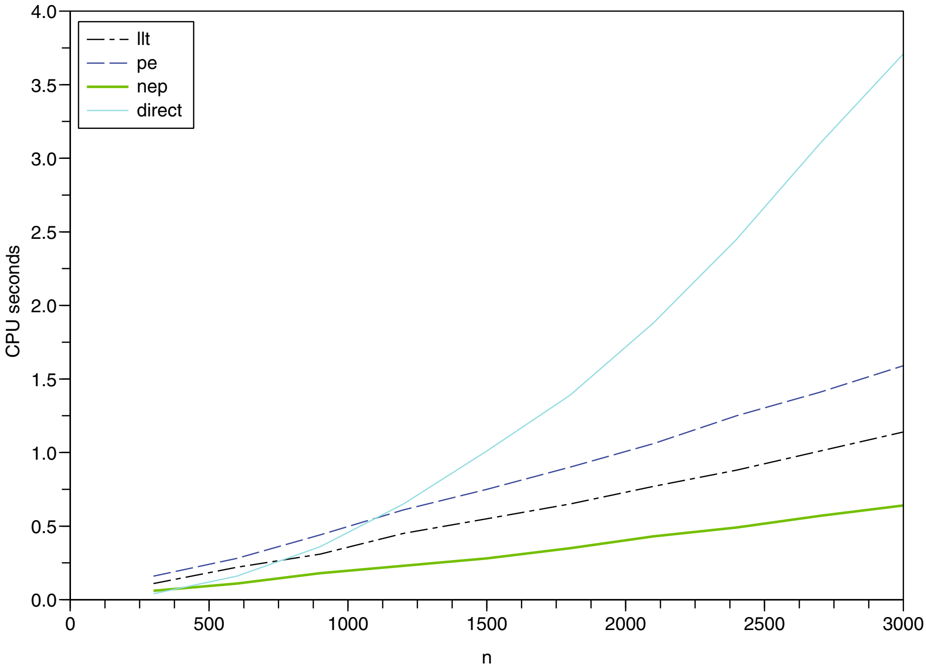
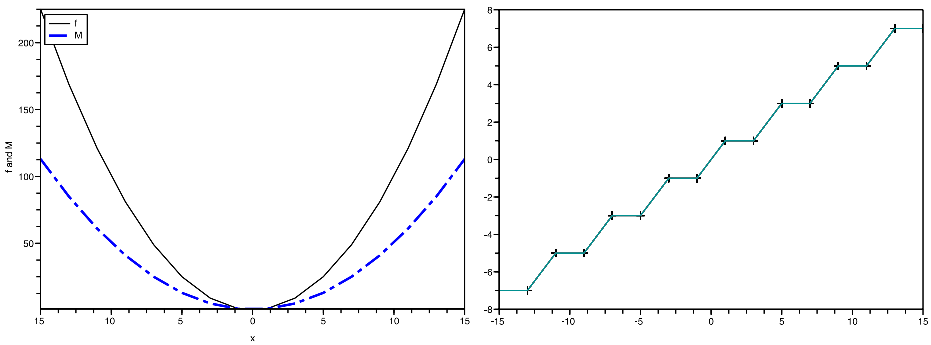


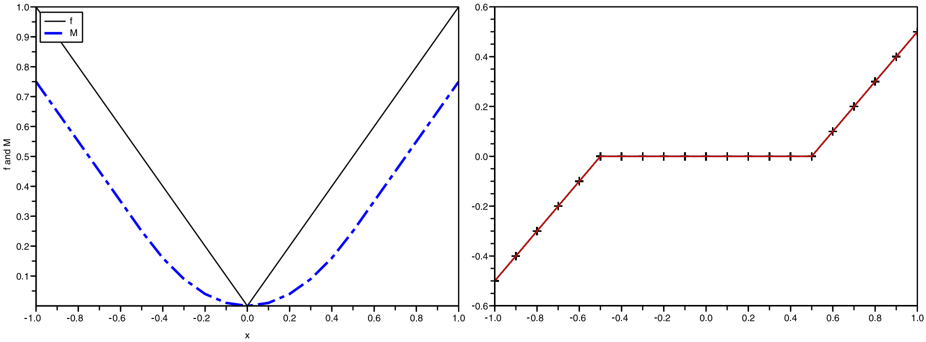
Fig. 4 Comparison of Fast Moreau Envelope Algorithms to compute the Moreau envelope of the function $f(x) = x^2$ on the grids $X = S = \{1, 2, \ldots, n\}$ with $n \in \{300, 600, \ldots, 3000\}$. Direct computation has a quadratic cost while all other algorithms run in linear time

$h = 1$, and $1 \leq i \leq n$. Clearly the fast algorithms perform better than the direct computation even taking into account Scilab optimization for matrix computation. In our implementation, NEP comes best, followed by LLT, PE, and then direct computation. Since NEP requires additional properties (it only works for convex



(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 5 Moreau envelope and Proximal mapping of the smooth convex function $f(x) = x^2$ on the grids $X = S = \{-15, -13, \ldots, 15\}$



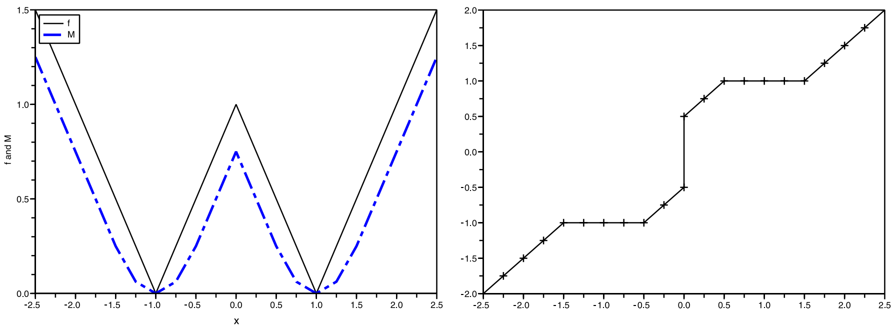
(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 6 Moreau envelope and Proximal mapping of the nonsmooth convex function $f(x) = |x|$ on the grids $X = S = \{-1, -0.9, \ldots, 1\}$

data), it is not surprising it performs better. (Note that we could have further tuned the LLT algorithm by skipping the computation of the convex hull since the function is convex.) The demo script time_me.sci (Option *1. Comparison of ME algorithms* in the CCA demos menu) generates (Fig. 4); see demo 1 of (Table 4).

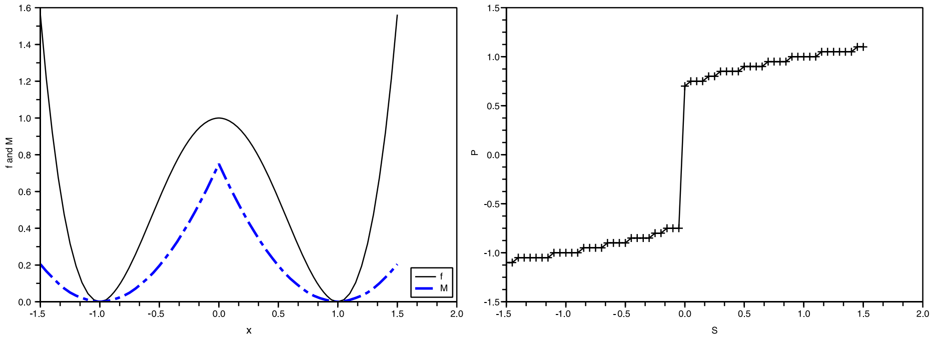
(Figure 5a) shows the smooth convex function $f(x) = x^2$ and its Moreau envelope while (Fig. 5b) shows the corresponding proximal mapping. The nonexpansiveness of the proximal mapping translates graphically as a step function where the height of the step cannot be greater than its width. Both figures were generated using the demo script sqr.sci (Option *3. ME of the square function* in the CCA demos menu; see demo 3 of (Table 4)).

(Figure 6) illustrates the regularization property of the Moreau envelope for the nonsmooth convex function $f(x) = |x|$. While the kink at the minimum is smoothed, the Moreau envelope and the function still share the same minimum, which corresponds to the fixed point of the proximal mapping shown in (Fig. 6b). The demo script abs.sci (Option *2. ME of the abs function* in the CCA demos menu; see demo 2 of (Table 4)) generates both figures.



(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 7 Moreau envelope and Proximal mapping of the nonsmooth nonconvex function $f(x) = ||x| - 1|$ on the grids $X = S = \{-2.5, -2.25, \ldots, 2.5\}$



(a) The function f and its Moreau envelope (b) Proximal mapping

Fig. 8 Moreau envelope and Proximal mapping of the smooth convex function $f(x) = (x^2 - 1)^2$ on the grids $X = S = \{-1.5, -1.45, \ldots, 1.5\}$

The last two examples illustrate nonconvex functions and the failure of the NEP algorithm. (Figure 7a) shows the graph of the function $f(x) = ||x| - 1|$ and its Moreau envelope. (Figure 7b) shows the associated proximal mapping. (both are generated by the demo script nonconvexabs.sci, Option *4. ME of a nonsmooth nonconvex function* in the CCA demos menu; see demo 4 of (Table 4). The jump at the origin indicates that the proximal mapping is not nonexpansive, so the NEP algorithm cannot be applied. Note that the jump is not related to the lack of regularity: (Fig. 8b) illustrates a similar jump for the smooth function $f(x) = (x^2 - 1)^2$ and (Fig. 8a) shows the corresponding Moreau envelope. Both figures are created by the demo script nonconvexsqr.sci (Option *5. ME of a smooth nonconvex function* in the CCA demos menu; see demo 5 of (Table 4)).

7 Conclusion

The principles used to build the fast algorithms presented can be applied in different settings.

The PE algorithm idea is to build the lower envelope of parabolas. It achieves a linear-time complexity because computing the intersection of two parabolas can be done in constant time $\Theta(1)$. The same principle applies to building the envelope of any family of functions provided the intersection of two functions can be computed in linear time. For example, computing the discrete Legendre transform using that principle amounts to computing the lower envelope of affine functions, which is essentially the same as applying the Beneath-Beyond algorithm.

The LLT algorithm relies on convexity. Namely it uses the Beneath-Beyond algorithm to compute the vertices of the lower convex envelope, and achieves linear time since the vertices are naturally sorted along one axis. Any algorithm benefiting from convexity can use the Beneath-Beyond algorithm as a pre-processing step to improve its speed.

Finally, the NEP algorithm principle is to select a regular grid and use the nonexpansiveness property of the proximal mapping. Any transform with a nonexpansive argmax can use the same strategy to achieve linear time.

Acknowledgements The author would like to thank Jeff Dicker and Andrew Tonner for implementing the initial version of the algorithms. The associated software package greatly benefited from several suggestions from the software editor. This work was partly supported by the Okanagan University College Grant in Aid program, and by the author NSERC Discovery grant. Part of this work begun while the author was a Postdoctoral Fellow at the Pacific Institute for the Mathematical Sciences (PIMS).

References

1. Attouch, H., Wets, R.J.-B.: A convergence theory for saddle functions. Trans. Am. Math. Soc. **280**, 1–41 (1983)
2. Aurell, E., Gurbatov, S.-G. N., Simdyankin, S.-G.I.: Numerical proof of self-similarity in Burgers' turbulence. Tech. Rep. http://arxiv.org/abs/patt-sol/9602005,arXiv.org (1996)
3. Bec, J., Frisch, U., Khanin, K.: Kicked Burger turbulence. J. Fluid Mech. **416**, 239–267 (2000)
4. Bernard, F., Thibault, L.: Prox-regularity of functions and sets in Banach spaces. Set-Valued Anal. **12**, 25–47 (2004)
5. Bernard, F., Thibault, L.: Prox-regular functions in Hilbert spaces. J. Math. Anal. Appl. **303**, 1–14 (2005)
6. Bernard, F., Thibault, L.: Uniform prox-regularity of functions and epigraphs in Hilbert spaces. Nonlinear Anal. **60**, 187–207 (2005)
7. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.A.: A family of variable metric proximal methods. Math. Program. **68**, 15–47 (1995)
8. Brenier, Y.: Un algorithme rapide pour le calcul de transformées de Legendre–Fenchel discrètes. C. R. Acad. Sci. Paris Sér. I Math. **308**, 587–589 (1989)
9. Burke, J.V., Qian, M.: A variable metric proximal point algorithm for monotone operators. Tech. Rep., Department of Mathematics, University of Washington, Seattle, Washington; and Department of Mathematics, California State University, Fullerton, CA (1992)
10. Burke, J.V., Qian, M.: On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating. Tech. Rep., Department of Mathematics, University of Washington, Seattle, Washington; and Department of Mathematics, California State University, Fullerton, CA (1996)
11. Burke, J.V., Qian, M.: On the local super-linear convergence of a matrix secant implementation of the variable metric proximal point algorithm for monotone operators. Tech. Rep., Department of Mathematics, University of Washington, Seattle, Washington; and Department of Mathematics, California State University, Fullerton, CA (1998)
12. Clarke, F.H., Stern, R.J., Wolenski, P.R.: Proximal smoothness and the lower-C^2 property. J. Convex Anal. **2**, 117–144 (1995)
13. Cominetti, R.: Coupling the proximal point algorithm with approximation methods. J. Optim. Theory Appl. **95**, 581–600 (1997)
14. Corrias, L.: Fast Legendre–Fenchel transform and applications to Hamilton–Jacobi equations and conservation laws. SIAM J. Numer. Anal. **33**, 1534–1558 (1996)
15. Deniau, L.: Proposition d'un opérateur géométrique pour l'analyse et l'identification de signaux et images. PhD thesis, Université de Paris-Sud, Centre d'Orsay (Dec. 1997)
16. Deniau, L., Blanc-Talon, J.: Fractal analysis with Hausdorff distance under affine transformations, Tech. Rep., ETCA-CREA-SP (1995)
17. Eckstein, J.: Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. Math. Oper. Res. **18**, 202–226 (1993)
18. Felzenszwalb, P.F., Huttenlocher, D.P.: Distance transforms of sampled functions, Tech. Rep. TR2004-1963, Cornell Computing and Information Science (Sept. 2004)
19. Frisch, U., Bec, J.: Burgulence. In: Lesieur, A.M., David, F. (eds.) Les Houches 2000: New Trends in Turbulence, pp. 341–383. Springer EDP-Sciences, France (2001)
20. Frisch, U., Bec, J., Villone, B.: Singularities and the distribution of density in the Burgers/adhesion model. Physica D, **152–153**, 620–635 (2001)
21. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)
22. Gurbatov, S.N., Simdyankin, S.I., Aurell, E., Frisch, U., Tóth, G.: On the decay of Burgers turbulence. J. Fluid Mech. **344**, 339–374 (1997)

23. Gurbatov, S.N., Troussov, A.V.: The decay of multiscale signals – deterministic model of the Burgers turbulence, Tech. Rep., arXiv.org (2000)
24. Hamdi, A.: A Moreau-Yosida regularization of a difference of two convex functions. Appl. Math. E-Notes **5**, 164–170 (2005) (electronic)
25. Hare, W.L., Sagastizabal, C.: Computing proximal points on nonconvex functions, Tech. Rep., Optimization Online Digest – June 2005 (2005)
26. Helluy, P.: Simulation numérique des écoulements multiphasiques : De la théorie aux applications, PhD thesis, Institut des Sciences de l'Ingenieur de Toulon et du Var, Laboratoire Modélisation Numérique et Couplages, BP 56, 83162 La Valette CEDEX, France Habilitation à Diriger des Recherches (Jan. 2005)
27. Hiriart-Urruty, J.B.: Lipschitz r-continuity of the approximate subdifferential of a convex function. Math. Scand. **47**, 123–134 (1980)
28. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex analysis and minimization algorithms. In: Grundlehren der Mathematischen Wissenschaften, vol. 305–306. [Fundamental Principles of Mathematical Sciences] Springer, Berlin Heidelberg New York (1993) (Vol. I: Fundamentals, Vol. II: Advanced theory and bundle methods)
29. Hisakado, T., Okumura, K., Vukadinovic, V., Trajkovic, L.: Characterization of a simple communication network using Legendre transform. In: Proc. IEEE Int. Symp. Circuits and Systems, vol. 3, pp. 738–741 (May 2003)
30. Iusem, A.N., Svaiter, B.F., Teboulle, M.: Entropy-like proximal methods in convex programming. Math. Oper. Res. **19**, 790–814 (1994)
31. Koopen, B.: Contact of bodies in 2D-space: Implementing the Discrete Legendre Transform. AI Master's thesis, Intelligent Autonomous Systems Group, University of Amsterdam (Feb. 2002)
32. Legras, B., Pisso, I., Berthet, G., Lefèvre, F.: Variability of the Lagrangian turbulent diffusion in the lower stratosphere. Atmos. Chem. Phys. **5**, 1605–1622 (2005)
33. Lehdili, N., Moudafi, A.: Combining the proximal algorithm and Tikhonov regularization. Optimization **37**, 239–252 (1996)
34. Lemaire, B.: The proximal algorithm. In: New Methods in Optimization and Their Industrial Uses (Pau/Paris, 1987) Internat, vol. 87, pp. 73–87. Schriftenreihe Numer. Math., Birkhäuser, Basel (1989)
35. Lemaréchal, C., Oustry, F.: Growth conditions and $\mathcal{U}$-Lagrangians. Set-Valued Anal. **9**, 123–129 (2001) (Wellposedness in optimization and related topics (Gargnano, 1999))
36. Lemaréchal, C., Oustry, F., Sagastizábal, C.: The $\mathcal{U}$-Lagrangian of a convex function. Trans. Am. Math. Soc. **352**, 711–729 (2000)
37. Lemaréchal, C., Sagastizábal, C.: More than first-order developments of convex functions: primal-dual relations. J. Convex Anal. **3**, 255–268 (1996)
38. Lemaréchal, C., Sagastizábal, C.: Practical aspects of the Moreau–Yosida regularization: theoretical preliminaries. SIAM J. Optim. **7**, 367–385 (1997)
39. Lemaréchal, C., Sagastizábal, C.: Variable metric bundle methods: from conceptual to implementable forms. Math. Program. **76**, 393–410 (1997)
40. Lucet, Y.: A fast computational algorithm for the Legendre–Fenchel transform. Comput. Optim. Appl. **6**, 27–57 (1996)
41. Lucet, Y.: Faster than the Fast Legendre Transform, the Linear-time Legendre Transform. Numer. Algor. **16**, 171–185 (1997)
42. Lucet, Y.: Fast Moreau envelope computation II: Applications. Tech. Rep., University of British Columbia Okanagan (2005)
43. Lucet, Y.: A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform. In: Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV 2005), Victoria BC, May 2005. IEEE Computer Society Press, Los Alamitos, CA
44. Meng, F., Sun, D., Zhao, G.: Semismoothness of solutions to generalized equations and the Moreau–Yosida regularization. Math. Program. (2005) (online first)
45. Meng, F., Zhao, G.: On second-order properties of the Moreau–Yosida regularization for constrained nonsmooth convex programs. Numer. Funct. Anal. Optim. **25**, 515–529 (2004)
46. Meng, F.W., Hao, Y.: Piecewise smoothness for Moreau–Yosida approximation to a piecewise C^2 convex function. Adv. Math. (China), **30**, 354–358 (2001)
47. Mifflin, R.: A quasi-second-order proximal bundle algorithm. Math. Program. **73**, 51–72 (1996)
48. Mifflin, R., Qi, L., Sun, D.: Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function. Math. Program. **84**, 269–281 (1999)
49. Mifflin, R., Sagastizábal, C.: $\mathcal{VU}$-decomposition derivatives for convex max-functions. In: Ill-posed Variational Problems and Regularization Techniques (Trier, 1998). Lecture Notes in

Economics and Mathematical Systems, vol. 477, pp. 167–186. Springer, Berlin Heidelberg New York (1999)
50. Mifflin, R., Sagastizábal, C.: Functions with primal-dual gradient structure and $\mathcal{U}$-Hessians. In: Nonlinear Optimization and Related Topics (Erice, 1998). Applied Optimization, vol. 36, pp. 219–233. Kluwer, Dordrecht (2000)
51. Mifflin, R., Sagastizábal, C.: On $\mathcal{VU}$-theory for functions with primal-dual gradient structure, SIAM J. Optim. **11**, 547–571 (2000) (electronic)
52. Mifflin, R., Sagastizábal, C.: Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions. SIAM J. Optim. **13**, 1174–1194 (2003) (electronic)
53. Mifflin, R., Sagastizábal, C.: On the relation between $\mathcal{U}$-Hessians and second-order epi-derivatives. Eur. J. Oper. Res. **157**, 28–38 (2004)
54. Mifflin, R., Sagastizábal, C.: $\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions. Optim. Methods Softw. **19**, 463–478 (2004)
55. Mifflin, R., Sagastizábal, C.: Relating $\mathcal{U}$-Lagrangians to second-order epi-derivatives and proximal-tracks. J. Convex Anal. **12**, 81–93 (2005)
56. Mifflin, R., Sagastizábal, C.: A $\mathcal{VU}$-algorithm for convex minimization. Math. Program. (2005)
57. Mifflin, R., Sun, D., Qi, L.: Quasi-Newton bundle-type methods for nondifferentiable convex optimization. SIAM J. Optim. **8**, 583–603 (1998) (electronic)
58. Moreau, J.-J.: Propriétés des applications “prox”. C. R. Acad. Sci. Paris **256**, 1069–1071 (1963)
59. Moreau, J.-J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
60. Moreau, J.-J.: Convexity and duality. In: Functional Analysis and Optimization, pp. 145–169. Academic, New York (1966)
61. Moudafi, A.: Coupling proximal methods and variational convergence. Z. Oper. Res. **38**, 269–280 (1993)
62. Moudafi, A.: Coupling proximal algorithm and Tikhonov method. Nonlinear Times Digest **1**, 203–209 (1994)
63. Noullez, A., Gurbatov, S.N., Aurell, E., Simdyankin, S.I.: The global picture of self-similar and not self-similar decay in Burgers turbulence. Tech. Rep. nlin.CD/0409022, arXiv.org eprint archive, (Sept. 2004)
64. Noullez, A., Vergassola, M.: A fast Legendre transform algorithm and applications to the adhesion model. J. Sci. Comput. **9**, 259–281 (1994)
65. Poliquin, R., Rockafellar, R.T.: Generalized Hessian properties of regularized nonsmooth functions. SIAM J. Optim. **6**, 1121–1137 (1996)
66. Poliquin, R., Rockafellar, R.T.: Prox-regular functions in variational analysis. Trans. Am. Math. Soc. **348**, 1805–1838 (1996)
67. Qi, L.: Second-order analysis of the Moreau-Yosida regularization. In: Nonlinear analysis and convex analysis (Niigata, 1998), pp. 16–25. World Sci., River Edge, NJ (1999)
68. Qi, L., Chen, X.: A preconditioning proximal Newton method for nondifferentiable convex optimization. Math. Program. **76**, 411–429 (1997)
69. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NY (1970)
70. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
71. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin Heidelberg New York (1998)
72. She, Z.-S., Aurell, E., Frisch, U.: The inviscid Burgers equation with initial data of Brownian type. Comm. Math. Phys. **148**, 623–641 (1992)
73. Spingarn, J.E.: Submonotone mappings and the proximal point algorithm. Numer. Funct. Anal. Optim. **4**, 123–150 (1981–1982)
74. Teboulle, M.: Entropic proximal mappings with applications to nonlinear programming. Math. Oper. Res. **17**, 670–690 (1992)
75. Vergassola, M., Dubrulle, B., Frisch, U., Noullez, A.: Burger’s equation, devil’s staircases and the mass distribution function for large-scale structures. Astron. Astrophys. **289**, 325–356 (1994)
76. Wei, Z., Qi, L.: Convergence analysis of a proximal Newton method. Numer. Funct. Anal. Optim. **17**, 463–472 (1996)
77. Yosida, K.: Functional Analysis, Classics in Mathematics. Springer, Berlin Heidelberg New York (1995) (Reprint of the sixth (1980) edition)

ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Image and Vision Computing xxx (2007) xxx–xxx

www.elsevier.com/locate/imavis

New sequential exact Euclidean distance transform algorithms based on convex analysis

Yves Lucet *

Computer Science, I. K. Barber School of Arts and Sciences, University of British Columbia Okanagan, 3333 University Way, Kelowna, BC, Canada V1V 1V7

Received 10 November 2005; received in revised form 24 August 2006; accepted 20 October 2006

Abstract

We present several sequential exact Euclidean distance transform algorithms. The algorithms are based on fundamental transforms of convex analysis: The Legendre Conjugate or Legendre–Fenchel transform, and the Moreau envelope or Moreau-Yosida approximate. They combine the separability of the Euclidean distance with convex properties to achieve an optimal linear-time complexity.

We compare them with a Parabolic Envelope distance transform, and provide several extensions. All the algorithms presented perform equally well in higher dimensions. They can naturally handle grayscale images, and their principles are generic enough to apply to other transforms.

Keywords: Distance transform; Euclidean distance; Feature transform; Fast Legendre transform; Legendre–Fenchel transform; Fenchel conjugate; Moreau envelope; Moreau–Yosida approximate; Computational convex analysis

1. Introduction

Consider an $n \times m$ binary image IMG stored as a 0–1 matrix. Its squared Euclidean distance transform is an $n \times m$ image where each pixel $p \in \{1, \ldots, n\} \times \{1, \ldots, m\}$ has the value

$$\mathrm{DT}^2(p) = \min_q \{\|p - q\|^2;\ \mathrm{IMG}(q) = 0\}.$$

In other words, the EDT computes a new image in which the value at each pixel is equal to the Euclidean distance from that pixel to the background. To avoid unnecessary floating point operations, the square EDT, which we denote DT^2, is usually computed. For example, the DT^2 of the binary image in Fig. 1 is displayed in Fig. 2. If you consider the upper right pixel (5,1), its value in the transformed image becomes its square distance to the background pixel (3,3): $(5-3)^2 + (1-3)^2 = 8$. The value of each pixel is computed similarly.

A distance transform algorithm takes advantage of the fact that the values for all the pixels have to be computed (and there are relations between adjacent pixels) to speed up the computation Fig. 3a shows a binary image and Fig. 3b shows its distance image.

Distance transforms have long been recognised for their importance [1–3], and their applications [4–7]. Starting from non-Euclidean metrics like the city-block distance, and Chamfer distances, several algorithms have been proposed. Computational algorithms for the exact Euclidean distance transform (EDT) appeared later [4,8–14], and several linear-time algorithms are now known. Current research focuses on providing simpler algorithms, now that we have a better understanding of the properties of the EDT, and on extending the distance transform to a more general setting [6,10,13–20].

The three algorithms presented are optimal. They do not rely on complex data structure like polygonal chain [14]

* Tel.: +1 250 807 9505; fax: +1 250 470 6001.
E-mail address: yves.lucet@ubc.ca

0262-8856/$ - see front matter
doi:10.1016/j.imavis.2006.10.011

$$\mathrm{IMG} = \begin{array}{|c|c|c|c|c|}\hline 1&1&1&1&1\\\hline 1&1&1&1&1\\\hline 1&1&0&1&1\\\hline 1&1&1&1&1\\\hline 1&1&1&1&1\\\hline\end{array}$$

Fig. 1. Example of a binary image IMG. The value 0 is associated with the background of the image.

$$\mathrm{DT}^2 = \begin{array}{|c|c|c|c|c|}\hline 8&5&4&5&8\\\hline 5&2&1&2&5\\\hline 4&1&0&1&4\\\hline 5&2&1&2&5\\\hline 8&5&4&5&8\\\hline\end{array}$$

Fig. 2. The Square Euclidean Distance transform of IMG. The value of each pixel is now its square Euclidean distance to the closest background pixel with value 0.

or on other concept like Voronoï diagram [8,10]. They manipulate one-dimensional arrays and are easily parallelizable since the computation on one row only depends on that row and not on values in adjacent rows contrary to for example [11].

Additionally, we consider an extension to the Euclidean distance transform to handle grayscale images instead of binary images. Our framework draws from convex analysis as developed in [21–23]. Following standard convex analysis formulations, we rewrite the constrained optimization problem defining DT^2 with an infinite penalization

$$\mathrm{DT}^2(p) = \min_q\{\|p-q\|^2 + I(q)\},$$

where $I(q) = 0$ if q is a background pixel ($\mathrm{IMG}(q) = 0$) and $+\infty$ otherwise (I is called the indicator function of the set $\{q; \mathrm{IMG}(q) = 0\}$). In that formulation, a natural extension is to consider a general function f instead of the indicator function I. The problem then becomes

$$\mathrm{DT}^2(p) = \min_q\{\|p-q\|^2 + f(q)\}. \qquad (1)$$

The function f could be naturally chosen as the graylevel of the image for example. The resulting extended function Eq. (1), which we still denote DT^2, is well known in convex analysis as the Moreau envelope (also named Moreau–Yosida approximate, or Moreau–Yosida regularisation). The difference between our first formulation of DT^2 and the Moreau envelope Eq. (1) is that the minimum is taken over all the $\mathbb{R}^2$ plane in the later while it is taken only over the pixel defining the image in the former.

Going back to the work of Yosida [24], the Moreau envelope regularisation properties have been studied extensively in convex and variational analysis [25–27,22]. Computing the Moreau envelope is equivalent to computing another fundamental transform in convex duality: The Legendre conjugate (also called Legendre–Fenchel conjugate, or Legendre–Fenchel transform)

$$f^*(s) = \sup_{x\in\mathbb{R}^2}[\langle s,x\rangle - f(x)], \qquad (2)$$

where $\langle\cdot,\cdot\rangle$ denotes the standard dot product in $\mathbb{R}^2$. Historically, its computation was first motivated by the study of Hamilton–Jacobi equations, and gave birth to several numerical algorithms [28–32], and later to a linear-time algorithm [33], with applications in numerous fields: numerical simulation of Burger's equation (see for example [34–37]), robotics [38], network communication [39], pattern recognition [40], numerical simulation of multiphasic flows [41], and analysis of the distribution of chemical compounds in the atmosphere [42].

The algorithms presented here rely on convex analysis properties to achieve a worst-case linear computation time. They either compute the Fenchel conjugate Eq. (2) or (equivalently) the Moreau envelope Eq. (1). They include the Linear-time Legendre Transform (LLT) [33], the parabolic envelope (PE) [12,13,20], and the non-expansive proximal mapping (NEP) [20]. The algorithms were imple-

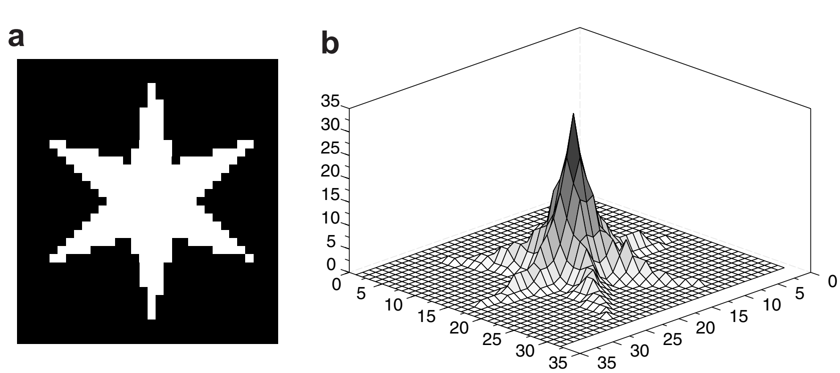


Fig. 3. 35×32 binary image of a star and its distance transform. (a) Binary image of a star. Background pixels are in black. (b) Distance transform of the image in (a).

mented in Scilab to allow for future comparison with known distance transform algorithms such as those available in the Scilab Image Processing toolbox (SIP) [43].

2. The LLT algorithm

We summarise the algorithm detailed in [40] to compute DT^2 the square EDT using the LLT algorithm. We write

$$DT^2(p) = \|p\|^2 - 2g^*(p) \quad (3)$$

where $g^*(p) = \max_q[\langle p,q\rangle - g(q)]$ is the Legendre conjugate of the function $g(q) = \|q\|^2/2 + I(q)$, and $\langle .,.\rangle$ is the standard scalar product. So the algorithm amounts to computing g for all pixels q, then applying the LLT algorithm to obtain g^* for all pixels p, and finally deducing DT^2 at all pixels p by Eq. (3).

The LLT algorithm uses convexity to obtain a linear running time. First, it factors the two-dimensional conjugate as several one-dimensional conjugates, which amounts to processing rows in a first pass, and columns in a second pass. Next, the computation of each one-dimensional transform involves computing the lower convex envelope of the function in linear time using the Beneath-Beyond algorithm, and then merging two increasing sequences. More precisely, to compute

$$g^*(p_1,p_2) = \max_{q_1}[p_1q_1 + \max_{q_2}[p_2q_2 - g(q_1,q_2)]]$$

in linear time, we only need to be able to compute u^* the conjugate of a convex real-valued function u in linear time, which can be achieved using the following property [33, Lemma 3]: Assume u is a convex univariate function, and define the finite difference slopes $c_i = (u(x_{i+1}) - (u(x_i))/(x_{i+1} - x_i)$. Then if $c_{i-1} < p_1 < c_i$, the maximum is attained at x_i, and if $c_i = p_1$, the maximum is attained at both x_i and x_{i+1}. Since u is convex the sequence c_i is non-decreasing, so computing u^* at all the x_i, amounts to merging the sequences c_i and x_i.

Figs. 4–9 illustrate the LLT algorithm and its partial computations. The binary image is transformed first into another (binary) image with pixel values in $\{0, +\infty\}$, then each background pixel (i,j) (pixels with value *0*) is set to the value $(i^2+j^2)/2$. The Fenchel conjugate is then applied to the resulting image first row by row, then column by column. Finally Formula Eq. (3) is used to deduced DT^2.

As noted in [40] there is no need to compute the lower convex envelope of the function g due to the particular

1	1	1	1	1
1	0	1	0	1
1	1	1	1	1
1	1	0	1	1
1	1	1	1	1

Fig. 4. The binary image IMG (the background pixels have value 0).

∞	∞	∞	∞	∞
∞	0	∞	0	∞
∞	∞	∞	∞	∞
∞	∞	0	∞	∞
∞	∞	∞	∞	∞

Fig. 5. The function f generated from IMG by replacing any non-background pixel with $+\infty$.

∞	∞	∞	∞	∞
∞	4	∞	10	∞
∞	∞	∞	∞	∞
∞	∞	12.5	∞	∞
∞	∞	∞	∞	∞

Fig. 6. The function g generated by replacing the value of any background pixel (i,j) with the value $(i^2+j^2)/2$.

∞	∞	∞	∞	∞
2	0	−2	−6	−10
∞	∞	∞	∞	∞
9.5	6.5	3.5	0.5	−2.5
∞	∞	∞	∞	∞

Fig. 7. The partial conjugate of g computed by applying the one-dimensional conjugate to the image row by row.

0	2	4	8	12
2	4	6	10	14
4	6	8.5	12	16
6.5	9.5	12.5	15.5	18.5
10.5	13.5	16.5	19.5	22.5

Fig. 8. The full conjugate of g obtained by applying the conjugacy operation to the partial conjugate column by column.

structure of g (it is the sum of a quadratic and an indicator function, so the vertices of its convex envelope are the points where g is finite). However, the convex envelope of the partial conjugate has to be computed along each columns, which can be done in linear time since the points $(i, g(i,j))_i$ are naturally sorted along the first coordinate.

The complexity of computing DT^2 using the LLT algorithm is linear. The LLT algorithm requires a two-pass scan of the image in addition to computing the function g and applying Eq. (3).

Please cite this article in press as: Y. Lucet, New sequential exact Euclidean distance transform algorithms based ..., Image Vis. Comput. (2007), doi:10.1016/j.imavis.2006.10.011

2	1	2	1	2
1	0	1	0	1
2	1	1	1	2
4	1	0	1	4
5	2	1	2	5

Fig. 9. The resulting square Euclidean distance transform DT^2 obtained using Eq. (3): multiply the image in Fig. 8 by −2, then at each pixel (i,j), add (i^2+j^2).

1	1	1	1	1
1	0	1	1	1
1	1	0	1	1
1	1	1	1	1
1	1	1	1	1

Fig. 10. Binary image IMG.

The LLT algorithm can be easily extended to non-indicator functions i.e. to compute the Moreau envelope Eq. (1) for any real-valued function f. For example, the function f can be defined as the gray level in a grayscale image. It also naturally extends to higher dimensions: When the data is d-dimensional, the LLT complexity is $O(d\,N)$, where N is the number of points in the grid ($N = n^2$ for an $n \times n$ image).

3. The NEP algorithm

The Non Expansive Proximal mapping (NEP) algorithm relies on the non-expansiveness[1] of the proximal mapping P defined by

$$P(p) = \mathrm{Argmin}_q\{\|p-q\|^2 + f(q)\}. \qquad (4)$$

The mapping P associates the closest background pixel to each pixel of the image (computing the distance between the two pixels gives the distance transform).

The non-expansiveness property holds when the function f is finite at some point, lower semi-continuous, and convex. When f is an indicator function of a binary image, it amounts to the object in the image (pixels with zero value) being convex. In that case, the NEP algorithm is extremely simple: Perform two scans first on the rows then on the columns, so we only compute P on $\{1,\ldots,n\}$ and not on the two-dimensional grid. On each scan, we perform a linear search to compute $P(1)$. Then a single loop returns all the values of P since under the assumptions above

$$0 \leqslant P(i+1) - P(i) \leqslant i+1-i = 1$$

[1] A function P is non-expansive if for any x, y, $\|P(x)-P(y)\| \leqslant \|x-y\|$.

∞	∞	∞	∞	∞
∞	0	∞	∞	∞
∞	∞	0	∞	∞
∞	∞	∞	∞	∞
∞	∞	∞	∞	∞

Fig. 11. The function f obtained by substituting the value $+\infty$ at each pixel with value 1.

5	6	6	6	6
2	2	2	2	2
3	3	3	3	3
5	6	6	6	6
5	6	6	6	6

Fig. 12. The partial transform Pp obtained by applying the operator P to the image in Fig. 11 row by row (only row 2 and 3 are relevant).

∞	∞	∞	∞	∞
1	0	1	4	9
4	1	0	1	4
∞	∞	∞	∞	∞
∞	∞	∞	∞	∞

Fig. 13. The partial distance transform associated with the partial transform in Fig. 12.

2	2	2	3	3
2	2	3	3	3
2	3	3	3	3
3	3	3	3	3
3	3	3	3	3

Fig. 14. The image resulting from applying the operator P to Fig. 13 column by column.

i.e. the value of P at the next pixel $i+1$ is either $P(i)$ or $P(i)+1$. The convexity assumption reduces the global search to a very narrow local search.

We illustrate the algorithm on Figs. 10–15. First Fig. 10 is converted to Fig. 11 by replacing the value of non-background pixels (pixels with value 1) with $+\infty$. Next each row of Fig. 11 is considered independently. The value of each pixel becomes the index in the row to the closest background pixel. For example, for the second row there is only one background pixel at column 2, so all the pixels in that

2	1	2	5	8
1	0	1	2	5
2	1	0	1	4
5	2	1	2	5
8	5	4	5	8

Fig. 15. The resulting DT^2 computed from Figs. 14 and 13 column by column.

row take the value 2. Note that rows 1, 4, and 5 do not have any background pixel. In that case, the algorithm assigns arbitrary values (when the image is considered column by column, there will always be a pixel with a lower value on that column and the arbitrary values will be superceded by a relevant value).

Finally, starting from the partial distance transform in Fig. 13, the algorithm is applied column by column. Again, the value of each pixel becomes the index of the closest background pixel in that column. Consider the second column. Pixel 2 is closest to pixel 1, so the value of pixel 1 becomes 2. Now the value of pixel 2 is either 2 or 3, and a simple comparison gives it the value 2. Similarly, the value of pixel 3 is either 2 or 3, and after comparison it is assigned value 3. The full DT^2 in Fig. 15 is obtained by reading Figs. 14 and 13 column by column and assigning to each pixel (i,j) with value $P_{i,j}$ in Fig. 14 and value $D_{i,j}$ in Fig. 13 the value $(P_{i,j} - i)^2 + D_{i,j}$.

When the data is not convex like the example in Fig. 4, the algorithm fails as Fig. 16 shows: Line 2 of the partial feature transform Pp cannot return the index *4* since its distance to the previous value is more than *1*. Hence DT^2 returns errors as shown by Fig. 17 (the right answer is shown in Fig. 9).

5	6	6	6	6
2	2	2	2	2
5	6	6	6	6
3	3	3	3	3
5	6	6	6	6

Fig. 16. Partial feature transform *Pp* of the image in Fig. 4.

2	1	2	5	10
1	0	1	4	9
2	1	2	5	10
5	4	5	8	13
10	9	10	13	18

Fig. 17. Erroneous DT^2 computed from Fig. 16. Fig. 9 shows the right answer.

The NEP principle extends to any transform having a non-expansive argmin or argmax. In effect, the NEP principle reduces a global search to a local search. As soon as the distance between the pixels realising the minimum is bounded by a constant, a linear-time algorithm exists.

4. The PE algorithm

To initiate comparison with other distance transform algorithms, we recall the parabolic envelope (PE) algorithm [12,13,20]. It reduces computation to one dimension and uses properties of parabolas to obtain a linear-time algorithm. Namely, given two parabolas $p_1 = (. - i)^2 + j$ and $p_2 = (. - i')^2 + j'$, we can compute their intersection in constant time $O(1)$. So computing the lower envelope of the family of parabolas $(. - i)^2 + j$ takes linear time. The algorithm is a two-pass scan. The first pass considers each row and computes the parabolic envelope, then evaluates the distance transform on that parabolic envelope. The second pass performs the same operations on the columns resulting from the first pass.

For the binary image shown in Fig. 4, Fig. 18 shows the results of the algorithm after the first pass: The parabolic envelope computed for each row is evaluated on the grid. Fig. 19 shows the resulting exact square EDT.

The PE algorithm runs in linear time. When the data are d dimensional, its complexity becomes $O(d\,N)$, for a grid containing N points (for an $n \times m$ binary picture $N = nm$). Its underlying principle is applicable to the computation of any family of functions provided the intersection of two such functions can be computed in constant time. Felzenszwalb [13] gave several examples of such

∞	∞	∞	∞	∞
1	0	1	0	1
∞	∞	∞	∞	∞
4	1	0	1	4
∞	∞	∞	∞	∞

Fig. 18. Partial DT^2 resulting from applying the PE algorithm to the rows of Fig. 4.

2	1	2	1	2
1	0	1	0	1
2	1	1	1	2
4	1	0	1	4
5	2	1	2	5

Fig. 19. DT^2 computed by applying the PE algorithm column by column to Fig. 18.

Please cite this article in press as: Y. Lucet, New sequential exact Euclidean distance transform algorithms based ..., Image Vis. Comput. (2007), doi:10.1016/j.imavis.2006.10.011

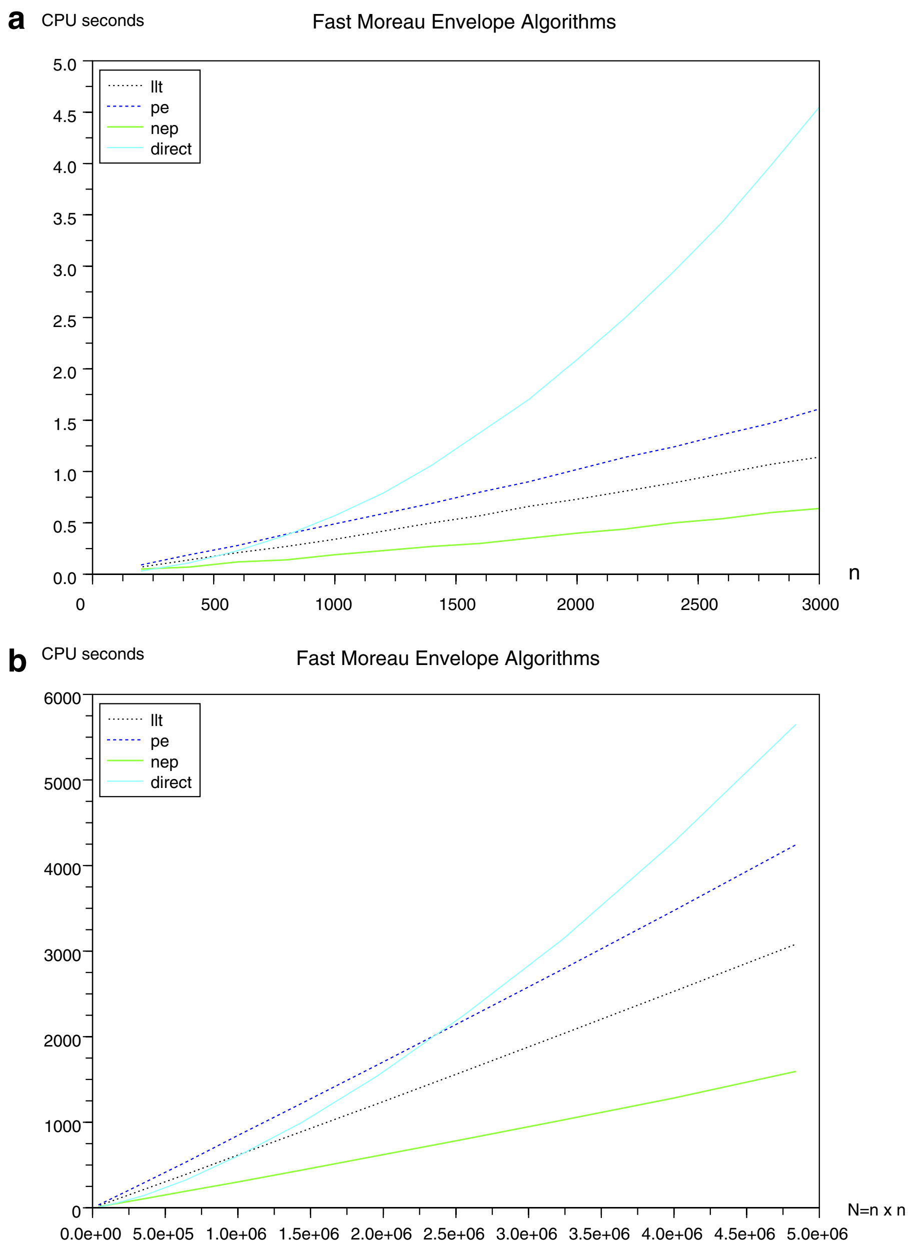


Fig. 20. Numerical validation of the linear-time complexity of the LLT, NEP, and PE algorithms. (a) Complexity for one-dimensional data. (b) Complexity for two-dimensional data such as distance transforms.

cases, including replacing the Euclidean distance with the l_1 distance.

