WHAT IS a Fenchel conjugate?

Heinz H. Bauschke and *Yves Lucet*

The ideas of duality and transforms are ubiquitous in mathematics, the most classical example being the Fourier transform in Harmonic Analysis. Convex Analysis, an area founded by W. Fenchel, J.-J. Moreau, and R.T. Rockfellar in the mid-20th century, concerns convex convex sets, convex functions, and their applications to optimization. The counterpart of the Fourier transform in Convex Analysis is the Fenchel conjugate. Suppose we have a real Hilbert space X and a function $f\colon X \to\,]-\infty,+\infty]$. We shall assume that f is *proper*, i.e., $\operatorname{dom} f = \big\{x \in X \;\big|\; f(x) \in \mathbb{R}\big\} \neq \varnothing$. Then the *Fenchel conjugate* f^* at $u \in X$ is

$$f^*(u) = \sup_{x\in X} \big(\langle x,u\rangle - f(x)\big).$$

An immediate consequence of the definition is the *Fenchel-Young inequality*

$$f(x) + f^*(u) \geq \langle x,u\rangle\,.$$

We also note that f^* is convex and lower semicontinuous because it is the supremum of the family of affine continuous functions $(\langle x,\cdot\rangle - f(x))_{x\in X}$. One has the beautiful duality

$$f(x) = f^{**}(x) \Leftrightarrow \begin{cases} f \text{ is convex and} \\ \text{lower semicontinuous,} \end{cases}$$

which shows that such a function f can be represented as a supremum of affine functions $x \mapsto \langle u,x\rangle - f^*(u)$, where $f^*(u)$ determines the constant term of the affine function with slope u.

Given a subset C of X, its *indicator function* ι_C is defined by $\iota_C(x) = 0$, if $x \in C$; $+\infty$, otherwise. As a first example, we compute that if $f(x) = \langle x,a\rangle$, where $a \in X$, then $f^* = \iota_{\{a\}}$. Thus, $+\infty$ is unavoidable and to be embraced in Convex Analysis. If f is convex and differentiable, then the supremum in the definition of $f^*(u)$ can be found by calculus and we obtain

$$f^*\big(\nabla f(x)\big) = \langle x, \nabla f(x)\rangle - f(x).$$

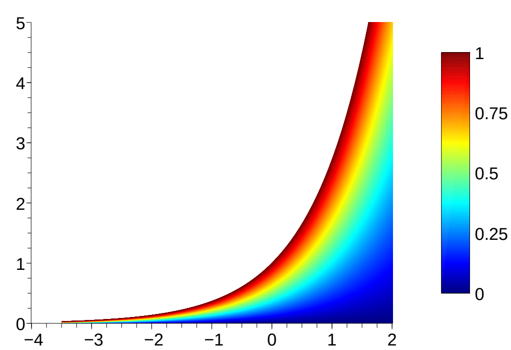
This formula explains not only why the Fenchel conjugate is also known as the Fenchel-Legendre transform but it also shows that the *energy* is self-dual; in fact,

$$f = f^* \Leftrightarrow f = \tfrac{1}{2}\|\cdot\|^2.$$

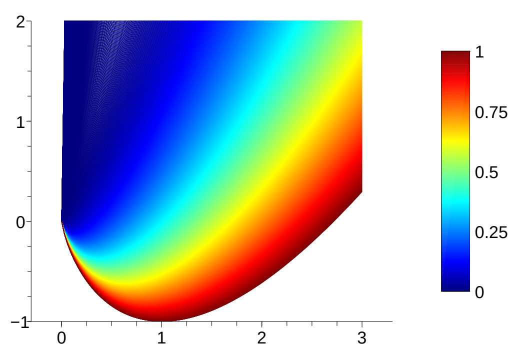
Given $\alpha > 0$, one also computes that $\alpha \exp$ and the following (scaled) negative entropy are conjugates of each other:

$$(\alpha\exp)^*(u) = \begin{cases} +\infty, & \text{if } u<0; \\ 0, & \text{if } u=0; \\ u\ln(u/\alpha) - u, & \text{if } u>0. \end{cases}$$

By associating each $\alpha \in\,]0,1]$ with a colour, we are able to display an entire family of conjugates.



The family $(\alpha\exp)_{\alpha\in]0,1]}$.



The associated family of conjugates.

Many more interesting pairs can be computed (in closed form, or at least numerically). For instance, if $1 < p < +\infty$, then

$$\left(\tfrac{1}{p}|\cdot|^p\right)^* = \tfrac{1}{q}|\cdot|^q, \text{ where } \tfrac{1}{p}+\tfrac{1}{q}=1$$

which, along with the Fenchel-Young inequality, leads to an elegant proof of Hölder's inequality.

The natural domain for Fenchel conjugation is Γ_X, the cone of functions that are convex, lower semicontinuous, and proper on X. One now wishes to obtain calculus rules for Fenchel conjugation. In Harmonic Analysis, one is led to discover convolution as crucial in describing the Fourier transform of a product. The counterpart in Convex Analysis is the *infimal convolution* $f \Box g$, defined by

$$(f \Box g)(x) = \inf_{y \in X} \big(f(y) + g(x-y)\big).$$

Under appropriate hypotheses, one has

$$(f \Box g)^* = f^* + g^* \text{ and } (f+g)^* = f^* \Box g^*.$$

Moreover, $(\alpha f)^*(u) = \alpha f^*(u/\alpha)$ where $\alpha > 0$. Closely tied to the Fenchel conjugate of $f \in \Gamma_X$ is the *subdifferential operator* ∂f. This is a *set-valued* mapping on X, i.e., it maps from X to the power set of X, and it is defined by

$$u \in \partial f(x) \;\Leftrightarrow\; (\forall h \in X)\; f(x) + \langle h, u\rangle \le f(x+h).$$

Now equality in the Fenchel-Young inequality characterizes *subgradients*, i.e., elements in the subdifferential, in the sense that

$$\begin{aligned} f(x) + f^*(u) = \langle x, u\rangle &\Leftrightarrow u \in \partial f(x) \\ &\Leftrightarrow x \in \partial f^*(u). \end{aligned}$$

When f is continuous at x, then differentiability of f at x is the same as requiring $\partial f(x)$ to be a singleton, in which case $\partial f(x) = \{\nabla f(x)\}$. When $\operatorname