Note that no assumption is made on the function f, all assumptions are on the distance function.

While the PE algorithm considers a family of parabolic functions, one can consider a family of affine functions and apply the same principle as the PE algorithm to compute the lower envelope. The resulting envelope is the lower convex envelope of the points. Hence, the LLT algorithm can be seen as following the same principle for the family of affine functions going through two consecutive pixels of the image. The only difference lies in how much computation is performed during each pass.

5. Numerical comparisons and complexity

Computing the EDT of an $n \times n$ binary image having $N = n^2$ pixels can be achieved by brute force at an $O(N^2) = O(n^4)$ cost as the following function shows.

```
function M = brute2d (Xr, Xc, f, Sr, Sc)
  [n1,n2]=size(f);
  m1=length(Sr);m2=length(Sc);
  M=zeros(m1,m2);
  for p1=1:m1
    for p2=1:m2
      t1 = (Xr-Sr(p1)).^2 * ones(1,n2);
  t2 = ones(n1,1) * ((Xc-Sc(p2)).^2)';
      t = t1 + t2 + f;
    M(p1,p2) = min(t);//O(n^2) cost
    end;
  end;
 endfunction
```

Using the fact the Euclidean distance is a separable function (it can be written as the sum of two one-dimensional functions) one can build a direct computation algorithm with complexity $O(N^{3/2}) = O(n^3)$. Compare the following with the `brute2d` function above (the function `me_direct`(X,f,S) is the one-dimensional brute force computation and runs in $O(n^2)$ when X and S have size n).

```
function M = direct2d (Xr, Xc, f, Sr, Sc)
  for i=1:length(Xr)
    F = me_direct(Xc, f(i,:)', Sc);
      r1(i,:) = F';
  end
  M = ones(size(Sr, 1), size(Sc, 1));
  for i=1:size(Sc, 1)
  M(:,i) = me_direct(Xr, r1(:,i), Sr);
  end
endfunction
```

Fig. 20(a) numerically validates the linear-time complexity of the algorithms for the univariate quadratic function $f(x) = x^2$ while Fig. 20(b) illustrates the same algorithms for two-dimensional data such as computing the EDT. In the later figure, the brute force algorithm is not represented since its cost is too prohibitive.

6. Conclusion

We have presented new exact Euclidean Distance Transform (EDT) algorithms for binary images which all share an optimal linear-time complexity of $\Theta(N) = \Theta(n^2)$ for an $n \times n$ image. By taking advantage of the separability of the Euclidean distance, and using convex properties, the algorithms reduce to simple calculations on a line. In addition to scaling to higher dimension and being trivially parallelizable, all the algorithms compute the more general Moreau envelope, of which the distance transform is a special case.

The principles used in each algorithm can be applied to different transforms. The LLT algorithms uses convex properties, the NEP algorithm uses non-expansiveness, and the PE algorithm uses a $O(1)$ intersection cost. The last two properties are easy to identify and are sufficient to guarantee a linear-time algorithm.

Our implementation of the algorithms show that the NEP algorithm is generally faster than the LLT algorithm, which is faster than the PE algorithm. So if one has an image with a convex object in the background, the NEP algorithm is indicated. In the absence of convexity, the LLT algorithm performs quite well even though it introduces intermediate steps. However, the intermediate steps are very quick to perform in a matrix-optimized language such as Scilab. Another implementation in a non matrix-optimized language will probably find the PE algorithm very competitive.

Acknowledgements

The present research was partly supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant (NSERC), and by the Okanagan University College Grant-in-Aid program. All computations were done on the mathematical software Scilab using the SIP (Scilab Image Processing) package on a Pentium IV 1.8 GHz computer. Our Scilab implementation of the LLT algorithm presented in the paper is available through the GPL license.

References

[1] A. Rosenfeld, J.L. Pfaltz, Sequential operations in digital picture processing, J. ACM 13 (4) (1966) 471–494.

[2] P.-E. Danielsson, Euclidean distance mapping, Comput. Graph. Image Process. 14 (1980) 227–248.

[3] G. Borgefors, Distance transformations in digital images, Comput. Vis. Graph. Image Process. 34 (3) (1986) 344–371.

[4] G. Borgefors, Distance transformations on hexagonal grids, PRL 9 (1989) 97–105.

[5] O. Cuisenaire, Distance transformations: fast algorithms and applications to medical image processing, Ph.D. thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgium (Oct. 1999).

[6] J. Toriwaki, K. Mori, Distance transformation and skeletonization of 3d pictures and their applications to medical images, in: R.K.E.G. Bertrand, A. Imiya (Eds.), Digital and Image Geometry: Advanced Lectures, Lecture Notes in Computer Science, vol. 2243/2001, Springer-Verlag Heidelberg, 2001, pp. 412–429.

[7] S. Prohaska, H.C. Hege, Fast visualization of plane-like structures in voxel data, in: VIS '02: Proceedings of the Conference on Visualization '02, IEEE Computer Society, Boston, MA, 2002, pp. 29–36.

[8] H. Breu, J. Gil, D. Kirkpatrick, M. Werman, Linear time Euclidean distance transform algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 17 (5) (1995) 529–533.

[9] L. Chen, H.Y.H. Chuang, A fast algorithm for Euclidean distance maps of a 2-d binary image, Inform. Process. Lett. 51 (1) (1994) 25–29.

[10] C.R. Maurer Jr., R. Qi, V.V. Raghavan, A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2) (2003) 265–270.

[11] F.Y. Shih, Y.-T. Wu, Fast Euclidean distance transformation in two scans using a 3×3 neighborhood, Comput. Vis. Image Underst. 93 (2) (2004) 195–205.

[12] L. Deniau, J. Blanc-Talon, Fractal analysis with Hausdorff distance under affine transformations, Tech. rep., ETCA-CREA-SP (1995).

[13] P.F. Felzenszwalb, D.P. Huttenlocher, Distance transforms of sampled functions, Tech. Rep. TR2004-1963, Cornell Computing and Information Science (Sep. 2004).

[14] M. Gavrilova, M.H. Alsuwaiyel, Two algorithms for computing the Euclidean distance transform, Tech. Rep. 2000-661-13, Computer Science Technical Reports, University of Calgary (2000).

[15] O. Cuisenaire, B. Macq, Fast and exact signed Euclidean distance transformation with linear complexity, in: Proc. of ICASSP'99, vol. 6, IEEE, Phoenix, AZ, USA, 1999, pp. 3293–3296, Int. Conference on Acoustics, Speech and Signal Processing.

[16] D.G. Bailey, An efficient Euclidean distance transform, in: R. Klette, J. Žunić (Eds.), Combinatorial Image Analysis: 10th International Workshop, IWCIA 2004, Auckland, New Zealand, December 1–3, 2004. Proceedings, Lecture Notes in Computer Science, vol. 3322, Springer-Verlag GmbH, 2004, pp. 394–408.

[17] W. Hesselink, A linear-time algorithm for euclidean feature transform sets, http://www.cs.rug.nl/~wim/pub/mans.html/ (May 2005).

[18] S. Mauch, A fast algorithm for computing the closest point and distance transform, http://www.acm.caltech.edu/seanm/projects/cpt/cpt.pdf/ (Dec. 2000).

[19] C. Sigg, R. Peikert, M. Gross, Signed distance transform using graphics hardware, in: Proceedings of IEEE Visualization '03, ETH Zürich, 2003.

[20] Y. Lucet, Fast Moreau envelope computation I: Numerical algorithms, Tech. rep., University of British Columbia Okanagan (2005).

[21] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer-Verlag, Berlin, 1998.

[22] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, New york, 1970.

[23] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms, vol. 305–306 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993, vol I: Fundamentals, vol II: Advanced theory and bundle methods.

[24] K. Yosida, Functional analysis, Classics in Mathematics, Springer-Verlag, Berlin, 1995, reprint of the sixth (1980) edition.

[25] J.-J. Moreau, Propriétés des applications "prox", C. R. Acad. Sci. Paris 256 (1963) 1069–1071.

[26] J.-J. Moreau, Proximité et dualité dans un espace Hilbertien, Bull. Soc. Math. France 93 (1965) 273–299.

[27] J.-J. Moreau, Convexity and duality, in: Functional Analysis and Optimization, Academic Press, New York, 1966, pp. 145–169.

[28] Y. Brenier, Un algorithme rapide pour le calcul de transformées de Legendre–Fenchel discrètes, C. R. Acad. Sci. Paris Sér. I Math. 308 (1989) 587–589.

[29] L. Corrias, Fast Legendre–Fenchel transform and applications to Hamilton–Jacobi equations and conservation laws, SIAM J. Numer. Anal. 33 (4) (1996) 1534–1558.

[30] Y. Lucet, A fast computational algorithm for the Legendre–Fenchel transform, Comput. Optimizat. Appl. 6 (1) (1996) 27–57.

[31] A. Noullez, M. Vergassola, A fast Legendre transform algorithm and applications to the adhesion model, J. Scient. Comput. 9 (3) (1994) 259–281.

[32] Z.-S. She, E. Aurell, U. Frisch, The inviscid Burgers equation with initial data of brownian type, Comm. Math. Phys. 148 (3) (1992) 623–641.

[33] Y. Lucet, Faster than the fast legendre transform, the linear-time legendre transform, Numer. Algorithms 16 (2) (1997) 171–185.

[34] J. Bec, U. Frisch, K. Khanin, Kicked Burger turbulence, J. Fluid Mech. 416 (2000) 239–267.

[35] U. Frisch, J. Bec, Burgulence, in: A.M. Lesieur, E.F. David (Eds.), Les Houches 2000: New Trends in Turbulence, Springer EDP-Sciences, 2001, pp. 341–383.

[36] U. Frisch, J. Bec, B. Villone, Singularities and the distribution of density in the Burgers/adhesion model, Phys. D 152-153 (2001) 620–635.

[37] A. Noullez, S.N. Gurbatov, E. Aurell, S.I. Simdyankin, The global picture of self-similar and not self-similar decay in Burgers turbulence, Tech. Rep. nlin.CD/0409022, arXiv.org eprint archive (Sep. 2004).

[38] B. Koopen, Contact of bodies in 2D-space: Implementing the Discrete Legendre Transform, AI Master's thesis, Intelligent Autonomous Systems Group, University of Amsterdam (Feb. 2002).

[39] T. Hisakado, K. Okumura, V. Vukadinovic, L. Trajkovic, Characterization of a simple communication network using Legendre transform, in: Proc. IEEE Int. Symp. Circuits and Systems, vol. 3, 2003, pp. 738–741.

[40] Y. Lucet, A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform, in: Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV 2005), IEEE Computer Society Press, Victoria BC, 2005.

[41] P. Helluy, Simulation numérique des écoulements multiphasiques : De la théorie aux applications, Ph.D. thesis, Institut des Sciences de l'Ingenieur de Toulon et du Var, Laboratoire Modélisation Numérique et Couplages, BP 56, 83162 La Valette CEDEX, France, habilitation à Diriger des Recherches (Jan. 2005).

[42] B. Legras, I. Pisso, G. Berthet, F. Lefvre, Variability of the Lagrangian turbulent diffusion in the lower stratosphere, Atmospheric Chem. Phys. 5 (2005) 1605–1622.

[43] R. Fabbri, Scilab image processing toolbox, 2005. http://siptoolbox.sourceforge.net/.

WHAT SHAPE IS YOUR CONJUGATE?
A SURVEY OF COMPUTATIONAL CONVEX ANALYSIS AND ITS APPLICATIONS

YVES LUCET

Abstract. Computational Convex Analysis algorithms have been rediscovered several times in the past by researchers from different fields. To further communications between practitioners, we review the field of Computational Convex Analysis, which focuses on the numerical computation of fundamental transforms arising from convex analysis. Current models use symbolic, numeric, and hybrid symbolic-numeric algorithms. Our objective is to disseminate widely the most efficient numerical algorithms in several fields such as image processing (distance transform, generalized distance transform, mathematical morphology), partial differential equations (solving Hamilton-Jacobi equations, and using differential equations numerical schemes to compute the convex envelope), max-plus algebra, multi-fractal analysis, and several others that span a wide range of applications in computer vision, robot navigation, phase transition in thermodynamics, electrical networks, medical imaging, network communications, discrete event systems, etc.

Contents

Introduction		1
1.	Fundamental Convex Transforms	3
2.	Computer-Aided Convex Analysis	6
3.	Antiderivatives, Network Flow, Phase Transition, Electrical Networks, and Robot Navigation	8
4.	Image Processing, Computer Vision, and Mathematical Morphology	11
5.	Partial Differential Equations	14
6.	Multifractal Analysis, Network Communication, and Extremal Algebra	16
7.	Conclusion	19
Acknowledgments		19
References		19

Introduction

The objective of the present paper is twofold. First, we summarize the state of the art in Computational Convex Analysis for researchers interested in computer-aided convex analysis to build their intuition, or generate nontrivial examples through a combination of convex transforms. Current algorithms allow symbolic, numerical, and hybrid symbolic-numeric computations, and have already been instrumental in discovering and illustrating several new results in Convex Analysis.

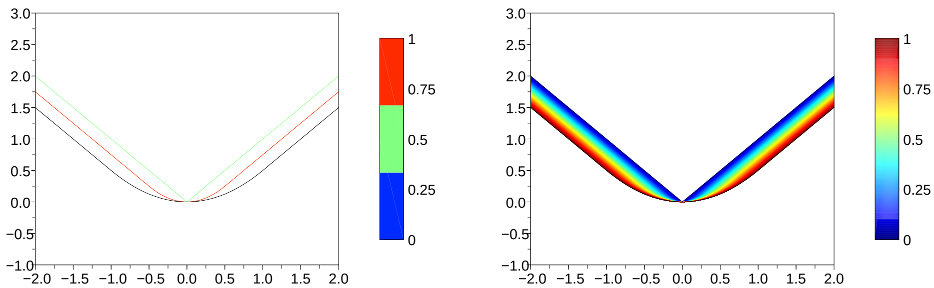
Then we present several applications benefiting from such efficient algorithms. Here we want to show Convex Analysis researchers the rich and varied set of applications they can contribute to. In addition, we want to connect the various specialized researchers with one another, by pointing out that they all use techniques related to Convex Analysis, often unknowingly, and encouraging them to consider the most recent algorithms in Computational Convex Analysis. We hope that the resulting awareness will result in new advances for all the fields involved.

While the impact of Convex Analysis in optimization is well-known, its applications to discrete problems are less understood. For example, the fact that Convex Analysis can be seen as operating on the max-plus algebra (instead of our usual plus-times algebra) in which the Fenchel conjugate plays a similar role as the

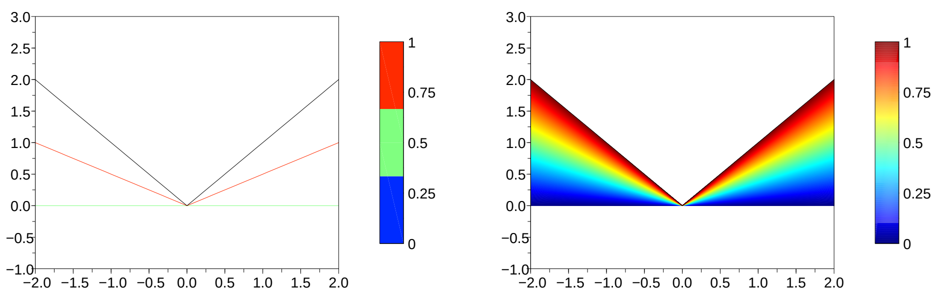
Date: March 28, 2008.
2000 *Mathematics Subject Classification.* 52B55,65D99.
This work was partly supported by the author NSERC Discovery grant.

FFT, is not widely known [82, p. 43]. Although they have a very wide range of applications, the most efficient numerical algorithms for computing convex transforms are still only familiar to Convex Analysis researchers, e.g. the Fast Legendre Transform is still widely used instead of the faster and simpler Linear-time Legendre Transform algorithm.

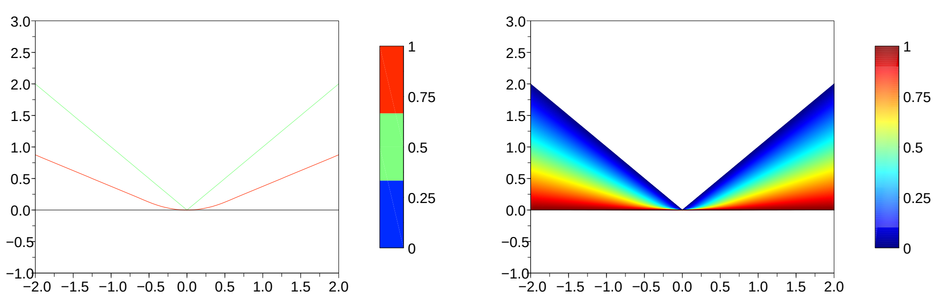
While the present article is concerned with the numerical computation of transforms like the Moreau envelope, contrary to [103] we do not consider computing its value at one point but instead we tackle the problem of computing the Moreau envelope on a grid. In other words, we are interested in computing the shape (or graph) of the Moreau envelope and other transforms. Figure 1 illustrates typical shapes: the graph of the operator is plotted for several values of a parameter.



(a) Moreau envelope of $|\cdot|$ for $\lambda \in \{0, 0.5, 1\}$. (b) Same as 1(a) for λ taking 512 values.



(c) Pasch-Hausdorff envelope of $|\cdot|$ for $\lambda \in \{0, 0.5, 1\}$. (d) Same as 1(c) for λ taking 512 values.



(e) Proximal Average of $|\cdot|$ and $g(x) := 0$ for $\lambda \in \{0, 0.5, 1\}$. (f) Same as 1(e) for λ taking 512 values.

FIGURE 1. Shapes of some operators of Convex Analysis applied to the function $f(x) = |x|$ with $0 \leq \lambda \leq 1$. The Moreau envelope and the proximal average are smooth while the Pasch-Hausdorff envelope is only Lipschitz.

The connection between Convex Analysis, image processing, differential calculus, and dynamical systems was noted by Maragos who named the resulting area differential morphology [168, 169]. Image Processing has long been using operators closely connected to Convex Analysis: the distance transform (a special case of the Moreau envelope [162, 164]), generalized distance transforms [81, 82] (regularization with nonquadratic kernels), and morphology operators like the dilation (resp. erosion) which corresponds to the inf-convolution (resp. deconvolution) operator of Convex Analysis [168, 169]. Partial Differential Equations (PDE) have also found applications in Image Processing e.g. the image segmentation with the Fast Marching and Level Set methods [234, 235]. The Lax and Hopf functions [119, 120, 231], which express the solution of a Hamilton-Jacobi PDE using Convex Analysis operators, are an example of the link between Convex Analysis and PDE. The computation of the convex envelope, motivated by the study of phase transition [106, 217, 180] and of the analysis of the distribution of chemical compounds [145, 146], is another example of how closely related these two fields are. Another well-known relation is the parallel between the Fourier transform and the Legendre conjugate [4, 47, 144, 155, 16, 98, 99, 59, 5]. In fact, the later plays the same role in a different algebra: the max-plus algebra. That framework has seen increased interest motivated by applications in network communication, neural networks, and discrete event systems. Classical linear and convex theory have been ported to the max-plus algebra [57, 59, 58] generating new results fundamentally related to Convex Analysis.

The applications presented in the present paper give a partial and personal overview of the wide range of fields benefiting from Computational Convex Analysis algorithms. In many instances, the same algorithm has been found independently by several authors working in different disciplines. One goal of the present paper is to point out the various connections so that future work can build on the present state-of-the-art instead of re-inventing existing algorithms.

The paper is organized as follow: Section 1 introduces the transforms: the Fenchel conjugate, inf-convolution and deconvolution operators, the Moreau envelope, the proximal average, and other related operators. Section 2 presents efficient algorithms to compute them: symbolic algorithms, numerical algorithms similar to the Fast Fourier Transform, and hybrid symbolic-numeric algorithms founded on piecewise linear-quadratic functions. Section 3 lists several applications in a wide variety of fields: Finite convex integration, network flow, phase transition, electrical networks, and robot navigation. Section 4 presents applications in image processing, computer vision, and differential morphology. Section 5 shows the link with Partial Differential Equations (PDE), while Section 6 puts the convex operators in the general framework of extremal algebra focusing on multifractal analysis, network communication, and discrete event systems. Finally, Section 7 concludes the paper.

1. Fundamental Convex Transforms

We first recall the most fundamental operators in Convex Analysis.

1.1. **The Fenchel Conjugate.**

The Fenchel conjugate (also named Legendre-Fenchel transform, Young-Fenchel transform, the maximum transform [29, 30, 32], or Legendre-Fenchel conjugate)

$$f^*(s) = \sup_{x \in \mathbb{R}^n} [\langle s, x \rangle - f(x)] \tag{1}$$

has long been studied in a wide range of fields for its duality properties.

Consider the following (Primal) optimization problem

$$p = \inf_{x \in \mathbb{R}^n} \{f(x) + g(Ax), \}$$

where $A \in \mathbb{R}^{mn}$, f (resp. g) is convex and lower semi-continuous on $\mathbb{R}^n$ (resp. on $\mathbb{R}^m$). Problem p is naturally associated, through Fenchel conjugation, to the dual problem

$$d = \sup_{z \in \mathbb{R}^m} \{-f^*(A^T z) - g^*(-z), \}$$

where A^T is the transpose of A. The Fenchel duality Theorem links both problems (see [40, Theorem 3.3.5], [228, Theorem 31.1], [25], [230, Example 11.41]).

Theorem (Fenchel's Duality Theorem)**.** *Assume $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$ with f, g and A as above. Then the following hold:*

(1) *Weak duality:* $p \geq d$.
(2) *Strong duality: If* $A(\operatorname{Dom} f) \cap \operatorname{int}\operatorname{Dom} g \neq \emptyset$, *then* $p = d$ *and the supremum defining* d *is attained.*
(3) *Primal solutions: If* z *is a solution to the dual, then the solutions to the primal are equal to the (possibly empty) set*
$$A^{-1}\partial g^*(z) \cap \partial f^*(A^T z),$$
where $\partial f(x) = \{s \in \mathbb{R}^m : \forall y \in \mathbb{R}^n,\ f(y) \geq f(x) + \langle s, y - x\rangle\}$ *is the convex subdifferential.*

Formulas to compute the conjugate for the main operations of Convex Analysis like addition, inf-convolution, maximum under a linear mapping, scalar multiplication, etc. have been investigated giving a complete conjugate calculus for Convex Analysis. Smoothness results linking the strict convexity to differentiability are also known, making the conjugate an invaluable tool in Convex Analysis. We refer to [228, 114] for general references on Convex Analysis (and to [230] for its generalization to variational analysis), the study of the Fenchel conjugate, and different formulations of Fenchel Duality Theorem.

Remark 1. *The name Legendre-Fenchel transform for the Fenchel conjugate above comes from the fact it is a generalization of the Legendre transform*
$$f^*(s) = \langle s, \nabla^{-1} f(s)\rangle - f(\nabla^{-1} f(s))$$
when the gradient of f *is invertible. When it is not, the Fenchel conjugate or its generalization: the slope transform (see references in Section 3.5), are used.*

The parallel between the Fenchel conjugate and the Fourier transform has long been known [47] (see Section 6.3 for more references).

1.2. **Inf-convolution and Deconvolution.** The inf-convolution [197, 201, 240, 241] (also called epi-addition) of two functions f and g is defined by
$$(f \oplus g)(x) := \inf_y [f(y) + g(x - y)].$$
It provides a very general transform giving rise to several regularization operators. Geometrically, it corresponds to the Minkowski addition of the epigraphs of the two functions. Under appropriate assumptions (convexity, lower semi-continuity (lsc), and properness), the infimal convolution reduces to several Fenchel conjugacy computations
$$(f \oplus g) = (f^* + g^*)^*. \tag{2}$$
The inverse of the inf-convolution operator is called the deconvolution [178, 113] of f by g and is defined by
$$(f \ominus g)(x) := \sup_y [f(x - y) - g(y)].$$
Under appropriate assumptions, the deconvolution of two convex functions reduces to computing several conjugates: $(f \ominus g) = (f^* - g^*)^*$. Mathematical Morphology has long been using erosion and dilation operators, which amounts to deconvolution and inf-convolution (see Section 4.4 for details).

The Pasch-Hausdorff envelope [230, Chapter 9], also called Lipschitz regularization, is a special case of inf-convolution with the norm function
$$(f \oplus c\|\cdot\|)(x) = \inf_y [f(y) + c\|x - y\|].$$
It has been studied for its Lipschitz regularization and Lipschitz extension properties [111, 112].

1.3. **Moreau Envelope.** The Moreau envelope of an extended real-valued function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$, (also called the Moreau–Yosida approximate, Yosida Approximate [13] or Moreau–Yosida regularization) corresponds to the inf-convolution with half the norm square
$$M_\lambda(x) := \left(f \oplus \frac{\|\cdot\|^2}{2\lambda}\right)(x) = \inf_{u \in \mathbb{R}^d} \left[f(u) + \frac{\|x - u\|^2}{2\lambda}\right]. \tag{3}$$
It has been studied extensively both theoretically and algorithmically for its regularization properties. Its origin goes back to the work of Yosida [257] on maximal monotone operators (it is also related to Tikhonov regularization [246]), and its behavior is well known in the field of convex analysis [198, 199, 200, 228]

and variational analysis [230, Chapter 12]. Under general conditions, M_λ is C^1 with Lipschitz continuous gradient, and critical points of f are fixed points of the proximal mapping

$$P_\lambda(x) := \underset{u\in\mathbb{R}^d}{\text{Argmin}}[f(u) + \frac{\|x-u\|^2}{2\lambda}]. \tag{4}$$

When f is convex lower semi-continuous and proper, the proximal mapping is a maximal monotone operator and its fixed points are the minimum of f. More precise smoothness of M_λ is known under various hypotheses on f [56, 102, 183, 187, 182, 216]. More recent developments have focused on extending the results to nonconvex functions through the notion of prox-regularity [20, 34, 33, 35, 213, 212, 181].

Considering that $M_\lambda(x)$ converges to $f(x)$ when λ decreases to 0, and shares the same critical points of f, the Moreau envelope is an attractive regularization transform. On the practical side, the proximal point algorithm exploits the fixed point property of the proximal mapping to converge to a minimum of f [229]. Its convergence properties are well known [147, 97], and variants have been introduced to speed up its convergence (see [48] and references therein). Extensions to non-quadratic kernels like entropy methods and Bregman distances have also been studied [75, 121, 244, 210, 41]. Bundle methods are intrinsically linked to the Moreau envelope (see [186], and [114, Chapter XV]). Recent developments in that direction focus on $\mathcal{VU}$-decomposition [150, 152, 151, 149, 148, 188, 189, 190, 191, 192, 193, 194, 195] to take advantage of both Newtonian and bundle algorithms.

We note that the computation of the Moreau envelope is equivalent to the computation of the Legendre–Fenchel conjugate as the following formulas shows [162]

$$M_\lambda f(x) = \frac{\|x\|^2}{2\lambda} - \frac{1}{\lambda}\left(\frac{\|\cdot\|^2}{2} + \lambda f\right)^*(x), \tag{5}$$

$$f^*(s) = \frac{\|s\|^2}{2} - \lambda M_\lambda\left(\frac{1}{\lambda}f - \frac{\|\cdot\|^2}{2\lambda}\right)(s), \tag{6}$$

where $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$, and $\lambda > 0$. So algorithms for computing one transform are trivially extended to compute the other.

1.4. **Other transforms.** The Lasry-Lions double envelope [142, 12] $h_{\mu,\lambda}$ is defined as several Moreau envelopes

$$h_{\mu,\lambda}(x) = -M_\mu(-M_\lambda(x)).$$

It is a smooth function [230, Proposition 12.62 p. 566]. Similarly the proximal hull (the proximal hull is different from the proximal mapping) can be written

$$g_\lambda(x) = h_{\lambda,\lambda}(x) = -M_\lambda(-M_\lambda(x)),$$

and so is also reducible to Moreau envelope computations.

More recently, the proximal average [24, 22, 23, 21, 165] of n functions $f_1,\dots,f_n$ is defined with combinations of Moreau envelopes

$$p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = -M_\mu(-(\lambda_1 M_\mu f_1 + \cdots + \lambda_n M_\mu f_n)),$$

where $\boldsymbol{f} = (f_1,\dots,f_n)$, $\boldsymbol{f}^* = (f_1^*,\dots,f_n^*)$, $\boldsymbol{\lambda} = (\lambda_1,\dots,\lambda_n)$, and $\mathfrak{q} = \frac{1}{2}\|\cdot\|^2$. It can also be computed as a combination of several Fenchel conjugates

$$p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \left(\lambda_1(f_1+\mu^{-1}\,\mathfrak{q})^* + \cdots + \lambda_n(f_n+\mu^{-1}\,\mathfrak{q})^*\right)^* - \mu^{-1}\,\mathfrak{q}.$$

Its key properties include been an homotopy between convex functions, and inheriting smoothness. It has been used to build counter-examples [24], and compute primal-dual symmetric antiderivative methods [23] (see also Section 3.1). The proximal average has also been generalized to a kernel average [26] with current research focusing on generalization to a Bregman average based on Bregman distances.

Generalization of the Fenchel conjugate such as the c-conjugate [178] can also be considered within our framework. Other generalizations involve considering different distances instead of the norm for the Pasch-Hausdorff envelope, or half the norm square for the Moreau envelope. For example, Bregman distances [43] $D(x,y) = f(x) - f(y) - (\nabla f(x), x-y)$ associated with some functions f, and divergence measures e.g. based on the Shannon entropy could be considered. Generalizations to quasi-convex or γ-convex functions fit also our framework.

2. Computer-Aided Convex Analysis

Introduction. While optimization algorithms avoid explicitly computing the conjugate, motivated by the study of some Hamilton-Jacobi partial differential equations, computational algorithms have been developed to compute it on grids. A log-linear algorithm named the Fast Legendre Transform (FLT for short, by analogy with the Fast Fourier Transform) was first introduced [44, 61, 159, 206, 236] to be subsequently improved by a linear-time algorithm: The Linear-time Legendre transform (LLT) [160]. Another linear-time algorithm, motivated by applications in image processing, was obtained by computing the Moreau envelope [69, 68, 80].

While fast algorithms have been the main strategy to compute convex transforms, different frameworks have also been investigated. A parametric framework was introduced in [115] and further expanded in [165]. It relies on the parametrization of the Fenchel conjugate to recover its graph up to affine parts. However, its restrictions led to the introduction of hybrid symbolic-numeric algorithms by considering the class of piecewise linear-quadratic (PLQ) functions [165].

Lately, a new strategy using graph-matrix calculus to compute only the graph of the transforms was introduced in [93] and further developed in [21]. For example, one can recover the graph of M_λ by quadrature from

$$\operatorname{gph} \nabla M_\lambda f = \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} \operatorname{gph} \partial f = \{(x+\lambda y, y) : (x,y) \in \operatorname{gph} \partial f\},$$

where I is the $n \times n$ identity matrix, $\operatorname{gph} \nabla M_\lambda f = \{(x, \nabla M_\lambda f(x)) : x \in \mathbb{R}^n\}$, $\operatorname{gph} \partial f = \{(x,y) : y \in \partial f(x)\}$, and

$$\partial f(x) = \{y \in \mathbb{R}^n : \forall x' \in \mathbb{R}^n, f(x') \geq f(x) + \langle y, x'-x \rangle\}$$

is the subdifferential of Convex Analysis. Whether graph-matrix calculus will provide efficient and competing algorithm is the subject of ongoing research.

We now recall what we consider the three main approaches to compute convex transforms: symbolic computation, fast algorithms, and PLQ-based algorithms.

2.1. **Symbolic Computation.** The natural strategy to compute the Fenchel conjugate is to differentiate the function under the supremum to obtain an equation satisfied by all the critical points. The difficulty resides in solving such an equation, which amounts to inverting the gradient of the function. For commonly used functions, symbolic computation software allows to perform some computation. Maple implementations were presented in [25] for the one-dimensional case, and in [39] for the multi-dimensional case. Large classes of functions can now be considered and some explicit formulas for the conjugate have been found using these packages. The packages offer a very efficient method to build some intuition, and to check one's computation.

However, the symbolic computation approach suffers from an intrinsic limitation: there may not be any closed form solution for the conjugate. Indeed, consider computing the conjugate of an even degree polynomial. If the degree is greater or equal to six, computing the conjugate involves finding the zeros of a polynomial of degree at least five, which may not admit a closed form. Moreover, in some cases, the explicit formula for the original function is not available e.g. the function is only available through a black box. So when the symbolic packages fail or are not applicable, ones turns to numerical computation, which is the subject of the next two subsections.

2.2. **Fast Algorithms.** The idea of a fast algorithm to compute the Fenchel conjugate was first formulated in [44], and later (independently) in [236]. It was subsequently investigated in [61, 206, 159] under the name Fast Legendre Transform. The complexity was subsequently improved in [160]. All subsequently developed algorithms focus on either computing the conjugate or the Moreau envelope. As we mentioned, both computations are equivalent.

The first step in any fast algorithm is to reduce computations to functions of one variable by noting that

$$M_\lambda(s_1,\ldots,s_d) = \inf_{x_1}[\frac{|s_1-x_1|^2}{2\lambda} + \cdots + \inf_{x_d}[\frac{|s_d-x_d|^2}{2\lambda} + f(x)]\ldots]. \tag{7}$$

A similar formula holds for the conjugate. Hence, all computations for functions in $\mathbb{R}^d$ can be reduced to computing several times transforms in $\mathbb{R}$.

The above "factorization" formula has been extended as a generalized distributive law to encompass various transforms beyond convex analysis [3]. In fact, by considering semi-rings instead of the usual $(\mathbb{R}, +, .)$ algebra, a common framework exists that encompasses the Fast Fourier Transform on any finite Abelian group, the fast Hadamard transform, Viterbi's algorithm, and Belief propagation algorithms [2, 137]. Among the many applications of such transforms, factor graphs have been applied to protein function [153] and, using the sum-product algorithm, to wireless communication [52].

2.2.1. *The Linear-time Legendre Transform (LLT) Algorithm.* The main idea behind the LLT algorithm is to note that computing the Fenchel conjugate is equivalent to computing the convex envelope (the convex envelope of a function f is the largest convex function that lies below f). While it is well-known that, for proper lsc convex functions, computing the conjugate of the conjugate gives the closed convex envelope, the LLT reverses the order: It first computes the convex envelope as a pre-processing step, and then computes the conjugate. More precisely, we first consider a discrete version of the transform

$$f_X^*(s) = \max_{x_i \in X}[s_j x_i - f(x_i)],$$

where the maximum is taken over $X = \{x_1, \dots, x_n\}$, and f_X^* is to be computed at all the slopes $s_j \in S = \{s_1, \dots, s_m\}$. The goal of the algorithm is to reduce the brute force computation of $O(nm)$ to $O(n+m)$. Since in practice we take $m = n$ to obtain a good numerical precision, the goal is to reduce the complexity from quadratic to linear.

Computing the lower convex envelope of the set of points $(x_i, f(x_i))$ in the plane can be achieved in linear time using the Beneath-Beyond algorithm [76, 161, 214], since the sequence x_i can be assumed sorted without any loss of generality: $x_i < x_{i+1}$. Now any point which is not a vertex of the convex hull, can be safely discarded since the maximum can never be attained at a point strictly in the interior of the epigraph, and vertices allow us to recover all points on the boundary of the epigraph. So it is sufficient to focuses on vertices of the convex hull.

After precomputation, we can assume the points $(x_i, f(x_i))$ are vertices of the convex hull. Hence the finite difference slopes $\frac{f(x_{i+1})-f(x_i)}{x_{i+1}-x_i}$ form an increasing sequence. Now computing the Fenchel conjugate amounts to merging the finite difference slopes with the slopes s_j, giving directly the point where the maximum is attained. More details on the LLT algorithm, including its proof of correctness, can be found in [160].

Note that no convexity assumption is made on the input data. (If the data is convex, the precomputation step can be skipped.) Convexity is explicitly introduced to speed up the computation, but the algorithm applies to nonconvex data.

Interestingly, the rank-one convex envelope computation also requires the computation of the convex envelope as an intermediate step [71].

2.2.2. *The Parabolic Envelope (PE) Algorithm.* The PE algorithm was introduced in [69] and later independently in [80]. It focuses on computing the discrete Moreau envelope

$$M_{\lambda,X}(s_j) = \min_{x_i \in X}[f(x_i) + \frac{\|x_i - s_j\|^2}{2\lambda}],$$

where as above $i = 1, \dots n$ and $j = 1, \dots m$. Assume $m = n$. The goal is again to reduce the quadratic brute force computation to linear.

The key step is to note that the computation amounts to finding the lower envelope of the family of parabola $s \mapsto f(x_i) + \frac{\|x_i - s\|^2}{2\lambda}$. Such envelope can be computed in linear time by adding parabola one at a time, since computing the intersection between two parabola can be done in constant time. See [163] for more details, comparison with other algorithms, and a Scilab [233] implementation.

2.3. **PLQ algorithms.** The PLQ algorithms were introduced specifically to compute composition of convex transforms such as the proximal average [165]. Such computation becomes very technical using fast transform algorithms since one has to keep track of the dual domain explicitly. Moreover, to obtain a reasonable numerical approximation of the result, one needs considerable knowledge of the dual domain of any intermediate transform. Such requirements make the fast algorithms cumbersome beyond a few compositions.

The key idea of PLQ algorithms is to explicitly represent convex functions. Fast algorithms manipulate points, so the underlying model is either a sample function, or a piecewise linear approximation. One reason the class of piecewise linear functions is not rich enough for our purpose, is the Moreau envelope of a piecewise linear function is no longer piecewise linear even for simple functions like the indicator of a single point. On the contrary, the class of piecewise linear-quadratic functions (functions whose domain can be expressed as the union of finitely many convex polyhedra, relative to each of which the function is at most quadratic) is closed under all major convex operations: addition, scalar multiplication, Fenchel conjugacy, and Moreau envelope. Hence, the computation of such transforms, or of compositions of such transforms, can be done symbolically. Moreover, there is no need to track the dual (or primal) domain of the function.

The PLQ algorithm to compute the conjugate amounts to matching each primal domain part with its dual counterpart, then computation is done symbolically. (See [165] for more details.) The price to pay for such simplicity is that we can no longer use the factorization formula, so computations beyond functions of one variable are the subject of active research.

2.4. **Nonconvex Extensions.** Several previously mentioned algorithms can handle nonconvex functions. The LLT and PE fast algorithms can be used to compute the conjugate and the Moreau envelope of nonconvex functions. In fact, considering that the conjugate is always a convex function that depends only on the convex envelope and using Formula (5), algorithms restricted to convex functions can be readily extended to nonconvex functions by first convexifying the function, then computing its conjugate (this is the principle of the LLT algorithm), and if needed its Moreau envelope. Hence, the PLQ algorithms can be extended to nonconvex functions as soon as one can compute the convex envelope of a PLQ function (which is a PLQ function) [248].

We now consider application areas benefiting from the previous framework.

3. Antiderivatives, Network Flow, Phase Transition, Electrical Networks, and Robot Navigation

3.1. **Finite Convex Integration.** Consider the following problem: given a finite set x_i^* of subgradients at points x_i, find a convex function f such that $x_i^* \in \partial f(x_i)$. The problem has been tackled in [141] under the name finite convex integration with links to linear programming. It can also be interpreted as a feasibility problem induced by a system of difference constraints [1, Section 4.5], which can be solved using shortest path algorithms.

Using tools from monotone operator theory, a solution with the additional constraint that the solution method should be symmetric with respect to convex duality was provided in [23] using the mid-point proximal average operator. We summarize their results to emphasize the role played by the PLQ algorithms in the numerical examples. (The availability of efficient algorithms also played a critical role in conjecturing the results.)

Assume x_i, x_i^* are given for $i = 1, \dots, n$. We say a function f is an antiderivative if $x_i^* \in \partial f(x_i)$ for $i = 1, \dots n$. The derivative is said *intrinsic* if in addition the function f does not depend on the order of the points x_i. A method m, which given a set $A = \{(x_i, x_i^*)\}$ produces an intrinsic antiderivative m_A is said to be *primal-dual symmetric* if m applied to the set $A^{-1} = \{(x_i^*, x_i)\}$ gives the conjugate of m applied to the set A:

$$m_{A^{-1}} = m_A^*. \tag{8}$$

The key idea behind primal-dual symmetric anti-derivative is that the input data is symmetric with respect to convex duality i.e. any antiderivative f satisfies $x_i^* \in \partial f(x_i)$ and $x_i \in \partial f(x_i^*)$. We would like the method to preserve that symmetry, which is the meaning of Formula (8).

While there are many antiderivatives, there is a priori no reason for primal-dual symmetric antiderivative methods to exist. However, it turns out that using the midpoint proximal average operator of two functions f_0 and f_1

$$\mathcal{P}(f_0, f_1) := \Big(\tfrac{1}{2}\big(f_0 + \tfrac{1}{2}\|\cdot\|^2\big)^* + \tfrac{1}{2}\big(f_1 + \tfrac{1}{2}\|\cdot\|^2\big)^*\Big)^* - \tfrac{1}{2}\|\cdot\|^2,$$

one creates primal-dual symmetric antiderivatives from any antiderivative using the fact that the midpoint proximal average of two antiderivatives is also an antiderivative. Given a method m producing intrinsic antiderivatives m_A for the set A, define the new method $\mathfrak{m}$ by

$$\mathfrak{m}_{\mathcal{A}} = \mathcal{P}\big(m_A, m^*_{A^{-1}}\big).$$

Then $\mathfrak{m}$ produces primal-dual symmetric antiderivatives [23].

The numerical computation of primal-dual symmetric antiderivative amounts to computing proximal averages, which can only be performed efficiently and robustly with PLQ algorithms.

3.2. **Network Flow.** The Linear Cost Network Flow on Series-Parallel Networks is another problem related to graph theory [250]. (We refer to references in [250] for its importance in combinatorial optimization.) Assume $\mathcal{G}$ is a strongly connected directed graph with vertex set $\mathcal{V}$, and edge set $\mathcal{E}$. Each edge $(i,j) \in \mathcal{E}$ is associated with a flow $x_{i,j}$, which is lower- and upper-bounded $-\infty < l_{i,j} \leq x_{i,j} \leq u_{i,j} < +\infty$, and a flow cost per unit $c_{i,j}$. The linear cost network flow problem is to minimize the total cost of the arc flows, subject to capacity and conservation constraints, in other words to solve

$$\begin{array}{lll} \text{minimize} & \sum_{(i,j)\in\mathcal{E}} c_{i,j}x_{i,j} & \\ \text{subject to} \quad \forall i \in \mathcal{V} & \sum_{\{j|(i,j)\in\mathcal{E}\}} x_{j,i} = \sum_{\{j|(i,j)\in\mathcal{E}\}} x_{i,j}, & \text{(Conservation condition)} \\ \forall (i,j) \in \mathcal{E} & l_{i,j} \leq x_{i,j} \leq u_{i,j}, & \text{(Capacity condition).} \end{array}$$

To solve the problem efficiently, nested sums and nested infimal convolutions are computed in [250]. The key idea to obtain an efficient algorithm is to sort grid nodes to compute the sum, and to sort the slopes to compute the inf-convolution. (A similar idea was used for the LLT algorithm except instead of inserting slopes, two sorted lists were merged.) The algorithm amounts to computing nested inf-convolution of piecewise-linear functions and is an alternative approach to the Fast Algorithms of Section 2.2 when nested operators are required. The resulting worst-case computation cost is $O(m \log m)$ where m is the number of arcs in the graph.

3.3. **Thermodynamics: Phase Transition.** In numerical simulation of multiphasic flows [106], a compressible flow with phase transition is considered. When the two different fluids are mixed, the mixture entropy is the sup-convolution of the entropies of the two phases, i.e. for positive pressures,

$$S(W) = \max_{W_1} S_1(W_1) + S_2(W - W_1),$$

where S (resp. S_1, S_2) is the entropy of the mixture (resp. of the first fluid, the second fluid), and $W = (M, V, E)$ is the vector of mass, volume, and energy for the mixture (W_1, W_2 correspond to the first and second fluid respectively). Assuming the entropies of the two fluids are known, the mixture entropy can be computed numerically using either a fast algorithm or the PLQ algorithms through Formula (2) since the functions W_1, and W_2 are concave. As mentioned in [106], such numerical computation is especially useful in the absence of closed form solutions.

Thermodynamics links to Convex Analysis run deeper than the above instance. The study of thermodynamic equilibrium is closely linked to the operation of convexification [217]. Consider the phase equilibrium problem at constant volume. It corresponds to

$$\min\{\sum_{i=1}^{q} \lambda_i E(d^i) \ : \ \sum_{i=1}^{q} \lambda_i d^i = d, \sum_{i=1}^{q} \lambda_i = 1, \lambda_i > 0\}$$

where E is a function associated with the Helmholtz free energy, $d = m/V$, m is the mole vector: $m^i > 0$ is the number of moles of the i^{th} fluid, and V is the volume. To recover the physical phases from the phase vector d, use $m^i = V_i d^i$, and $V_i = \lambda_i V$. The solution to the optimization problem is the convex envelope of E. (The convex envelope is the largest convex function upper bounded by E.) At a point d, the optimal solution d^i satisfies $\nabla E(d^i) = \nabla E(d^j)$, which represents the equality of the chemical potentials. It also satisfies

$$E(d^i) - \langle \nabla E(d^i), d^i \rangle = E(d^j) - \langle \nabla E(d^j), d^j \rangle,$$

which expresses the equality of pressure in each phase. The phase equilibrium at constant pressure problem consists in minimizing the Gibbs free energy instead of the Helmholtz free energy. The former is obtained as the Legendre transform of the later. Global minimization of the Gibbs free energy to solve the chemical and phase transition problem was studied in [180].

Another application explored the analysis of the distribution of chemical compounds in the atmosphere. In [145], a measure of roughness is defined, and is further applied in [146]. It consists in smoothing noisy data by rolling a parabola from above, then rolling another parabola from below, and considering the area between the parabolas as the measure of roughness. The efficient computation of the measure is performed with the LLT algorithm. (Intuitively, smoothing with a parabola corresponds to computing a Moreau envelope, which is equivalent to computing the Legendre conjugate.)

3.4. **Electrical Networks.** The study of a *mechanical system* consisting of two springs in series can be performed by computing the total potential energy of the system, which is the inf-convolution of the potential energy of each spring. Such systems with series and/or parallel strings are similar to electrical networks. In fact, the study of electrical circuits motivated the definition of the parallel addition and parallel subtraction operators, which corresponds to the inf-convolution and deconvolution of quadratic functions. Anderson [6, 7, 8] defined the parallel addition operator, and Mazure [174, 177, 175, 178, 176, 116] studied its properties from a Convex Analysis perspective (some of her results also apply for nonconvex functions). Consider [114, Example IV.2.3.8 p. 165]: an electrical circuit is made up of two generalized resistors A_1 and A_2 connected in parallel, and we want to find the equivalent resistor. By Maxwell's variational principle, a given current-vector $i \in \mathbb{R}^n$ is distributed among the two branches such that the dissipated power $\langle A_1 i_1, i_1\rangle + \langle A_2 i_2, i_2\rangle$ is minimal. So the real current distribution $i = \bar{i_1} + \bar{i_2}$ satisfies

$$\langle A_1\bar{i_1}, \bar{i_1}\rangle + \langle A_2\bar{i_2}, \bar{i_2}\rangle = \inf_{i_1+i_2=i}\{\langle A_1 i_1, i_1\rangle + \langle A_1 i_2, i_2\rangle\}.$$

When the matrices A_1 and A_2 are positive definite, the solution corresponds to the inf-convolution of two quadratic forms $f_j(x) = \langle A_j x, x\rangle/2$ for $j = 1, 2$. The result $(f_1 \oplus f_2)$ is the quadratic form associated with $A_{1,2} := (A_1^{-1} + A_2^{-1})^{-1}$. Similarly, the parallel subtraction corresponds to replacing a resistor with an equivalent circuit using two resistors in parallel.

A short history of parallel sum and shorted operators related to electrical networks is provided in the introduction to [10]. Applications to network connections are explored in [9, 196]. Extensions of the parallel sum have also been considered e.g. the quasi-projection operator [73] $!(A, B) = 2A(A + B)^+B$, where $+$ denotes the Moore-Penrose inverse, reduces to the harmonic mean $!(A, B) = 2(A^{-1} + B^{-1})^{-1}$ when A, and B are invertible. The parallel addition was also defined as the limit of the sequence $(x^{-1} + b_n^{-1})^{-1}$ when $b_n \to b$ in [17], and of the sequence $((A + \varepsilon I)^{-1} + (B + \varepsilon I)^{-1})^{-1}$ when $\varepsilon \downarrow 0$ in [138]. A generalization to connections through an axiomatic approach is given in [139] (including an interpretation of series-parallel networks).

The relation between the Moore-Penrose generalized inverse of the sum of two matrices and their parallel sum can be found in [84]. The variational characterization using the inf-convolution was investigated in [203] while the parallel sum of k matrices was studies in [245]. (See also [209] for further studies of the parallel sum.) A new regularization process based on parallel addition was studied in [211]. See also [222] for a generalization to monotone operators, and [77, 78] for another generalization. Parallel sum have also found applications in quantum effects [90].

3.5. **Robot Navigation.** Building from related work on the slope transform [72, 104, 105], the Legendre-Fenchel transform has been investigated to navigate a robot in a 2D space [135]. The LLT algorithm was adapted to handle discrete convex, concave, and nonconvex functions, then extended to polygons. In that context, the key property of the Legendre-Fenchel transform is its ability to detect contact between bodies using slopes. Another important property used in [135] is Formula (2) to reduce inf-convolution of convex functions to Legendre-Fenchel transforms.

While the framework of [135] focuses on piecewise linear functions and polygons and as such relies on results first established for the LLT algorithm, it could be extended to PLQ functions, which would make the addition operator trivial instead of explicitly generating the domain of the conjugate of the sum as the union of the domain of each conjugate. It involves extending the PLQ framework to nonconvex functions [248] and considering piecewise quadratic approximation of objects instead of polygons.

Robot navigation has long been performed using distance transforms. See for example [243] for a fast distance transform based heuristic path planning algorithm, [107, 108] for robot manipulator path planning. Both build from the work in [126] further developed in [123, 124]. See also the Jarvis' previous work on collision free path planning [125, 122].

Extensions of the orginal robot navigation problem include covert robotic [172] (move a robot while escaping sentinels' notice), real-time detection and navigation [256], outdoor robot navigation using vision [55], robot exploration with industrial applications [259, 260, 261], and multidimensional alignment [136].

An interesting link between distance transform and Hamilton-Jacobi equations is illustrated in [242], in which the robot path planning problem is solved by considering an Hamilton-Jacobi-Bellman equation instead of computing distance transforms.

We now turn our attention to applications arising from image science.

4. Image Processing, Computer Vision, and Mathematical Morphology

4.1. **Medical imaging.** Many image reconstruction methods rely on the Radon transform to reconstruct an image. Then the problem of detecting singularities in the image becomes important since those correspond to a crack in a solid e.g. an aircraft wing or an engine, or a rupture in a tissue in medical diagnosis. It turns out that the singularities of the radon transform of a function f are related to the singularities of the function f through the Legendre transform [219]: if a curve S is the graph of a smooth function $y = g(x)$, then the dual curve S^* in the appropriate coordinates (β, q) is the graph of the function $q = h(\beta)$, where $h = L(g)$ is the Legendre transform of g.

The Legendre transform is defined when the gradient is invertible by

$$L(f)(s) := \langle s, \nabla^{-1} f(s)\rangle - f(\nabla^{-1}(s)).$$

It coincides with the Legendre-Fenchel transform when in addition the function f is convex. When the gradient is not invertible, the Legendre transform may be multi-valued, and it has been generalized accordingly (see [219, Definition 1] and the slope transform [72, 105, 167] and references therein).

While the computation of the Legendre transform may be ill-posed, this is not the case for the Legendre-Fenchel transform, see [219, Section 4.3], which also lists various methods to compute the Legendre transform numerically (at a single point contrary to the fast algorithms of Section 2.2). The stable computation of the generalized Legendre transform is investigated in [218]. See also [220, 258] for further results on that topic.

The problem is generalized in [221], which considers the X-ray transform of a function f as the function which associates to each straight line l in $\mathbb{R}^3$, the integral of f over l with respect to the Lebesgue measure on l. The Radon transform uses planes in $\mathbb{R}^3$ instead of straight lines. The general case involves considering linear subspaces of arbitrary dimensions. As already mentioned, the main application of such investigation is computerized tomography when one looks for boundary of bones, or for holes in solids.

4.2. **Image Processing: Distance Transforms.** In image processing, distance transforms have been investigated for decades [232, 64] due to their diverse applications (see for example [38, 63, 247, 215] and references therein). For a binary image B defined as an application from $\{1,\dots,n\} \times \{1,\dots,m\}$ to $\{0,1\}$, the distance transform is the mapping that associates B with an $\{1,\dots,n\} \times \{1,\dots,m\}$ array D defined as follow. Assume B has at least one pixel p with $B[i] = 0$. For each pixel p in B, $D[i]$ contains the Euclidean distance to the closest pixel in B containing the value 0. In practice, to restrict computation to integer arithmetic, the square Euclidean distance transform is computed. For example, the squared distance transform of

$$\mathrm{B} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \text{ is } \mathrm{D}^2 = \begin{bmatrix} 2 & 1 & 2 & 5 & 8 \\ 1 & 0 & 1 & 4 & 5 \\ 2 & 1 & 2 & 1 & 2 \\ 5 & 4 & 1 & 0 & 1 \\ 8 & 5 & 2 & 1 & 2 \end{bmatrix}.$$

Several algorithms [38, 45, 53, 63, 173, 238, 109] were introduced to compute the Euclidean Distance Transform (EDT), and recent research focuses on simplifying the algorithms while still achieving linear-time complexity. Different properties were used to achieve linear-time. For example, the fact that the Euclidean distance transform computation is equivalent to computing the lower envelope of quadratic functions was exploited in [69, 68, 80] to achieve a simple linear-time algorithm. Other algorithms based on monotonicity or neighborhood properties also managed to achieve linear complexity [88, 238]. Most recently the relationship between the Moreau envelope and the Legendre conjugate was exploited in [162] to reduce the core of the distance transform computation to the LLT algorithm as follow. The squared distance transform is the application D^2 from $\{1, \dots, n\} \times \{1, \dots, m\}$ to the set of non-negative integers defined by

$$D^2(p) = \min_{q \in O} \|p - q\|^2,$$

where $O = \{q; B(q) = 0\}$ is the set of pixels with value 0 in B. Using the indicator function $I(p) = 0$ if $B(p) = 0$ and $+\infty$ otherwise, we find that the square distance transform is the Moreau envelope of I:

$$D^2(p) = \min_q [\|p - q\|^2 + I(q)].$$

Hence, distance transform algorithms are particular cases of discrete Moreau envelope algorithms.

Applications of distance transforms to pattern recognition, by computing the Hausdorff distance of two objects, have been considered, see for example [118] for a link with Voronoi surfaces, [156] for efficient algorithms, [89] for image matching, [127] for face detection, [262] for face structure extraction and recognition, and [42] for two-dimensional discrete morphing. Note also that the computation of Hausdorff distances motivated the PE algorithm [69].

4.3. **Image Processing: Generalized Distance Transforms.** We detail two contributions in computer vision and object recognition that rely on efficient algorithms for the generalized distance transform.

4.3.1. *Efficient Belief Propagation for Early Vision* [82]. Early vision problems such as stereo and image restoration have been solved using Markov Random Field (MRF) models. Since the resulting problems are NP hard, approximation techniques based on graph cuts and belief propagation have been used with high accuracy results in practice. However, both approaches are still computationally expensive especially compared with local methods that are faster but produce poorer results.

A general framework consists of finding a labeling function $f : p \mapsto f_p$ from the set of pixels $\mathcal{P}$ to the set of labels $\mathcal{L}$ (labels may correspond to disparities or intensities) by minimizing an energy function

$$E(f) = \sum_{p \in \mathcal{P}} D_p(f_p) + \sum_{(p,q) \in \mathcal{N}} W(f_p, f_q),$$

where $\mathcal{N}$ is the set of edges in the four-connected image grid graph. The general form of the function W is restricted to the particular case of $W(f_p, f_q) = V(f_p - f_q)$.

The max-product belief propagation (BP) algorithm can be used to find a labeling. It is an iterative algorithm that works by passing messages in parallel around the graph. Denoting $m^t_{p \to q}$ the message that node p sends to a neighboring node q at iteration t, it can be summarized as follow:

(1) Initialize $m^0_{p \to q}$ to 0.
(2) At each iteration t ($t = 1$ to T) compute

$$m^t_{p \to q}(f_q) = \min_{f_p} \left(V(f_p - f_q) + D_p(f_p) + \sum_{s \in \mathcal{N}(p) \setminus q} m^{t-1}_{s \to p}(f_p) \right).$$

(3) After T iterations, compute the belief vector

$$b_q(f_q) = D_q(f_q) + \sum_{p \in \mathcal{N}(q)} m^T_{p \to q}(f_q).$$

(4) Finally compute

$$f^*_q = \operatorname*{Argmin}_{f_q} b_q(f_q).$$

The key step in the algorithm for our purposes is Step (2): It requires computing a min convolution at each iteration. Using a fast algorithm, the quadratic computation cost is reduced to linear for specific functions V which, coupled with other optimization techniques, reduces the BP algorithm cost of $O(nk^2T)$ to $O(nk)$, where $n = |\mathcal{P}|$ is the number of pixels in the image, $k = |\mathcal{L}|$ is the number of possible labels for each pixel, and T is the number of iterations.

Felzenszwalb considers several models for the function V. The Potts model consists of a piecewise constant function $V(x) = 0$ when $x = 0$ and d otherwise. A direct approach leads to a linear time algorithm. For the linear model $V(x) = c|x|$, and the truncated linear model $V(x) = \min(c|x|, d)$, the computation is similar to computing a distance transform since it amounts to computing the min convolution with a linear cost. Note that the min convolution corresponds to a Pasch-Haussdorf regularization of Section 1.4.

Finally a quadratic model and a truncated quadratic model are considered. Both are equivalent to a Moreau envelope (Euclidean distance transform) and can also be computed in linear time. Felzenszwalb explains the algorithm and refers to [80, 70] for the fast algorithm.

More general distances could be used while still keeping a linear cost, e.g., any function V for which the intersection between two translations of its graph can be computed in constant time results in a linear time algorithm without making any convexity assumption. If the labeling function is a discretization of a convex function, then any convex function V could be used, since the LLT algorithm coupled with Formula (2) gives a linear-time algorithm.

4.3.2. *Pictorial Structures for Object Recognition* [81]. The paper focuses on recognizing generic objects in an image, and on learning how to recognize from example images. The best match is obtained by minimizing an energy function that measures a match cost for each part and a deformation cost for each pair of connected parts. Related problems include maximum a posteriori probability (MAP). It amounts to solving

$$L^* = \operatorname*{Argmin}_L \left(\sum_{i=1}^n m_i(l_i) + \sum_{(v_i,v_j)\in E} d_{ij}(l_i, l_j) \right).$$

While the minimization for arbitrary graphs $G = (V, E)$ and arbitrary functions m_i, d_{ij} is NP-hard, special cases can be solved efficiently, e.g. when the graph is a chain, a dynamic programming solution runs in $O(h^2 n)$, where n is the number of parts of the model and h the number of possible locations of each part.

By restricting the d_{ij} to the Mahalanobis distance between transformed locations

$$d_{ij}(l_i, l_j) = (T_{ij}(l_i) - T_{ji}(l_j))^T M_{ij}^{-1} T_{ij}(l_i) - T_{ji}(l_j)),$$

a minimization algorithm can be obtained that runs in $O(h'n)$, where h' is the number of grid locations in a discretization of the space of transformed locations given by T_{ij} and T_{ji}.

More precisely for an acyclic graph $G = (V, E)$, pick v_r an arbitrary node as the root of a tree. Denote by d_i the depth level of node v_i (the depth level of v_r is 0). For any vertex $v_j \neq v_r$, the best location given a location for its parent v_i is

$$B_j(l_i) = \min_{l_j} \left(m_j(l_j) + d_{ij}(l_i, l_j) + \sum_{v_c \in C_j} B_c(l_j) \right),$$

where C_j is the set of children of node v_j. Consequently a dynamic programming approach (computing B_j from the bottom up and then tracing the solution to get the argmin) gives a $O(nh^2)$ algorithm. Using generalized distance transforms as introduced in [80, 162, 164], the computation is reduced to a $O(nh)$ cost.

See also [62] for the application of fast algorithms in that context, and [51, 60] for a Convex Analysis point of view.

4.4. **Differential Morphology.** Image processing has long been using morphological operators to compute various transformations. The core operators are the dilation and the erosion operators

$$\begin{aligned} (f \oplus g)(x) &= \sup\nolimits_{y\in B}[f(y) + g(x - y)], \\ (f \ominus g)(x) &= \inf\nolimits_{y\in B}[f(y) - g(x - y)], \end{aligned}$$

which correspond to the inf-convolution and deconvolution operators of convex analysis. See for instance [251] on the Minkowski addition operators for sets. Composition of these give smoothing filters like the opening $f \mapsto ((f \ominus g) \oplus g)$ and the closing $f \mapsto ((f \oplus g) \ominus g)$ operators. From dilation, one can define the morphological gradient, which is so important in edge detection for image segmentation. The link between mathematical morphology in image processing, and Convex Analysis was noticed by Maragos in [168] who also noted the connection with partial differential equations like the Hamilton-Jacobi and the Eikonal equation. Maragos also made the connection with the Legendre-Fenchel transform through the slope transform [167, 104]. He further investigated the link with PDEs in [170] and in [171], which makes the link between distance transforms and PDEs, using level set methods, while [162, 164] took the reverse view of using fast algorithms from Section 2.2 to compute distance transforms. The connection to Hamilton-Jacobi equations was also made in [11, 239], and to the Eikonal equation in [128].

More traditional algorithms to compute dilation and erosion were presented in [254] for binary images, and in [255] in a broader context. Other efficient algorithms for morphological operators were presented in [91] (see also [74, 204]). A technique to compute the erosion using the FFT was presented in [251]. More connection between morphology and Convex Analysis were used in [37] with an explanation of the relationship using the max-plus algebra in [47].

5. Partial Differential Equations

While links between Convex Analysis and Partial Differential Equations (PDE) are well-known, recent work focused on using efficient numerical methods in one field to solve a problem in the other. In this section, we first explain how Convex Analysis helps finding solution to an Hamilton-Jacobi PDE. Conversely, we then explain how efficient PDE solvers help computing a fundamental Convex Analysis transform: the convex envelope.

5.0.1. *Lax-Hopf Formula.* The Lax and the Hopf functions are explicit solutions of

$$\begin{cases} \frac{\partial u}{\partial t} + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0, \infty), \\ u(\cdot, 0) = g(\cdot) & \text{in } \mathbb{R}^n, \end{cases}$$

when either H or g is convex (where Du stands for the derivative of u with respect to the space variable x). They are defined as follow.

$$u_{\text{Lax}}(x,t) = \inf_{y \in \mathbb{R}^n} \sup_{q \in \mathbb{R}^n} [g(x-y) + \langle y, q \rangle - tH(q)] = (g \oplus (tH)^*)(x),$$
$$u_{\text{Hopf}}(x,t) = \sup_{q \in \mathbb{R}^n} \inf_{y \in \mathbb{R}^n} [g(x-y) + \langle y, q \rangle - tH(q)] = (g^* + tH)^*(x).$$

The study of their properties using tools from Convex Analysis was performed in [119] (see also [120]). The formulas were extended further in [231]. The extension to quasiconvex functions was performed in [19] while Bardi et al. [18] considered the nonconvex nonconcave case. The use of fast algorithms to compute the solutions numerically was investigated in [61]. Considering there are numerous results on Hamilton-Jacobi-Bellman equations, we refer to [119] for an introduction from the point of view of Convex Analysis.

While the Hopf function can be computed in linear time as several conjugates, none of the current algorithms allows the computation of the inf-convolution in linear time. (We cannot use Formula (2) since the function u_0 is not assumed convex.) A nonlinear-time algorithm was proposed in [61] but it does not scale well with the dimension.

The FLT and the faster LLT algorithms have also been used in efficient numerical simulations of the Burgers equation. For example, in [252] an adhesion model is investigated and numerical simulations (using the FLT) are performed to compare theories on mass distribution in the universe. The tools used are the Fenchel conjugate, the convex envelope, and other Convex Analysis arguments. The same algorithm is key to numerous numerical simulations for the Burgers' equation [14, 27, 86, 87, 100, 101, 205].

5.0.2. *Convexification.* For a locally bounded function $u_0 : \mathbb{R}^N \to \mathbb{R}$, the system

$$\begin{cases} \frac{\partial u}{\partial t} = \sqrt{1 + \|Du\|^2} F(Du, D^2 u) & \text{for } (t,x) \in (0,\infty) \times \mathbb{R}^N, \\ u(0,x) = u_0(x) & \text{for } x \in \mathbb{R}^N, \end{cases}$$

models the motion of the graph of the solution $u(t,\cdot)$ in the normal direction at each point, with speed $F(Du, D^2u)$. Using $F(Du, D^2u) = \min(0, \lambda_{\min}(D^2u))$, where $\lambda_{\min}$ denotes the smallest eigenvalue of D^2u, and under the appropriate assumptions, the solution $u(t,\cdot)$ converges to the convex envelope of u_0 when $t \to \infty$ (see [253]). While finite difference methods were used to compute the convex hull, the reverse could also be done: using computational geometry algorithms to compute the solution to the partial differential equation above.

More recently, the convex envelope was found to be the solution of a nonlinear obstacle problem. The convex envelope u of the function $g : \mathbb{R}^n \to \mathbb{R}$ is a viscosity solution of

$$\max(u(x) - g(x), -\lambda_1[u](x)) = 0,$$

where $\lambda_1[u](x)$ is the smallest eigenvalue of the Hessian $D^2u(x)$ [208]. That formulation was further studied in [207] to obtain a PDE-based numerical algorithm to compute the convex envelope.

The convex envelope is also the solution of

$$\min \int_{[a,b]} \sqrt{1 + \dot{u}^2(s)}ds$$

under the constraints $u \in W^{1,1}[a,b]$, $u \le f$ on $[a,b]$, $u(a) = f(a)$, and $u(b) = f(b)$. The problem can then be discretized and, assuming the initial function u is usc on $[a,b]$, its solution converges uniformly to the convex envelope [129].

In [46], the convex envelope of a function φ is computed as the solution to the problem

$$\varphi^{**}(\alpha) = \inf_{v \in W_0^{1,\infty}(\Omega)} \frac{1}{|\Omega|} \int_\Omega \varphi(\alpha + \nabla v(x))dx,$$

while the convex envelope of a function f is approximated in [110] as the solution of

$$\min \frac{1}{2} \int_\Omega (u - f)^2$$

with the constraints $u \in BV^2(\Omega)$, $u = \hat{f}$ on $\partial\Omega$, and $u \le f$ on $\partial\Omega$. (BV^2 is the space of bounded second variation, and $\hat{f}$ is the set of Dirichlet data on $\partial\Omega$, which is assumed known a priori.)

Other recent work on computing the convex envelope has focused on polynomials for which the computation of the convex envelope can be transformed into a minimization problem on a set of probability measures [184]. The later can be reduced to a semidefinite programming problem corresponding to the Hamiltonian of a convex formulation of the problem. See also [185] for another application of the method of moments and its relation with the convex envelope.

PDE have also been used for global optimization through a smoothing method linked with the convex envelope. More precisely, a cost function f of an unconstrained global optimization problem is smoothed with a function $u : [0,T) \times \mathbb{R}^n \to \mathbb{R}$ satisfying

$$\frac{\partial u}{\partial t}(t,x) = \Delta u(t,x) - \max(0, \Delta u(t,x)),$$

with $0 < t < T$ and $u(0,\cdot) = f$ (Δu denotes the Laplacian of u with respect to x). Under suitable assumption, the function $u(t,\cdot)$ converges to the convex envelope of f. Using the fact that global minima of the convex envelope are the same as the original function, an algorithm is devised to compute the global minimum [143].

5.0.3. *Interface Propagation.* The search for numerical methods to solve Hamilton-Jacobi equations has given rise to very efficient numerical schemes to compute curve evolutions and interface propagations. The Fast Marching method, the Level Set method, and the Fast Sweeping method are examples of such methods with a wide range of applications [235] (see also [249] for the application of level set methods to image science). In our context, these methods have been considered to compute distance transforms in image processing, which are a particular case of Moreau envelope. They could also be used to compute the convex envelope. Note that the main advantage of such methods is not the speed of computation, since they are outperformed by computational geometry and fast algorithms, but their potential ability to build a nonuniform grid on which the convex envelope is approximated.

Moreover, recent investigation into the fast sweeping methods for static Hamilton-Jacobi equations require the computation of the Legendre transform [130], which is performed symbolically or numerically using the fast transform algorithms. The Fenchel conjugate allows the transformation of the evolutive Hamilton-Jacobi equation of the first order into a Bellman equation with a finite horizon control problem [79]. The Hamiltonian can then be computed using the fast algorithms. (Both articles refer to the original Fast Legendre Transform algorithm, which has since been superceded by the Linear-time Legendre Transform algorithm.)

6. Multifractal Analysis, Network Communication, and Extremal Algebra

6.1. **Multifractal Analysis.** Fractal processes have allowed significant advances in a variety of fields, e.g. turbulence theory [237], stock market modeling, image processing, medical data, geophysics, network modelings [36, 225] (and in particular TCP traffic [154]), computer worms in network [54], analysis of paleoclimatic records [131], etc. (see also references in [223]).

6.1.1. *Multifractal Processes.* We give a rough introduction to the functions of interest in multifractal analysis leading to the definition of the Legendre spectrum below. The interested reader is referred to [223] for the details. Properties of the Legendre conjugate of interest to multifractal analysis are introduced in [224].

A fractal process $Y(t)$ has a non-integer degree of differentiability formalized as its local Hölder exponent. More formally, $Y \in C_t^h$ if there is a polynomial P_t with

$$|Y(u) - P_t(u)| \leq C|u-t|^h,$$

for u sufficiently close to t. Then the degree of local Hölder regularity of Y at t is $H(t) := \sup\{h \mid Y \in C_t^h\}$. If the Taylor polynomial of degree $\lfloor H(t) \rfloor$ exists, then it is equal to P. When the approximating polynomial is a constant: $P_t(u) = Y(t)$, $H(t)$ can be computed by introducing

$$h(t) := \liminf_{\epsilon \to 0} \frac{1}{\log_2(2\epsilon)} \log_2 \sup_{|u-1|<\epsilon} |Y(u) - Y(t)|,$$

with the convention $\log(0) = -\infty$. We always have $h(t) \leq H(t)$, and when $h(t)$ is not a positive integer and P_t is a constant, we have $h(t) = H(t)$.

To study the geometry or local regularity of Y, the local analysis focuses on quantifying which values of $H(t)$ appear on a given path of the process Y. For each a, define the sets $E_h^{[a]} := \{t \mid H(t) = a\}$. The function Y is said to have a *right multifractal structure* if the sets $E^{[a]}$ are highly interwoven. The function $a \mapsto dim(E^{[a]})$ gives an insight into the structure of Y, and has been named the *multifractal spectrum* of Y.

A global analysis of Y is performed using the concept of box-dimension. Introduce

$$h_k^{(n)} := -\frac{1}{n} \log_2 \sup\{|Y(s) - Y(t)| \ : \ (k-1)2^{-n} \leq s \leq t \leq (k+2)2^{-n}\}$$

to define the *grain (multifractal) spectrum* as

$$f(a) := \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log N^{(n)}(a, \epsilon)}{n \log 2},$$

where $N^{(n)}(a,\epsilon) = \#\{k \ : \ |h_k^{(n)} - a| < \epsilon\}$. Using Large Deviation Principles, we can interpret the coarse spectrum f by studying the scaling of "sample moments" through the partition function

$$\tau_h(q) := \liminf_{n \to \infty} \frac{\log S_n^{(n)}(q)}{-n \log 2} \text{ where } S_n^{(n)}(q) := \sum_{k=0}^{2^n-1} 2^{-nqh_k^{(n)}},$$

defined for all $q \in \mathbb{R}$. Under the appropriate assumptions, one can prove that τ is the (concave) Legendre-Fenchel transform of f

$$\tau_h(q) = f^*(a) := \inf_a (qa - f(a)).$$

The function τ^* is referred to as the Legendre spectrum.

A fast Legendre algorithm is included in the toolbox FracLab [85], which focuses on multifractal analysis under Scilab and Matlab.

6.1.2. *Object detection through the Legendre spectrum.* Using multifractal analysis, the detection of artificial objects within natural environments was studied by noting that artificial objects have a wider Legendre spectrum than natural ones. Hence, a given image is subdivided into several subareas for which the Legendre spectrum is computed. The results are then compared to locate any artificial object [49]. While fast algorithms can be used to compute the Legendre spectrum, the authors of [49] used a combination of numerical summations and limits coupled with the explicit Legendre conjugate formula for the function considered.

6.2. **Network Communication.** In [117], a simple communication network with a single input function $x(k)$, and a single output function $y(k)$ is considered. The output is computed as the infimal convolution $y = (h \oplus x)$ of the input with the response characteristic function, which is network and protocol dependent. The goal is to recover the response characteristic by deconvolution: $h = (y \ominus x)$. Since the functions may not be convex, the Legendre transform is extended to handle nonconvex/nonconcave data as the slope transform

$$\mathcal{L}[x](s) = \{x(u^*) - su^* \mid s = \frac{dx}{du}(u^*)\},$$

which is set-valued. Then specialized computation are performed to evaluate the deconvolution operation associated with the slope transform. Note that the formulation of the extended Legendre transform is close to the parametric Legendre transform algorithm introduced and studied in [115].

Network calculus is a theory of deterministic queuing systems found in computer networks. It is the equivalent of system theory for which the usual $(\mathbb{R}, +. \cdot)$ algebra has been replaced with the commutative dioid $(\mathbb{R} \cup \{+\infty\}, \min, +)$. The usual convolution operation now becomes the inf-convolution (called min-plus convolution), its dual the deconvolution (called min-plus deconvolution), while the equivalent of the Fourier transform is the Fenchel conjugate [144]. The resulting theory has very practical applications, e.g. it covers the TCP protocol [15]. While some authors have noted the connection between Network calculus and Convex Analysis, it does not appear that the full power of Convex Analysis, e.g. support functions and subadditive functions [114], has been fully exploited yet.

We summarize the presentation in [83]. The foundation of network calculus are the min-plus convolution (inf-convolution) and the min-plus deconvolution

$$(f \oplus g)(t) = \inf_u f(t-u) + g(u), \quad (f \ominus g)(t) = \sup_u f(t+u) - g(u).$$

Network calculus also assumes $t \geq u \geq 0$, $t \geq 0$, and $u \geq 0$. We recall the characteristics of a network that can be computed with network calculus.

Arrival curves $\alpha(t)$ give upper bounds on arrival function. An arrival function $F(t)$ is said to conforms to an arrival curve $\alpha(t)$, if for all $t \geq 0$, and for all $s \in [0, t]$, $\alpha(t-s) \geq F(t) - F(s)$. The leaky-bucket algorithm defines a typical constraint on incoming flows by enforcing the arrival curve $\alpha(t) = 0$ for $t = 0$, and $\alpha(t) = b + rt$ for $t > 0$.

A lossless network element with input arrival function $F(t)$ and output arrival function $F'(t)$ offers a *service curve* $\beta(t)$ if for all $t \geq 0$, there is $s \in [0, t]$ with $F'(t) - F(t) \geq \beta(t-s)$.

The min-plus algebra gives bounds on these quantities. The service curve $\beta(t)$ of the *concatenation* of n service elements with service curves $\beta_i(t)$ is $\beta(t) = \oplus_{i=1}^n \beta_i(t)$. A service element $\beta(t)$ with input bounded by $\alpha(t)$ admits a bound on its output $\alpha'(t)$ given by $\alpha' = (\alpha \ominus \beta)$.

Diving deeper in min-plus algebra, one finds that eigenfunctions with respect to min-plus deconvolution are the affine functions admitting the Fenchel conjugate as eigenvalue. In particular, it transforms inf-convolutions in additions and deconvolutions in subtractions. As such it becomes very advantageous to work in the Legendre domain. For example, the conjugate service curve $B(s)$ of the concatenation of service elements is the sum of the individual conjugate service curves $B(s) = \sum_{i=1}^n B_i(s)$. This is Formula (2). Interestingly, Fenchel's duality Theorem gives a backlog bound in the Legendre domain [83].

6.3. **Max-Plus, Tropical, Idempotent, and Extremal Algebras.** By replacing the usual arithmetic operations with new operations with idempotent property one obtains a rich structure with numerous applications. Important semi-rings are the max-plus algebra $\mathbb{R}_{\max} := (\mathbb{R} \cup \{-\infty\}, \max, +)$, and the min-plus algebra $\mathbb{R}_{\min} := (\mathbb{R} \cup \{+\infty\}, \min, +)$; although many other semi-rings have been studied e.g. $(\mathcal{C}, \oplus, \odot)$ where $\mathcal{C}$ is the set of all convex compact subsets of $\mathbb{R}^d$ equiped with the Minkowski operations: $A \oplus B = \mathrm{co}(A \cup B)$

is the convex hull of the union, and $A \odot B = \{x \,:\, x = a + b, \text{ where } a \in A, b \in B\}$. The new addition $\oplus$ is idempotent: for all x, $x \oplus x = x$. Note that the terminology varies: idempotent semi-rings are sometimes called tropical semi-rings, idempotent semi-fields, minimax algebra, or extremal algebras.

The idempotent semi-rings can be seen as the limit of the usual algebra under various transforms, e.g. $u \oplus_h v = h \ln(\exp(u/h) + \exp(v/h))$ gives $u \oplus_h v \to \max(u, v)$ as $h \to 0$. The passage from $\mathbb{R}$ to $\mathbb{R}_{\max}$ (or min) is sometimes called the Maslov dequantization or Cole-Hopf transformation [155]. In the new algebra, the role of the Fourier transform and the convolution operators are played by the Legendre-Fenchel transform and the inf-convolution operator.

More result on idempotent calculus, in particular with links to Hamilton-Jacobi equations, can be found in [67, 66]. While we refer to the recent survey [155] for more information and further references, and to [16] (especially Section 9.4) for an introduction in the context of discrete event systems (see also the introduction to a nonlinear theory for discrete event systems based on Max-plus algebra in [98, 99]), let us emphasize the following points related to our context.

The area of application of idempotent semi-rings is wide ranging: discrete mathematics, computer science, computer languages, linguistic problems, finite automata, optimization problems on graphs, discrete event systems and Petri nets, stochastic systems, evaluation of computer performance, computational problems, mathematical economics, etc. See references in [155]. Specific research has also focused on car-traffic laws [157, 158].

Many equivalent results from the usual algebra have been obtained, sometimes simplified due to the idempotent property. Numerous concepts have been investigated: the equivalent of the Riemann sum allows one to define a corresponding measure theory, while abstract convex sets have given rise to global algorithms for Lipschitz functions by extending the cutting plane algorithm [28]. A strong motivation to study such semi-rings comes from the fact that some nonlinear equations in the regular algebra become linear in the idempotent semi-ring, e.g. the Hamilton-Jacobi equation is linear over $\mathbb{R}_{\max}$. An abstract linear algebra theory has been studied within which the properties of the Legendre-Fenchel transform allow to hugely reduce the cost of computation for Hamilton-Jacobi equations [179]. More recent work has focused on geometry properties like convexity [59]. Algorithms also have their counterpart, and generic implementations over abstract semi-rings have been studied, while generalized linear algebra methods like the Jacobi Gauss-Seidel, and Gauss-Jordan method correspond to path-finding problems [50]. More details on semirings can be found in the monographs [92, 94, 95] while [96] gives an historial perspective with precise naming of the various semi-rings.

The counterpart to the Fourier transform is the Legendre-Fenchel transform, which strongly highlights the importance of the fast algorithms of Section 2.2, especially the LLT which can be seen as a counterpart of the FFT. The huge importance of the Fourier transform in signal analysis is directly translated to the critical importance of the Legendre-Fenchel transform in Convex Analysis and Optimization. The link between the Fourier, Legendre, and Cramer transformed was studied in [47] making a connection between linear and morphological system theory, the later been seen as linear system theory in the max-plus algebra. Connection between the Fourier Transform and the Fenchel conjugate were also made in [166] through the slope transform.

One original motivation to introduce the maximum transform was to transform inf-convolution into addition [30] with view toward applications in resource allocation [132]. It was later applied to nonlinear knapsack problems for which Formula (2) mitigated the curse of dimensionality [202].

Morphology Neural Networks is another field benefiting from nonstandard algebra. In a classical neural net, each node combines information by multiplying output values and corresponding weights, and summing. However, in a morphology neural net, values and corresponding weights are added, and then the maximum value is taken. The effect is to perform computations in the $\mathbb{R}_{\max}$ algebra instead. In essence, the linear operations are performed in a different algebra. Morphology neural networks arose from applications in image processing, which is not surprising considering the relation between classical dilation and erosion operators in image processing and the $\mathbb{R}_{\max}$ algebra. We refer to [65] for an introduction to morphology neural networks with some initial applications, and to [227] for another introduction with a computation of the capabilities of such neural network.

7. Conclusion

We presented core convex transforms, Computational Convex Analysis algorithms to compute them, and a wide range of application areas using them. While additional applications can be found in the literature [241], let us emphasize a few more fields. Economics has long been using convex analysis tools. Consider the following simple economic example: a person buys quantities of a product from two manufacturers, with prices depending on the quantities bought. Minimizing the total cost amounts to computing the inf-convolution of the costs. The general dynamic programming (DP) model for linear transition benefits from conjugate duality, which reduces the curse of dimensionality by reformulating the problem as a recursive sequence of inf-convolutions, and computing its dual in which inf-convolutions are reduced to additions [133]. Similarly a general discrete resource allocation problem reduces to a sequence of inf-convolutions, which in the dual become additions [134]. Note that the reduction of inf-convolution to addition was already noted in [31] in the context of the Maximum Transform. A family of energy minimization problems that take advantage of fast algorithms to compute distance transforms include the Viterbi algorithm for Markov models, and max-product belief propagation that have been used for vision problems[80].

We note the following directions for future research.

New Convex Analysis transforms, like the kernel average [26], have been recently introduced that require extension of current algorithms. While some transforms like the Moreau envelope and the Fenchel conjugate can be computed efficiently for convex and nonconvex functions, the efficient computation for others is limited to convex functions e.g. the inf-convolution. There are also important applications that require the computation of the closest closed convex function of a nonconvex function i.e. to project on the cone of closed convex functions [140], which is closely linked to the problem of shape-preserving interpolation and approximation.

From the application perspective, researchers should use the most efficient algorithms e.g. the LLT algorithm instead of the FLT algorithm, and use the power of the Convex Analysis machinery. Rifkin and Lippert contribution to Machine Learning [226] is one such example. It finds new results by using Fenchel duality, instead of the classical Lagrangian duality, coupled with Tikhonov regularization.

Acknowledgments

The author would like to thank Pr. J.-B. Hiriart-Urruty for a careful reading of the manuscript which greatly improved the paper, Pr. W. Hare for pointing out several typos, and Pr. H. H. Bauschke for his continuous support and encouragements.

The paper was initially announced as a companion paper of [163] under the title *Fast Moreau Envelope Computation II: Applications*.

References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network flows*, Prentice Hall Inc., Englewood Cliffs, NJ, 1993. Theory, algorithms, and applications.

[2] S. M. Aji and R. J. Mceliece, *The generalized distributive law*, Information Theory, IEEE Transactions on, 46 (2000), pp. 325–343.

[3] S. M. Aji and R. J. McEliece, *The generalized distributive law and free energy minimization*, in Proc. 39th Annu. Allerton Conf. Communication, Control, and Computing, Oct. 2001, pp. 672–681. Monticello, IL.

[4] M. Akian, J.-P. Quadrat, and M. Viot, *Bellman processes*, in 11th International Conference on Analysis and Optimization of Systems : Discrete Event Systems, S. Verlag, ed., vol. 199 of Lecture notes in control and information sciences, 1994.

[5] M. Akian, J.-P. Quadrat, and M. Viot, *Duality between probability and optimization*, in Idempotency (Bristol, 1994), vol. 11 of Publ. Newton Inst., Cambridge Univ. Press, Cambridge, 1998, pp. 331–353.

[6] W. N. Anderson, Jr. and R. J. Duffin, *Series and parallel addition of matrices*, J. Math. Anal. Appl., 26 (1969), pp. 576–594.

[7] W. N. Anderson, Jr., R. J. Duffin, and G. E. Trapp, *Parallel subtraction of matrices (Hermitian semidefinite)*, Proc. Nat. Acad. Sci. U.S.A., 69 (1972), pp. 2530–2531.

[8] W. N. Anderson, Jr., T. D. Morley, and G. E. Trapp, *Characterization of parallel subtraction*, Proc. Nat. Acad. Sci. U.S.A., 76 (1979), pp. 3599–3601.

[9] W. N. Anderson, Jr. and G. E. Trapp, *Inequalities for the parallel connection of resistive n-port networks*, J. Franklin Inst., 209 (1975), pp. 305–313.

[10] J. Antezana, G. Corach, and D. Stojanoff, *Bilateral shorted operators and parallel sums*, Linear Algebra Appl., 414 (2006), pp. 570–588.

[11] A. B. Arehart, L. Vincent, and B. B. Kimia, *Mathematical morphology: The Hamilton-Jacobi connection*, in Computer Vision, 1993. Proceedings., Fourth International Conference on, 1993, pp. 215–219.

[12] H. Attouch and D. Azé, *Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry-Lions method*, Ann. Inst. H. Poincaré. Anal. Non Linéaire, 10 (1993), pp. 289–312.

[13] H. Attouch and R. J.-B. Wets, *A convergence theory for saddle functions*, Trans. Amer. Math. Soc., 280 (1983), pp. 1–41.

[14] E. Aurell, S.-g. N. Gurbatov, and S.-g. I. Simdyankin, *Numerical proof of self-similarity in Burgers' turbulence*, Tech. Rep. http://arxiv.org/abs/patt-sol/9602005, arXiv.org, 1996.

[15] F. Baccelli and D. Hong, *Tcp is max-plus linear and what it tells us on its throughput*, in SIGCOMM '00: Proceedings of the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, New York, NY, USA, 2000, ACM Press, pp. 219–230.

[16] F. L. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and linearity*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester, 1992. An algebra for discrete event systems.

[17] H. Bae and Y. Lim, *Parallel addition and fixed points of compressions on symmetric cones*, Math. Nachr., 246/247 (2002), pp. 20–30.

[18] M. Bardi and S. Faggian, *Hopf-type estimates and formulas for nonconvex nonconcave Hamilton-Jacobi equations*, SIAM J. Math. Anal., 29 (1998), pp. 1067–1086 (electronic).

[19] E. N. Barron, R. Jensen, and W. Liu, *Explicit solution of some first-order PDE's*, J. Dynam. Control Systems, 3 (1997), pp. 149–164.

[20] H. H. Bauschke, P. L. Combettes, and D. Noll, *Joint minimization with alternating Bregman proximity operators*, Pac. J. Optim., 2 (2006), pp. 401–424.

[21] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, *The proximal average: Further properties*, tech. rep., University of British Columbia, 2007.

[22] H. H. Bauschke, Y. Lucet, and M. Trienis, *How to transform one convex function continuously into another*, SIAM Review, (2006). Accepted for publication.

[23] H. H. Bauschke, Y. Lucet, and X. Wang, *Primal-dual symmetric intrinsic methods for finding antiderivatives for cyclically monotone operators*, SIAM J. Control Optim., (2006). Accepted for publication.

[24] H. H. Bauschke, E. Matoušková, and S. Reich, *Projection and proximal point methods: Convergence results and counterexamples*, Nonlinear Anal., 56 (2004), pp. 715–738.

[25] H. H. Bauschke and M. von Mohrenschildt, *Symbolic computation of Fenchel conjugates*, SIGSAM Bull., 40 (2006).

[26] H. H. Bauschke and X. Wang, *The kernel average of two convex functions and its application to the extension and representation of monotone operators*, Submitted, (2007).

[27] J. Bec, Frisch, U., and K. Khanin, *Kicked Burgers turbulence*, J. Fluid Mech., 416 (2000), pp. 239–267.

[28] G. Beliakov, *A review of applications of the cutting angle methods*, in Continuous optimization, vol. 99 of Appl. Optim., Springer, New York, 2005, pp. 209–248.

[29] R. Bellman and W. Karush, *On a new functional transform in analysis: the maximal transform*, Bull. Amer. Math. Soc., 67 (1961), pp. 501–503.

[30] R. Bellman and W. Karush, *Mathematical programming and the maximum transform*, Journal of the Society for Industrial and Applied Mathematics, 10 (1962), pp. 550–567.

[31] R. Bellman and W. Karush, *Functional equations in the theory of dynamic programming. XII. An application of the maximum transform*, J. Math. Anal. Appl., 6 (1963), pp. 155–157.

[32] ———, *On the maximum transform*, J. Math. Anal. Appl., 6 (1963), pp. 67–74.

[33] F. Bernard and L. Thibault, *Prox-regularity of functions and sets in Banach spaces*, Set-Valued Anal., 12 (2004), pp. 25–47.

[34] ———, *Prox-regular functions in Hilbert spaces*, J. Math. Anal. Appl., 303 (2005), pp. 1–14.

[35] ———, *Uniform prox-regularity of functions and epigraphs in Hilbert spaces*, Nonlinear Anal., 60 (2005), pp. 187–207.

[36] G. Bianchi, F. Vieira, and L. L. Ling, *A novel network traffic predictor based on multifractal traffic characteristic*, in Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE, vol. 2, Dec. 2004, pp. 680–684.

[37] R. V. D. Boomgaard and H. J. A. M. Heijmans, *Mathematical Morphology and its Applications to Image and Signal Processing*, vol. 18 of Computational Imaging and Vision, Springer Netherlands, 2002, ch. Morphological Scale-Space Operators: An Algebraic Framework, pp. 282–290.

[38] G. Borgefors, *Distance transformations on hexagonal grids*, PRL, 9 (1989), pp. 97–105.

[39] J. M. Borwein and C. H. Hamilton, *Symbolic Fenchel conjugation*, Math. Programming, (2007). Online.

[40] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3, Springer-Verlag, New York, 2000. Theory and examples.

[41] M. Bougeard, J.-P. Penot, and A. Pommellet, *Towards minimal assumptions for the infimal convolution regularization*, J. Approx. Theory, 64 (1991), pp. 245–270.

[42] I. Boukhriss, S. Miguet, and L. Tougne, *Two-dimensional discrete morphing*, in Combinatorial Image Analysis: 10th International Workshop, IWCIA 2004, Auckland, New Zealand, December 1-3, 2004. Proceedings, J. u. Reinhard Klette, ed., vol. 3322 of Lecture Notes in Computer Science, Springer-Verlag GmbH, 2004, pp. 409–420.

[43] L. M. Bregman, *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming*, Ž. Vyčisl. Mat. i Mat. Fiz., 7 (1967), pp. 620–631.

[44] Y. Brenier, *Un algorithme rapide pour le calcul de transformées de Legendre–Fenchel discrètes*, C. R. Acad. Sci. Paris Sér. I Math., 308 (1989), pp. 587–589.

[45] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman, *Linear time Euclidean distance transform algorithms*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17 (1995), pp. 529–533.

[46] B. Brighi and M. Chipot, *Approximated convex envelope of a function*, SIAM J. Numer. Anal., 31 (1994), pp. 128–148.

[47] B. Burgeth and J. Weickert, *An explanation for the logarithmic connection between linear and morphological system theory*, International Journal of Computer Vision, 64 (2005), pp. 157–169(13).

[48] J. V. Burke and M. Qian, *On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating*, Math. Program., 88 (2000), pp. 157–181.

[49] Y. Caron, P. Makris, and N. Vincent, *A method for detecting objects using Legendre transform*, in RFAI team publication, Maghrebian Conference on Computer Science MCSEAI, Annaba (Algeria), May 2002, pp. 219–225.

[50] B. A. Carré, *An algebra for network routing problems*, J. Inst. Math. Appl., 7 (1971), pp. 273–294.

[51] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, *A forward-backward algorithm for image restoration with sparse representations*, in Signal Processing with Adaptative Sparse Structured Representations (SPARS'05), Nov. 2005, pp. 49–52.

[52] J.-C. Chen, Y.-C. Wang, and J.-T. Chen, *A novel broadcast scheduling strategy using factor graphs and the sum-product algorithm*, Wireless Communications, IEEE Transactions on, 5 (2006), pp. 1241–1249.

[53] L. Chen and H. Y. H. Chuang, *A fast algorithm for Euclidean distance maps of a 2-d binary image*, Inf. Process. Lett., 51 (1994), pp. 25–29.

[54] Y. Chen, Y. Dong, D. Lu, and Y. Pan, *Intelligence and Security Informatics*, vol. 3495 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2005, ch. The Multi-fractal Nature of Worm and Normal Traffic at Individual Source Level, pp. 505–510.

[55] Y. T. Chin, H. Wang, L. P. Tay, H. Wang, and W. Y. C. Soh, *Vision guided agv using distance transform*, in Proceedings of the 32nd ISR (International Symposium on Robotics) 19-21 April 2001, 2001.

[56] F. H. Clarke, R. J. Stern, and P. R. Wolenski, *Proximal smoothness and the lower-C^2 property*, J. Convex Anal., 2 (1995), pp. 117–144.

[57] G. Cohen, S. Gaubert, and J.-P. Quadrat, *Duality and separation theorems in idempotent semimodules*, Linear Algebra Appl., 379 (2004), pp. 395–422. Tenth Conference of the International Linear Algebra Society.

[58] ———, *Projection and aggregation in maxplus algebra*, in Current trends in nonlinear systems and control, Systems Control Found. Appl., Birkhäuser Boston, Boston, MA, 2006, pp. 443–454.

[59] G. Cohen, S. Gaubert, J.-P. Quadrat, and I. Singer, *Max-plus convex sets and functions*, in Idempotent mathematics and mathematical physics, vol. 377 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2005, pp. 105–129.

[60] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, SIAM Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[61] L. Corrias, *Fast Legendre–Fenchel transform and applications to Hamilton–Jacobi equations and conservation laws*, SIAM J. Numer. Anal., 33 (1996), pp. 1534–1558.

[62] D. Crandall, P. F. Felzenszwalb, and D. P. Huttenlocher, *Object recognition by combining appearance and geometry.*, in Toward Category-Level Object Recognition, 2006, pp. 462–482.

[63] O. Cuisenaire, *Distance Transformations: Fast Algorithms and Applications to Medical Image Processing*, PhD thesis, Université Catholique de Louvain, Louvain-la-Neuve, Belgium, Oct. 1999.

[64] P.-E. Danielsson, *Euclidean distance mapping*, Computer Graphics and Image Processing, 14 (1980), pp. 227–248. CGIP, Academic Press.

[65] J. L. Davidson and F. Hummer, *Morphology neural networks: an introduction with applications*, Circuits Syst. Signal Process., 12 (1993), pp. 177–210.

[66] P. Del Moral and M. Doisy, *Maslov idempotent probability calculus. I*, Teor. Veroyatnost. i Primenen., 43 (1998), pp. 735–751.

[67] ———, *Maslov idempotent probability calculus. II*, Teor. Veroyatnost. i Primenen., 44 (1999), pp. 384–400.

[68] L. Deniau, *Proposition d'un opérateur géométrique pour l'analyse et l'identification de signaux et images*, PhD thesis, Université de Paris-Sud, Centre d'Orsay, Dec. 1997.

[69] L. Deniau and J. Blanc-Talon, *Fractal analysis with Hausdorff distance under affine transformations*, tech. rep., ETCA-CREA-SP, 1995.

[70] O. Devillers and M. J. Golin, *Incremental algorithms for finding the convex hulls of circles and the lower envelopes of parabolas*, Inf. Process. Lett., 56 (1995), pp. 157–164.

[71] G. Dolzmann, *Numerical computation of rank-one convex envelopes*, SIAM J. Numer. Anal., 36 (1999), pp. 1621–1635 (electronic).

[72] L. Dorst and R. van den Boomgaard, *Morphological signal processing and the slope transform*, Signal Processing, 38 (1994), pp. 79–98.

[73] K. R. Driessel and A. Gerisch, *Zero-preserving iso-spectral flows based on parallel sums*, Linear Algebra Appl., 421 (2007), pp. 69–84.

[74] M. V. Droogenbroeckb and H. Talbot, *Fast computation of morphological operations with arbitrary structuring elements*, Pattern Recognition Letters, 17 (1996), pp. 1451–1460.

[75] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res., 18 (1993), pp. 202–226.

[76] H. Edelsbrunner, *Algorithms in Combinatorial Geometry*, EATC Monographs on Theoretical Computer Science, Springer-Verlag, 1987.

[77] S.-L. Eriksson-Bique and H. Leutwiler, *A generalization of parallel addition*, Aequationes Math., 38 (1989), pp. 99–110.

[78] ———, *Minimal operators from a potential-theoretic viewpoint [MR1266217 (95a:47012)]*, in ICPT '91 (Amersfoort, 1991), Kluwer Acad. Publ., Dordrecht, 1994, pp. 47–59.

[79] M. Falcone and R. Ferretti, *Semi-Lagrangian schemes for Hamilton-Jacobi equations, discrete representation formulae and Godunov methods*, J. Comput. Phys., 175 (2002), pp. 559–575.

[80] P. F. Felzenszwalb and D. P. Huttenlocher, *Distance transforms of sampled functions*, Tech. Rep. TR2004-1963, Cornell Computing and Information Science, Sept. 2004.

[81] ———, *Pictorial structures for object recognition*, Int. J. Comp. Vision, 61 (2005), pp. 55–79.

[82] ———, *Efficient belief propagation for early vision*, Int. J. Comp. Vision, 70 (2006), pp. 41–54.

[83] M. Fidler and S. Recker, *Conjugate network calculus: a dual approach applying the Legendre transform*, Comput. Networks, 50 (2006), pp. 1026–1039.

[84] J. A. Fill and D. E. Fishkind, *The Moore-Penrose generalized inverse for sums of matrices*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 629–635 (electronic).

[85] *Fraclab: A fractal analysis toolbox for signal and image processing*. http://complex.inria.fr/, 2001.

[86] U. Frisch and J. Bec, *Burgulence*, in Les Houches 2000: New Trends in Turbulence, A. M. Lesieur and e. F. David, eds., Springer EDP-Sciences, 2001, pp. 341–383.

[87] U. Frisch, J. Bec, and B. Villone, *Singularities and the distribution of density in the Burgers/adhesion model*, Physica D, 152-153 (2001), pp. 620–635.

[88] M. Gavrilova and M. H. Alsuwaiyel, *Two algorithms for computing the Euclidean distance transform*, Tech. Rep. 2000-661-13, Computer Science Technical Reports, University of Calgary, 2000.

[89] A. Ghafoor, R. N. Iqbal, and S. Khan, *Image Analysis*, vol. 2749, Springer Berlin / Heidelberg, 2003, ch. Image Matching Using Distance Transforms, pp. 212–229.

[90] A. Gheondea and S. Gudder, *Sequential product of quantum effects*, Proc. Amer. Math. Soc., 132 (2004), pp. 503–512 (electronic).

[91] J. Gil and R. Kimmel, *Efficient dilation, erosion, opening, and closing algorithms*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (2002).

[92] K. Głazek, *A guide to the literature on semirings and their applications in mathematics and information sciences*, Kluwer Academic Publishers, Dordrecht, 2002. With complete bibliography.

[93] R. Goebel, *Self-dual smoothing of convex and saddle functions*, tech. rep., Department of Mathematics, University of Washington, Seattle, USA, 2006.

[94] J. S. Golan, *Power algebras over semirings*, vol. 488 of Mathematics and its Applications, Kluwer Academic Publishers, Dordrecht, 1999. With applications in mathematics and computer science.

[95] ———, *Semirings and their applications*, Kluwer Academic Publishers, Dordrecht, 1999. Updated and expanded version of *The theory of semirings, with applications to mathematics and theoretical computer science* [Longman Sci. Tech., Harlow, 1992; MR1163371 (93b:16085)].

[96] ———, *Some recent applications of semiring theory*, in International Conference on Algebra in Memory of Kostia Beidar, National Cheng Kung University, Tainan, Mar. 2005.

[97] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., 29 (1991), pp. 403–419.

[98] J. Gunawardena, *An introduction to idempotency*, in Idempotency (Bristol, 1994), vol. 11 of Publ. Newton Inst., Cambridge Univ. Press, Cambridge, 1998, pp. 1–49.

[99] ———, *From max-plus algebra to nonexpansive mappings: a nonlinear theory for discrete event systems*, Theoret. Comput. Sci., 293 (2003), pp. 141–167. Max-plus algebras.

[100] S. N. Gurbatov, S. I. Simdyankin, E. Aurell, U. Frisch, and G. Tth, *On the decay of Burgers turbulence*, J. Fluid Mech., 344 (1997), pp. 339–374.

[101] S. N. Gurbatov and A. V. Troussov, *The decay of multiscale signals - deterministic model of the Burgers turbulence*, tech. rep., arxiv.org, 2000.

[102] A. Hamdi, *A Moreau-Yosida regularization of a difference of two convex functions*, Appl. Math. E-Notes, 5 (2005), pp. 164–170 (electronic).

[103] W. Hare and C. Sagastizábal, *Benchmark of some nonsmooth optimization solvers for computing nonconvex proximal points*, Pac. J. Optim., 2 (2006), pp. 545–573.

[104] H. J. Heijmans and P. Maragos, *Lattice calculus of the morphological slope transform*, Signal Processing, 59 (May 1997), pp. 17–42(26).

[105] H. J. Heijmans and I. S. Molchanov, *Morphology on convolution lattices with applications to the slope transform and random set theory*, Journal of Mathematical Imaging and Vision, 8 (1998), pp. 199–214.

[106] P. Helluy and N. Seguin, *Relaxation models of phase transition flows*, M2AN Math. Model. Numer. Anal., 40 (2006), pp. 331–352.

[107] B. J. Hendrey and R. A. Jarvis, *Robot manipulator path planning*, in TENCON '92. Technology Enabling Tomorrow : Computers, Communications and Automation towards the 21st Century. 1992 IEEE Region 10 International Conference, 11-13 Nov 1992, Melbourne, Australia, vol. 2, 1992, pp. 865–870.

[108] ———, *A path planning heuristic for robotic manipulators*, in Intelligent Robots and Systems '93, IROS '93. Proceedings of the 1993 IEEE/RSJ International Conference on, 26-30 Jul 1993, Yokohama, Japan, vol. 2, 1993, pp. 1417–1423.

[109] W. H. Hesselink, *A linear-time algorithm for Euclidean feature transform sets*, Inf. Process. Lett., 102 (2007), pp. 181–186.

[110] W. Hinterberger and O. Scherzer, *Variational methods on the space of functions of bounded Hessian for convexification and denoising*, Computing, 76 (2006), pp. 109–133.

[111] J.-B. Hiriart-Urruty, *Extension of Lipschitz functions*, J. Math. Anal. Appl., 77 (1980), pp. 539–554.

[112] ———, *Lipschitz r-continuity of the approximate subdifferential of a convex function*, Math. Scand., 47 (1980), pp. 123–134.

[113] ———, *The deconvolution operation in convex analysis: an introduction*, Kibernet. Sistem. Anal., (1994), pp. 97–104, 190.

[114] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, vol. 305–306 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993. Vol I: Fundamentals, Vol II: Advanced theory and bundle methods.

[115] J.-B. Hiriart-Urruty and Y. Lucet, *Parametric computation of the Legendre–Fenchel conjugate*, J. Convex Anal., 14 (2007), pp. 657–666.

[116] J.-B. Hiriart-Urruty and M.-L. Mazure, *Formulations variationnelles de l'addition parallèle et de la soustraction parallèlle d'opérateurs semi-définis positifs*, C. R. Acad. Sci. Paris Sér. I Math., 302 (1986), pp. 527–530.

[117] T. Hisakado, K. Okumura, V. Vukadinovic, and L. Trajkovic, *Characterization of a simple communication network using Legendre transform*, in Proc. IEEE Int. Symp. Circuits and Systems, vol. 3, May 2003, pp. 738–741.

[118] D. P. Huttenlocher, K. Kedem, and M. Sharir, *The upper envelope of Voronoĭ surfaces and its applications*, Discrete Comput. Geom., 9 (1993), pp. 267–291.

[119] C. Imbert, *Analyse non lisse : - Fonction d'appui de la Jacobienne généralisée de Clarke et de son enveloppe plénière - Quelques applications aux équations de Hamilton-Jacobi du premier ordre (fonctions de Hopf-Lax, Hamiltoniens diff. convexes, solutions sci)*, PhD thesis, Laboratoire de Mathématiques pour l'Industrie et la Physique, Université Paul Sabatier - Toulouse III, May 2000.

[120] ———, *Convex analysis techniques for Hopf-Lax formulae in Hamilton-Jacobi equations*, J. Nonlinear Convex Anal., 2 (2001), pp. 333–343.

[121] A. N. Iusem, B. F. Svaiter, and M. Teboulle, *Entropy-like proximal methods in convex programming*, Math. Oper. Res., 19 (1994), pp. 790–814.

[122] R. Jarvis, *Distance transform based path planning for robot navigation*, in Recent Trends in Mobile Robots, Y. Zheng, ed., vol. 11 of Robotics and Automated Systems, World Scientific, 1993, ch. 1, pp. 3–31.

[123] R. Jarvis, *An articulated six wheel drive robot for very rough terrain navigation*, in Proc. 2002 Australian Conference on Robotics and Automation, 27-29 November 2002, Auckland, 2002, pp. 133–139.

[124] ———, *Distance transform based visibility measures for covert path planning in known but dynamic environments*, in 2nd International Conference on Autonomous Robots and Agents December 13-15, 2004 Palmerston North, New Zealand, Dec. 2004, pp. 396–400.

[125] R. A. Jarvis, *Collision-free trajectory planning using distance transforms*, Trans. Inst. Eng. Aust. Mech. Eng., ME10 (1985), pp. 187–191.

[126] R. A. Jarvis, *Collision-free path planning in time-varying environments*, in Intelligent Robots and Systems '89. (IROS '89) 'The Autonomous Mobile Robots and Its Applications'. , Proceedings. IEEE/RSJ International Workshop on, 4-6 Sep 1989, Tsukuba, Japan, 1989, pp. 99–106.

[127] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz, *Robust Face Detection Using the Hausdorff Distance*, in Audio- and Video-Based Person Authentication - AVBPA 2001, J. Bigun and F. Smeraldi, eds., vol. 2091 of Lecture Notes in Computer Science, Halmstad, Sweden, 2001, Springer, pp. 90–95.

[128] S. K., B. S., T. A., and Z. S.W., *Hamilton-jacobi skeletons*, International Journal of Computer Vision, 48 (July 2002), pp. 215–231(17).

[129] F. Kadhi and A. Trad, *Characterization and approximation of the convex envelope of a function*, J. Optim. Theory Appl., 110 (2001), pp. 457–466.

[130] C.-Y. Kao, S. Osher, and Y.-H. Tsai, *Fast sweeping methods for static Hamilton-Jacobi equations*, SIAM J. Numer. Anal., 42 (2005), pp. 2612–2632 (electronic).

[131] L. Karimovaa, Y. Kuandykova, N. Makarenkoa, M. Novakb, and S. Helamac, *Fractal and topological dynamics for the analysis of paleoclimatic records*, Physica A: Statistical and Theoretical Physics, 373 (2007).

[132] W. Karush, *A general algorithm for the optimal distribution of effort*, Management Science, 9 (1962), pp. 50–72.

[133] C. M. Klein, *Conjugate duality and its implications in dynamic programming*, Math. Comput. Modelling, 14 (1990), pp. 151–154.

[134] C. M. Klein and T. L. Morin, *Conjugate duality and the curse of dimensionality*, European Journal of Operational Research, 50 (1991), pp. 220–228.

[135] B. Koopen, *Contact of bodies in 2D-space: Implementing the Discrete Legendre Transform*, AI Master's thesis, Intelligent Autonomous Systems Group, University of Amsterdam, Feb. 2002.

[136] D. Kozinska, O. J. Tretiak, J. Nissanov, and C. Ozturk, *Multidimensional alignment using the Euclidean distance transform*, Graphical Models and Image Processing, 59 (1997), pp. 373–387.

[137] F. R. Kschischang, B. Frey, and H.-A. Loeliger, *Factor graphs and the sum-product algorithm*, IEEE Trans. Inform. Theory, 47 (2001), pp. 498–519.

[138] F. Kubo, *Conditional expectations and operations derived from network connections*, J. Math. Anal. Appl., 80 (1981), pp. 477–489.

[139] F. Kubo and T. Ando, *Means of positive linear operators*, Math. Ann., 246 (1979/80), pp. 205–224.

[140] T. Lachand-Robert and É. Oudet, *Minimizing within convex bodies using a convex hull method*, SIAM J. Optim., 16 (2005), pp. 368–379 (electronic).

[141] D. Lambert, J.-P. Crouzeix, V. H. Nguyen, and J.-J. Strodiot, *Finite convex integration*, J. Convex Anal., 11 (2004), pp. 131–146.

[142] J.-M. Lasry and P.-L. Lions, *A remark on regularization in Hilbert spaces*, Israel J. Math., 55 (1986), pp. 257–266.

[143] M. S. K. Lau and C. P. Kwong, *A smoothing method of global optimization that preserves global minima*, J. Global Optim., 34 (2006), pp. 369–398.

[144] J.-Y. Le Boudec and P. Thiran, *Network calculus*, vol. 2050 of Lecture Notes in Computer Science, Springer-Verlag, Berlin, 2001. A theory of deterministic queuing systems for the internet.

[145] B. Legras, B. Joseph, and F. Lefèvre, *Vertical diffusivity in the lower stratosphere from lagrangian back-trajectory reconstructions of ozone profiles*, Journal of Geophysical Research-Atmospheres, 108 (2003), pp. 148–227.

[146] B. Legras, I. Pisso, G. Berthet, and F. Lefvre, *Variability of the Lagrangian turbulent diffusion in the lower stratosphere*, Atmospheric Chemistry and Physics, 5 (2005), pp. 1605–1622.

[147] B. Lemaire, *The proximal algorithm*, in New methods in Optimization and their industrial uses (Pau/Paris, 1987), vol. 87 of Internat. Schriftenreihe Numer. Math., Birkhäuser, Basel, 1989, pp. 73–87.

[148] C. Lemaréchal and F. Oustry, *Growth conditions and $\mathcal{U}$-Lagrangians*, Set-Valued Anal., 9 (2001), pp. 123–129. Wellposedness in optimization and related topics (Gargnano, 1999).

[149] C. Lemaréchal, F. Oustry, and C. Sagastizábal, *The $\mathcal{U}$-Lagrangian of a convex function*, Trans. Amer. Math. Soc., 352 (2000), pp. 711–729.

[150] C. Lemaréchal and C. Sagastizábal, *More than first-order developments of convex functions: primal-dual relations*, J. Convex Anal., 3 (1996), pp. 255–268.

[151] ———, *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*, SIAM J. Optim., 7 (1997), pp. 367–385.

[152] ———, *Variable metric bundle methods: from conceptual to implementable forms*, Math. Programming, 76 (1997), pp. 393–410.

[153] S. Letovsky and S. Kasif, *Predicting protein function from protein/protein interaction data: a probabilistic approach.*, Bioinformatics, 19 Suppl 1 (2003).

[154] J. Lévy Véhel and B. Sikdar, *A multiplicative multifractal model for tcp traffic*, iscc, 00 (2001), p. 0714.

[155] G. L. Litvinov, *Maslov dequantization, idempotent and tropical mathematics: A brief introduction*, Journal of Mathematical Sciences, 140 (2007), pp. 426–444.

[156] B. Llanas, *Efficient computation of the Hausdorff distance between polytopes by exterior random covering*, Comput. Optim. Appl., 30 (2005), pp. 161–194.

[157] P. Lotito, E. Mancinelli, and J.-P. Quadrat, *Traffic assignment & Gibbs-Maslov semirings*, in Idempotent mathematics and mathematical physics, vol. 377 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2005, pp. 209–219.

[158] P. A. Lotito, E. M. Mancinelli, and J.-P. Quadrat, *A min-plus derivation of the fundamental car-traffic law*, IEEE Trans. Automat. Control, 50 (2005), pp. 699–705.

[159] Y. Lucet, *A fast computational algorithm for the Legendre–Fenchel transform*, Comput. Optim. Appl., 6 (1996), pp. 27–57.

[160] ———, *Faster than the Fast Legendre Transform, the Linear-time Legendre Transform*, Numer. Algorithms, 16 (1997), pp. 171–185.

[161] ———, *The Legendre-Fenchel conjugate: Numerical computation*, tech. rep., CECM, 1998.

[162] ———, *A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform*, in Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV 2005), Victoria BC, May 2005, IEEE Computer Society Press.

[163] ———, *Fast Moreau envelope computation I: Numerical algorithms*, Numer. Algorithms, 43 (2006), pp. 235–249.

[164] ———, *New sequential exact Euclidean distance transform algorithms based on convex analysis*, Image and Vision Computing, (2006). online.

[165] Y. Lucet, H. H. Bauschke, and M. Trienis, *The piecewise linear-quadratic model for computational convex analysis*, Comput. Optim. Appl., (2006). Accepted for publication.

[166] P. Maragos, *Morphological systems: Slope transforms and max-min difference and differential equations*, Signal Processing, 38 (1994), pp. 57–77.

[167] P. Maragos, *Slope transforms: theory and application to nonlinear signal processing*, Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on], 43 (1995), pp. 864–877. 1053-587X.
[168] P. Maragos, *Differential morphology and image processing*, IEEE Transactions on Image Processing, 5 (1996), pp. 922–927.
[169] ———, *Differential morphology*, in Nonlinear Image Processing, S. K. Mitra and G. L. Sicuranza, eds., Academic Press Series in Communications, Networking, and Multimedia, Academic Press, 2001, ch. 10, pp. 289–329.
[170] ———, *Handbook of Image and Video Processing*, Elsevier, second ed., 2005, ch. PDEs for Morphological Scale-Spaces and Eikonal Applications, pp. 587–612.
[171] P. Maragos and M. A. Butt, *Curve evolution, differential morphology, and distance transforms applied to multiscale and Eikonal problems*, Fundamenta Informaticae, 41 (2000), pp. 91–129.
[172] M. Marzouqi and R. A. Jarvis, *Covert robotics: hiding in known environments*, in Robotics, Automation and Mechatronics, 2004 IEEE Conference on, 1-3 Dec. 2004, Singapore, vol. 2, Dec. 2004, pp. 804–809.
[173] C. R. Maurer Jr, R. Qi, and V. V. Raghavan, *A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions*, IEEE Trans. Pattern Anal. Mach. Intell., 25 (2003), pp. 265–270.
[174] M.-L. Mazure, *L'addition parallèle d'opérateurs interprétée comme inf-convolution de formes quadratiques convexes*, RAIRO Modél. Math. Anal. Numér., 20 (1986), pp. 497–515.
[175] M.-L. Mazure, *La soustraction parallèle d'opérateurs interprétée comme déconvolution de formes quadratiques convexes*, Optimization, 18 (1987), pp. 465–484.
[176] ———, *Shorted operators through convex analysis*, in Trends in mathematical optimization (Irsee, 1986), vol. 84 of Internat. Schriftenreihe Numer. Math., Birkhäuser, Basel, 1988, pp. 139–152.
[177] M.-L. Mazure, *Équations de convolution et formes quadratiques*, Ann. Mat. Pura Appl. (4), 158 (1991), pp. 75–97.
[178] M.-L. Mazure and M. Volle, *Équations inf-convolutives et conjugaison de Moreau-Fenchel*, Ann. Fac. Sci. Toulouse Math. (5), 12 (1991), pp. 103–126.
[179] W. M. McEneaney, *Combining Legendre/Fenchel transformed operators on max-plus spaces for nonlinear control solution*, in Decision and Control, 2004. CDC. 43rd IEEE Conference on, vol. 2, Dec. 2004, pp. 1152–1157.
[180] K. McKinnon and M. Mongeau, *A generic global optimization algorithm for the chemical and phase equilibrium problem*, J. Global Optim., 12 (1998), pp. 325–351.
[181] F. Meng, D. Sun, and G. Zhao, *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*, Math. Program., 104 (2005), pp. 561–581.
[182] F. Meng and G. Zhao, *On second-order properties of the Moreau-Yosida regularization for constrained nonsmooth convex programs*, Numer. Funct. Anal. Optim., 25 (2004), pp. 515–529.
[183] F. W. Meng and Y. Hao, *Piecewise smoothness for Moreau-Yosida approximation to a piecewise C^2 convex function*, Adv. Math. (China), 30 (2001), pp. 354–358.
[184] R. J. Meziat, D. P. no, and P. Pedregal, *An alternative approach for non-linear optimal control problems based on the method of moments*, Computational Optimization and Applications, 38 (2007), pp. 147–171.
[185] R. J. Meziat and J. Villalobos, *Analysis of microstructures and phase transistion phenomena in one-dimensional, non-linear elasticity by convex optimization*, Struct. Multidisc. Optim., 32 (2006), pp. 507–519.
[186] R. Mifflin, *A quasi-second-order proximal bundle algorithm*, Math. Programming, 73 (1996), pp. 51–72.
[187] R. Mifflin, L. Qi, and D. Sun, *Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function*, Math. Program., 84 (1999), pp. 269–281.
[188] R. Mifflin and C. Sagastizábal, *$\mathcal{VU}$-decomposition derivatives for convex max-functions*, in Ill-posed variational problems and regularization techniques (Trier, 1998), vol. 477 of Lecture Notes in Econom. and Math. Systems, Springer, Berlin, 1999, pp. 167–186.
[189] ———, *Functions with primal-dual gradient structure and $\mathcal{U}$-Hessians*, in Nonlinear optimization and related topics (Erice, 1998), vol. 36 of Appl. Optim., Kluwer Acad. Publ., Dordrecht, 2000, pp. 219–233.
[190] ———, *On $\mathcal{VU}$-theory for functions with primal-dual gradient structure*, SIAM J. Optim., 11 (2000), pp. 547–571 (electronic).
[191] ———, *Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions*, SIAM J. Optim., 13 (2003), pp. 1174–1194 (electronic).
[192] ———, *On the relation between $\mathcal{U}$-Hessians and second-order epi-derivatives*, European J. Oper. Res., 157 (2004), pp. 28–38.
[193] ———, *$\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions*, Optim. Methods Softw., 19 (2004), pp. 463–478.
[194] ———, *Relating $\mathcal{U}$-Lagrangians to second-order epi-derivatives and proximal-tracks*, J. Convex Anal., 12 (2005), pp. 81–93.
[195] R. Mifflin and C. Sagastizábal, *A $\mathcal{VU}$-algorithm for convex minimization*, Mathematical Programming, (2005).
[196] S. K. Mitra and K. Manjunatha Prasad, *The regular shorted matrix and the hybrid sum*, Adv. in Appl. Math., 18 (1997), pp. 403–422.
[197] J.-J. Moreau, *Inf-convolution des fonctions numériques sur un espace vectoriel*, C. R. Acad. Sci. Paris, 256 (1963), pp. 5047–5049.
[198] ———, *Propriétés des applications "prox"*, C. R. Acad. Sci. Paris, 256 (1963), pp. 1069–1071.
[199] ———, *Proximité et dualité dans un espace Hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[200] ———, *Convexity and duality*, in Functional Analysis and Optimization, Academic Press, New York, 1966, pp. 145–169.

[201] ———, *Inf-convolution, sous-additivité, convexité des fonctions numériques*, J. Math. Pures Appl. (9), 49 (1970), pp. 109–154.
[202] T. L. Morin and R. E. Marsten, *An algorithm for nonlinear knapsack problems*, Management Science, 22 (1976), pp. 1147–1158.
[203] T. D. Morley, *Parallel summation, Maxwell's principle and the infimum of projections*, J. Math. Anal. Appl., 70 (1979), pp. 33–41.
[204] N. Nikopoulos and I. Pitas, *A fast implementation of 3-D binary morphological transformations*, IEEE Transactions on Image Processing, 9 (2000).
[205] A. Noullez, S. N. Gurbatov, E. Aurell, and S. I. Simdyankin, *The global picture of self-similar and not self-similar decay in Burgers turbulence*, Tech. Rep. nlin.CD/0409022, arXiv.org eprint archive, Sept. 2004.
[206] A. Noullez and M. Vergassola, *A fast Legendre transform algorithm and applications to the adhesion model*, J. Sci. Comput., 9 (1994), pp. 259–281.
[207] A. M. Oberman, *Computing the convex envelope using a nonlinear partial differential equation*, Mathematical Models and Methods in Applied Sciences, (2007). Accepted.
[208] ———, *The convex envelope is the solution of a nonlinear obstacle problem*, Proc. Amer. Math. Soc., 135 (2007), pp. 1689–1694 (electronic).
[209] È. L. Pekarev and J. L. Šmul′jan, *Parallel addition and parallel subtraction of operators*, Izv. Akad. Nauk SSSR Ser. Mat., 40 (1976), pp. 366–387, 470.
[210] J.-P. Penot, *Proximal mappings*, J. Approx. Theory, 94 (1998), pp. 203–221.
[211] J.-P. Penot and R. Ratsimahalo, *On the Yosida approximation of operators*, Proc. Roy. Soc. Edinburgh Sect. A, 131 (2001), pp. 945–966.
[212] R. Poliquin and R. T. Rockafellar, *Generalized Hessian properties of regularized nonsmooth functions*, SIAM J. Optim., 6 (1996), pp. 1121–1137.
[213] ———, *Prox-regular functions in variational analysis*, Trans. Amer. Math. Soc., 348 (1996), pp. 1805–1838.
[214] F. P. Preparata and M. I. Shamos, *Computational Geometry*, Texts and monographs in computer science, Springer-Verlag, third ed., 1990.
[215] S. Prohaska and H. C. Hege, *Fast visualization of plane-like structures in voxel data*, in VIS '02: Proceedings of the conference on Visualization '02, IEEE Computer Society, 2002, pp. 29–36. Boston, Massachusetts.
[216] L. Qi, *Second-order analysis of the Moreau-Yosida regularization*, in Nonlinear analysis and convex analysis (Niigata, 1998), World Sci. Publishing, River Edge, NJ, 1999, pp. 16–25.
[217] P. J. Rabier and A. Griewank, *Generic aspects of convexification with applications to thermodynamic equilibrium*, Arch. Rational Mech. Anal., 118 (1992), pp. 349–397.
[218] A. G. Ramm, A. M. STeinberg, and A. I. Zaslavsky, *Stable calculation of the Legendre transform*, Journal of Mathematical Analysis and Applications, 178 (1993), pp. 592–602.
[219] A. G. Ramm and A. I. Zaslavsky, *Reconstructing singularities of a function from its Radon transform*, Math. Comput. Modelling, 18 (1993), pp. 109–138.
[220] ———, *Singularities of the Radon transform*, Bull. Amer. Math. Soc. (N.S.), 28 (1993), pp. 109–115.
[221] ———, *X-ray transform, the Legendre transform, and envelopes*, J. Math. Anal. Appl., 183 (1994), pp. 528–546.
[222] J. P. Revalski and M. Théra, *Enlargements and sums of monotone operators*, Nonlinear Anal., 48 (2002), pp. 505–519.
[223] R. H. Riedi, *Multifractal processes*, in Theory and applications of long-range dependence, Birkhäuser Boston, Boston, MA, 2003, pp. 625–716.
[224] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk, *A multifractal wavelet model with application to network traffic*, IEEE Transactions on Information Theory, 45 (1999), pp. 992–1018.
[225] R. H. Riedi and J. L. Vhel, *Multifractal properties of TCP traffic: a numerical study*, Tech. Rep. 3129, INRIA, Rocquencourt, Mar. 1997.
[226] R. M. Rifkin and R. A. Lippert, *Value regularization and Fenchel duality*, J. Mach. Learn. Res., 8 (2007), pp. 441–479.
[227] G. Ritter and P. Sussner, *An introduction to morphological neural networks*, in ICPR '96: Proceedings of the International Conference on Pattern Recognition (ICPR '96) Volume IV-Volume 7472, Washington, DC, USA, 1996, IEEE Computer Society, pp. 709–717.
[228] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New york, 1970.
[229] ———, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optimization, 14 (1976), pp. 877–898.
[230] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[231] R. T. Rockafellar and P. R. Wolenski, *Convexity in Hamilton-Jacobi theory. II. Envelope representations*, SIAM J. Control Optim., 39 (2000), pp. 1351–1372 (electronic).
[232] A. Rosenfeld and J. L. Pfaltz, *Sequential operations in digital picture processing*, J. ACM, 13 (1966), pp. 471–494.
[233] *Scilab*. http://www.scilab.org, 1994.
[234] J. A. Sethian, *Fast marching methods*, SIAM Review, 41 (1999), pp. 199–235.
[235] ———, *Level Set Methods and Fast Marching Methods*, vol. 3 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, second ed., 1999. Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science.
[236] Z.-S. She, E. Aurell, and U. Frisch, *The inviscid Burgers equation with initial data of brownian type*, Comm. Math. Phys., 148 (1992), pp. 623–641.

[237] B. Shi, B. Vidakovic, and G. G. Katul, *Quantifying the effects of atmospheric stability on the multifractal spectrum of turbulence*, tech. rep., Georgia Institute of Techology, School of Industrial and Systems Engineering, Statistics Group, 2004.

[238] F. Y. Shih and Y.-T. Wu, *Fast Euclidean distance transformation in two scans using a 3×3 neighborhood*, Comput. Vis. Image Underst., 93 (2004), pp. 195–205.

[239] K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker, *The Hamilton-Jacobi skeleton*, Seventh International Conference on Computer Vision (ICCV'99), 02 (1999), p. 828.

[240] T. Strömberg, *Representation formulae for infimal convolution with applications*, in Analysis, algebra, and computers in mathematical research (Luleå, 1992), vol. 156 of Lecture Notes in Pure and Appl. Math., Dekker, New York, 1994, pp. 319–334.

[241] ———, *The operation of infimal convolution*, Dissertationes Math. (Rozprawy Mat.), 352 (1996).

[242] S. Sundar and Z. Shiller, *Optimal obstacle avoidance based on the Hamilton-Jacobi-Bellman equation*, IEEE Transactions on Robotics and Automation, 13 (1997), pp. 305–310.

[243] K. W. Tang and R. A. Jarvis, *A fast algorithm to plan a collision-free path in cluttered 2d environments*, in Robotics, Automation and Mechatronics, 2004 IEEE Conference on, 1-3 Dec. 2004, Singapore, vol. 2, 2004, pp. 786–791.

[244] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Math. Oper. Res., 17 (1992), pp. 670–690.

[245] Y. Tian, *How to express a parallel sum of k matrices*, J. Math. Anal. Appl., 266 (2002), pp. 333–341.

[246] A. N. Tikhonov and V. A. Morozov, *Methods for the regularization of ill-posed problems*, Vychisl. Metody i Programmirovanie, (1981), pp. 3–34.

[247] J. Toriwaki and K. Mori, *Distance transformation and skeletonization of 3d pictures and their applications to medical images*, in Digital and Image Geometry: Advanced Lectures, R. K. E. G. Bertrand, A. Imiya, ed., vol. 2243 / 2001 of Lecture Notes in Computer Science, Springer-Verlag Heidelberg, 2001, pp. 412–429.

[248] M. Trienis, *Computational convex analysis: From continuous deformation to finite convex integration*, Master's thesis, University of British Columbia, 2007.

[249] Y.-H. R. Tsai and S. Osher, *Total variation and level set methods in image science*, Acta Numer., 14 (2005), pp. 509–573.

[250] P. Tseng and Z.-Q. Luo, *On computing the nested sums and infimal convolutions of convex piecewise-linear functions*, Journal of Algorithms, 21 (1996), pp. 240–266.

[251] A. V. Tuzikov, G. L. Margolin, and A. I. Grenov, *Convex set symmetry measurement via Minkowski addition*, J. Math. Imaging Vision, 7 (1997), pp. 53–68.

[252] M. Vergassola, B. Dubrulle, U. Frisch, and N. A., *Burgers's equation, devil's staircases and the mass distribution function for large-scale structures*, Astronomy and Astrophysics, 289 (1994), pp. 325–356.

[253] L. A. Vese, *A method to convexify functions via curve evolution*, Comm. Partial Differential Equations, 24 (1999), pp. 1573–1591.

[254] L. Vincent, *Morphological transformations of binary images with arbitrary structuring elements*, Signal Processing, 22 (1991), pp. 3–23.

[255] ———, *Mathematical Morphology in Image Processing*, Marcel-Dekker, New York, Sept. 1992, ch. Morphological Algorithms, pp. 255–288.

[256] E. T. P. Wong and R. Jarvis, *Real time obstacle detection and navigation planning for a humanoid robot in an indoor environment*, in Proceedings of the 2004 IEEE Conference on Robotics, Automation and Mechatronics, 1-3 December 2004, Singapore, Dec. 2004, pp. 693–699.

[257] K. Yosida, *Functional analysis*, Classics in Mathematics, Springer-Verlag, Berlin, 1995. Reprint of the sixth (1980) edition.

[258] A. I. Zaslavsky, *Legendre transforms and dual varieties*, J. Math. Anal. Appl., 183 (1994), pp. 670–684.

[259] A. Zelinsky, *A mobile robot exploration algorithm*, IEEE Transactions on Robotics and Automation, 8 (1992), pp. 707–717.

[260] A. Zelinsky, *A navigation algorithm for industrial mobile robots*, in Emerging Technologies and Factory Automation, 1992. IEEE International Workshop on, 11-14 Aug 1992, 1992, pp. 525–530.

[261] A. Zelinsky, *Using path transforms to guide the search for findpath in 2d*, International Journal of Robotics Research, 13 (1994), pp. 315–325.

[262] C. Zhang and F. Cohen, *3-d face structure extraction and recognition from images using 3-d morphing and distance mapping*, IEEE Transactions on Image Processing, 11 (2002), pp. 1249–1259.

Computer Science, I. K. Barber School of Arts & Sciences Unit # 4, University of British Columbia Okanagan, 3333 University Way, Kelowna BC V1V 1V7, Canada

E-mail address: `yves.lucet@ubc.ca`

URL: `http://people.ok.ubc.ca/ylucet/`

Comput Optim Appl
DOI 10.1007/s10589-007-9124-y

The piecewise linear-quadratic model for computational convex analysis

Yves Lucet · Heinz H. Bauschke · Mike Trienis

Received: 18 July 2006 / Revised: 21 June 2007

Abstract A new computational framework for computer-aided convex analysis is proposed and investigated. Existing computational frameworks are reviewed and their limitations pointed out. The class of piecewise linear-quadratic functions is introduced to improve convergence and stability. A stable convex calculus is achieved using symbolic-numeric algorithms to compute all fundamental transforms of convex analysis. Our main result states the existence of efficient (linear time) algorithms for the class of piecewise linear-quadratic functions. We also recall that such class is closed under convex transforms. We illustrate the results with numerical examples, and validate numerically the resulting computational framework.

Keywords Computational convex analysis · Proximal average · Legendre-Fenchel transform · Fenchel conjugate · Moreau envelope · Moreau-Yosida approximate · Convex analysis

1 Introduction

Computational convex analysis focuses on developing efficient tools to compute fundamental transforms arising in convex analysis. Symbolic computation tools have been developed [4, 5, 14], and have allowed more insight into the calculation of the

Y. Lucet (✉) · H.H. Bauschke · M. Trienis
University of British Columbia—Okanagan, 3333 University Way, Kelowna BC V1V 1V7, Canada
e-mail: yves.lucet@ubc.ca

H.H. Bauschke
e-mail: heinz.bauschke@ubc.ca

M. Trienis
e-mail: mjtrieni@interchange.ubc.ca

Legendre-Fenchel conjugate (also called Fenchel conjugate, convex conjugate, or—in the context of convex analysis—conjugate) and related transforms. When such tools are not applicable *e.g.* when there is no closed form, fast transform algorithms perform numerical computation efficiently.

Although the early idea of efficient numerical computation of convex transforms can be traced back to [29], the development of efficient algorithms started with the note [6] on the Fast Legendre Transform (FLT), which was subsequently investigated in [7, 23]. The FLT algorithm was also independently presented in [32, 37]. Its log-linear worst-case time complexity was subsequently improved with the Linear-time Legendre transform (LLT) algorithm [24] (see also [25] for the description of a numerical implementation of the LLT algorithm).

Recently, several new linear fast algorithms have been investigated [18, 27]. They take advantage of the equivalence between the computation of the conjugate and of the Moreau envelope [35, Example 11.26c] also called the Moreau-Yosida approximate, Yosida Approximate or Moreau-Yosida regularization. The Moreau envelope goes back to the work of Yosida [40] on maximal monotone operators, and its behaviour is well known in the fields of convex analysis [28–30, 33] and variational analysis [35, Chap. 12]. Its associated transform, the proximal mapping, has been extensively studied through the analysis of the proximal point algorithm [34], whose convergence properties are well known [1, 13, 22].

The fast algorithms are not limited to computing the conjugate or the Moreau envelope. For example, Moreau [29] already noted that the set of proximal mappings is convex, although an explicit formula for the convex function whose proximal mapping is the convex combination of two proximal mappings was only recently presented [3]. The resulting transform, named the proximal average in [2], combines several fundamental operations of convex analysis: addition, scalar multiplication, conjugation, and regularization with the Moreau envelope. Other transforms related to the Moreau envelope include the Lasry-Lions double envelope, the proximal hull, the inf-convolution of convex functions, and the deconvolution of convex functions [17, 35]. They are all computable by combining fast algorithms.

Beyond computer-aided convex analysis, the FLT algorithm and the faster LLT algorithm have been used in efficient numerical simulations of the Burger's equation, see for example [11, 31]. The LLT has also found applications in robotics [20], network communication [19], pattern recognition [26], numerical simulation of multiphasic flows [15], and analysis of the distribution of chemical compounds in the atmosphere [21]. The Moreau envelope is an extension of the distance transform encountered in image processing, and several authors have investigated fast algorithms in that context [8–10, 12, 38] (see also [26] for the explicit application of the LLT algorithm to the computation of the distance transform). The inf-convolution and addition of convex functions are also related to the Linear Cost Network Flow problem on Series-Parallel Networks [39].

We summarize the intrinsic framework of fast algorithms, and point out their limitations in Sect. 2. We then consider the parametric framework introduced in [18], which implicitly models the domain of the conjugate in Sect. 3. The underlying models are explicited in Sect. 4, and the class of piecewise linear-quadratic (PLQ) functions is introduced in Sect. 5 to remedy shortcomings in the implicit and parametric frameworks. In contrast to these frameworks, the PLQ functions are closed under standard convex operations, and can be manipulated with linear time algorithms. We present numerical examples in Sect. 6 while Sect. 7 introduces our numerical package, and validates the algorithms numerically. We conclude the paper in Sect. 8.

Unless otherwise stated, we restrict our framework to lower semicontinuous (lsc) proper extended-valued convex functions on the real line. (Future extensions to functions of several variables will be considered in Sect. 8.) Note that all algorithms for univariate functions extend straightforwardly to separable functions. So checking that $(\|\cdot\|^2/2)^* = \|\cdot\|^2/2$ can be performed in our framework. We denote I_S the indicator function of a set S: $I_S(x) = 0$ when $x \in S$, and $I_S(x) = +\infty$ otherwise.

2 Discrete convex transform

In this section, we recall the computational framework for fast algorithms and point out their limitations for computing composition of convex transforms. We recall that the domain of a lower semicontinuous (lsc) proper extended-valued convex function $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is defined by $\operatorname{Dom} f := \{x \in \mathbb{R} | f(x) < +\infty\}$. (A function is proper if its domain is nonempty.)

Computing fundamental transforms of convex analysis is reduced to a discrete optimization problem. For example, the computation of the conjugate or Legendre-Fenchel transform

$$f^*(s) := \sup_{x \in \mathbb{R}} [sx - f(x)] \tag{1}$$

is approximated with the discrete Legendre transform

$$f_n^*(s_j) := \max_i [s_j x_i - f(x_i)] \tag{2}$$

where the function f is only evaluated at the points x_i, $i = 1 \dots n$, and the conjugate is approximated at the slopes s_j, $j = 1 \dots m$. The key point we are interested in is *evaluating the transform on a grid*, and not at a single point. By evaluating the transform on the full grid, we can take advantage of evaluations at other points to speed up computation.

We focus on the efficient computation of (2) in terms of worst-case time complexity. While a brute force computation has an $O(nm)$ complexity (evaluate a maximum over n points for m slopes), an efficient (linear-time) algorithm like the LLT algorithm [24] evaluates (2) in $O(n + m)$.

Convergence results [7, 23] allow us to approximate (1) from (2) as follows. When the function or its transform have an unbounded domain, we enlarge the grid to obtain a more accurate numerical approximation of its domain *i.e.* we consider intervals $[a, b]$ with $a \to -\infty$ and $b \to +\infty$. Then we fix a and b, and increase the number of points in the grid by decreasing the grid stepsize.

For the sake of completeness we recall the main convergence facts.

Fact 2.1 Assume $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is proper.

(i) Convergence on a bounded domain: (see for example [7, Theorems 2.1, 2.3], [23, Propositions 2, 3], or [24, Proposition 1]).
   (i) If f is upper semi-continuous on $[a, b]$, then f_n^* converges pointwise to $f^*_{[a,b]}(s) := \sup_{x\in[a,b]}[sx - f(x)]$.
   (ii) If f is twice continuously differentiable on an open interval containing $[a, b]$, then

$$\max_{[a,b]} |f^*_{[a,b]} - f_n^*| \le \frac{1}{2}\frac{(b-a)^2}{n^2}\max_{[a,b]} f''.$$

(ii) Convergence on unbounded domains: (see for example [16, Proposition 2.3], [23, Proposition 5], or [24, Proposition 2]).
The following equivalence holds for any $s \in \mathbb{R}$, and any $a > 0$

$$\partial f^*(s) \cap [-a, a] \neq \emptyset \Leftrightarrow f^*_{[-a,a]}(s) = f^*(s),$$

where ∂f is the subdifferential in the sense of convex analysis

$$\partial f(x) := \{s \in \mathbb{R} : f(y) \ge f(x) + s(y - x) \text{ for all } y\}.$$

Figure 1a illustrates an approximation of the domain: increasing the size of the grid for x_i gives a better approximation of the conjugate (the slopes converge to the two vertical lines), which is the indicator function of the interval $[-1, 1]$. Figure 1b shows an approximation of the function within the domain: reducing the grid stepsize gives a better approximation (the approximation is the piecewise linear function; it converges to the exact conjugate, which is the quadratic function).

We will see in Sect. 5 how the approximation of the domain can be avoided altogether by using a more general class of functions than the piecewise linear functions. The convergence on the domain will also be improved by using piecewise quadratic functions instead of piecewise linear functions.

Similarly, the computation of the Moreau envelope

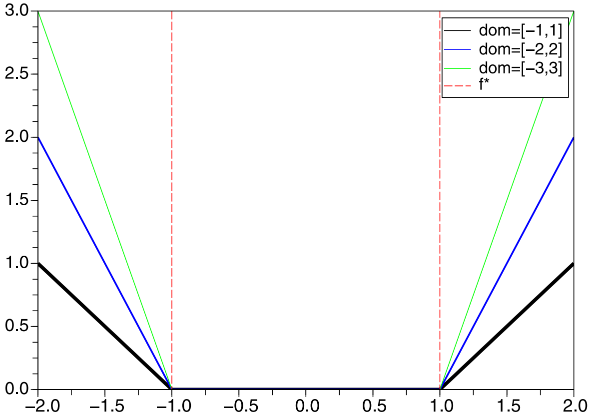
$$M_\lambda(s) := \inf_{x\in\mathbb{R}}\left[\frac{(s-x)^2}{2\lambda} + f(x)\right]$$

for all values of s in an interval $[a, b]$ ($\lambda > 0$ is a fixed parameter) is approximated with the discrete Moreau envelope

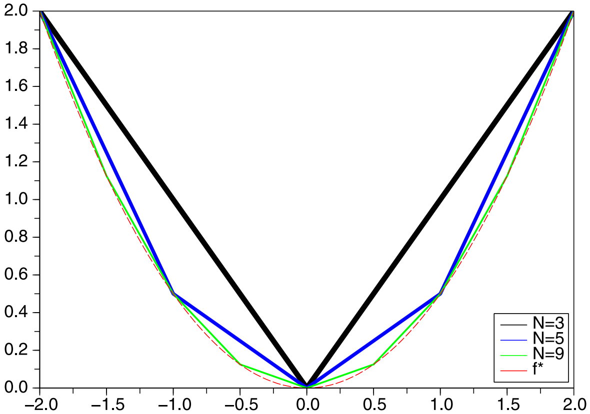
$$M_{n,\lambda}(s_j) := \min_i\left[\frac{(s_j - x_i)^2}{2\lambda} + f(x_i)\right]$$

where $i = 1 \dots n$, and $j = 1 \dots m$. Note that the Moreau envelope is a special case of the inf-convolution operator

$$f \,\square\, g(x) := \inf_y[f(x-y) + g(y)].$$



(a) Convergence by enlarging the domain: the conjugate of $f(x) := |x|$ converges to $I_{[-1,1]}(x)$, the indicator function of the interval $[-1, 1]$



(b) Convergence by decreasing the grid stepsize: the discrete conjugate of the function $f(x) := x^2/2$ converges to $f^* = f$

Fig. 1 Convergence of the discrete Legendre transform

Throughout the paper we will use the following facts.

Fact 2.2 [17, X.2.1.3] Assume f and g are both lsc convex proper functions and $\operatorname{Dom} f^* \cap \operatorname{Dom} g^*$ is nonempty. Then $(f \,\Box\, g)^* = (f^* + g^*)^*$.

Fact 2.3 [35, 11.26c] and [27, Proposition 3] The following formula holds for any function f

$$M_\lambda(s) = \frac{s^2}{2\lambda} - \frac{1}{\lambda}\left(\lambda f + \frac{|\cdot|^2}{2}\right)^*(s). \tag{3}$$

Formula (3) implies that as far as computational algorithms are concerned, computing the conjugate is equivalent to computing the Moreau envelope. So any efficient (linear-time) algorithm to compute one transform can be used to compute the other.

The framework is the same for all the fast transform algorithms previously considered [27]: the computation is restricted to a grid of points, and convergence results are employed to obtain a numerical approximation of the transform under consideration.

Unfortunately, the fast algorithm approach becomes awkward when we consider transforms like the proximal average

$$\mathcal{P}(f_0,\lambda,f_1) := \left[(1-\lambda)\left(f_0+\frac{1}{2}|\cdot|^2\right)^* + \lambda\left(f_1+\frac{1}{2}|\cdot|^2\right)^*\right]^* - \frac{1}{2}|\cdot|^2.$$

While it decomposes in several transforms, functions resulting from intermediate computations have unbounded domains. In other words, even when the input function requires only few grid points and we only want a coarse approximation of the transform, we may have to compute on large grids with a prohibitive number of points during intermediate computations.

Consider the following example.

Example 2.4 Let $f_0(x) := -x$, $f_1(x) := x$, and compute their proximal average $f := \mathcal{P}(f_0, 1/2, f_1)$. The conjugate of f_0 (resp. f_1) is $f_0^*(s) = I_{\{-1\}}(s)$ (resp. $f_1^*(s) = I_{\{1\}}(s)$). The Moreau envelope of f_0^* (resp. f_1^*) is $f_0^* \square 1/2|\cdot|^2 = (x+1)^2/2$ (resp. $f_1^* \square 1/2|\cdot|^2 = (x-1)^2/2$). Since the domain of both Moreau envelopes is $\mathbb{R}$, computing their average requires modeling the domain: taking a large number of points to numerically approximate the domain. Failing to do that results in a poor numerical approximation of the sum, which propagates to a poor approximation of the conjugate, and eventually to a poor approximation of the proximal average.

In fact, assume the approximation is exact on the grid, and only consider the convergence required due to unbounded domains (Fact 2.1(ii)). Then the combination of fast discrete algorithms to compute the proximal average—add the energy $x^2/2$, apply the LLT, build the convex combination, apply the LLT, and subtract the energy—results in the following function

$$\begin{cases} +\infty & \text{if } |x| > b,\\ (2\lambda-1)x - 2\lambda(1-\lambda) & \text{if } 2(1-\lambda)-b \le x \le b-2\lambda,\\ \frac{\lambda}{2(1-\lambda)}x^2 + \frac{\lambda+\lambda b-1}{1-\lambda}x + \frac{\lambda b(4\lambda+b-4)}{2(1-\lambda)} & \text{if } -b \le x \le 2(1-\lambda)-b,\\ \frac{1-\lambda}{2\lambda}x^2 + \frac{\lambda-b+\lambda b}{\lambda}x + \frac{b(4\lambda^2+b-\lambda b-4\lambda)}{2\lambda} & \text{if } b-2\lambda \le x \le b \end{cases}$$

where f_0 (resp. f_1) is approximated with $-x + I_{[-b,b]}(x)$ (resp. $x + I_{[-b,b]}(x)$), and $b > 1$.

Even though the functions f_0 and f_1 are linear, the fast algorithm framework requires an (unnecessary) technical knowledge: in addition to guessing the domain on which to calculate the convex combination for each of the intermediate operations, we now need to increase b to $+\infty$ to guarantee accurate approximations. When b is chosen large enough, we finally have to decrease the grid stepsize to achieve our desired approximation.

To use the fast numerical algorithms to compute the proximal average, we have to keep track of four sets: The domain of f_0, the domain of f_1, the set that approximates the domain of $(1-\lambda)(f_0+|\cdot|^2/2)^* + \lambda(f_1+|\cdot|^2/2)^*$, and the domain on which we want to compute $\mathcal{P}(f_0,\lambda,f_1)$. Section 5 presents a new model that does not require such tracking. It thus greatly simplify the computational framework, and the technical knowledge required to compute compositions of convex operators such as the proximal average.

The next section considers another existing computational framework which has a better modeling of the domain.

3 Parametric convex transforms

We summarize the recent introduction of a parametric algorithm [18] to compute the conjugate (which using (3) can also be used to compute the Moreau envelope), generalize the approach to a full convex calculus, and point out the limitations of such a framework.

Recently, the numerical computation of the conjugate was investigated using the parametrization

$$\begin{cases} s \in \partial f(x), \\ f^*(s) = sx - f(x), \end{cases}$$

for $x \in \mathbb{R}$. The term parametrization comes from the description of the conjugate: instead of returning a function f^*, the algorithm returns a parametric description of the planar curve $(s, f^*(s))$. Note that we no longer have access to $f^*(s)$ at any slope s, we only obtain f^* on the range of ∂f. The idea relies on the geometric characterization of the epigraph of f^* whose extreme points are obtained using the parametrization, then the epigraph is completed by affine parts to recover the full graph of f^*. Consider the following example.

Example 3.1 Take $f(x) := |x|$. Then $\partial f(x) = \{-1\}$ when $x<0$, $\partial f(x) = [-1,1]$ when $x=0$, and $\partial f(x) = \{1\}$ when $x>0$. So the parametrization above gives

- For $x<0$, $s=-1$ and $f^*(s) = sx - f(x) = -x-|x| = 0$.
- For $x=0$, $s\in[-1,1]$ and $f^*(s)=0$.
- For $x>0$, $s=1$ and $f^*(s) = x-|x| = 0$.

Consequently, $f^*(s)=0$ for $s\in[-1,1]$. For any s not in the range of ∂f, which is $[-1,1]$, $f^*(s)=+\infty$. We obtain the right answer: $f^* = I_{[-1,1]}$.

While the framework is different from Sect. 2, the fast algorithms can be formalized in the parametric framework as follows. Their computation of the conjugate amounts to computing the function

$$(x_i, f_i, s_j) \mapsto (s_j, f_j^*),$$

where x_i is a primal grid point, f_i is an approximation of $f(x_i)$, s_j is a dual grid point, and f_j^* is an approximation of $f^*(s_j)$. Note that the algorithms really return f_j^* since

s_j is part of the input, but we write the pair (s_j, f_j^*) to emphasize the comparison with the parametric framework of the current section. The points (s_j, f_j^*) are an approximation of the graph of f^* in the plane.

To simplify our presentation, we assume the function f is twice differentiable for the remainder of this section only. Note that contrary to second order results which do not hold when f is not twice differentiable, the first order results still hold by replacing the derivative f' with the subgradient ∂f when f is not differentiable.

In the parametric framework, given a discretization $f_i \approx f(x_i)$ of a function f and of its derivatives $g_i \approx f'(x_i)$, $h_i \approx f''(x_i)$ on a grid x_i, we define the discretization of the conjugate and its derivatives by

$$\mathrm{Conj} : (x_i, f_i, g_i, h_i) \mapsto \left(g_i, x_i g_i - f_i, x_i, \frac{1}{h_i}\right). \tag{4}$$

Proposition 3.2 *The function* Conj *provides an approximation of the conjugate and of its first two derivatives*:

$$\mathrm{Conj}(x_i, f_i, g_i, h_i) \approx (g_i, f^*(g_i), (f^*)'(g_i), (f^*)''(g_i)).$$

Proof The proof that Conj is an approximation of the conjugate is contained in [18] except for the second order approximation, which is proven in [17, X.4.2.9] and also in [35, 13.21 and p. 605] (we are using the convention $1/0 := +\infty$). □

Remark 3.3 Note that formula (4) implies the discrete equivalent of the Biconjugation Theorem, which states that $f^{**} = f$. Indeed, apply Conj twice, to obtain

$$\mathrm{Conj}(\mathrm{Conj}(x_i, f_i, g_i, h_i)) = \mathrm{Conj}\left(g_i, x_i g_i - f_i, x_i, \frac{1}{h_i}\right) = (x_i, f_i, g_i, h_i).$$

So we recover our original discrete data set (x_i, f_i, g_i, h_i).

In addition to providing approximations to the derivatives of the conjugate, the advantage of the parametrization is two-fold: the implementation is very simple in matrix-like languages like Matlab, Scilab, or GNU Octave, and the domain of the conjugate is automatically obtained without any a priori knowledge of the original function. In effect, the domain of the conjugate is implicitly modeled using the range of ∂f.

The first price we pay for such advantages is that we only obtain vertices on the convex envelope of the graph of the conjugate, not the full graph. We have to complete the result by linear interpolation. Consider the following example.

Example 3.4 Consider Example 3.1 with $x_1 := -1$, $x_2 := 0$, and $x_3 := 1$. Then $\mathrm{Conj}(x_1, f_1, g_1) = \mathrm{Conj}(-1, 1, -1) = (-1, 0, -1)$, $\mathrm{Conj}(x_2, f_2, g_2) = \mathrm{Conj}(0, 0, 0) = (0, 0, 0)$, and $\mathrm{Conj}(x_3, f_3, g_3) = \mathrm{Conj}(1, 1, 1) = (1, 1, 1)$. The values of f^* between the grid values are recovered by linear interpolation. The values outside the grid must be computed by extrapolation (see [18] for more details).

See Example 3.6 below for the equivalent effect on computing the Moreau envelope (in that case quadratic interpolation is needed).

The second price we pay is we require more than just a black box returning the values of the function, we also need an approximation of the derivative of f or of its subgradients when such derivative does not exist.

To obtain a similar parametrization for the Moreau envelope, we use formula (3) to deduce (see [18] for details)

$$\begin{cases} z \in x + \lambda \partial f(x), \\ M_\lambda(z) = f(x) + \frac{(z-x)^2}{2\lambda}. \end{cases}$$

The equivalent to the discrete parametrization Conj is then the discrete parametrization Me defined by

$$\text{Me} : (x_i, f_i, g_i, h_i) \mapsto \left(x_i + \lambda g_i, f_i + \frac{\lambda}{2} g_i^2, g_i, \frac{h_i}{1 + \lambda h_i} \right), \tag{5}$$

where, as for Conj, appropriate modifications are made when f is not twice differentiable.

Proposition 3.5 *The function* Me *provides an approximation of the Moreau envelope and, when they exist, of its first two derivatives.*

Proof The discretization (5) can be obtained either from (4) and (3) or in our present convex framework from (4) and using $M_\lambda = (f^* + \frac{\lambda}{2}|\cdot|^2)^*$. The result follows using Proposition 3.2. □

The next example from [18] illustrates the fact that while the graph of the conjugate has to be completed by linear interpolation, the graph of the Moreau envelope has to be completed by quadratic interpolation.

Example 3.6 Consider the function f defined as the convex envelope of $x \mapsto ||x| - 1|$. We have $f(x) = -x - 1$ when $x \le -1$, $f(x) = 0$ when $-1 \le x \le 1$, and $f(x) = x - 1$ when $1 < x$. Its Moreau envelope is

$$M_\lambda(x) = \begin{cases} -x - 1 - \frac{\lambda}{2} & \text{when } x \le -\lambda - 1, \\ \frac{(x+1)^2}{2\lambda} & \text{when } -\lambda - 1 \le x \le -1, \\ 0 & \text{when } -1 \le x \le 1, \\ \frac{(x-1)^2}{2\lambda} & \text{when } 1 \le x \le \lambda + 1, \\ x - 1 - \frac{\lambda}{2} & \text{when } 1 + \lambda < x. \end{cases}$$

However, the parametrization formula gives

$$\begin{cases} z = x - \lambda \text{ and } M_\lambda(z) = -x - 1 + \frac{\lambda}{2} & \text{when } x < -1, \\ z = x \text{ and } M_\lambda(z) = 0 & \text{when } -1 < x < 1, \\ z = x + \lambda \text{ and } M_\lambda(z) = x - 1 + \frac{\lambda}{2} & \text{when } 1 < x. \end{cases}$$

So we deduce

$$M_\lambda(z) = \begin{cases} -z - \frac{\lambda}{2} - 1 & \text{when } z < -\lambda - 1, \\ 0 & \text{when } -1 < z < 1, \\ z - \frac{\lambda}{2} - 1 & \text{when } \lambda + 1 < z. \end{cases}$$

The two remaining segments when x is in $[-\lambda - 1, -1]$ and $[1, \lambda + 1]$, have to be recovered by quadratic interpolation.

We finish the section by pointing out that in addition to conjugation, two further elementary operations are needed to compute transforms such the proximal average: scalar multiplication and addition.

The multiplication by a positive scalar λ (the result is true for $\lambda \in \mathbb{R}$ although convexity is preserved only when $\lambda \geq 0$) trivially gives the following transform

$$\text{scalar} : (x_i, f_i, g_i, h_i) \mapsto (x_i, \lambda f_i, \lambda g_i, \lambda h_i).$$

So scalar multiplication integrates perfectly in our parametric framework.

Unfortunately, the addition operator is not so easy, since adding two discretizations is not straightforward when the functions do not have identical grids. In the extreme case when the functions have no grid point in common, the pointwise addition results in the function that is identically equal to $+\infty$. So the underlying functions have to be modeled between grid points, *e.g.* as piecewise linear functions, to compute their sum. The next section considers several such models

4 The zeroth- and first-order models

In this section, we consider several models to recover a (convex) continuous function from a discrete set of points. We aim to provide an answer to the previous section in terms of an addition operator for the parametric framework. In addition, our results motivate the introduction of the class of piecewise linear-quadratic functions in Sect. 5.

Given a discretization of a function f on a grid x_i and possibly of its derivatives, we say that the function $\check{f}$ is *a model of the function* f if it interpolates the function on the grid *i.e.* $f(x_i) = \check{f}(x_i)$. It is a zeroth-order model if only the value of the function is taken into account, a *first-order model* if both the value of the function and of its first derivative are interpolated *i.e.* $f(x_i) = \check{f}(x_i)$ and $f'(x_i) = \check{f}'(x_i)$, and a second order model if in addition $f''(x_i) = \check{f}''(x_i)$. Again, we do not assume f is differentiable. At any point where it is not, substitute ∂f for f', make the appropriate modifications, and discard the statement for f''. We only write f' to simplify the exposition and clarify the main ideas.

A natural zeroth-order model to consider for convex transforms is the piecewise linear model

$$\check{f}_0(x) := \max_i [f_i + s_i(x - x_i)],$$

where s_i corresponds to increasing slopes in $\mathbb{R}$ (*i.e.* $s_i < s_{i+1}$), not necessarily to the derivative of f at x_i. Since the function f is convex, under a reasonable choice of s_i, we have $\check{f}_0(x_i) = f_i = f(x_i)$, and the function $\check{f}_0$ is convex because it is the maximum of convex functions. Additionally, its conjugate is also piecewise linear [35, 11.14(a)] with bounded domain. This is the implicit model used by the fast transform algorithms. In particular, the LLT algorithm relies explicitly on the finite difference slopes $s_i := (f_{i+1} - f_i)/(x_{i+1} - x_i)$.

The main drawback of this model is its absence of any modeling of the domain of the function f (see Example 2.4). Even if we only consider piecewise linear functions, their conjugate is piecewise linear but with a bounded domain. So its domain has to be approximated using convergence results. Moreover, the lack of second order term means that any quadratic function is approximated by piecewise linear functions. Unfortunately, the Moreau envelope gives rise to quadratic functions for the simplest convex functions: indicators of a point.

We now turn to the parametric framework of Sect. 3. The PLT algorithm uses the following first order model

$$\check{f}_1(x) := \max_i [f_i + g_i(x - x_i)]$$

with $g_i := f'(x_i)$ (or $g_i \in \partial f(x_i)$ when f is not differentiable at x_i). We have $\check{f}_1(x_i) = f(x_i)$, and $\check{f}_1'(x_i) = g_i = f'(x_i)$. Hence $\check{f}_1$ is a first order model. Additionally, formula (4) provides a nice duality relation for the class of piecewise linear functions.

However, $\check{f}_1$ suffers from the same drawbacks as $\check{f}_0$: the class of piecewise linear functions is not closed under standard convex operators. For example, computing the conjugate of a linear function results in an indicator function whose Moreau envelope is a quadratic function. Consequently, convergence results have to be used to obtain a good numerical approximation even for linear, indicator, or quadratic functions.

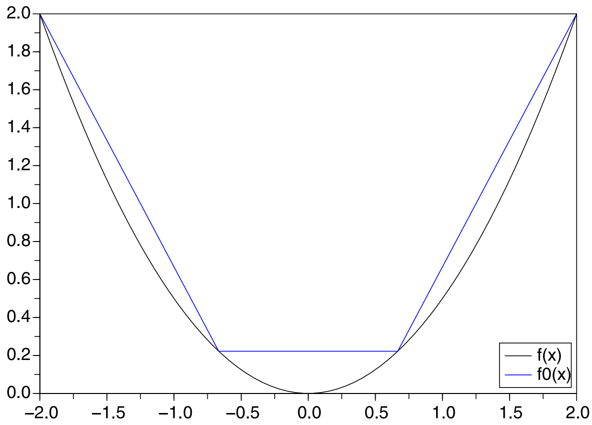
Figure 2a illustrates the zeroth order model $\check{f}_0$ to approximate a quadratic function, while Fig. 2b shows the first order model $\check{f}_1$ to also approximate a quadratic. Taylor approximation justifies the better fitting of the first order model.

Both models can be used to build a discrete addition operator for the parametric framework of the previous section as follows. Consider two functions to be added on two potentially disjoint grids. Consider the union of both grids, and compute any missing value for each function by linear interpolation. Then just add the values at each grid point. While the idea is simple and provides a solution to the parametric framework, it is an ad hoc solution that does not solve the intrinsic issues of a piecewise linear model, namely, the class of piecewise linear functions is not closed under the standard convex operators.

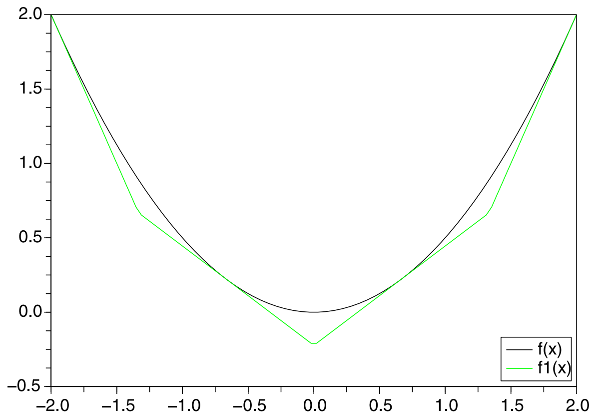
To remedy this shortcoming we consider second order models in the next section using the class of piecewise linear-quadratic functions.

5 Piecewise linear-quadratic (PLQ) functions

Our objective is to work with a class of functions that is closed under the standard convex operators (conjugation, addition, regularization, scalar multiplication).



(a) Zeroth order interpolation as used in the LLT algorithm



(b) First order interpolation as used in the PLT algorithm

Fig. 2 Interpolation schemes for discrete transforms

The class should be rich enough to approximate any convex function, and offer efficient numerical algorithms for all fundamental convex transforms namely addition, scalar multiplication, conjugation, regularization (by taking the Moreau envelope), and combinations.

Let $\mathcal{F}$ be the class of all convex lower-semicontinuous proper extended-valued functions with a piecewise linear subdifferential mapping $\partial f : \mathbb{R} \to 2^{\mathbb{R}}$. The set $\mathcal{F}$ contains the piecewise linear functions, the piecewise quadratic functions, and the sum of any such function with an indicator function. In [35, 10.20, p. 440], $\mathcal{F}$ is defined as the set of convex lower-semicontinuous proper extended-valued functions with piecewise linear domain for which the function is either linear or quadratic on each piece of its domain. A function in $\mathcal{F}$ is called a piecewise linear-quadratic (PLQ) function. In the present paper, we keep the same name even though for functions of one variable the class is usually called convex piecewise quadratic functions (the classes differ when considering functions of more than one variable). (Note that in the present paper, a PLQ function is always convex.) The PLQ functions provide a natural class of functions for convex calculus.

Within that framework, we have the following properties.

Proposition 5.1 (Closedness under convex transforms) *The class of PLQ functions is closed under positive scalar multiplication, addition, conjugation, and taking the Moreau envelope.*

Proof It is clear that the class is closed under (positive) scalar multiplication and addition. Using formula (3), it will be closed by taking Moreau envelope as soon as it is closed for conjugation.

So we only need to prove that the conjugate of a PLQ function f is a PLQ function. The fact that f^* is PLQ comes from [35, 11.14, p. 484]. Alternatively, one can notice that the graph of ∂f^* is piecewise linear as it is the symmetric with respect to the line $y = x$ of the graph of f. So f^* is PLQ.

An alternate proof is to note that the proposition is contained in [35, 10.22, 11.14, 11.32, 11.33]. □

Proposition 5.2 (Efficient Algorithms) *All fundamental convex operators: addition, scalar multiplication, conjugation, and taking the Moreau envelope, can be computed in linear time and space within the class of PLQ functions.*

Proof Assume f, f_0, f_1 are PLQ functions.

First for scalar multiplication, take any $\alpha \in \mathbb{R}$. The function αf is defined on the same grid as the function f (only its values are multiplied by α). So we can compute it in linear time and space by multiplying each coefficients of f by α.

Next we consider the addition of f_0 and f_1. The function $f_0 + f_1$ is PLQ and can be computed in linear time and space *i.e.* in $O(n+m)$, where n and m are the number of intervals partitioning the domain of f_0 and of f_1, respectively. Indeed, elementary computations show that the function $f_0 + f_1$ is PLQ on the grid $\{x_i\} \cup \{y_j\}$, where x_i (resp. y_j) is the grid of f_0 (resp. f_1).

For the conjugation operation, we first provide the details for a representative special case. Consider the PLQ function f defined by

$$f(x) := \begin{cases} \varphi_0(x) := a_0 x^2 + b_0 x + c_0 & \text{if } x \le x_0, \\ \varphi_1(x) := a_1 x^2 + b_1 x + c_1 & \text{otherwise.} \end{cases}$$

It is convex if, and only if,

$$\begin{cases} \varphi_0(x_0) = \varphi_1(x_0) & \text{(continuity condition),} \\ \varphi_0'(x_0) \le \varphi_1'(x_0) & \text{(convexity condition).} \end{cases}$$

Then its conjugate is directly computed as

$$f^*(s) = \begin{cases} \varphi_0^*(s) & \text{if } s \le \varphi_0'(x_0), \\ \bar{s}(s - \varphi_0'(x_0)) + \varphi_0^*(\varphi_0'(x_0)) & \text{if } \varphi_0'(x_0) \le s \le \varphi_1'(x_0), \\ \varphi_1^*(s) & \text{if } s \ge \varphi_1'(x_0), \end{cases}$$

where

$$\bar{s} := \frac{\varphi_1'(x_0)x_0 - \varphi_1(x_0) - \varphi_0'(x_0)x_0 + \varphi_0(x_0))}{\varphi_1'(x_0) - \varphi_0'(x_0)}.$$

In fact, the middle case, which only arises when the function f' is not continuous at x_0, is an affine function bridging the graph of φ_0^* with φ_1^*. Now computing f^* can be done in constant time and space since the computation of φ_0^* and φ_1^* can be performed explicitly directly.

The result follows by considering a general (convex) PLQ function as a sequence of quadratic (or linear) functions from left to right, taking boundary points of the domain into account. Applying variants of the above special case to each interval in the domain of f individually, we can compute the coefficients of that part in constant time and space. Hence, computing f^* takes linear time and space in the number of intervals that partition its domain.

Finally, the existence of a linear time and space algorithm for the Moreau envelope is a direct consequence of formula (3) and of the above algorithms. □

Corollary 5.3 *The proximal average of two convex PLQ functions is a convex PLQ function that can be computed in linear time and space.*

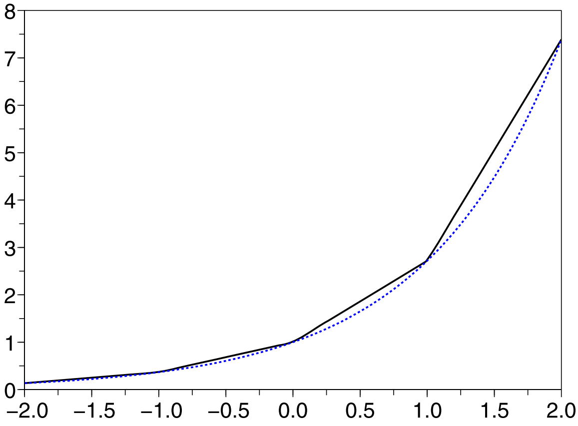
Proof The proximal average decomposes into conjugation, addition, and scalar multiplication, so it is a closed operation on the class of PLQ functions and can be computed in linear time and space. □

The key results in the present section are not only that the class of convex PLQ functions is closed under the operations of addition, positive scalar multiplication, conjugation, and regularization, but also that for each operation we can explicitly compute the resulting transform in linear time in the size of the grid; in other words, each transform has an associated efficient algorithmic implementation.

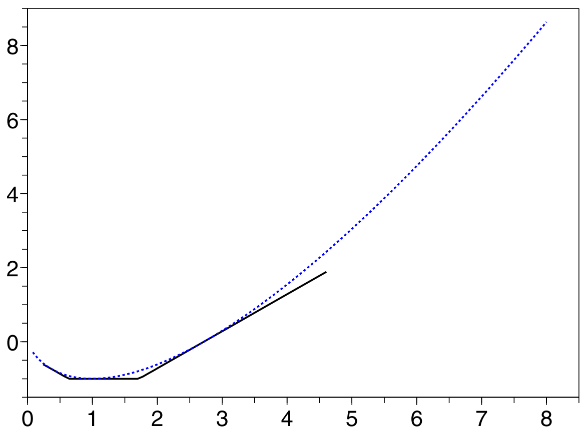
One can view the PLQ algorithms as hybrid symbolic-numeric algorithms: they provide exact results on the class of PLQ functions. Hence, computing the transform of any convex function decomposes into two steps: a numerical approximation step which returns a PLQ function, and a symbolic computation step which returns the transform of that PLQ function. The later step is exact up to floating point arithmetic while the former relies on convergence.

Note also that the domain of a PLQ function is explicitly modeled, and that PLQ functions form an explicit second order model, which resolve the issues pointed out in Sect. 4. Consider again Example 2.4. The PLQ algorithms compute the (exact) proximal average

$$\mathcal{P}(f_0, \lambda, f_1) = (2\lambda - 1)x - 2\lambda(1 - \lambda).$$



(a) Zeroth order approximation



(b) Conjugate of the zeroth-order model

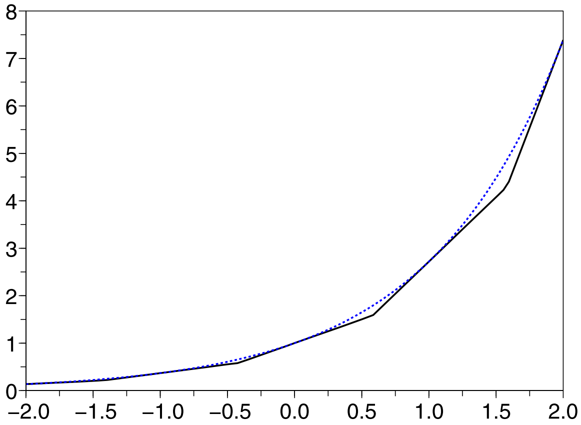
Fig. 3 Approximation of the exponential by a zeroth- and first-order models with the corresponding conjugates

Compare the simplicity of that answer with the one provided by fast algorithms in Example 2.4: no convergence is necessary to approximate an intermediate function, or its domain.

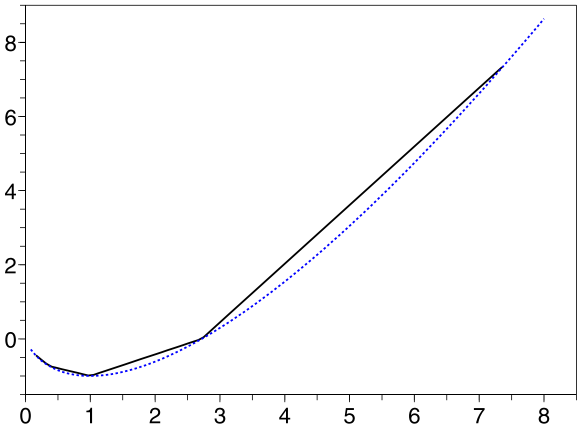
6 Numerical examples

Several examples are presented here to illustrate the powerful convex calculus provided through the PLQ class.

The transform of any function belonging to the PLQ class is computed symbolically. So we immediately obtain the conjugate of the absolute value $(|\cdot|)^* = I_{[-1,1]}$,



(c) First order approximation

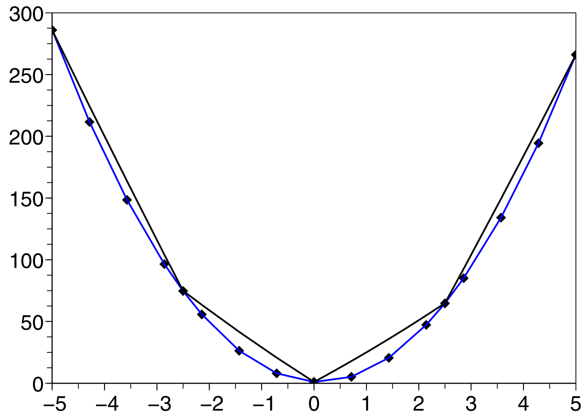


(d) Conjugate of the first-order model

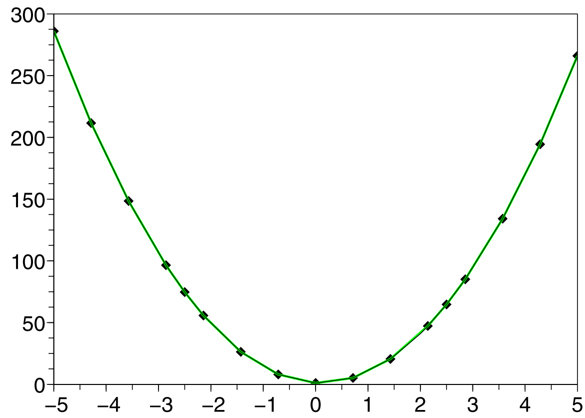
Fig. 3 (*Continued*)

the Moreau envelope of the absolute value, and that the energy function is self-conjugate: $(0.5|\cdot|^2)^* = 0.5|\cdot|^2$. We also obtain that for any linear function $f(x) := ax$, its conjugate is the indicator function $f^* = I_{\{a\}}$. Hence, we do not need to resort to symbolic computation packages like [14] to build simple examples with piecewise quadratic functions.

To illustrate Sect. 4, consider Fig. 3a, which shows a zeroth order model approximating the exponential function. Figure 3b shows the corresponding conjugate: $f^*(x) = x\ln(x) - x$. Compare with Fig. 3c (resp. Fig. 3d) which shows a first-order model approximating the exponential (resp. which shows the conjugate of the first-order model).



(a) Fast algorithms (using piecewise linear functions)

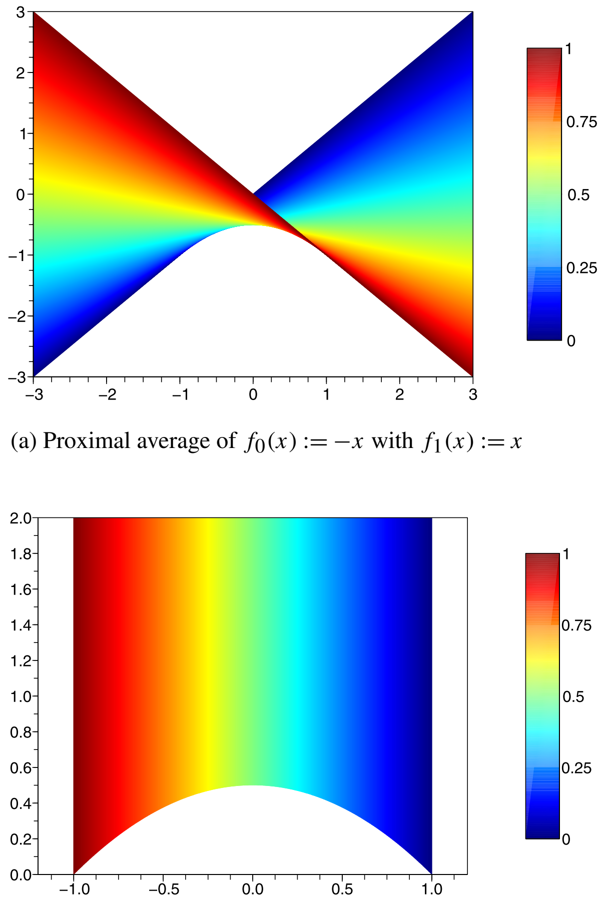


(b) PLQ (using piecewise quadratic functions)

Fig. 4 Comparing the addition of functions $f_1(x) := (x-1)^2$ and $f_2(x) := x^2$

To compare the PLQ framework of Sect. 5 with the Fast Algorithm framework of Sect. 2, consider Fig. 4. The addition of two quadratic functions on disjoint grids (15 points equispaced for function f_1 but 5 points equispaced for function f_2) results in a visible error for the fast algorithms but no visible difference for PLQ functions. (Recall that addition is one of the missing operators for the parametric framework of Sect. 3, which motivated the models of Sect. 4.)

Finally, Fig. 5a shows the proximal average for Example 2.4 computed using the PLQ algorithm. The colour scheme corresponds to the value of the parameter λ in $\mathcal{P}(f_0, \lambda, f_1)$. Figure 5b illustrates another example for which the PLQ algorithm is



(a) Proximal average of $f_0(x) := -x$ with $f_1(x) := x$

(b) Proximal average of $f_0 := I_{\{-1\}}$ with $f_1 := I_{\{1\}}$

Fig. 5 Proximal averages computed exactly by the PLQ algorithm

exact: the proximal average of the indicator function of a single point with the indicator function of (another) single point

$$\mathcal{P}(I_{\{-1\}}, \lambda, I_{\{1\}}) = I_{\{2\lambda-1\}} + 2\lambda(1-\lambda),$$

see [2, Example 7.6] for the general formula.

7 The PLQ toolbox

The PLQ toolbox is a collection of functions to manipulate PLQ functions implemented in Scilab v4.0 [36]. A PLQ function is stored as an $n \times 4$ matrix *e.g.* the

Table 1 Functions provided in the PLQ package

Function name	Description
`plq_eval(plqf,X)`	Evaluate a PLQ function on the grid X
`plq_build(x,f,df)`	Build a first-order model of a function
`plq_add(plqf1,plqf2)`	Addition
`plq_me(plqf,λ)`	Moreau envelope
`plq_scalar(plqf,λ)`	Scalar multiplication.
`plq_pa(plqf1,plqf2,λ)`	Proximal average
`plq_lft(plqf)`	Conjugation

function

$$f(x) := \begin{cases} a_0x^2 + b_0x + c_0 & \text{if } x \le x_0, \\ a_1x^2 + b_1x + c_1 & \text{if } x_0 < x \le x_1, \\ \vdots & \vdots \\ a_{n-1}x^2 + b_{n-1}x + c_{n-1} & \text{if } x_{n-1} < x \le x_{n-1}, \\ a_nx^2 + b_nx + c_n & \text{otherwise} \end{cases}$$

is stored as the matrix

$$\texttt{plqf} := \begin{bmatrix} x_0 & a_0 & b_0 & c_0 \\ x_1 & a_1 & b_1 & c_1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{n-1} & a_{n-1} & b_{n-1} & c_{n-1} \\ +\infty & a_n & b_n & c_n \end{bmatrix},$$

with bounded domains stored by setting $a_0 := b_0 := 0$, $c_0 := +\infty$, and $a_n := b_n := 0$, $c_n := +\infty$. The indicator function $f(x) := I_{\{x_0\}}(x)$ of a single point x_0 is stored, as a special case, as the row vector $\texttt{plqf} := [x_0 \quad 0 \quad 0 \quad +\infty]$.

The toolbox provides the functions listed on Table 1. For example, $\texttt{plq\_lft}([+\infty, 1/2, 0, 0]) = [+\infty, 1/2, 0, 0]$ means the conjugate of the energy $x^2/2$ is the energy itself. Similarly, the fact the conjugate of the absolute value is the indicator function of $[-1, 1]$ is written as

$$\texttt{plq\_lft}\left(\begin{bmatrix} 0 & 0 & -1 & 0 \\ +\infty & 0 & 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} -1 & 0 & 0 & +\infty \\ 1 & 0 & 0 & 0 \\ +\infty & 0 & 0 & +\infty \end{bmatrix},$$

and the Moreau envelope of the absolute value is computed by

$$\texttt{plq\_me}\left(\begin{bmatrix} 0 & 0 & -1 & 0 \\ +\infty & 0 & 1 & 0 \end{bmatrix}, \frac{1}{2}\right) = \begin{bmatrix} -\frac{1}{2} & 0 & -1 & -\frac{1}{4} \\ \frac{1}{2} & 1 & 0 & 0 \\ +\infty & 0 & 1 & -\frac{1}{4} \end{bmatrix}.$$

Note that all these computations do not involve numerical approximation or convergence—they are exact!

```
1 plqf = plq_build([-2, -1, 0, 0.5], exp, exp)
2 plqfstar = plq_lft(plqf)
3 x = linspace(-2, 2)';
4 y = plq_eval(plqf, x);
5 s = linspace(0,2)'; s = s(2:$);
6 ystar = plq_eval(plqfstar, s);
7 scf(0); clf(); scf(1); clf();
8 plot2d(x, y); plot2d(x, exp(x), style=2);
9 plot2d(s, ystar); plot2d(s, s.*log(s)-s, style=2)
```

Listing 1 Computing the conjugate of the exponential

Table 2 Computational convex analysis algorithm comparison

Algorithm	Transform	Convex only	Scales	Order	Complexity ($\mathbb{R}^d$)
`Brute`	Both	No	Yes	0	N^2
`Direct`	Both	No	Yes	0	$N^{1+1/d}$
`FLT`	Conjugate	No	Yes	0	$N^d \log N$
`LLT`	Conjugate	No	Yes	0	dN
`PE`	Moreau	No	Yes	0	dN
`NEP`	Moreau	Yes	Yes	0	dN
`PLT`	Conjugate	Yes	No	1	N
`PLQ`	Conjugate	Yes	No	2	N

Listing 1 illustrates how one can quickly investigate the conjugate of a given function. Line 1 computes a PLQ approximation of the exponential function, and Line 2 computes the conjugate of the exponential: $s \ln(s) - s$. The remaining lines plot the PLQ function along with their exact counterparts. Line 3 builds a uniform grid x of 100 points between -2 and 2. Line 4 evaluates the PLQ function on the grid x, and Line 8 plots both the exponential function and its PLQ approximation. Line 5 builds a grid for the dual space (the singular value 0 is removed), and the conjugate function is plotted along with its PLQ approximation on Line 9.

As for numerical performance, the PLQ package is compared with previous computational convex analysis algorithms in Table 2. The Transform column indicates which transform (Moreau or Conjugate) the core algorithm computes, the Scales column indicates whether the algorithm scales immediately to higher dimension, and the Order column indicates up to what derivative the algorithm takes into account (zeroth order for incorporating function values, first order for using also the values of the first derivative, etc.). The algorithms are fully detailed in [27]. Briefly, `Brute` and `Direct` are provided for comparison only (`Brute` stands for brute force while `Direct` uses the separability of the dot product to speed up the computations in higher dimensions). The `FLT` is historically the first algorithms considered and runs in log-linear time (provided for comparison only, see [23]). The `LLT`, `NEP`, and `PE` belong to the fast algorithm framework of Sect. 2 and are presented in [24], [27] and [10] respectively. The `PLT` belongs to the parametric framework of Sect. 3 and is detailed in [18]. The `PLQ` column stands for algorithms in the PLQ framework.

Table 3 Numerical comparison of the fast algorithms, parametric and PLQ framework for computing the conjugate of the function $f(x) := x^2/2$ on the interval $[-n/2, n/2]$. The given results are in seconds

n	PLQ	PLTc	PLTi	LLTc	LLTi
10,000	0.04	0.00	0.29	0.02	2.20
20,000	0.10	0.01	0.57	0.03	8.61
30,000	0.15	0.01	0.86	0.05	18.75
40,000	0.20	0.02	1.20	0.06	33.16
50,000	0.29	0.02	1.50	0.09	49.71
60,000	0.31	0.03	1.79	0.10	79.84
70,000	0.41	0.03	2.21	0.13	106.56
80,000	0.49	0.05	2.54	0.17	146.57
90,000	0.54	0.05	2.92	0.20	174.68
100,000	0.61	0.06	3.22	0.23	223.60
110,000	0.66	0.06	3.57	0.21	270.26
120,000	0.73	0.08	3.73	0.21	330.63

Table 4 Numerical comparison of the fast algorithms, parametric and PLQ framework for computing the Moreau envelope of the function $f(x) := x^2/2$ on the interval $[-n/2, n/2]$. The results are in seconds

n	PLQ	PLTc	PLTi	LLTc
10,000	0.05	0.01	0.58	0.03
20,000	0.13	0.02	1.32	0.06
30,000	0.19	0.02	1.99	0.15
40,000	0.26	0.03	2.76	0.23
50,000	0.33	0.04	3.28	0.33
60,000	0.38	0.04	3.91	0.45
70,000	0.46	0.06	4.69	0.59
80,000	0.45	0.05	5.19	0.75
90,000	0.50	0.07	6.01	0.83
100,000	0.67	0.08	6.66	0.98
110,000	0.74	0.09	7.28	1.19
120,000	0.80	0.08	8.03	1.57

The current state-of-the-art as summarized in the table indicates that one should use the PLQ framework to compute with univariate functions, and the fast algorithms for functions of more than one variable.

Table 3 provides a running time comparison of the algorithms. The test were run on an IBM Thinkpad T60 with Intel dual core CPU at 1.8 GHz under Linux Mandriva 2007 running Scilab v4.0. The table confirms the linear running time of our implementations.

Table 3 columns `PLQ`, `PLTc`, `PLTi`, `LLTc`, `LLTi` correspond to the PLQ algorithm for computing the conjugate (using interpreted Scilab syntax), the PLT algorithm (Parametric Legendre Transform) using vectorized Scilab syntax, the PLT algorithm using interpreted Scilab syntax, the LLT algorithm (Linear-time Legendre Transform) using Scilab vectorized syntax, and the LLT algorithm using interpreted Scilab syntax respectively. Note that the c in `LLTc` and `PLTc` stands for compiled since these implementations are almost as efficient as compiled functions.

Although the ranking favours `PLTc` then `LLTc`, the comparison with `PLTi` and `LLTi` shows that most of the difference comes from taking advantage of Scilab optimized matrix operations. In contrast, the PLQ algorithm for computing conjugates cannot be straightforwardly put into matrix operations. So it should really be compared with `PLTi` and `LLTi`. In that context, the PLQ algorithm is very competitive, achieving an order of magnitude similar to matrix-optimized codes like `LLTc` and `PLTc`. In addition, keep in mind that the PLT needs to be completed by linear interpolation, and the LLT algorithm requires both a priori knowledge of the domain of the conjugate, and convergence properties to achieve similar results as the PLQ algorithm. In other words, the PLQ algorithm has many significant structural advantages, while enjoying similar computational efficiency as the LLT and PLQ algorithms.

Remark 7.1 If one compares Table 3 with [18, Table 1], one needs to take into account that we computed the conjugate instead of the Moreau envelope, we also assumed the data is convex and did not include the computation of the convex hull with the timing of the LLT algorithm, and finally we used vectorized versions of all three algorithms. Vectorization accounts for the difference between our timings of the LLT and the timings reported in [18].

Table 4 shows computational results for the Moreau envelope. Note that the `LLTi` implementation was not included since it takes a prohibitive amount of time to run for these values of n. The conclusions are the same as for computing the conjugate: the PLQ algorithm for the Moreau envelope is competitive with compiled versions of the PLT and LLT algorithms (in this case, `PLQ` even outperforms `LLTc` for higher values of n) while providing significant structural advantages.

8 Conclusion

To conclude, we reviewed the two existing frameworks in computer-aided convex analysis: fast algorithms, and parametric algorithms. We pointed out their intrinsic limitations, and presented a new framework, based on PLQ functions, that allows for hybrid symbolic-numeric algorithms, and enjoys the following advantages:

- The class of PLQ functions is closed under the standard operations of convex analysis. As such it offers a natural framework for investigation.
- The PLQ functions give rise to natural algorithms with a linear worst-case time (and space) complexity. Moreover, the algorithms are easy to implement: they require only a single loop and no advanced data structures.
- The algorithms do not require a priori knowledge of the domain of the conjugate. Furthermore, there is no need to enlarge the domain of the function, which would be unavoidable when relying on convergence results. Both properties simplify significantly the manipulation of functions when one computes advanced convex operators like the proximal average.
- The PLQ functions offer a flexible framework in which a function can be initially numerically approximated using a zeroth-order model, a first-order model, or a second-order model.

While the parametric framework of Sect. 3 is limited to functions of one variable, fast algorithms extend to functions of several variables using the key fact that computing multivariate conjugates amounts to computing several univariate conjugates. This key property named factorization applies to all convex operators, and has been used in all fast algorithms [23, 24, 27].

The existence of efficient algorithms for the PLQ framework for multivariate functions remains a challenging topic of future research. For example, bivariate PLQ functions have a domain which is the intersection of linear functions, in other words it corresponds to a triangulation of the plane.

Acknowledgements Yves Lucet was partially supported by the Natural Sciences and Engineering Research Council of Canada. Heinz Bauschke was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chair Program. Mike Trienis was partially supported by the Pacific Institute for the Mathematical Sciences, and UBC Okanagan Internal Grant program.

The authors would like to thank the referees for their valuable comments and suggestions which have greatly improved the quality and the presentation of the paper.

References

1. Bauschke, H.H., Burke, J.V., Deutsch, F.R., Hundal, H.S., Vanderwerff, J.D.: A new proximal point iteration that converges weakly but not in norm. Proc. Am. Math. Soc. **133**, 1829–1835 (2005) (electronic)
2. Bauschke, H.H., Lucet, Y., Trienis, M.: How to transform one convex function continuously into another. Technical report, University of British Columbia, July 2006. Accepted for publication in SIAM Review
3. Bauschke, H.H., Matoušková, E., Reich, S.: Projection and proximal point methods: Convergence results and counterexamples. Nonlinear Anal. **56**, 715–738 (2004)
4. Bauschke, H.H., von Mohrenschildt, M.: Symbolic computation of Fenchel conjugates. SIGSAM Bull. **40** (2006)
5. Borwein, J.M., Hamilton, C.H.: Symbolic Fenchel conjugation. Math. Program. (2007). doi:10.1007/s10107-007-0134-4
6. Brenier, Y.: Un algorithme rapide pour le calcul de transformées de Legendre-Fenchel discrètes. C. R. Acad. Sci. Paris Sér. I Math. **308**, 587–589 (1989)
7. Corrias, L.: Fast Legendre-Fenchel transform and applications to Hamilton-Jacobi equations and conservation laws. SIAM J. Numer. Anal. **33**, 1534–1558 (1996)
8. Deniau, L.: Proposition d'un opérateur géométrique pour l'analyse et l'identification de signaux et images. Ph.D. thesis, Université de Paris-Sud, Centre d'Orsay (1997)
9. Deniau, L., Blanc-Talon, J.: Fractal analysis with Hausdorff distance under affine transformations. Technical report, ETCA-CREA-SP (1995)
10. Felzenszwalb, P.F., Huttenlocher, D.P.: Distance transforms of sampled functions. Technical report, TR2004-1963, Cornell Computing and Information Science (2004)
11. Frisch, U., Bec, J.: Burgulence. In: Lesieur, A.M., David, E.F. (eds.) Les Houches 2000: New Trends in Turbulence, pp. 341–383. Springer EDP-Sciences, Berlin (2001)
12. Gavrilova, M., Alsuwaiyel, M.H.: Two algorithms for computing the Euclidean distance transform. Technical report, 2000-661-13, Computer Science Technical Reports, University of Calgary (2000)
13. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)
14. Hamilton, C.H.: Symbolic convex analysis. Master's thesis, Simon Fraser University (2005)
15. Helluy, P.: Simulation numérique des écoulements multiphasiques: de la théorie aux applications. Ph.D. thesis, Institut des Sciences de l'Ingenieur de Toulon et du Var, Laboratoire Modélisation Numérique et Couplages, BP 56, 83162 La Valette CEDEX, France, Jan. 2005. Habilitation à Diriger des Recherches

16. Hiriart-Urruty, J.-B.: Lipschitz r-continuity of the approximate subdifferential of a convex function. Math. Scand. **47**, 123–134 (1980)
17. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vols. 305–306. Springer, Berlin (1993). Vol I: Fundamentals, Vol II: Advanced theory and bundle methods
18. Hiriart-Urruty, J.-B., Lucet, Y.: Parametric computation of the Legendre-Fenchel conjugate with application to the computation of the Moreau envelope. J. Convex Anal. **14**(3), 657–666 (2007)
19. Hisakado, T., Okumura, K., Vukadinovic, V., Trajkovic, L.: Characterization of a simple communication network using Legendre transform. In: Proc. IEEE Int. Symp. Circuits and Systems, vol. 3, pp. 738–741 (2003)
20. Koopen, B.: Contact of bodies in 2D-space: implementing the discrete Legendre transform. AI Master's thesis, Intelligent Autonomous Systems Group, University of Amsterdam (2002)
21. Legras, B., Pisso, I., Berthet, G., Lefïvre, F.: Variability of the Lagrangian turbulent diffusion in the lower stratosphere. Atmos. Chem. Phys. **5**, 1605–1622 (2005)
22. Lemaire, B.: The proximal algorithm. In: New Methods in Optimization and their Industrial Uses (Pau/Paris, 1987). Internat. Schriftenreihe Numer. Math., vol. 87, pp. 73–87. Birkhäuser, Basel (1989)
23. Lucet, Y.: A fast computational algorithm for the Legendre-Fenchel transform. Comput. Optim. Appl. **6**, 27–57 (1996)
24. Lucet, Y.: Faster than the fast Legendre transform, the Linear-time Legendre transform. Numer. Algorithms **16**, 171–185 (1997)
25. Lucet, Y.: The Legendre-Fenchel conjugate: Numerical computation. Technical report, CECM (1998)
26. Lucet, Y.: A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform. In: Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV 2005), Victoria BC, May 2005. IEEE Computer Society, Los Alamitos (2005)
27. Lucet, Y.: Fast Moreau envelope computation I: Numerical algorithms. Numer. Algorithms **43**, 235–249 (2006)
28. Moreau, J.-J.: Propriétés des applications "prox". C. R. Acad. Sci. Paris **256**, 1069–1071 (1963)
29. Moreau, J.-J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
30. Moreau, J.-J.: Convexity and duality. In: Functional Analysis and Optimization, pp. 145–169. Academic, New York (1966)
31. Noullez, A., Gurbatov, S.N., Aurell, E., Simdyankin, S.I.: The global picture of self-similar and not self-similar decay in Burgers turbulence. Technical report, nlin.CD/0409022, arXiv.org eprint archive (2004)
32. Noullez, A., Vergassola, M.: A fast Legendre transform algorithm and applications to the adhesion model. J. Sci. Comput. **9**, 259–281 (1994)
33. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
34. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
35. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
36. Scilab. http://www.scilab.org (1994)
37. She, Z.-S., Aurell, E., Frisch, U.: The inviscid Burgers equation with initial data of Brownian type. Commun. Math. Phys. **148**, 623–641 (1992)
38. Shih, F.Y., Wu, Y.-T.: Fast Euclidean distance transformation in two scans using a 3×3 neighborhood. Comput. Vis. Image Underst. **93**, 195–205 (2004)
39. Tseng, P., Luo, Z.-Q.: On computing the nested sums and infimal convolutions of convex piecewise-linear functions. J. Algorithms **21**, 240–266 (1996)
40. Yosida, K.: Functional Analysis. Classics in Mathematics. Springer, Berlin (1995). Reprint of the 6th edn. (1980)

SIAM J. CONTROL OPTIM.
Vol. 41, No. 4, pp. 1315–1319

ERRATUM: SENSITIVITY ANALYSIS OF THE VALUE FUNCTION FOR OPTIMIZATION PROBLEMS WITH VARIATIONAL INEQUALITY CONSTRAINTS*

YVES LUCET† AND JANE J. YE‡

Abstract. In our paper [*SIAM J. Control Optim.*, 40 (2001), pp. 699–723], due to an error in the proof, an additional assumption is needed for the conclusion of Theorem 3.6 to hold. In this erratum, we restate and prove Theorem 3.6 and correct other related mistakes accordingly.

PII. S036301290139926X

In our paper [1], due to an error in the proof, an additional assumption is needed for the conclusion of Theorem 3.6 to hold. As a consequence, Theorem 4.2 does not hold, each of Theorems 4.4, 4.8, 4.11, and 4.13 requires an additional assumption, and the last two lines on page 701 and the first two lines on page 702 should be changed to

$$M^1 = M^1_{CD}(\Sigma), M^1_C(\Sigma), M^1_S(\Sigma),$$
$$M^0 = M^0_{CD}(\Sigma), M^0_C(\Sigma), M^0_S(\Sigma).$$

We first correct Theorem 3.6 by adding the additional assumption (0.1) as follows.

THEOREM 3.6. *In addition to the basic assumption* (BH), *assume that there exists* $\delta > 0$ *such that the set*

$$\{(x,y) \in C \ : \ \Psi(x,y,\bar\alpha) \le p, H(x,y,\bar\alpha) = q, r \in F(x,y,\bar\alpha) + N_\Omega(y), f(x,y,\bar\alpha) \le M, (p,q,r) \in B(0;\delta)\}$$

is bounded for each M *and the following assumption holds:*

$$(0.1) \qquad (\gamma,\beta,\eta,0) \in M^0(\bar x,\bar y,\bar\alpha) \text{ implies } \gamma = 0,\ \beta = 0,\ \eta = 0.$$

Then the value function $V(\alpha)$ *is lower semicontinuous near* $\bar\alpha$, *and*

$$\partial V(\bar\alpha) \subseteq \bigcup_{(\bar x,\bar y)\in\Sigma(\bar\alpha)} \{-\zeta : (\gamma,\beta,\eta,\zeta) \in M^1(\bar x,\bar y,\bar\alpha)\},$$
$$\partial^\infty V(\bar\alpha) \subseteq \bigcup_{(\bar x,\bar y)\in\Sigma(\bar\alpha)} \{-\zeta : (\gamma,\beta,\eta,\zeta) \in M^0(\bar x,\bar y,\bar\alpha)\},$$

where $M^\lambda(\bar x,\bar y,\bar\alpha)$ *is the set of index* λ *multipliers for problem* GP(p,q,r,α) *at* $(0,0,0,\bar\alpha)$, *i.e., vectors* $(\gamma,\beta,\eta,\zeta)$ *in* $R^d \times R^l \times R^m \times R$ *satisfying*

$$\begin{cases} 0 \in \lambda\partial f(\bar x,\bar y,\bar\alpha) + \partial\langle\Psi,\gamma\rangle(\bar x,\bar y,\bar\alpha) + \partial\langle H,\beta\rangle(\bar x,\bar y,\bar\alpha) + \partial\langle F,\eta\rangle(\bar x,\bar y,\bar\alpha) \\ +\{0\} \times D^*N_\Omega(\bar y, -F(\bar x,\bar y,\bar\alpha))(\eta) \times \{0\} + \{(0,0,\zeta)\} + N_C(\bar x,\bar y) \times \{0\}, \\ \gamma \ge 0 \text{ and } \langle\Psi(\bar x,\bar y,\bar\alpha),\gamma\rangle = 0, \end{cases}$$

*Received by the editors December 6, 2001; accepted for publication (in revised form) May 20, 2002; published electronically December 3, 2002. This work was partly supported by an NSERC research grant.
http://www.siam.org/journals/sicon/41-4/39926.html

†Center for Experimental and Constructive Mathematics, Simon Fraser University, 8888 University Dr., Burnaby, BC, Canada, V5A 1S6 (lucet@cecm.sfu.ca).
‡Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3P4 (janeye@Math.UVic.CA).

and $\Sigma(\bar\alpha)$ is the set of solutions of problem $GP(\bar\alpha)$.

We now make the correct statements for Theorems 4.4, 4.8, 4.11, and 4.13 by translating assumption (0.1) to the case of CD, C, P, and S multipliers, respectively. Unless otherwise indicated, we denote by $\nabla f(x,y,\alpha)$ the gradient of function f with respect to (x,y,α) and not the gradient of f with respect to (x,y) as in section 4 of [1].

THEOREM 4.4. *Assume that there exists $\delta > 0$ such that the set*

$$\{(x,y)\in C : (p,q,r)\in B(0;\delta), \Psi(x,y,\bar\alpha)\le p, H(x,y,\bar\alpha)=q,\\ y\ge 0,\ F(x,y,\bar\alpha)\ge r, \langle y, F(x,y,\bar\alpha)-r\rangle = 0, f(x,y,\bar\alpha)\le M\}$$

is bounded for each M. Assume also that

$$\begin{aligned}
&0\in \nabla\Psi(\bar x,\bar y,\bar\alpha)^\top\gamma + \nabla H(\bar x,\bar y,\bar\alpha)^\top\beta + \nabla F(\bar x,\bar y,\bar\alpha)^\top\eta + (0,\xi,0) + N_C(\bar x,\bar y)\times\{0\},\\
&\gamma\ge 0 \text{ and } \langle\Psi(\bar x,\bar y,\bar\alpha),\gamma\rangle = 0,\\
&\xi_i = 0 \qquad \text{if } \bar y_i > 0 \text{ and } F_i(\bar x,\bar y,\bar\alpha)=0,\\
&\eta_i = 0 \qquad \text{if } \bar y_i = 0 \text{ and } F_i(\bar x,\bar y,\bar\alpha)>0,\\
&\text{either } \xi_i<0, \eta_i<0 \text{ or } \xi_i\eta_i = 0 \qquad \text{if } \bar y_i = 0 \text{ and } F_i(\bar x,\bar y)=0
\end{aligned}$$

implies that $\gamma = 0$, $\beta = 0$, $\eta = 0$. Then the value function V is lower semicontinuous near $\bar\alpha$, and

$$\partial V(\bar\alpha)\subseteq \bigcup_{(\bar x,\bar y)\in\Sigma(\bar\alpha)} \{\nabla_\alpha f(\bar x,\bar y,\bar\alpha) + \nabla_\alpha\Psi(\bar x,\bar y,\bar\alpha)^\top\gamma + \nabla_\alpha H(\bar x,\bar y,\bar\alpha)^\top\beta \\ +\nabla_\alpha F(\bar x,\bar y,\bar\alpha)^\top\eta : (\gamma,\beta,\eta)\in M^1_{CD}(\bar x,\bar y)\}, \tag{0.2}$$

$$\partial^\infty V(\bar\alpha)\subseteq \bigcup_{(\bar x,\bar y)\in\Sigma(\bar\alpha)} \{\nabla_\alpha\Psi(\bar x,\bar y,\bar\alpha)^\top\gamma + \nabla_\alpha H(\bar x,\bar y,\bar\alpha)^\top\beta \\ +\nabla_\alpha F(\bar x,\bar y,\bar\alpha)^\top\eta : (\gamma,\beta,\eta)\in M^0_{CD}(\bar x,\bar y)\}. \tag{0.3}$$

If the set in the right-hand side of inclusion (0.3) *contains only the zero vector, then the value function V is Lipschitz near $\bar\alpha$. If the set in the right-hand side of inclusion* (0.3) *contains only the zero vector and the set in the right-hand side of inclusion* (0.2) *is a singleton, then the value function is strictly differentiable at $\bar\alpha$.*

THEOREM 4.8. *Assume that there exists $\delta > 0$ such that the set*

$$\{(x,y)\in C : (p,q,q^m)\in B(0;\delta), \Psi(x,y,\bar\alpha)\le p, H(x,y,\bar\alpha)=q,\\ \min\{y_i, F_i(x,y,\bar\alpha)\} = q_i^m, i=1,\dots,m, f(x,y,\bar\alpha)\le M\}$$

is bounded for each M. Assume also that

$$\begin{aligned}
&0\in \nabla\Psi(\bar x,\bar y,\bar\alpha)^\top\gamma + \nabla H(\bar x,\bar y,\bar\alpha)^\top\beta + \nabla F(\bar x,\bar y,\bar\alpha)^\top\eta + (0,\xi,0) + N_C(\bar x,\bar y)\times\{0\},\\
&\gamma\ge 0, \langle\Psi,\gamma\rangle(\bar x,\bar y,\bar\alpha) = 0,
\end{aligned}$$

where

$$\begin{aligned}
&\eta_i = 0 \quad \forall i\in I_+,\\
&\xi_i = 0 \quad \forall i\in L,\\
&\eta_i = r_i(1-\bar t_i), \xi_i = r_i\bar t_i \text{ for some } \bar t_i\in[0,1], \quad \forall i\in I_0
\end{aligned}$$

implies that $\gamma = 0$, $\beta = 0$, $\eta = 0$, $r_i = 0$, $i = 1, \dots, m$. *Then the value function* V *is lower semicontinuous near* $\bar{\alpha}$, *and*

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta + \nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in M_C^1(\bar{x},\bar{y})\}, \tag{0.4}$$

$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta + \nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in M_C^0(\bar{x},\bar{y})\}. \tag{0.5}$$

If the set in the right-hand side of inclusion (0.5) *contains only the zero vector, then the value function* V *is Lipschitz near* $\bar{\alpha}$. *If the set in the right-hand side of inclusion* (0.5) *contains only the zero vector and the set in the right-hand side of inclusion* (0.4) *is a singleton, then the value function is strictly differentiable at* $\bar{\alpha}$.

THEOREM 4.11. *Assume that there exists* $\delta > 0$ *such that, for* $(\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})$ *and each index set* $\sigma \subseteq I_0(\bar{x},\bar{y})$, *the set in Proposition* 4.10 *is bounded for each* M *and*

$$\begin{cases} 0 = \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top \beta + \nabla F(\bar{x},\bar{y},\bar{\alpha})^\top \eta + (0,\xi,0) + N_C(\bar{x},\bar{y}) \times \{0\}, \\ \gamma_{J(\Psi)} = 0, \eta_{I_+} = 0, \xi_L = 0, \xi_\sigma \le 0, \eta_{I_0\setminus\sigma} \le 0, \end{cases}$$

implies that $\gamma = 0$, $\beta = 0$, $\eta = 0$. *Then the value function* V *is lower semicontinuous near* $\bar{\alpha}$, *and*

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta + \nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in \cup_{\sigma\subseteq I_0} M_\sigma^1(\bar{x},\bar{y})\}, \tag{0.6}$$

$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta + \nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in \cup_{\sigma\subseteq I_0} M_\sigma^0(\bar{x},\bar{y})\}. \tag{0.7}$$

If the set in the right-hand side of inclusion (0.7) *contains only the zero vector, then the value function* V *is Lipschitz near* $\bar{\alpha}$. *If the set in the right-hand side of inclusion* (0.7) *contains only the zero vector and the set in the right-hand side of inclusion* (0.6) *is a singleton, then the value function is strictly differentiable at* $\bar{\alpha}$.

THEOREM 4.13. *In addition to the assumptions of Theorem* 4.11, *assume that* $C = R^n \times R^a \times R^b$ *and, for all* $(\bar{x},\bar{z},\bar{u}) \in \Sigma(\bar{\alpha})$, *the partial MPEC linear independence constraint qualification is satisfied; i.e.,*

$$\begin{cases} 0 = \nabla_{x,y}\Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_{x,y} H(\bar{x},\bar{y},\bar{\alpha})^\top \beta + \nabla_{x,y} F(\bar{x},\bar{y},\bar{\alpha})^\top \eta + (0,\xi), \\ \gamma_{J(\Psi)} = 0, \eta_{I_+} = 0, \xi_L = 0, \end{cases}$$

implies that $\eta_{I_0} = 0, \xi_{I_0} = 0$, *where* $J(\Psi) := \{i : \Psi_i(\bar{x},\bar{y},\bar{\alpha}) < 0\}$. *Further assume that*

$$\begin{cases} 0 = \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top \beta + \nabla F(\bar{x},\bar{y},\bar{\alpha})^\top \eta + (0,\xi,0), \\ \gamma_{J(\Psi)} = 0, \eta_{I_+} = 0, \xi_L = 0, \eta_{I_0} \le 0, \xi_{I_0} \le 0, \end{cases}$$

implies that $\gamma = 0$, $\beta = 0$, $\eta = 0$. *Then the value function* V *is lower semicontinuous near* $\bar{\alpha}$, *and*

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta$$
$$+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in M_S^1(\bar{x},\bar{y})\},$$
$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta$$
$$+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in M_S^0(\bar{x},\bar{y})\}.$$

Note that the additional assumption (0.1) and its corresponding assumptions in Theorems 4.4., 4.8, 4.11, and 4.13 are automatically satisfied in the case in which the perturbation is additive. In the case of nonadditive perturbations, they are needed even in the case of nonlinear programming, i.e., when $\Omega = R^m$ in Theorem 3.6.

The main error occurs in the proof of Theorem 3.6 when we applied [1, Proposition 2.6] to obtain the partial subdifferentials from the subdifferentials of the fully perturbed value function. The positions of vectors ζ and 0 were switched by mistake. Instead of proving that $(\zeta, 0) \in \partial^\infty \widetilde{V}(0,\bar{\alpha})$ implies $\zeta = 0$, we proved that $(0,\zeta) \in \partial^\infty \widetilde{V}(0,\bar{\alpha})$ implies $\zeta = 0$. Hence, on page 709 in lines 13–18, "For any $(0,0,0,\zeta) \in \partial^\infty \widetilde{V}(0,0,0,\bar{\alpha})$, we have $(0,0,0,\zeta) \in -M^0(\bar{x},\bar{y},\bar{\alpha})$ for some point $(\bar{x},\bar{y},\bar{\alpha}) \in \Sigma(0,0,0,\bar{\alpha})$. Therefore,

$$(0,0,\zeta) \in N_C(\bar{x},\bar{y}) \times \{0\},$$

which implies that $\zeta = 0$" should be changed to "For any $(-\gamma,-\beta,-\eta,0) \in \partial^\infty \widetilde{V}(0,0,0,\bar{\alpha})$, we have $(-\gamma,-\beta,-\eta,0) \in -M^0(\bar{x},\bar{y},\bar{\alpha})$ for some point $(\bar{x},\bar{y},\bar{\alpha}) \in \Sigma(0,0,0,\bar{\alpha})$. Hence $(\gamma,\beta,\eta,0) \in M^0(\bar{x},\bar{y},\bar{\alpha})$, which implies $\gamma = 0$, $\beta = 0$, $\eta = 0$ by assumption (0.1)."

Consider the nonlinear programming formulation of (OPCC) in [1, section 4.1]. Assumption (0.1) amounts to the nonexistence of a nonzero vector $(\gamma,\beta,r^F,r^y,\mu)$ such that

$$\begin{aligned}
&0 \in \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top \beta \\
&-\nabla F(\bar{x},\bar{y},\bar{\alpha})^\top r^F - \{(0,r^y,0)\} + \mu\nabla\langle y,F\rangle(\bar{x},\bar{y},\bar{\alpha}) + N_C(\bar{x},\bar{y}) \times \{(0)\}, \\
&\gamma \geq 0, \langle \gamma, \Psi(\bar{x},\bar{y},\bar{\alpha})\rangle = 0, \\
&r^F \geq 0, r^y \geq 0, \langle r^F, F(\bar{x},\bar{y},\bar{\alpha})\rangle = 0, \langle r^y, \bar{y}\rangle = 0.
\end{aligned}$$

However, using [1, Proposition 4.16] with x replaced by (x,α), the above assumption will never be satisfied, and hence [1, Theorem 4.2] does not hold. Consider the following example, which is the example in [1] with the extra constraint $(x,y) \in [-1,1] \times [-1,1]$:

$$\begin{aligned}
\text{minimize} \quad & -y \\
\text{subject to} \quad & x - y = 0, \\
& x \geq 0, y \geq 0, xy = 0, (x,y) \in [-1,1] \times [-1,1].
\end{aligned}$$

Note that the growth hypothesis holds since the set $[-1,1] \times [-1,1]$ is compact. The normal multiplier set $M^1_{NLP}(0,0) = \emptyset$. So [1, Theorem 4.2] is not true for this example.

That is, the nonlinear programming multipliers may not be useful in the sensitivity analysis.

Acknowledgments. The authors would like to thank the anonymous referees, in particular, referee 2, and the associate editor who suggested listing the statements that should be corrected in the beginning of the erratum.

REFERENCE

[1] Y. Lucet and J. J. Ye, *Sensitivity analysis of the value function for optimization problems with variational inequality constraints*, SIAM J. Control Optim., 40 (2001), pp. 699–723.

SIAM J. CONTROL OPTIM.
Vol. 40, No. 3, pp. 699–723

SENSITIVITY ANALYSIS OF THE VALUE FUNCTION FOR OPTIMIZATION PROBLEMS WITH VARIATIONAL INEQUALITY CONSTRAINTS*

YVES LUCET† AND JANE J. YE‡

Abstract. In this paper we perform sensitivity analysis for optimization problems with variational inequality constraints (OPVICs). We provide upper estimates for the limiting subdifferential (singular limiting subdifferential) of the value function in terms of the set of normal (abnormal) coderivative (CD) multipliers for OPVICs. For the case of optimization problems with complementarity constraints (OPCCs), we provide upper estimates for the limiting subdifferentials in terms of various multipliers. An example shows that the other multipliers may not provide useful information on the subdifferentials of the value function, while the CD multipliers may provide tighter bounds. Applications to sensitivity analysis of bilevel programming problems are also given.

Key words. sensitivity analysis, optimization problems, variational inequality constraints, complementarity constraints, limiting subdifferentials, value functions, bilevel programming problems

AMS subject classification. 49K40

PII. S0363012999361718

1. Introduction.

In this paper, we consider the sensitivity analysis for the following optimization problem with variational inequality constraints (OPVIC):

$$\text{(OPVIC)}\qquad \begin{aligned} &\text{minimize} && f(x,y)\\ &\text{subject to} && \Psi(x,y)\le 0, H(x,y)=0, (x,y)\in C,\\ &&& y\in\Omega, \langle F(x,y), y-z\rangle \le 0 \quad \forall z\in\Omega, \end{aligned} \tag{1}$$

where the following basic assumptions are satisfied:

(BA) The functions $f : R^{n+m} \to R$, $\Psi : R^{n+m} \to R^d$, $H : R^{n+m} \to R^l$, and $F : R^{n+m} \to R^m$ are Lipschitz near any given point of C; C is a closed subset of R^{n+m}, and Ω is a closed convex subset of R^m. Note that the OPVIC is also called the mathematical program with equilibrium constraints (MPEC).

By definition of a normal cone in the sense of convex analysis, the variational inequality (1) is equivalent to saying that $y \in \Omega$ and the vector $-F(x,y)$ is in the normal cone of the convex set Ω at y. Hence the OPVIC can be rewritten as an optimization problem with a generalized equation constraint:

$$\text{(GP)}\qquad \begin{aligned} &\text{minimize} && f(x,y)\\ &\text{subject to} && \Psi(x,y)\le 0, H(x,y)=0, (x,y)\in C,\\ &&& 0\in F(x,y)+N_\Omega(y), \end{aligned} \tag{2}$$

where

$$N_\Omega(y) := \begin{cases} \text{the normal cone of } \Omega \text{ if } y\in\Omega,\\ \emptyset \text{ otherwise} \end{cases}$$

*Received by the editors September 15, 1999; accepted for publication (in revised form) April 2, 2001; published electronically September 7, 2001. This work was partly supported by the Pacific Institute for the Mathematical Sciences and by an NSERC research grant.
http://www.siam.org/journals/sicon/40-3/36171.html

†Centre for Experimental and Constructive Mathematics, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia, Canada V5A 1S6 (lucet@cecm.sfu.ca).

‡Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3P4 (janeye@Math.UVic.CA).

is the normal cone operator.

Let $(\bar{x}, \bar{y})$ be an optimal solution of the OPVIC. If $N_\Omega(y)$ is single-valued and smooth, then the generalized equation constraint (2) would reduce to an ordinary equation $0 = F(x, y) + N_\Omega(y)$. Moreover, if all problem data are smooth and there is no abstract constraint, then the Fritz John necessary optimality condition can be stated as follows. There exist scalar $\lambda \geq 0$ and the vectors (γ, β, η) not all zero such that

$$\begin{cases} 0 = \lambda \nabla f(\bar{x}, \bar{y}) + \nabla \Psi(\bar{x}, \bar{y})^\top \gamma + \nabla H(\bar{x}, \bar{y})^\top \beta + \nabla F(\bar{x}, \bar{y})^\top \eta + \{0\} \times \nabla N_\Omega(\bar{y})^\top \eta, \\ \gamma \geq 0, \text{ and } \langle \Psi(\bar{x}, \bar{y}), \gamma \rangle = 0, \end{cases}$$

where ∇ denotes the usual gradient and $A^\top$ denotes the transpose of a matrix A. In general, however, the map $y \rightrightarrows N_\Omega(y)$ is a set-valued map. Naturally, the usual gradient $\nabla N_\Omega(\bar{y})$ has to be replaced by some kinds of derivatives of set-valued maps.

The Kuhn–Tucker-type necessary conditions with the transpose of the usual gradient ∇N_Ω replaced by the Mordukhovich coderivative $D^* N_\Omega$ were first derived in Ye and Ye [24] under the so-called pseudo-upper-Lipschitz condition for the case of no inequality, no equality constraints, and an abstract constraint in x only. They were further studied under the strong regularity condition in the sense of Robinson and the generalized Mangasarian–Fromovitz constraint qualifications by Outrata in [14] in the case of complementarity constraints and constraints in x only. The first order theory including the necessary optimality conditions involving the Mordukhovich coderivative, various constraint qualifications and their relationships for the general setting of this paper was given in Ye [23]. (Although the equality constraint $H(x, y) = 0$ was not considered explicitly there, the general results under the presence of an equality constraint still hold without any difficulty.) In Ye [22] the Kuhn–Tucker-type necessary conditions with the proximal coderivative for the case of optimization problems with complementarity constraints (OPCCs) were also studied . For recent developments and references on other optimality conditions and computational algorithms, the reader is referred to recent monographs of Luo, Pang, and Ralph [8] and Outrata, Kočvara, and Zowe [15].

In this paper we continue the study by considering the value function $V(p, q, r)$ associated with the right-hand side perturbations

$$\text{(3)} \qquad \mathrm{GP}(p, q, r) \qquad \begin{array}{ll} \text{minimize} & f(x, y) \\ \text{subject to} & \Psi(x, y) \leq p, H(x, y) = q, (x, y) \in C, \\ & r \in F(x, y) + N_\Omega(y), \end{array}$$

i.e.,

$$V(p, q, r) := \inf\{f(x, y) : \begin{array}{l} \Psi(x, y) \leq p, H(x, y) = q, (x, y) \in C \\ r \in F(x, y) + N_\Omega(y)\}, \end{array}$$

where by convention $\inf \emptyset := +\infty$.

Our main result shows that as in sensitivity analysis for ordinary nonlinear programming (NLP) problems, under certain growth hypotheses, the value function V is lower semicontinuous near 0, and the limiting subdifferentials of the value functions are contained in the negative of the multiplier sets, i.e.,

$$\partial V(0) \subseteq -M^1(\Sigma), \tag{4}$$

$$\partial^\infty V(0) \subseteq -M^0(\Sigma). \tag{5}$$

where Σ is the set of solutions of GP and $M^\lambda(\Sigma)$ is the set of index λ CD multipliers for problem GP, which is the set of vectors (γ, β, η) satisfying the Fritz John necessary condition stated above with the transpose of the usual gradient ∇N_Ω replaced by the Mordukhovich coderivative $D^* N_\Omega$ in the case of smooth problem data and no abstract constraints.

In the case of $M^0(\Sigma) = \{0\}$, (5) implies that the singular limiting subgradient $\partial^\infty V(0)$ contains only the zero vector, and hence the value function is Lipschitz continuous near 0. Moreover, if the optimal solution is unique, if the set of abnormal multipliers $M^0(\Sigma)$ contains only the zero vector, and if the set of Kuhn–Tucker multipliers $M^1(\Sigma)$ is a singleton ζ, then inclusion (4) implies that the value function is smooth and $\nabla V(0) = -\zeta$.

In the case where $\Omega = R^m_+, C = R^{n+m}$, OPVIC reduces to the following OPCC.

$$\text{(OPCC)} \qquad \begin{array}{ll} \text{minimize} & f(x,y), \\ \text{subject to} & \Psi(x,y) \leq 0, H(x,y) = 0, \\ & y \geq 0,\ F(x,y) \geq 0, \langle y, F(x,y)\rangle = 0. \end{array}$$

In this case (when all functions involved are smooth), an index λ CD multiplier set corresponding to a feasible solution $(\bar{x}, \bar{y})$ denoted by $M^\lambda_{CD}(\bar{x}, \bar{y})$ consists of $(\gamma, \beta, \eta) \in R^d \times R^l \times R^m$ such that

$$0 = \lambda \nabla f(\bar{x}, \bar{y}) + \nabla \Psi(\bar{x}, \bar{y})^\top \gamma + \nabla H(\bar{x}, \bar{y})^\top \beta + \nabla F(\bar{x}, \bar{y})^\top \eta + (0, \xi), \tag{6}$$

$$\gamma \geq 0 \text{ and } \langle \Psi(\bar{x}, \bar{y}), \gamma \rangle = 0, \tag{7}$$

$$\xi_i = 0 \qquad \text{if } \bar{y}_i > 0 \text{ and } F_i(\bar{x}, \bar{y}) = 0, \tag{8}$$

$$\eta_i = 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x}, \bar{y}) > 0, \tag{9}$$

and

$$\text{either } \xi_i < 0,\ \eta_i < 0, \text{or } \xi_i \eta_i = 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x}, \bar{y}) = 0.$$

We call vectors $(\gamma, \beta, \eta) \in R^d \times R^l \times R^m$ satisfying (6)–(9) and

$$\xi_i \eta_i \geq 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x}, \bar{y}) = 0$$

an index λ C-multiplier set and denote it by $M^\lambda_C(\bar{x}, \bar{y})$, and we call those satisfying (6)–(9) and

$$\xi_i \leq 0, \eta_i \leq 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x}, \bar{y}) = 0$$

an index λ S-multiplier set and denote it by $M^\lambda_S(\bar{x}, \bar{y})$.

Under certain growth hypotheses, we show that the value function

$$\begin{aligned} V(p,q,r) := \{f(x,y) : &\Psi(x,y) \leq p, H(x,y) = q, \\ & y \geq 0, F(x,y) - r \geq 0, \langle y, F(x,y) - r \rangle = 0\} \end{aligned}$$

is lower semicontinuous near 0 and

$$\partial V(0) \subseteq -M^1 \quad \partial^\infty V(0) \subseteq -M^0,$$

where

$$\begin{aligned} M^1 &= M^1_{CD}(\Sigma), M^1_C(\Sigma), M^1_S(\Sigma), \text{ or } \{(\gamma, \beta, \mu\bar{y} - r^F) : (\gamma, \beta, r^F, r^y, \mu) \in M^1_{NLP}(\Sigma)\}, \\ M^0 &= M^0_{CD}(\Sigma), M^0_C(\Sigma), M^0_S(\Sigma), \text{ or } \{(\gamma, \beta, \mu\bar{y} - r^F) : (\gamma, \beta, r^F, r^y, \mu) \in M^0_{NLP}(\Sigma)\}, \end{aligned}$$

where $M^{\lambda}_{NLP}(\bar{x},\bar{y})$ is the set of index λ ordinary NLP multipliers when the OPCC is treated as an ordinary NLP problem.

Moreover, we show that the above multiplier sets can be ordered as follows:

$$\{(\gamma,\beta,\mu\bar{y}-r^F):(\gamma,\beta,r^F,r^y,\mu)\in M^{\lambda}_{NLP}(\Sigma)\}\subseteq M^{\lambda}_{S}(\Sigma)\subseteq M^{\lambda}_{CD}(\Sigma)\subsetneq M^{\lambda}_{C}(\Sigma).$$

It is obvious that one should use the smallest multiplier sets as possible. However, the smaller multiplier sets may be empty and hence may not provide any information on the properties of the value function. We show that under reasonable constraint qualifications such as the generalized Mangasarian–Fromovitz constraint qualification and the strongly regular constraint qualification, the abnormal CD multiplier set contains only the zero vector, and the set of normal CD multipliers is nonempty. An example is given to show that in sensitivity analysis the CD multipliers may provide more useful information than the other multipliers. In this example, the value function is Lipschitz, and the limiting subdifferentials of the value function coincide with the set of negative CD multipliers, while the limiting subdifferentials are contained strictly in the set of negative C multipliers and the set of P multipliers, NLP multipliers, and S multipliers are empty. Applications to the bilevel programming problem are also given.

In this paper we deal only with the sensitivity analysis of the optimal values. For the sensitivity analysis of the optimal solutions, the reader is referred to Scheel and Scholtes [19].

The following notations are used throughout the paper: B denotes the open unit ball; $B(\bar{z};\delta)$ denotes the open ball centered at $\bar{z}$ with radius $\delta>0$. For a set E, co$E$ denotes the convex hull of E, and int$E$ and cl$E$ denote the interior and the closure of E, respectively. The notation $\langle a,b\rangle$ denotes the inner products of vectors a and b. For a differentiable function f, $\nabla f(\bar{x})$ denotes the gradient of f at $\bar{x}$. For a vector $a\in R^n$, a_i denotes the ith component of a. For an m by n matrix A and index sets $I\subseteq\{1,2,\dots,m\}$, $J\subseteq\{1,2,\dots,n\}$, A_I and $A_{I,J}$ denote the submatrix of A with rows specified by I and the submatrix of A with rows and columns specified by I and J, respectively. $A^\top$ denotes the transpose of a matrix A. For a vector $d\in R^m$, d_I is the subvector composed from the components $d_i, i\in I$.

2. Preliminaries. The purpose of this section is to provide the background material on nonsmooth analysis which will be used later. We give only concise definitions and facts that will be needed in the paper. For more detailed information on the subject, our references are Clarke [3], Loewen [7], Rockafellar and Wets [18], and Mordukhovich [10, 12, 13].

First we give some definitions for various subdifferentials and normal cones.

DEFINITION 2.1. *Let $f:R^n\to R\cup\{+\infty\}$ be lower semicontinuous and finite at $\bar{x}\in R^n$. The proximal subdifferential of f at $\bar{x}$ is the set defined by*

$$\partial^\pi f(\bar{x})=\{v\in R^n:\exists M>0,\delta>0 \text{ s.t. } f(x)\geq f(\bar{x})+\langle v,x-\bar{x}\rangle+M\|x-\bar{x}\|^2\ \forall x\in\bar{x}+\delta B\},$$

the limiting subdifferential of f at $\bar{x}$ is the set defined by

$$\partial f(\bar{x}):=\left\{v\in R^n:v=\lim_{\nu\to\infty}v^\nu \text{ with } v^\nu\in\partial^\pi f(x^\nu) \text{ and } x^\nu\to\bar{x}\right\},$$

the singular limiting subdifferential of f at $\bar{x}$ is the set defined by

$$\partial^\infty f(\bar{x}):=\left\{v\in R^n:v=\lim_{\nu\to\infty}\lambda^\nu v^\nu \text{ with } v^\nu\in\partial^\pi f(x^\nu) \text{ and } \lambda^\nu\downarrow 0, x^\nu\to\bar{x}\right\}.$$

Let $f : R^n \to R$ be Lipschitz near $\bar{x} \in R^n$. The Clarke generalized gradient of f at $\bar{x}$ is the set

$$\partial_C f(\bar{x}) := clco\partial f(\bar{x}).$$

For set-valued maps, the definition for a limiting normal cone leads to the definition of the coderivative of a set-valued map introduced by Mordukhovich in [9].

DEFINITION 2.2. *For a closed set $C \subset R^n$ and $\bar{x} \in C$, the proximal normal cone to C at $\bar{x}$ is defined by*

$$N_C^\pi(\bar{x}) := \{v \in R^n : \exists M > 0 \text{ s.t. } \langle v, x - \bar{x}\rangle \leq M\|x - \bar{x}\|^2 \quad \forall x \in C\},$$

and the limiting normal cone to C at $\bar{x}$ is defined by

$$N_C(\bar{x}) := \left\{\lim_{\nu\to\infty} v^\nu : v^\nu \in N_C^\pi(x^\nu), x^\nu \to \bar{x}\right\}.$$

DEFINITION 2.3. *Let $\Phi : R^n \rightrightarrows R^q$ be a set-valued map. Let $(\bar{x}, \bar{p}) \in clGph\Phi$, where $Gph\Phi := \{(x,p) \ : \ p \in \Phi(x)\}$ is the graph of the set-valued map Φ. The set-valued map $D^*\Phi(\bar{x},\bar{p})$ from R^q into R^n, defined by*

$$D^*\Phi(\bar{x},\bar{p})(\eta) := \{\xi \in R^n \ : \ (\xi, -\eta) \in N_{Gph\Phi}(\bar{x},\bar{p})\},$$

is called the Mordukhovich coderivative of Φ at $(\bar{x},\bar{p})$.

In general, we have the following inclusions, which may be strict:

$$\partial^\pi f(\bar{x}) \subseteq \partial f(\bar{x}) \subseteq \partial_C f(\bar{x}).$$

In the case where f is a convex function, all subdifferentials coincide with the subdifferentials in the sense of convex analysis, i.e.,

$$\partial^\pi f(\bar{x}) = \partial f(\bar{x}) = \partial_C f(\bar{x}) = \{\zeta : f(x) - f(\bar{x}) \geq \langle \zeta, x - \bar{x}\rangle \quad \forall x\}.$$

In the case where f is strictly differentiable (see the definition, e.g., in Clarke [2]), we have

$$\partial f(\bar{x}) = \partial_C f(\bar{x}) = \{\nabla f(\bar{x})\}.$$

The following facts about the subdifferentials are well known.

PROPOSITION 2.4.

(i) *A function $f : R^n \to R$ is Lipschitz near $\bar{x}$ and $\partial f(\bar{x}) = \{\zeta\}$ if and only if f is strictly differentiable at $\bar{x}$ and the gradient of f at $\bar{x}$ equals ζ.*

(ii) *A function $f : R^n \to R$ is Lipschitz near $\bar{x}$ if and only if $\partial^\infty f(\bar{x}) = \{0\}$.*

(iii) *If a function $f : R^n \to R$ is Lipschitz near $\bar{x}$ with positive constant L_f, then $\partial f(\bar{x}) \subseteq L_f clB$.*

The following calculus rules will be useful and can be found in the references given in the beginning of this section.

PROPOSITION 2.5 (see, e.g., [7, Proposition 5A.4]). *Let $f : R^n \to R$ be Lipschitz near $\bar{x}$, and let $g : R^n \to R \cup \{+\infty\}$ be lower semicontinuous and finite at $\bar{x}$. Then*

$$\partial(f+g)(\bar{x}) \subseteq \partial f(\bar{x}) + \partial g(\bar{x}),$$
$$\partial^\infty(f+g)(\bar{x}) \subseteq \partial^\infty g(\bar{x}).$$

PROPOSITION 2.6 (see, e.g., [7, Lemma 5A.3]). *Let $f : R^n \times R^m \to R \cup \{+\infty\}$ be lower semicontinuous and finite at $(\bar{x}, \bar{y})$. If $(\zeta, 0) \in \partial^\infty f(\bar{x}, \bar{y})$ implies that $\zeta = 0$, then*

$$\partial_y f(\bar{x}, \bar{y}) \subseteq \{\eta : (\zeta, \eta) \in \partial f(\bar{x}, \bar{y}) \quad \text{for some } \zeta\},$$
$$\partial_y^\infty f(\bar{x}, \bar{y}) \subseteq \{\eta : (\zeta, \eta) \in \partial^\infty f(\bar{x}, \bar{y}) \quad \text{for some } \zeta\}.$$

PROPOSITION 2.7 (see, e.g., [13, Theorem 7.6]). *Let the minimum function be*

$$(\wedge f_j)(x) := \min\{f_j(x) | j = 1, 2, \dots, m\},$$

where $f_j : R^n \to R \cup \{+\infty\}$. Assume that f_j are lower semicontinuous around $\bar{x}$ for $j \in J(\bar{x})$ and lower semicontinuous at $\bar{x}$ for $j \notin J(\bar{x})$, where

$$J(x) := \{j | f_j(x) = \wedge f_j(x)\}.$$

Then the minimum function $\wedge f_j(x)$ is lower semicontinuous around $\bar{x}$ and

$$\partial(\wedge f_j)(\bar{x}) \subseteq \bigcup\{\partial f_j(\bar{x}) | j \in J(\bar{x})\},$$
$$\partial^\infty(\wedge f_j)(\bar{x}) \subseteq \bigcup\{\partial^\infty f_j(\bar{x}) | j \in J(\bar{x})\}.$$

Classical results on the value function can be found in [2, 4, 7, 11, 18], while the results we quote are from [7].

PROPOSITION 2.8 (see [7, (b) and (d) of Theorem 5A.2]). *Let $g : R^n \times R^m \to R \cup \{+\infty\}$ be lower semicontinuous everywhere and finite at $(\bar{z}, \bar{\alpha})$. Suppose g is bounded below on some set $E \times O$, where E is a compact neighborhood of $\bar{z}$ and O is an open set containing $\bar{\alpha}$. Define the value function $V : R^m \to R \cup \{+\infty\}$ and the set of minimizers Σ as follows:*

$$V(\alpha) := \inf\{g(z, \alpha) : z \in E\},$$
$$\Sigma(\alpha) := \{z \in E : g(z, \alpha) = V(\alpha)\}.$$

If $\Sigma(\bar{\alpha}) \subseteq \text{int}E$, then the value function V is lower semicontinuous on O, and the subdifferentials of V satisfy these estimates:

$$\partial V(\bar{\alpha}) \subseteq \{\eta \in R^m : (0, \eta) \in \partial g(z, \bar{\alpha}) \quad \text{for some } z \in \Sigma(\bar{\alpha})\},$$
$$\partial^\infty V(\bar{\alpha}) \subseteq \{\eta \in R^m : (0, \eta) \in \partial^\infty g(z, \bar{\alpha}) \quad \text{for some } z \in \Sigma(\bar{\alpha})\}.$$

Our results are stated using the limiting subdifferentials. Alternatively, they could be derived by using the Fréchet subdifferentials instead of the proximal subdifferentials. (Both lead to the same limiting subdifferentials in finite dimensional spaces.) In [18] arguments are given in favor of the former (called there the regular subdifferentials). In the present paper we use the proximal subdifferentials to provide the same framework as in [23].

3. Main results.

Let $(\bar{x}, \bar{y})$ be a feasible solution of the OPVIC and let λ be a nonnegative number. We define $M^\lambda(\bar{x}, \bar{y})$, the index λ CD multiplier set corresponding to $(\bar{x}, \bar{y})$, to be the set of vectors (γ, β, η) in $R^d \times R^l \times R^m$ satisfying the Fritz John-type necessary optimality condition involving the Mordukhovich coderivatives for GP, that is, the vectors (γ, β, η) such that

$$\begin{cases} 0 \in \lambda \partial f(\bar{x}, \bar{y}) + \partial \langle \Psi, \gamma \rangle (\bar{x}, \bar{y}) + \partial \langle H, \beta \rangle (\bar{x}, \bar{y}) + \partial \langle F, \eta \rangle (\bar{x}, \bar{y}) \\ \qquad + \{0\} \times D^* N_\Omega(\bar{y}, -F(\bar{x}, \bar{y}))(\eta) + N_C(\bar{x}, \bar{y}), \\ \gamma \geq 0, \text{ and } \langle \Psi(\bar{x}, \bar{y}), \gamma \rangle = 0. \end{cases}$$

Then by Ye [23, Theorem 3.1], the Fritz John-type necessary optimality condition involving the Mordukhovich coderivatives can be rephrased as follows.

PROPOSITION 3.1. *Under the basic assumption* (BA), *if* $(\bar{x}, \bar{y})$ *is a local solution of* OPVIC, *then either the set of normal* CD *multipliers is nonempty or there is a nonzero abnormal* CD *multiplier, i.e.,*

$$M^1(\bar{x}, \bar{y}) \cup (M^0(\bar{x}, \bar{y}) \setminus \{0\}) \neq \emptyset.$$

Note that by the definition of the Mordukhovich coderivative,

$$\xi \in D^* N_\Omega(\bar{y}, -F(\bar{x}, \bar{y}))(\eta) \text{ if and only if } (\xi, -\eta) \in N_{GphN_\Omega}(\bar{y}, -F(\bar{x}, \bar{y})).$$

In the case where $\Omega = \{0\}$, OPVIC reduces to an ordinary mathematical programming problem with equality, inequality, and abstract constraints. The term $D^* N_\Omega(\bar{y}, -F(\bar{x}, \bar{y}))(\eta)$ vanishes, and the above Fritz John condition can be considered as a limiting subdifferential version of the generalized Lagrange multiplier rule as found in Clarke [2, Theorem 6.1.1] and was obtained by Mordukhovich [9, Theorem 1(b)].

In the case where $\Omega = R^m_+$, (1) reduces to a complementarity constraint,

$$y \geq 0, \;\; F(x, y) \geq 0, \;\; \langle F(x, y), y \rangle = 0,$$

and the coderivative of the normal cone to the set R^m_+ can be calculated using the following lemma whose proof follows from [22, Proposition 2.7] and the definition of the limiting normal cones.

LEMMA 3.2. *For any* $(\bar{u}, -\bar{v}) \in GphN_{R^m_+}$,

$$\begin{aligned} N_{GphN_{R^m_+}}(\bar{u}, -\bar{v}) = \{(\xi, -\eta) \in R^{2m} : & \xi_i = 0 \textit{ if } \bar{u}_i > 0, \bar{v}_i = 0, \\ & \eta_i = 0 \textit{ if } \bar{u}_i = 0, \bar{v}_i > 0, \\ & \textit{either } \xi_i \eta_i = 0 \textit{ or } \xi_i < 0 \textit{ and } \eta_i < 0 \textit{ if } \bar{u}_i = 0, \bar{v}_i = 0\}. \end{aligned}$$

In the case where Ω is a polyhedral convex set, one can calculate the Mordukhovich coderivative of the normal cone to the set Ω by using the formula of the limiting normal cone to the graph of the normal cone to the set Ω, which was first given in the proof of Dontchev and Rockafellar [5, Theorem 2] and stated in Poliquin and Rockafellar [16, Proposition 4.4].

We first consider the following additively (right-hand side) perturbed GP:

$$\begin{array}{lll} \text{GP}(p, q, r) & \text{minimize} & f(x, y) \\ & \text{subject to} & \Psi(x, y) \leq p, H(x, y) = q, (x, y) \in C, \\ & & r \in F(x, y) + N_\Omega(y), \end{array}$$

with the solution set denoted by $\Sigma(p, q, r)$.

In order to obtain useful information on the subdifferentials of the value function at $(\bar{p}, \bar{q}, \bar{r})$, some hypotheses are usually made for GP(p, q, r), where (p, q, r) are sufficiently close to the point of interest $(\bar{p}, \bar{q}, \bar{r})$ (see, for example, [4, Growth Hypothesis 3.1.1], [2, Hypothesis 6.5.1], [18, Definition 1.8]). In this paper, we make the following growth hypothesis [7, Theorem 5A.2]:

(GH) at $(\bar{p}, \bar{q}, \bar{r})$: There exists $\delta > 0$ such that the set

$$\begin{aligned} \{(x, y) \in C \; : \; & \Psi(x, y) \leq p, H(x, y) = q, r \in F(x, y) + N_\Omega(y), f(x, y) \leq M, \\ & (p, q, r) \in B(\bar{p}, \bar{q}, \bar{r}; \delta)\} \end{aligned}$$

is bounded for each M.

In order to apply Proposition 2.8, we rewrite the value function in the following form:

$$V(p,q,r) = \inf g(x,y,p,q,r),$$

where g is the extended-value function defined by

$$g(x,y,p,q,r) := f(x,y) + I_{(Gph\Phi)\cap(C\times R^{d+l+m})}(x,y,p,q,r)$$

with I_E being the indicator function of a set E defined by

$$I_E(x) := \begin{cases} 0 \text{ if } x \in E, \\ \infty \text{ if } x \notin E \end{cases}$$

and Φ being the set-valued map defined by

$$\Phi(x,y) = (\Psi(x,y), H(x,y), F(x,y)) + R^d_+ \times \{0\} \times N_\Omega(y).$$

The growth hypothesis (GH) amounts to saying the function g is level-bounded in (x,y) uniformly for any $(p,q,r) \in B(\bar{p},\bar{q},\bar{r};\delta)$. Hence by virtue of [18, Theorem 1.9], $\bigcup_{(p,q,r)\in B(\bar{p},\bar{q},\bar{r};\delta)} \Sigma(p,q,r)$ is a compact set and for all $(p,q,r) \in B(\bar{p},\bar{q},\bar{r};\delta)$,

$$V(p,q,r) = \inf\{g(x,y,p,q,r) : (x,y) \in E\},$$

where E is a compact set with interior containing $\bigcup_{(p,q,r)\in B(\bar{p},\bar{q},\bar{r};\delta)} \Sigma(p,q,r)$. It is clear that g is lower semicontinuous everywhere and finite at any $(x,y,p,q,r) \in (Gph\Phi) \cap (C \times R^{d+l+m})$. Since f is Lipschitz on E, g is bounded below on $E \times B(\bar{x},\bar{y};\epsilon)$. The following result then follows immediately by applying Proposition 2.8.

PROPOSITION 3.3. *Under the basic assumption* (BA) *and the growth hypothesis* (GH) *at* $(\bar{p},\bar{q},\bar{r})$ *the value function* V *is lower semicontinuous on* $B(\bar{p},\bar{q},\bar{r};\delta)$ *and*

$$\partial V(\bar{p},\bar{q},\bar{r}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{p},\bar{q},\bar{r})} \{(u,v,w) : (0,0,u,v,w) \in \partial g(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r})\}, \tag{10}$$

$$\partial^\infty V(\bar{p},\bar{q},\bar{r}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{p},\bar{q},\bar{r})} \{(u,v,w) : (0,0,u,v,w) \in \partial^\infty g(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r})\}. \tag{11}$$

We now prove that the set in the right-hand side of (10) (respectively, (11)) is included in the normal multiplier set M^1 (respectively, the abnormal multiplier set M^0).

By the sum rule (see Proposition 2.5) and the fact that for any closed set E with $\bar{z} \in E$

$$\partial I_E(\bar{z}) = \partial^\infty I_E(\bar{z}) = N_E(\bar{z}),$$

we have

$$\partial g(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r}) \subset \partial f(\bar{x},\bar{y}) \times \{(0,0)\} + N_{(Gph\Phi)\cap(C\times R^{d+l+m})}(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r}),$$
$$\partial^\infty g(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r}) \subset N_{(Gph\Phi)\cap(C\times R^{d+l+m})}(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r}).$$

Hence we need only to compute the normal cone.

LEMMA 3.4. *If* $(s_x,s_y,s_p,s_q,s_r) \in N_{(Gph\Phi)\cap(C\times R^{d+l+m})}(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r})$, *then*

$$\begin{cases} (s_x,s_y) \in \partial\langle\Psi,-s_p\rangle(\bar{x},\bar{y}) + \partial\langle H,-s_q\rangle(\bar{x},\bar{y}) + \partial\langle F,-s_r\rangle(\bar{x},\bar{y}) + N_C(\bar{x},\bar{y}) \\ \qquad +\{0\} \times D^*N_\Omega(\bar{y},\bar{r}-F(\bar{x},\bar{y}))(-s_r). \\ s_p \geq 0, \text{ and } \langle\Psi(\bar{x},\bar{y})-\bar{p},s_p\rangle = 0. \end{cases}$$

Proof. Step 1. Let $(\tilde{x},\tilde{y},\tilde{p},\tilde{q},\tilde{r})$ be any point in a neighborhood of $(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r})$ on which Ψ, H, and F are Lipschitz continuous and

$$(s_x,s_y,s_p,s_q,s_r) \in N^{\pi}_{(Gph\Phi)\cap(C\times R^{d+l+m})}(\tilde{x},\tilde{y},\tilde{p},\tilde{q},\tilde{r}).$$

By definition of the proximal normal cone, there is $M>0$ such that for all $(x,y,p,q,r) \in (Gph\Phi)\cap(C\times R^{d+l+m})$,

$$\langle (s_x,s_y,s_p,s_q,s_r),(x,y,p,q,r)-(\tilde{x},\tilde{y},\tilde{p},\tilde{q},\tilde{r})\rangle \le M\|(x,y,p,q,r)-(\tilde{x},\tilde{y},\tilde{p},\tilde{q},\tilde{r})\|^2.$$

In other words, $(\tilde{x},\tilde{y},\tilde{p},\tilde{q},\tilde{r})$ is a solution to the optimization problem

$$\begin{array}{ll} \text{minimize} & \langle -(s_x,s_y,s_p,s_q,s_r),(x,y,p,q,r)\rangle + M\|(x,y,p,q,r)-(\tilde{x},\tilde{y},\tilde{p},\tilde{q},\tilde{r})\|^2 \\ \text{subject to} & \Psi(x,y)\le p,\ H(x,y)=q,\ (x,y)\in C, \\ & r\in F(x,y)+N_\Omega(y). \end{array}$$

We now prove that the only abnormal CD multiplier for the above problem is the zero vector. Indeed, the set of abnormal CD multipliers at $(\tilde{x},\tilde{y},\tilde{p},\tilde{q},\tilde{r})$ for the above problem are the vectors (γ,β,η) satisfying

$$\begin{cases} 0\in \partial\langle\Psi,\gamma\rangle(\tilde{x},\tilde{y})\times\{(-\gamma,0,0)\}+\partial\langle H,\beta\rangle(\tilde{x},\tilde{y})\times\{(0,-\beta,0)\}+\partial\langle F,\eta\rangle(\tilde{x},\tilde{y}) \\ \times\{(0,0,-\eta)\}+\{0\}\times D^*N_\Omega(\tilde{y},\tilde{r}-F(\tilde{x},\tilde{y}))(\eta)\times\{(0,0,0)\}+N_C(\tilde{x},\tilde{y}) \\ \times\{(0,0,0)\},\ \ \gamma\ge 0,\ \text{ and } \langle\Psi(\tilde{x},\tilde{y})-\tilde{p},\gamma\rangle=0, \end{cases}$$

which obviously coincides with the set $\{(0,0,0)\}$. Applying Proposition 3.1, we conclude that the set of normal CD multipliers for the above problem must be nonempty. That is, there are vectors $\eta\in R^m, \beta\in R^l$, and $\gamma\in R^d$ such that

$$\begin{cases} 0\in -\{(s_x,s_y,s_p,s_q,s_r)\}+\partial\langle\Psi,\gamma\rangle(\tilde{x},\tilde{y})\times\{(-\gamma,0,0)\}+\partial\langle H,\beta\rangle(\tilde{x},\tilde{y})\times\{(0,-\beta,0)\} \\ \qquad +\partial\langle F,\eta\rangle(\tilde{x},\tilde{y})\times\{(0,0,-\eta)\}+\{0\}\times D^*N_\Omega(\tilde{y},\tilde{r}-F(\tilde{x},\tilde{y}))(\eta)\times\{(0,0,0)\} \\ \qquad\qquad +N_C(\tilde{x},\tilde{y})\times\{(0,0)\}, \\ \gamma\ge 0,\ \text{ and } \langle\Psi(\tilde{x},\tilde{y})-\tilde{p},\gamma\rangle=0. \end{cases}$$

That is,

$$\begin{cases} (s_x,s_y)\in\partial\langle\Psi,-s_p\rangle(\tilde{x},\tilde{y})+\partial\langle H,-s_q\rangle(\tilde{x},\tilde{y}) \\ \qquad +\partial\langle F,-s_r\rangle(\tilde{x},\tilde{y})+\{0\}\times D^*N_\Omega(\tilde{y},\tilde{r}-F(\tilde{x},\tilde{y}))(-s_q)+N_C(\tilde{x},\tilde{y}), \\ s_p\ge 0,\ \text{ and } \langle\Psi(\tilde{x},\tilde{y})-\tilde{p},s_p\rangle=0. \end{cases}$$

Step 2. Now take any $(s_x,s_y,s_p,s_q,s_r)\in N_{(Gph\Phi)\cap(C\times R^{d+l+m})}(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r})$. Then by definition of limiting normal cones, there are sequences $(x^\nu,y^\nu,p^\nu,q^\nu,r^\nu)\to(\bar{x},\bar{y},\bar{p},\bar{q},\bar{r})$ and $(s_x^\nu,s_y^\nu,s_p^\nu,s_q^\nu,s_r^\nu)\to(s_x,s_y,s_p,s_q,s_r)$ with

$$(s_x^\nu,s_y^\nu,s_p^\nu,s_q^\nu,s_r^\nu)\in N^{\pi}_{(Gph\Phi)\cap(C\times R^{d+l+m})}(x^\nu,y^\nu,p^\nu,q^\nu,r^\nu).$$

By virtue of step 1,

$$\begin{cases} (s_x^\nu,s_y^\nu)\in\partial\langle\Psi,-s_p^\nu\rangle(x^\nu,y^\nu)+\partial\langle H,-s_q^\nu\rangle(x^\nu,y^\nu)+\partial\langle F,-s_r^\nu\rangle(x^\nu,y^\nu) \\ \qquad +\{0\}\times D^*N_\Omega(y^\nu,r^\nu-F(x^\nu,y^\nu))(-s_r^\nu)+N_C(x^\nu,y^\nu), \\ s_p^\nu\ge 0,\ \text{ and } \langle\Psi(x^\nu,y^\nu)-p^\nu,s_p^\nu\rangle=0. \end{cases}$$

Since Ψ is Lipschitz near $(\bar{x}, \bar{y})$, we have

$$\begin{aligned}\partial\langle\Psi, -s_p^\nu\rangle(x^\nu, y^\nu) &\subseteq \partial\langle\Psi, -s_p\rangle(x^\nu, y^\nu) + \partial\langle\Psi, s_p - s_p^\nu\rangle(x^\nu, y^\nu) \text{ by Proposition 2.5} \\ &\subseteq \partial\langle\Psi, -s_p\rangle(x^\nu, y^\nu) + \|s_p^\nu - s_p\| L_\Psi clB \text{ by Proposition 2.4,}\end{aligned}$$

where L_Ψ is the Lipschitz constant of Ψ. Similarly,

$$\begin{aligned}\partial\langle H, -s_q^\nu\rangle(x^\nu, y^\nu) &\subseteq \partial\langle H, -s_q\rangle(x^\nu, y^\nu) + \|s_q^\nu - s_q\| L_H clB, \\ \partial\langle F, -s_r^\nu\rangle(x^\nu, y^\nu) &\subseteq \partial\langle F, -s_r\rangle(x^\nu, y^\nu) + \|s_r^\nu - s_r\| L_F clB,\end{aligned}$$

where L_H, L_F are the Lipschitz constants of F and H. Hence we have

$$\begin{cases}(s_x^\nu, s_y^\nu) \in \partial\langle\Psi, -s_p\rangle(x^\nu, y^\nu) + \partial\langle H, -s_q\rangle(x^\nu, y^\nu) + \partial\langle F, -s_r\rangle(x^\nu, y^\nu) \\ \qquad + (\|s_p^\nu - s_p\| + \|s_q^\nu - s_q\| + \|s_r^\nu - s_r\|)(L_\Psi + L_H + L_F) clB \\ \qquad + \{0\} \times D^* N_\Omega(y^\nu, r^\nu - F(x^\nu, y^\nu))(-s_r^\nu) + N_C(x^\nu, y^\nu), \\ s_p^\nu \geq 0, \text{ and } \langle\Psi(x^\nu, y^\nu) - p^\nu, s_p^\nu\rangle = 0.\end{cases}$$

Taking limits as $\nu \to \infty$ and using the definitions of the limiting normal cone and the limiting subdifferentials completes the proof. □

Remark. As is pointed out by referee 1, alternatively, Lemma 3.4 can also be proved by formulating the constraints in the form of [12, equation (6.19)] and applying [12, Theorem 6.10].

All in all, we proved the following result.

THEOREM 3.5. *Assume* (GH) *and* (BA) *hold. Then the value function V is lower semicontinuous on $B(\bar{p}, \bar{q}, \bar{r}; \delta)$ and*

$$\partial V(\bar{p}, \bar{q}, \bar{r}) \subset \bigcup_{(\bar{x}, \bar{y}) \in \Sigma(\bar{p}, \bar{q}, \bar{r})} -M^1(\bar{x}, \bar{y}) \text{ and } \partial^\infty V(\bar{p}, \bar{q}, \bar{r}) \subset \bigcup_{(\bar{x}, \bar{y}) \in \Sigma(\bar{p}, \bar{q}, \bar{r})} -M^0(\bar{x}, \bar{y}).$$

We now consider the value function $V(\alpha)$ associated with the following perturbed GP:

$$\begin{array}{lll}\text{GP}(\alpha) & \text{minimize} & f(x, y, \alpha) \\ & \text{subject to} & \Psi(x, y, \alpha) \leq 0, H(x, y, \alpha) = 0, (x, y) \in C, \\ & & 0 \in F(x, y, \alpha) + N_\Omega(y),\end{array}$$

i.e.,

$$\begin{aligned}V(\alpha) := \inf\{f(x, y, \alpha) :\ & \Psi(x, y, \alpha) \leq 0, H(x, y, \alpha) = 0, (x, y) \in C, \\ & 0 \in F(x, y, \alpha) + N_\Omega(y)\},\end{aligned}$$

where the following basic assumptions are satisfied:

(BH) The functions $f : R^{n+m+c} \to R, \Psi : R^{n+m+c} \to R^d, H : R^{n+m+c} \to R^l$, and $F : R^{n+m+c} \to R^m$ are locally Lipschitz near any points in $C \times R^c$; C is a closed subset of R^{n+m}; and Ω is a closed convex subset of R^m.

It is easy to see that we can turn the nonadditive perturbations into additive perturbations by adding an auxiliary variable:

$$\begin{array}{lll}\text{GP}(\alpha) & \text{minimize} & f(x, y, z) \\ & \text{subject to} & \Psi(x, y, z) \leq 0, H(x, y, z) = 0, (x, y, z) \in C \times R^c, \\ & & 0 \in F(x, y, z) + N_\Omega(y), \\ & & z = \alpha,\end{array}$$

which is the partially perturbed problem of the fully perturbed problem

$$\begin{array}{lll} \mathrm{GP}(p,q,r,\alpha) & \text{minimize} & f(x,y,z) \\ & \text{subject to} & \Psi(x,y,z) \le p, H(x,y,z) = q, (x,y,z) \in C \times R^c, \\ & & r \in F(x,y,z) + N_\Omega(y), \\ & & z = \alpha. \end{array}$$

By Theorem 3.5, if the fully perturbed problem $\mathrm{GP}(p,q,r,\alpha)$ satisfies the growth hypothesis (GH) at $(0,0,0,\bar{\alpha})$, then the value function $\widetilde{V}(p,q,r,\alpha)$ defined by

$$\widetilde{V}(p,q,r,\alpha) := \inf\{f(x,y,z) : \Psi(x,y,z) \le p, H(x,y,z) = q, (x,y,z) \in C \times R^c, \\ r \in F(x,y,z) + N_\Omega(y), z = \alpha\}$$

is lower semicontinuous on $B(0,0,0,\bar{\alpha};\delta)$ and

$$\partial \widetilde{V}(0,0,0,\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y},\bar{\alpha})\in\Sigma(0,0,0,\bar{\alpha})} -M^1(\bar{x},\bar{y},\bar{\alpha}),$$

$$\partial^\infty \widetilde{V}(0,0,0,\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y},\bar{\alpha})\in\Sigma(0,0,0,\bar{\alpha})} -M^0(\bar{x},\bar{y},\bar{\alpha}).$$

For any $(0,0,0,\zeta) \in \partial^\infty \widetilde{V}(0,0,0,\bar{\alpha})$, we have $(0,0,0,\zeta) \in -M^0(\bar{x},\bar{y},\bar{\alpha})$ for some point $(\bar{x},\bar{y},\bar{\alpha}) \in \Sigma(0,0,0,\bar{\alpha})$. Therefore,

$$(0,0,\zeta) \in N_C(\bar{x},\bar{y}) \times \{0\},$$

which implies that $\zeta = 0$. By Proposition 2.6, we have

$$\partial_\alpha \widetilde{V}(0,0,0,\bar{\alpha}) \subseteq \{-\zeta : -(\gamma,\beta,\eta,\zeta) \in \partial \widetilde{V}(0,0,0,\bar{\alpha}) \text{ for some } (\gamma,\beta,\eta)\},$$

$$\partial_\alpha^\infty \widetilde{V}(0,0,0,\bar{\alpha}) \subseteq \{-\zeta : -(\gamma,\beta,\eta,\zeta) \in \partial^\infty \widetilde{V}(0,0,0,\bar{\alpha}) \text{ for some } (\gamma,\beta,\eta)\}.$$

Moreover, since all functions involved are continuous, it suffices to fix α at $\bar{\alpha}$ in the growth hypothesis (GH) at $(0,0,0,\bar{\alpha})$ for the fully perturbed problem $\mathrm{GP}(p,q,r,\alpha)$. Consequently, noticing that $V(\alpha) = \widetilde{V}(0,0,0,\alpha)$, we have proved the following theorem.

THEOREM 3.6. *In addition to the basic assumption* (BH), *assume that there exists $\delta > 0$ such that the set*

$$\{(x,y) \in C \ : \ \Psi(x,y,\bar{\alpha}) \le p, H(x,y,\bar{\alpha}) = q, r \in F(x,y,\bar{\alpha}) + N_\Omega(y), f(x,y,\bar{\alpha}) \le M, \\ (p,q,r) \in B(0;\delta)\}$$

is bounded for each M. Then the value function $V(\alpha)$ is lower semicontinuous near $\bar{\alpha}$ and

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{-\zeta : (\gamma,\beta,\eta,\zeta) \in M^1(\bar{x},\bar{y},\bar{\alpha})\},$$

$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{-\zeta : (\gamma,\beta,\eta,\zeta) \in M^0(\bar{x},\bar{y},\bar{\alpha})\},$$

where $M^\lambda(\bar{x},\bar{y},\bar{\alpha})$ is the set of index λ multipliers for problem GP*(p,q,r,α) at* $(0,0,0,\bar{\alpha})$, *i.e., vectors $(\gamma,\beta,\eta,\zeta)$ in $R^d \times R^l \times R^m \times R$ satisfying*

$$\begin{cases} 0 \in \lambda\partial f(\bar{x},\bar{y},\bar{\alpha}) + \partial\langle\Psi,\gamma\rangle(\bar{x},\bar{y},\bar{\alpha}) + \partial\langle H,\beta\rangle(\bar{x},\bar{y},\bar{\alpha}) + \partial\langle F,\eta\rangle(\bar{x},\bar{y},\bar{\alpha}) \\ \qquad +\{0\} \times D^*N_\Omega(\bar{y},-F(\bar{x},\bar{y},\bar{\alpha}))(\eta) \times \{0\} + \{(0,0,\zeta)\} + N_C(\bar{x},\bar{y}) \times \{0\}, \\ \gamma \ge 0, \text{ and } \langle\Psi(\bar{x},\bar{y},\bar{\alpha}),\gamma\rangle = 0, \end{cases}$$

and $\Sigma(\bar{\alpha})$ is the set of solutions of problem GP($\bar{\alpha}$).

The above estimates may not be useful in the case where $\partial V(\bar{\alpha})$ is empty. The following consequence of Theorem 3.6 and Proposition 2.4 provides conditions which rule out this possibility.

COROLLARY 3.7. *Under the assumption of Theorem 3.6, if the set of ζ components of the abnormal CD multiplier set contains only the zero vector, i.e.,*

$$\bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{-\zeta : (\gamma,\beta,\eta,\zeta) \in M^0(\bar{x},\bar{y},\bar{\alpha})\} = \{0\},$$

then $V(\bar{\alpha})$ is finite and Lipschitz near $\bar{\alpha}$ with

$$\emptyset \neq \partial V(\bar{\alpha}) \subset \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{-\zeta : (\gamma,\beta,\eta,\zeta) \in M^1(\bar{x},\bar{y},\bar{\alpha})\}.$$

In addition to the above assumptions, if the ζ components of the normal CD multiplier set are unique, i.e.,

$$\bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{-\zeta : (\gamma,\beta,\eta,\zeta) \in M^1(\bar{x},\bar{y},\bar{\alpha})\} = \{-\zeta\},$$

then V is strictly differentiable at $\bar{\alpha}$ and $\nabla V(\bar{\alpha}) = -\zeta$.

In the case where all functions are smooth, the estimates have the following simple expression.

COROLLARY 3.8. *In addition to the assumptions in Theorem 3.6, assume that f,Ψ,H,F are C^1 at each $(\bar{x},\bar{y},\bar{\alpha})$, where $(\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})$; then the value function V is lower semicontinuous near $\bar{\alpha}$, and*

$$\begin{aligned}
\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{&\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta \\
&+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : \qquad (\gamma,\beta,\eta) \in M^1(\bar{x},\bar{y})\},\\
\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y},\bar{z})\in\Sigma(\bar{\alpha})} \{&\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta \\
&+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : \qquad (\gamma,\beta,\eta) \in M^0(\bar{x},\bar{y})\},
\end{aligned}$$

where $M^\lambda(\bar{x},\bar{y})$ is the set of index λ CD multipliers for problem GP($\bar{\alpha}$).

Note that in the case where there are no variational inequality constraints, the CD multipliers are the ordinary NLP multipliers, and the above results recover the well-known results in the sensitivity analysis of NLP.

4. Applications to OPCCs. In this section, we apply our main results to the following perturbed OPCC:

$$\text{(OPCC)}(\alpha) \qquad \begin{array}{ll} \text{minimize} & f(x,y,\alpha), \\ \text{subject to} & \Psi(x,y,\alpha) \le 0, H(x,y,\alpha) = 0, (x,y) \in C, \\ & y \ge 0,\ F(x,y,\alpha) \ge 0, \\ & \langle y, F(x,y,\alpha)\rangle = 0, \end{array}$$

which is GP(α) with $\Omega = R^m_+$.

For easier exposition, we assume in this section that all problem data f,Ψ,H,F are C^1. We denote by $\nabla f(x,y,\alpha)$ the gradient of function f with respect to (x,y).

For $(\bar{x}, \bar{y})$, a feasible solution of (OPCC)($\bar{\alpha}$), we define the index sets

$$L := L(\bar{x},\bar{y}) := \{1 \le i \le m : \bar{y}_i > 0, F_i(\bar{x},\bar{y},\bar{\alpha}) = 0\},$$
$$I_+ := I_+(\bar{x},\bar{y}) := \{1 \le i \le m : \bar{y}_i = 0, F_i(\bar{x},\bar{y},\bar{\alpha}) > 0\},$$
$$I_0 := I_0(\bar{x},\bar{y}) := \{1 \le i \le m : \bar{y}_i = 0, F_i(\bar{x},\bar{y},\bar{\alpha}) = 0\}.$$

4.1. Sensitivity analysis of the value function via NLP multipliers. Let $(\bar{x}, \bar{y})$ be a local optimal solution for (OPCC)($\bar{\alpha}$). Treating (OPCC)($\bar{\alpha}$) as an ordinary NLP problem with inequality constraints

$$\Psi(x,y,\bar{\alpha}) \le 0, y \ge 0, F(x,y,\bar{\alpha}) \ge 0,$$

equality constraints

$$H(x,y,\bar{\alpha}) = 0, \langle y, F(x,y,\bar{\alpha}\rangle = 0,$$

and the abstract constraint $(x,y) \in C$, it is easy to see that the Fritz John optimality condition implies the existence of $\lambda \ge 0, \gamma \in R^d, \beta \in R^l, r^F \in R^m, r^y \in R^m, \mu \in R$, not all zero, such that

$$\begin{aligned}
&0 \in \lambda \nabla f(\bar{x},\bar{y},\bar{\alpha}) + \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top \beta \\
&\qquad - \nabla F(\bar{x},\bar{y},\bar{\alpha})^\top r^F - \{(0, r^y)\} + \mu \nabla \langle y, F\rangle(\bar{x},\bar{y},\bar{\alpha}) + N_C(\bar{x},\bar{y}),\\
&\gamma \ge 0, \langle \gamma, \Psi(\bar{x},\bar{y},\bar{\alpha})\rangle = 0,\\
&r^F \ge 0, r^y \ge 0, \langle r^F, F(\bar{x},\bar{y},\bar{\alpha})\rangle = 0, \langle r^y, \bar{y}\rangle = 0.
\end{aligned}$$

Using the sum and product rules, we have

$$\nabla\langle y, F\rangle(\bar{x},\bar{y},\bar{\alpha}) = \{(0, F(\bar{x},\bar{y},\bar{\alpha}))\} + \nabla F(\bar{x},\bar{y},\bar{\alpha})^\top \bar{y}.$$

Therefore, the Fritz John necessary condition becomes

$$\begin{aligned}
&0 \in \lambda \nabla f(\bar{x},\bar{y},\bar{\alpha}) + \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top \beta \\
&\qquad + \nabla F(\bar{x},\bar{y},\bar{\alpha})^\top (\mu\bar{y} - r^F) + \{(0, \mu F(\bar{x},\bar{y},\bar{\alpha}) - r^y)\} + N_C(\bar{x},\bar{y}),\\
&\gamma \ge 0, \langle \gamma, \Psi(\bar{x},\bar{y},\bar{\alpha})\rangle = 0,\\
&r^F \ge 0, r^y \ge 0, \text{ and } \langle r^F, F(\bar{x},\bar{y},\bar{\alpha})\rangle = 0, \langle r^y, \bar{y}\rangle = 0.
\end{aligned}$$

DEFINITION 4.1 (NLP multipliers). *We call all vectors $(\gamma, \beta, r^F, r^y, \mu) \in R^d \times R^l \times R^m \times R^m \times R$ satisfying the above Fritz John necessary condition for any $\lambda \ge 0$ the index λ NLP multipliers for OPCC($\bar{\alpha}$) and denote the set by $M^\lambda_{NLP}(\bar{x},\bar{y})$.*

Since we treat OPCC(α) as an ordinary NLP problem, $\Omega = \{0\}$ in the corresponding problem GP(α). Hence the CD multipliers for the corresponding $GP(\alpha)$ are the NLP multipliers defined above. Applying Corollary 3.8 and Proposition 2.4, we derive the following upper estimates of the limiting subdifferentials of the value function in terms of the NLP multipliers.

THEOREM 4.2. *Assume that there exists $\delta > 0$ such that the set*

$$\begin{aligned}
\{(x,y) \in C :&(p,q,p_y,p_F,q_\mu) \in B(0;\delta), \Psi(x,y,\bar{\alpha}) \le p, H(x,y,\bar{\alpha}) = q,\\
&y \ge p_y,\ F(x,y,\bar{\alpha}) \le p_F, \langle y, F(x,y,\bar{\alpha})\rangle = q_\mu, f(x,y,\bar{\alpha}) \le M\}
\end{aligned}$$

is bounded for each M. Then the value function V is lower semicontinuous near $\bar{\alpha}$ and

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta$$
$$+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top (\mu\bar{y} - r^F) : (\gamma,\beta,r^F,r^y,\mu) \in M^1_{NLP}(\bar{x},\bar{y})\}, \tag{12}$$

$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta$$
$$+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top (\mu\bar{y} - r^F) : (\gamma,\beta,r^F,r^y,\mu) \in M^0_{NLP}(\bar{x},\bar{y})\}. \tag{13}$$

If the set in the right-hand side of inclusion (13) *contains only the zero vector, then the value function V is Lipschitz near $\bar{\alpha}$. If the set in the right-hand side of inclusion* (13) *contains only the zero vector and the set in the right-hand side of inclusion* (12) *is a singleton, then the value function is strictly differentiable at $\bar{\alpha}$.*

4.2. Sensitivity analysis of the value function via CD multipliers. Since OPCC($\bar{\alpha}$) is OPVIC($\bar{\alpha}$) with $\Omega = R^m_+$, the following expression of CD multiplers follows immediately from Lemma 3.2.

PROPOSITION 4.3. *For* OPCC($\bar{\alpha}$)*, an index λ* CD *multiplier corresponding to a feasible solution $(\bar{x},\bar{y})$ is a vector $(\gamma,\beta,\eta) \in R^d \times R^l \times R^m$ such that*

$$0 \in \lambda\nabla f(\bar{x},\bar{y},\bar{\alpha}) + \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top\beta$$
$$+\nabla F(\bar{x},\bar{y},\bar{\alpha})^\top\eta + (0,0,\xi) + N_C(\bar{x},\bar{y}), \tag{14}$$
$$\gamma \geq 0 \text{ and } \langle \Psi(\bar{x},\bar{y},\bar{\alpha}),\gamma\rangle = 0, \tag{15}$$
$$\xi_i = 0 \qquad \text{if } \bar{y}_i > 0 \text{ and } F_i(\bar{x},\bar{y},\bar{\alpha}) = 0, \tag{16}$$
$$\eta_i = 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x},\bar{y},\bar{\alpha}) > 0, \tag{17}$$
$$\text{either } \xi_i < 0,\ \eta_i < 0,\ \text{or } \xi_i\eta_i = 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x},\bar{y}) = 0. \tag{18}$$

Corollary 3.8 and Proposition 2.4 now lead to the following result.

THEOREM 4.4. *Assume that there exists $\delta > 0$ such that the set*

$$\{(x,y) \in C : (p,q,r) \in B(0;\delta), \Psi(x,y,\bar{\alpha}) \leq p, H(x,y,\bar{\alpha}) = q,$$
$$y \geq 0,\ F(x,y,\bar{\alpha}) \geq r, \langle y, F(x,y,\bar{\alpha}) - r\rangle = 0, f(x,y,\bar{\alpha}) \leq M\}$$

is bounded for each M. Then the value function V is lower semicontinuous near $\bar{\alpha}$ and

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta$$
$$+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in M^1_{CD}(\bar{x},\bar{y})\}, \tag{19}$$

$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top \beta$$
$$+\nabla F(\bar{x},\bar{y},\bar{\alpha})^\top \eta : (\gamma,\beta,\eta) \in M^0_{CD}(\bar{x},\bar{y})\}. \tag{20}$$

If the set in the right-hand side of inclusion (20) *contains only the zero vector, then the value function V is Lipschitz near $\bar{\alpha}$. If the set in the right-hand side of inclusion* (20) *contains only the zero vector and the set in the right-hand side of inclusion* (19) *is a singleton, then the value function is strictly differentiable at $\bar{\alpha}$.*

We say that the generalized Mangasarian–Fromovitz constraint qualification for OPCC($\bar{\alpha}$) is satisfied at $(\bar{x},\bar{y})$ if $C = D \times R^m$ and

(i) for every partition of I_0 into sets P, Q, R with $R \neq \emptyset$, there exist vectors $k \in \text{int}T_C(\bar{x}, D), h \in R^m$ such that $h_{I_+} = 0, h_Q = 0, h_R \geq 0$,

$$\begin{aligned}
&\nabla_x \Psi_{I(\Psi)}(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y \Psi_{I(\Psi)}(\bar{x},\bar{y},\bar{\alpha})h \leq 0,\\
&\nabla_x H(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y H(\bar{x},\bar{y},\bar{\alpha})h = 0,\\
&\nabla_x F_{L\cup P}(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y F_{L\cup P}(\bar{x},\bar{y},\bar{\alpha})h = 0,\\
&\nabla_x F_R(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y F_R(\bar{x},\bar{y},\bar{\alpha})h \geq 0,
\end{aligned}$$

and either $h_i > 0$ or

$$\nabla_x F_i(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y F_i(\bar{x},\bar{y},\bar{\alpha})h > 0 \text{ for some } i \in R;$$

(ii) for every partition of I_0 into the sets P, Q, the matrix

$$\begin{bmatrix} \nabla_x H(\bar{x},\bar{y},\bar{\alpha}) & \nabla_y H_{A,L\cup P}(\bar{x},\bar{y},\bar{\alpha}) \\ \nabla_x F_{L\cup P}(\bar{x},\bar{y},\bar{\alpha}) & \nabla_y F_{L\cup P,L\cup P}(\bar{x},\bar{y},\bar{\alpha}) \end{bmatrix}$$

has full row rank and there exist vectors $k \in \text{int}T_C(\bar{x}, D), h \in R^m$ such that

$$\begin{aligned}
&h_{I_+} = 0, h_Q = 0,\\
&\nabla_x \Psi_{I(\Psi)}(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y \Psi_{I(\Psi)}(\bar{x},\bar{y},\bar{\alpha})h < 0,\\
&\nabla_x H(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y H(\bar{x},\bar{y},\bar{\alpha})h = 0,\\
&\nabla_x F_{L\cup P}(\bar{x},\bar{y},\bar{\alpha})k + \nabla_y F_{L\cup P}(\bar{x},\bar{y},\bar{\alpha})h = 0,
\end{aligned}$$

where $A := \{1, \ldots, l\}$, $T_C(\bar{x}, D)$ denotes the Clarke tangent cone of D at $\bar{x}$, and $I(\Psi) := \{i : \Psi_i(\bar{x}, \bar{y}) = 0\}$ is the index set of the binding inequality constraints.

In [23, Proposition 4.5] it was proved that the generalized Mangasarian–Fromovitz constraint qualification implies that the only abnormal CD multiplier is the zero vector. Hence Theorem 4.4 has the following consequence.

COROLLARY 4.5. *In addition to the assumptions of Theorem* 4.4, *if the generalized Mangasarian–Fromovitz constraint qualification as defined above is satisfied for* OPCC$(\bar{\alpha})$, *then* $V(\alpha)$ *is finite and Lipschitz near* $\bar{\alpha}$.

Another sufficient condition for $M^0_{CD}(\Sigma(\bar{\alpha})) = \{0\}$ is the strong regularity condition in the sense of Robinson [17]. For OPCC $(\bar{\alpha})$, the strong regularity condition has the following form according to [17, Theorem 3.1].

COROLLARY 4.6. *In addition to the assumptions of Theorem* 4.4, *assume that* $C = D \times R^m$ *for some* $D \subseteq R^n$, *that there are no inequality constraints, and that the following conditions are satisfied:*

(i) *the matrix*

$$\begin{bmatrix} \nabla_y H_{A,L}(\bar{x},\bar{y},\bar{\alpha}) \\ \nabla_y F_{L,L}(\bar{x},\bar{y},\bar{\alpha}) \end{bmatrix}$$

is nonsingular, where $A := \{1, \ldots, l\}$;

(ii) *the Schur complement of the above matrix in the matrix*

$$\begin{bmatrix} \nabla_y H_{A,L}(\bar{x},\bar{y},\bar{\alpha}) & \nabla_y H_{A,I_0}(\bar{x},\bar{y},\bar{\alpha}) \\ \nabla_y F_{L,L}(\bar{x},\bar{y},\bar{\alpha}) & \nabla_y F_{L,I_0}(\bar{x},\bar{y},\bar{\alpha}) \\ \nabla_y F_{I_0,L}(\bar{x},\bar{y},\bar{\alpha}) & \nabla_y F_{I_0,I_0}(\bar{x},\bar{y},\bar{\alpha}) \end{bmatrix}$$

has positive principle minors;

then $V(\alpha)$ *is finite and Lipschitz near* $\bar{\alpha}$.

4.3. Sensitivity analysis of the value function via C multipliers. It is easy to see that OPCC ($\bar{\alpha}$) can be formulated as the following optimization problem with a nonsmooth equation:

$$\text{(21)}\quad \begin{array}{lll} \text{OPCC}(\bar{\alpha}) & \text{minimize} & f(x,y,\alpha) \\ & \text{subject to} & \Psi(x,y,\alpha) \le 0, H(x,y,\alpha)=0, (x,y)\in C, \\ & & \min\{y_i, F_i\}(x,y,\alpha)=0, \qquad i=1,2,\dots,m. \end{array}$$

It can be shown as in Scheel and Scholtes [19, Lemma 1] that a solution of the OPCC is C stationary defined as follows.

DEFINITION 4.7 (C multipliers). *Let $(\bar{x},\bar{y})$ be a feasible point of the* OPCC. *The point $(\bar{x},\bar{y})$ is C stationary if there exist vectors $(\gamma,\beta,\eta,\xi)\in R^d\times R^l\times R^m\times R^m$ satisfying* (14)–(17) *and*

$$\xi_i\eta_i \ge 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x},\bar{y},\bar{\alpha})=0.$$

The set of vectors (γ,β,η) satisfying the above condition for some ξ is called the index λ C multiplier set and is denoted by $M_C^\lambda(\bar{x},\bar{y})$.

THEOREM 4.8. *Assume that there exists $\delta>0$ such that the set*

$$\{(x,y)\in C : (p,q,q^m)\in B(0;\delta), \Psi(x,y,\bar{\alpha})\le p, H(x,y,\bar{\alpha})=q, \\ \min\{y_i, F_i(x,y,\bar{\alpha})\} = q_i^m, i=1,\dots,m, f(x,y,\bar{\alpha})\le M\}$$

is bounded for each M. Then the value function V is lower semicontinuous near $\bar{\alpha}$ and

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta \\ +\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta,\eta)\in M_C^1(\bar{x},\bar{y})\}, \tag{22}$$

$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta \\ +F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta,\eta)\in M_C^0(\bar{x},\bar{y})\}. \tag{23}$$

If the set in the right-hand side of inclusion (23) *contains only the zero vector, then the value function V is Lipschitz near $\bar{\alpha}$. If the set in the right-hand side of inclusion* (23) *contains only the zero vector and the set in the right-hand side of inclusion* (22) *is a singleton, then the value function is strictly differentiable at $\bar{\alpha}$.*

Proof. By Theorem 3.6, since the growth assumption is satisfied, the value function is lower semicontinuous near $\bar{\alpha}$ and

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{-\zeta : (\gamma,\eta,\zeta)\in M^1(\bar{x},\bar{y},\bar{\alpha})\},$$
$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{-\zeta : (\gamma,\eta,\zeta)\in M^0(\bar{x},\bar{y},\bar{\alpha})\},$$

where $M^\lambda(\bar{x},\bar{y},\bar{\alpha})$ is the set of vectors $(\gamma,\beta,r,\zeta)\in R^{d+l+m+c}$ such that

$$0\in\lambda\nabla f(\bar{x},\bar{y},\bar{\alpha}) + \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top\beta \\ +\partial\sum_{i=1}^m r_i\min\{y_i,F_i\}(\bar{x},\bar{y},\bar{\alpha}) + \{(0,0,\zeta)\} + N_C(\bar{x},\bar{y})\times\{0\}, \\ \gamma\ge 0, \langle\gamma,\Psi\rangle(\bar{x},\bar{y},\bar{\alpha})=0.$$

Note that, in the above, ∇f denotes the gradient of a function f with respect to (x, y, α). Since

$$\partial \sum_{i=1}^{m} r_i \min\{y_i, F_i\}(\bar{x}, \bar{y}, \bar{\alpha}) \subseteq \sum_{i=1}^{m} r_i \partial_C \min\{y_i, F_i\}(\bar{x}, \bar{y}, \bar{\alpha})$$

and

$$\partial_C \min\{y_i, F_i\}(\bar{x}, \bar{y}, \bar{\alpha}) = \begin{cases} (0, e_i, 0) & \forall i \in I_+, \\ \nabla F_i(\bar{x}, \bar{y}, \bar{\alpha}) & \forall i \in L, \\ \{t(0, , e_i, 0) + (1-t)\nabla F_i(\bar{x}, \bar{y}, \bar{\alpha}) : t \in [0,1]\} & \forall i \in I_0, \end{cases}$$

where e_i is the unit vector whose ith component is 1 and those other components are zero, there exist γ, β, η such that

$$\zeta = \lambda \nabla_\alpha f(\bar{x}, \bar{y}, \bar{\alpha}) + \nabla_\alpha \Psi(\bar{x}, \bar{y}, \bar{\alpha})^\top \gamma + \nabla_\alpha H(\bar{x}, \bar{y}, \bar{\alpha})^\top \beta + \nabla_\alpha F(\bar{x}, \bar{y}, \bar{\alpha})^\top \eta$$

and

$$\begin{aligned} 0 \in\ & \lambda \nabla f(\bar{x}, \bar{y}, \bar{\alpha}) + \nabla \Psi(\bar{x}, \bar{y}, \bar{\alpha})^\top \gamma + \nabla H(\bar{x}, \bar{y}, \bar{\alpha})^\top \beta + \nabla F(\bar{x}, \bar{y}, \bar{\alpha})^\top \eta \\ & + (0, \xi) + N_C(\bar{x}, \bar{y}), \\ \gamma \geq\ & 0, \langle \Psi, \gamma \rangle(\bar{x}, \bar{y}, \bar{\alpha}) = 0, \end{aligned}$$

where

$$\begin{array}{ll} \eta_i = 0 & \forall i \in I_+, \\ \xi_i = 0 & \forall i \in L, \\ \eta_i = r_i(1 - \bar{t}_i), \xi_i = r_i \bar{t}_i \text{ for some } \bar{t}_i \in [0,1], & \forall i \in I_0. \end{array}$$

It is then easy to see that

$$\forall i \in I_0, \eta_i \xi_i \geq 0.$$

Hence (γ, β, η) is a C multiplier, and the proof of the theorem is complete. □

4.4. Sensitivity analysis via P multipliers and S multipliers. Taking the "piecewise programming" approach, for any given index set $\nu \subseteq I := \{1, \ldots, m\}$, we consider the subproblem associated with ν:

$$\begin{array}{lll} \text{OPCC}(\alpha)_\nu & \text{minimize} & f(x, y, \alpha) \\ & \text{subject to} & \Psi(x, y, \alpha) \leq 0, H(x, y, \alpha) = 0, (x, y) \in C, \\ & & y_i \geq 0,\ F_i(x, y, \alpha) = 0 \quad \forall i \in \nu \\ & & y_i = 0,\ F_i(x, y, \alpha) \geq 0 \quad \forall i \in I \backslash \nu. \end{array}$$

As suggested by referee 2, since the value function is the minimum of the value functions for the subproblems, i.e.,

$$V(\alpha) = \min_{\nu \subset I} V_\nu(\alpha)$$

and

$$V(\bar{\alpha}) = V_\nu(\bar{\alpha})\ \forall \nu = L(\bar{x}, \bar{y}) \cup \sigma, \sigma \subseteq I_0(\bar{x}, \bar{y}), (\bar{x}, \bar{y}) \in \Sigma(\bar{\alpha}),$$

applying the calculus for the minimum functions in Proposition 2.7, we conclude that the value function V is lower semicontinuous if each $V_\nu(\alpha), \nu = L(\bar{x},\bar{y}) \cup \sigma, \sigma \subseteq I_0(\bar{x},\bar{y}), (\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})$, is lower semicontinuous and the following inclusion holds:

$$\partial^\infty V(\bar{\alpha}) \subseteq \{\partial^\infty V_\nu(\bar{\alpha}) : \nu = L(\bar{x},\bar{y}) \cup \sigma, \sigma \subseteq I_0(\bar{x},\bar{y}), (\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})\}, \tag{24}$$

$$\partial V(\bar{\alpha}) \subseteq \{\partial V_\nu(\bar{\alpha}) : \nu = L(\bar{x},\bar{y}) \cup \sigma, \sigma \subseteq I_0(\bar{x},\bar{y}), (\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})\}. \tag{25}$$

The Fritz John condition for the subproblem $\mathrm{OPCC}(\bar{\alpha})_\nu$ with

$$\nu = L(\bar{x},\bar{y}) \cup \sigma, \sigma \subseteq I_0(\bar{x},\bar{y}), (\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})$$

implies the existence of vectors $(\gamma,\beta,\eta,\xi) \in R^d \times R^a \times R^b \times R^b$ satisfying (14)–(17) and

$$\xi_\sigma \leq 0, \eta_{I_0\setminus\sigma} \leq 0. \tag{26}$$

DEFINITION 4.9 (P multipliers). *The set of all vectors (γ,β,η) satisfying the above Fritz John condition at $(\bar{x},\bar{y})$ is denoted by $M_\sigma^\lambda(\bar{x},\bar{y})$, and $\bigcup_{\sigma\subseteq I_0} M_\sigma^\lambda(\bar{x},\bar{y})$ is called the set of P multipliers.*

Applying Corollary 3.8, we have the following result.

PROPOSITION 4.10. *For any $(\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})$ and any given index set $\sigma \subseteq I_0(\bar{x},\bar{y})$, assume that there exists $\delta > 0$ such that the set*

$$\begin{aligned}\{(x,y) \in C :&(p,q,q^y,q^F,) \in B(0;\delta), \Psi(x,y,\bar{\alpha}) \leq p, H(x,y,\bar{\alpha}) = q,\\ &y_i \geq q_i^y,\ F_i(x,y,\bar{\alpha}) = q_i^F \quad \forall i \in \nu := \sigma \cup L(\bar{x},\bar{y}),\\ &y_i = q_i^y,\ F_i(x,y,\bar{\alpha}) \geq q_i^F \quad \forall i \in I\setminus\nu, f(x,y,\bar{\alpha}) \leq M\}\end{aligned}$$

is bounded for each M. Then the value function for subproblem $\mathrm{OPCC}(\bar{\alpha})_\nu$ with $\nu = L(\bar{x},\bar{y}) \cup \sigma$ is lower semicontinuous near $\bar{\alpha}$ and

$$\begin{aligned}\partial V_\nu(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma_\nu(\bar{\alpha})} &\{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta\\ &+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta,\eta) \in M_\sigma^1(\bar{x},\bar{y})\},\\ \partial^\infty V_\nu(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma_\nu(\bar{\alpha})} &\{\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta\\ &+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta,\eta) \in M_\sigma^0(\bar{x},\bar{y})\},\end{aligned}$$

where $\Sigma_\nu(\bar{\alpha})$ denotes the set of solutions for the subproblem OPCC $(\alpha)_\nu$.

We have the following estimates for the value function in terms of P multipliers.

THEOREM 4.11. *Assume that there exists $\delta > 0$ such that for $(\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})$ and each index set $\sigma \subseteq I_0(\bar{x},\bar{y})$, the set in Proposition 4.10 is bounded for each M. Then the value function V is lower semicontinuous near $\bar{\alpha}$ and*

$$\begin{aligned}\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} &\{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta\\ &+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta.\eta) \in \cup_{\sigma\subseteq I_0} M_\sigma^1(\bar{x},\bar{y})\}, \quad (27)\\ \partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} &\{\nabla_\alpha \Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta\\ &+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta,\eta) \in \cup_{\sigma\subseteq I_0} M_\sigma^0(\bar{x},\bar{y})\}. \quad (28)\end{aligned}$$

If the set in the right-hand side of inclusion (28) *contains only the zero vector, then the value function V is Lipschitz near $\bar{\alpha}$. If the set in the right-hand side of inclusion* (28) *contains only the zero vector and the set in the right-hand side of inclusion* (27) *is a singleton, then the value function is strictly differentiable at $\bar{\alpha}$.*

DEFINITION 4.12 (S multipliers). *The set of index λ S multipliers, denoted by $M_S^\lambda(\bar{x},\bar{y})$, is the set of all vectors $(\gamma,\beta,\eta) \in R^d \times R^a \times R^b$ satisfying* (14)–(17) *and*

$$\xi_i \le 0, \eta_i \le 0 \qquad \text{if } \bar{y}_i = 0 \text{ and } F_i(\bar{x},\bar{y},\bar{\alpha}) = 0.$$

In the following theorem, we give a condition under which the set of P multipliers and S multipliers coincide, and so we have the estimates in terms of the S multipliers.

THEOREM 4.13. *In addition to the assumptions of Theorem* 4.11*, assume that $C = R^n \times R^a \times R^b$ and for all $(\bar{x},\bar{z},\bar{u}) \in \Sigma(\bar{\alpha})$, the partial MPEC linear independence constraint qualification is satisfied, i.e.,*

$$\begin{cases} 0 = \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top\beta + \nabla F(\bar{x},\bar{y},\bar{\alpha})^\top\eta + (0,0,\xi), \\ \gamma_{J(\Psi)} = 0, \ \eta_{I_+} = 0, \ \xi_L = 0, \end{cases}$$

implies that $\eta_{I_0} = 0, \xi_{I_0} = 0$, where $J(\Psi) := \{i : \Psi_i(\bar{x},\bar{y},\bar{\alpha}) < 0\}$. Then the value function V is lower semicontinuous near $\bar{\alpha}$ and

$$\partial V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha f(\bar{x},\bar{y},\bar{\alpha}) + \nabla_\alpha\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta$$
$$+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta,\eta) \in M_S^1(\bar{x},\bar{y})\},$$
$$\partial^\infty V(\bar{\alpha}) \subseteq \bigcup_{(\bar{x},\bar{y})\in\Sigma(\bar{\alpha})} \{\nabla_\alpha\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla_\alpha H(\bar{x},\bar{y},\bar{\alpha})^\top\beta$$
$$+\nabla_\alpha F(\bar{x},\bar{y},\bar{\alpha})^\top\eta : (\gamma,\beta,\eta) \in M_S^0(\bar{x},\bar{y})\}.$$

Remark. As in the proof of [22, Theorem 3.2], it is easy to see that under the partial MPEC linear independence constraint qualification, all multipliers including the S multiplier, the CD multiplier, the C multiplier, and the P multiplier coincide.

Recently, the MPEC linear independence constraint qualifications have received a lot of attention. It is known that under the MPEC linear independence constraint qualification, the computation of the OPCC is much easier and more efficient (see, e.g., Scholtes [20]). Furthermore, it was shown in Scholtes [21] that the MPEC linear independence constraint qualification is a generic condition for the OPCC. Here we prove the importance of the MPEC linearly independence constraint qualification from the aspect of the sensitivity analysis: the value function is Lipschitz continuous, and it is even strictly differentiable in the case where the optimal solution set is unique. Note that the MPEC linear independence constraint qualification is stronger than the partial MPEC linear independence constraint qualification.

COROLLARY 4.14. *In addition to the assumptions of Theorem* 4.11*, assume that the MPEC linear independence constraint qualifications are satisfied at all $(\bar{x},\bar{y}) \in \Sigma(\bar{\alpha})$, i.e.,*

$$\begin{cases} 0 = \nabla\Psi(\bar{x},\bar{y},\bar{\alpha})^\top\gamma + \nabla H(\bar{x},\bar{y},\bar{\alpha})^\top\beta + \nabla F(\bar{x},\bar{y},\bar{\alpha})^\top\eta + (0,0,\xi), \\ \gamma_{J(\Psi)} = 0, \eta_{I_+} = 0, \xi_L = 0, \end{cases}$$

implies that $\gamma = 0, \beta = 0, \eta = 0, \xi = 0$. Then the value function is Lipschitz continuous near $\bar{\alpha}$. Furthermore, if the set of optimal solutions $\Sigma(\bar{\alpha})$ is a singleton, then the value function V is strictly differentiable at $\bar{\alpha}$.

Proof. The MPEC linear independence constraint qualification obviously implies that $M_S^0(\bar{x},\bar{y}) = \{0\}$ and $M_S^1(\bar{x},\bar{y})$ is a singleton. Hence the conclusion follows from Theorem 4.13 and Proposition 2.4. □

4.5. Relationships between the multipliers for the OPCC. Applying the definitions, it is clear that

$$M_S^\lambda(\bar{x},\bar{y}) \subseteq M_{CD}^\lambda(\bar{x},\bar{y}) \subseteq M_C^\lambda(\bar{x},\bar{y}), \qquad M_S^\lambda(\bar{x},\bar{y}) \subseteq M_P^\lambda(\bar{x},\bar{y}). \tag{29}$$

It is not possible to compare the set of NLP multipliers directly with the other multipliers since the spaces they belong to have different dimensions. However, the following interesting relationships can be obtained.

PROPOSITION 4.15 (relationship between an NLP multiplier and an S multiplier).

$$\{(\gamma,\beta,\mu\bar{y}-r^F) : (\gamma,\beta,r^F,r^y,\mu) \in M_{NLP}^\lambda(\bar{x},\bar{y})\} \subseteq M_S^\lambda(\bar{x},\bar{y})$$

for all $\lambda \geq 0$.

Proof. Let $(\gamma,\beta,r^F,r^y,\mu) \in M_{NLP}^\lambda(\bar{x},\bar{y})$. We consider the following cases.

Case $\bar{y}_i > 0, F_i(\bar{x},\bar{y}) = 0$. Then $r_i^y = 0$. So $\xi_i := \mu F_i - r_i^y = 0$.

Case $\bar{y}_i = 0, F_i(\bar{x},\bar{y}) > 0$. Then $r_i^F = 0$. So $\eta_i = \mu\bar{y}_i - r_i^F = 0$.

Case $\bar{y}_i = 0, F_i(\bar{x},\bar{y}) = 0]$. Then $\xi_i = \mu F_i(\bar{x},\bar{y}) - r_i^y = -r_i^y$ and $\eta_i = \mu\bar{y}_i - r_i^F = -r_i^F$. So $\xi_i = -r_i^y \leq 0$ and $\eta_i = -r_i^F \leq 0$.

Hence (γ,β,η), where $\eta := \mu\bar{y}-r^F$, is an S multiplier, and the proof of the proposition is complete. □

The above relationship indicates that one can arrange the upper estimates of the limiting subdifferentials in Theorems 4.2, 4.13, 4.4, and 4.8 from the smallest to the largest in the order of NLP multipliers, S multipliers, CD multipliers, and C multipliers.

One may try to use the smallest multiplier set in sensitivity analysis. However, the smaller multipler sets tend to require stronger constraint qualifications and hence may be empty. In such a case, where the smaller multipler set is empty, one may have to use the larger multiplier set.

We now use the following example to show that in some cases the smaller multiplier sets such as the NLP and the S multiplier sets may be empty while the CD multiplier provides the tightest bound.

Example. Consider the OPCC

$$(P) \qquad \begin{array}{rl} \text{minimize} & -y \\ \text{subject to} & x - y = 0, \\ & x \geq 0, y \geq 0, xy = 0, \end{array}$$

where $x \in R$ and $y \in R$, and its perturbed problem

$$P(q,r) \qquad \begin{array}{rl} \text{minimize} & -y \\ \text{subject to} & x - y = q, \\ & x - r \geq 0, y \geq 0, (x-r)y = 0, \end{array}$$

which is OPCC (α) with $\alpha = (q,r), f = -y, H = x - y - q, F = x - r$. Let $\bar{\alpha} = (0,0)$. It is clear that the only feasible solution for problem $(P) = P(0,0)$ is $(0,0)$. Hence the only optimal solution for *(P)* is $(0,0)$. The set of index λ NLP multipliers (β, r^y, r^F, μ)

at $(0,0)$ satisfy

$$\begin{cases} 0 = \lambda(0,-1) + \beta(1,-1) - (r^F, 0) - (0, r^y) + \mu(0,0), \\ r^F, r^y \geq 0. \end{cases}$$

It is clear that any $(\beta, r^y, r^F, \mu) = (0,0,0,\mu)$ with $\mu \neq 0$ is a nonzero NLP abnormal multiplier and there is no NLP normal multiplier. Hence $M^0_{NLP}(0,0) = \{(0,0,0)\} \times (-\infty, +\infty) \neq \{(0,0,0,0)\}$ and $M^1_{NLP}(0,0) = \emptyset$.

Since $\bar{y} = 0$ and $F(\bar{x}, \bar{y}, 0) = \bar{x} = 0$, the index λ CD multipliers (β, η) at $(0,0)$ satisfy

$$0 = \lambda(0,-1) + \beta(1,-1) + \eta(1,0) + (0,\xi),$$
$$\text{either } \xi < 0, \eta < 0, \text{ or } \xi\eta = 0.$$

When $\lambda = 0$, the above condition implies that $\beta = \eta = \xi = 0$, while when $\lambda = 1$, either $\eta = 1, \beta = -1, \xi = 0$, or $\beta = \eta = 0, \xi = 1$. So $M^1_{CD}(0,0) = \{(0,0)\} \cup \{(-1,1)\}$ and $M^0_{CD}(0,0) = \{(0,0)\}$.

The set of index λ C multipliers (β, η) at $(0,0)$ satisfy

$$0 = \lambda(0,-1) + \beta(1,-1) + \eta(1,0) + (0,\xi),$$
$$\xi\eta \geq 0.$$

When $\lambda = 0$, the above condition implies that $\beta = \eta = \xi = 0$, while for $\lambda = 1$, $-\beta = \eta \in [0,1]$. So $M^1_C(0,0) = \{(\beta,\eta) : \eta = -\beta \in [0,1]\}$ and $M^0_C(0,0) = \{(0,0)\}$.

Since the optimal solution for (P) is $(\bar{x}, \bar{y}) = (0,0)$, $(0,0)$ is also optimal for the subproblem associated with $\nu = \{1\}$,

$$(P_1) \qquad \begin{array}{rl} \text{minimize} & -y \\ \text{subject to} & x - y = 0, \\ & y \geq 0, x = 0, \end{array}$$

and the subproblem associated with $\nu = \emptyset$,

$$(P_2) \qquad \begin{array}{rl} \text{minimize} & -y \\ \text{subject to} & x - y = 0, \\ & y = 0, x \geq 0. \end{array}$$

The index λ multiplier set for (P_1) consists of vectors (β, η) satisfying

$$\begin{cases} 0 = \lambda(0,-1) + \beta(1,-1) + \eta(1,0) + (0,\xi), \\ \xi \leq 0, \end{cases}$$

and the index λ multiplier set for (P_2) consist of vectors (β, η) satisfying

$$\begin{cases} 0 = \lambda(0,-1) + \beta(1,-1) + \eta(1,0) + (0,\xi), \\ \eta \leq 0. \end{cases}$$

Therefore, the abnormal P multiplier set is

$$\begin{aligned} M^0_P(0,0) = M^0_1(0,0) \cup M^0_2(0,0) &= \{(\beta,\eta) : \beta = -\eta \leq 0\} \cup \{(\beta,\eta) : \beta = -\eta \geq 0\} \\ &= \{(\beta,\eta) : \beta = -\eta\}, \end{aligned}$$

and the normal P multiplier set is

$$\begin{aligned} M_P^1(0,0) = M_1^1(0,0) \cup M_2^1(0,0) &= \{(\beta,\eta) : \beta = -\eta \leq -1\} \cup \{(\beta,\eta) : \beta = -\eta \geq 0\} \\ &= \{(\beta,\eta) : \beta = -\eta \in (-\infty,-1] \cup [0,\infty)\}. \end{aligned}$$

The index λ S multiplier set consists of vectors (β,η) satisfying

$$\begin{cases} 0 = \lambda(0,-1) + \beta(1,-1) + \eta(1,0) + (0,\xi), \\ \xi \leq 0, \eta \leq 0, \end{cases}$$

i.e.,

$$\begin{aligned} &\beta = -\eta, \ \beta = -\lambda + \xi, \\ &\xi \leq 0, \ \eta \leq 0. \end{aligned}$$

That is, $M_S^0(0,0) = \{0\}$, and $M_S^1(0,0) = \emptyset$.

Consider the value function

$$V(q,r) := \inf\{-y : x - r \geq 0, y \geq 0, (x-r)y = 0, x - y = q\}.$$

Then by Theorem 4.4, since the only abnormal CD multiplier is the zero vector, we conclude that the value function is Lipschitz near $(0,0)$, and

$$\begin{aligned} \emptyset \neq \partial V(0,0) &\subseteq \{\beta(-1,0) + \eta(0,-1) : (\beta,\eta) \in M_{CD}^1(0,0)\} \\ &= -M_{CD}^1(0,0) = \{(0,0)\} \cup \{(1,-1)\}. \end{aligned}$$

In fact, we can easily find the expression for the value function for this simple example since the feasible set of the perturbed problem $P(q,r)$ still reduces to one point. Indeed, we have

$$\begin{cases} \Sigma(q,r) = \{(r, r-q)\} \text{ and } V(q,r) = q - r & \text{if } q < r, \\ \Sigma(q,r) = \{(q,0)\} \text{ and } V(q,r) = 0 & \text{if } q \geq r. \end{cases}$$

So $V(q,r) = \min(0, q-r)$, which is Lipschitz continuous everywhere. By definition of the limiting subdifferentials, it is easy to see that

$$\begin{aligned} \partial V(0,0) &= \{(0,0)\} \cup \{(1,-1)\}, \\ \partial^\infty V(0,0) &= \{(0,0)\}. \end{aligned}$$

Therefore, the inclusions in Theorem 4.4 are actually equalities here, i.e.,

$$\begin{aligned} \partial V(0,0) &= \{(0,0)\} \cup \{(1,-1)\} = -M_{CD}^1(0,0), \\ \partial^\infty V(0,0) &= \{(0,0)\} = -M_{CD}^0(0,0). \end{aligned}$$

Using Theorem 4.8, since the only abnormal C multiplier is the zero vector, one also concludes that the value function is Lipschitz. However, the upper estimate for the limiting subdifferentials of the value function in terms of the C multiplier set is a strict inclusion here:

$$\begin{aligned} \partial V(0,0) &= \{(0,0)\} \cup \{(1,-1)\} \subset \{(\beta,\eta) : \beta = -\eta \in [0,1]\} = -M_C^1(0,0), \\ \partial^\infty V(0,0) &= \{(0,0)\} = -M_{CD}^0(0,0). \end{aligned}$$

The upper estimate for both the limiting and the singular limiting subdifferentials of the value function in Theorem 4.11 are both strict:

$$\begin{aligned}\partial V(0,0) &= \{(0,0)\} \cup \{(1,-1)\}\\ &\subset \{(\beta,\eta) : \beta = -\eta \in (-\infty,0] \cup (1,\infty)\} = -M_P^1(0,0),\\ \partial^\infty V(0,0) &= \{(0,0)\}\\ &\subset \{(\beta,\eta) : \beta = -\eta\} = -M_P^0(0,0).\end{aligned}$$

These inclusions are not very helpful since the Lipschitz continuity of the value function cannot be detected and the upper estimate is unbounded.

Since there is no S multiplier for this problem, the limiting subdifferential of the value function cannot be estimated in terms of the S multiplier. In fact, the assumptions in Theorem 4.13 are not satisfied for this problem. Indeed,

$$(0,0) = \beta(1,-1) + \eta(1,0) + (0,\xi)$$

does not imply that $\eta = 0, \xi = 0$.

Note that by Theorem 4.2, if the growth hypotheses were satisfied, then

$$\begin{aligned}\partial V(0,0) &\subseteq \{\beta(-1,0) - r^F(0,-1) : (\beta, r^F, r^y, \mu) \in M_{NLP}^1(\Sigma)\},\\ \partial^\infty V(0,0) &\subseteq \{\beta(-1,0) - r^F(0,-1) : (\beta, r^F, r^y, \mu) \in M_{NLP}^0(\Sigma)\}.\end{aligned}$$

But this is not possible since $M_{NLP}^1(\Sigma) = \emptyset$. Indeed, (GH) is not satisfied for this example.

In the above example, $M_{NLP}^0(\Sigma) \neq \{0\}$, while $M_{CD}^0(\Sigma) = \{0\}$. In fact, it is not just a coincidence that $M_{NLP}^0(\Sigma) \neq \{0\}$. In general, the Mangasarian–Fromovitz constraint qualification satisfying at a feasible solution $(\bar{x}, \bar{z}, \bar{u})$ implies that $M_{NLP}^0((\bar{x}, \bar{z}, \bar{u})) = \{0\}$, and in the case of no abstract constraint, the two conditions are equivalent (see, e.g., [6] and [25, Proposition 4.5] for details). It is well known that in the case of no abstract constraint, the Mangasarian–Fromovitz constraint qualification fails to hold at every feasible point of the OPCCs. (The proof for the case where the complementarity constraint comes from the KKT condition of a lower level quadratic programming problem was given in Chen and Florian [1, Lemma 3.1], and the proof for the general case was given in [25, Proposition 1.1].) We now prove that even for the case when the abstract constraint set C is present, there always exist nonzero abnormal NLP multipliers for the OPCC.

PROPOSITION 4.16. *Let $(\bar{x}, \bar{y}) \in R^{n+m}$ be any feasible solution of the* OPCC. *Then $M_{NLP}^0(\bar{x}, \bar{y}) \backslash \{0\} \neq \emptyset$.*

Proof. The point $(\bar{x}, \bar{y})$ is obviously a solution to the following optimization problem:

$$\begin{aligned}&\text{minimize} && \langle y, F(x,y)\rangle\\ &\text{subject to} && y \geq 0, F(x,y) \geq 0.\end{aligned}$$

By the multiplier rule, there exists $\mu \geq 0, r^y \in R_+^m, r^F \in R_+^m$ not all zero such that

$$\begin{aligned}&0 = \mu \nabla \langle y, F\rangle(\bar{x}, \bar{y}) - (0, r^y) - \nabla F(\bar{x}, \bar{y})^\top r^F,\\ &\langle \bar{y}, r^y\rangle = 0, \langle r^F, F(\bar{x}, \bar{y}) = 0.\end{aligned}$$

Therefore, taking $\gamma = 0, \beta = 0$ ($\gamma = 0, \beta = 0, r^F, r^y, \mu$) is a nonzero NLP abnormal multiplier of the OPCC. ☐

4.6. Applications to the bilevel programming problem. One of the motivations to consider OPCCs is to solve the following bilevel programming problem:

$$\text{(BLPP)} \qquad \begin{array}{ll} \text{minimize} & f(x,z) \\ \text{subject to} & z \in S(x), \Psi(x,z) \le 0, (x,z) \in C, \end{array} \tag{30}$$

where $S(x)$ is the solution of the lower-level problem

$$P_x \qquad \begin{array}{ll} \text{minimize} & g(x,z) \\ \text{subject to} & \psi(x,z) \le 0, \end{array} \tag{31}$$

where $f : R^{n+a} \to R, g : R^{n+a} \to R, \psi : R^{n+a} \to R^b, \Psi : R^{n+a} \to R^d$. Under suitable convexity assumptions, we can replace the lower problem by its KKT conditions. As in [24], we find that any (x,z) is solution of (BLPP) if and only if there is u such that (x,z,u) is solution of the problem

$$\begin{array}{ll} \text{minimize} & f(x,z) \\ \text{subject to} & \psi(x,z) \le 0 \text{ and } u \ge 0, \\ & \langle \psi(x,z), u \rangle = 0, \\ & \nabla_z g(x,z) + \nabla_z \psi(x,z)^{\top} u = 0, \\ & \Psi(x,z) \le 0, (x,z) \in C, \end{array} \tag{32}$$

which is an OPCC.

Consider the perturbed bilevel programming problem

$$\text{BLPP}(\alpha) \quad \begin{array}{ll} \text{minimize} & f(x,z,\alpha) \\ \text{subject to} & z \in S(x,\alpha), \Psi(x,z,\alpha) \le 0, (x,z) \in C, \end{array} \tag{33}$$

where $S(x,\alpha)$ is the solution of the lower-level problem

$$\begin{array}{ll} \text{minimize} & g(x,z,\alpha) \\ \text{subject to} & \psi(x,z,\alpha) \le 0. \end{array} \tag{34}$$

Under suitable assumptions, BLPP(α) is equivalent to

$$\begin{array}{ll} \text{minimize} & f(x,z,\alpha) \\ \text{subject to} & \psi(x,z,\alpha) \le 0 \text{ and } u \ge 0, \\ & \langle \psi(x,z,\alpha), u \rangle = 0, \\ & \nabla_z g(x,z,\alpha) + \nabla_z \psi(x,z,\alpha)^T u = 0, \\ & \Psi(x,z,\alpha) \le 0, (x,z) \in C. \end{array} \tag{35}$$

Hence the results in this section allow us to derive the properties of the value function and compute the upper estimates of the limiting subdifferentials of V by the various kinds of multipliers for the above problem. For example, we can conclude that the value function is Lipschitz continuous when the strong second order sufficient condition and the linear independence of the binding constraints hold for the lower level problem. Indeed, in this case the corresponding generalized equation is strongly regular; hence the set of abnormal CD multipliers contains only the zero vector (see Ye [23, Theorem 5.1]).

Acknowledgments. The authors would like to thank the anonymous referees whose constructive suggestions led to the better presentation of the results and the consideration of the upper estimates in terms of various multipliers other than the CD multipliers in the last section.

REFERENCES

[1] Y. CHEN AND M. FLORIAN, *The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions*, Optimization, 32 (1995), pp. 193–209.

[2] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, 2nd ed., Classics in Appl. Math. 5, SIAM, Philadelphia, 1990.

[3] F. H. CLARKE, *Methods of Dynamic and Nonsmooth Optimization*, CBNS-NSF Regional Conf. Ser. in Appl. Math. 57, SIAM, Philadelphia, 1989.

[4] F. H. CLARKE, YU. S. LEDYAEV, R. J. STERN, AND P. R. WOLENSKI, *Nonsmooth Analysis and Control Theory*, Grad. Texts in Math. 178, Springer-Verlag, New York, 1998.

[5] A. DONTCHEV AND R. T. ROCKAFELLAR, *Characterizations of strong regularity for variational inequalities over polyhedral convex sets*, SIAM J. Optim., 6 (1996), pp. 1087–1105.

[6] A. JOURANI, *Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems*, J. Optim. Theory Appl., 81 (1994), pp. 533–548.

[7] P. D. LOEWEN, *Optimal Control via Nonsmooth Analysis*, AMS, Providence, RI, 1993.

[8] Z. Q. LUO, J. S. PANG, AND D. RALPH, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, New York, 1996.

[9] B. S. MORDUKHOVICH, *Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems*, Soviet Math. Dokl., 22 (1980), pp. 526–530.

[10] B. S. MORDUKHOVICH, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988.

[11] B. S. MORDUKHOVICH, *Sensitivity analysis in nonsmooth optimization*, in Theoretical Aspects of Industrial Design, D. A. Field and V. Komkov, eds., SIAM, Philadelphia, 1992, pp.32–46.

[12] B. S. MORDUKHOVICH, *Generalized differential calculus for nonsmooth and set-valued mappings*, J. Math. Anal. Appl., 183 (1994), pp. 250–288.

[13] B. S. MORDUKHOVICH, *Nonsmooth sequential analysis in Asplund spaces*, Trans. Amer. Math. Soc., 348 (1996), pp. 1235–1280.

[14] J. V. OUTRATA, *Optimality conditions for a class of mathematical programs with equilibrium constraints*, Math. Oper. Res., 24 (1999), pp. 627–644.

[15] J. V. OUTRATA, M. KOČVARA, AND J. ZOWE, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Kluwer, Dordrecht, The Netherlands, 1998.

[16] R. A. POLIQUIN AND R. T. ROCKAFELLAR, *Tilt stability of a local minimum*, SIAM J. Optim., 8 (1998), pp. 287–299.

[17] S. M. ROBINSON, *Strongly regular generalized equation*, Math. Oper. Res., 5 (1980), pp. 43–62.

[18] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

[19] H. SCHEEL AND S. SCHOLTES, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22.

[20] S. SCHOLTES, *Convergence Properties of a Regularization Scheme for Mathematical Programs with Complementarity Constraints*, Working paper, Judge Institute of Management Studies, University of Cambridge, Cambridge, UK, 1999.

[21] S. SCHOLTES, *How Stringent is the Linear Independence Assumption for Mathematical Programs with Stationary Constraints?*, Working paper, Judge Institute of Management Studies, University of Cambridge, Cambridge, UK, 1999.

[22] J. J. YE, *Optimality conditions for optimization problems with complementarity constraints*, SIAM J. Optim., 9 (1999), pp. 374–387.

[23] J. J. YE, *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints*, SIAM J. Optim., 10 (2000), pp. 943–962.

[24] J. J. YE AND X. Y. YE, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res., 22 (1997), pp. 977–997.

[25] J. J. YE, D. L. ZHU, AND Q. J. ZHU, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim., 7 (1997), pp. 481–507.

Bibliographie

[1] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang. The proximal average : Basic theory. *SIAM J. Optim.*, Mar. 2008. Accepté pour publication.

[2] H. H. Bauschke, Y. Lucet, and M. Trienis. How to transform one convex function continuously into another. *SIAM Review*, 50(1) :115–132, 2008.

[3] H. H. Bauschke, Y. Lucet, and X. Wang. Primal-dual symmetric intrinsic methods for finding antiderivatives of cyclically monotone operators. *SIAM J. Control Optim.*, 46(6) :2031–2051, 2007.

[4] J. Borwein, Y. Lucet, and B. Mordukhovich. Compactly epi-lipschitzian convex sets and functions in normed spaces. *J. Convex Anal.*, 7(2) :375–393, 2000.

[5] J.-B. Hiriart-Urruty and Y. Lucet. Parametric computation of the Legendre–Fenchel conjugate. *J. Convex Anal.*, 14(3) :657–666, Aug. 2007.

[6] Y. Lucet. A linear Euclidean distance transform algorithm based on the Linear-time Legendre Transform. In *Proceedings of the Second Canadian Conference on Computer and Robot Vision*, 262–267, 2005. IEEE Computer Society Press.

[7] Y. Lucet. Fast Moreau envelope computation I : Numerical algorithms. *Numer. Algorithms*, 43(3) :235–249, Nov. 2006.

[8] Y. Lucet. New sequential exact Euclidean distance transform algorithms based on convex analysis. *Image and Vision Computing*, Dec. 2006. Accepté pour publication.

[9] Y. Lucet. What shape is your conjugate ? A survey of computational convex analysis and its applications. Manuscrit soumis en 2007.

[10] Y. Lucet, H. H. Bauschke, and M. Trienis. The piecewise linear-quadratic model for computational convex analysis. *Comput. Optim. Appl.*, July 2006. Accepté pour publication.

[11] Y. Lucet and J. Ye. Sensitivity analysis for the value function for optimization problems with variational inequalities constraints. *SIAM J. Control Optim.*, 40(3) :699–723, Sept. 2002.

[11 bis] Y. Lucet and J. Ye. Erratum : Sensitivity analysis of the value function for optimization problems with variational inequality constraints. *SIAM J. Control Optim.*, 41(4) :1315–1319, Dec. 2002.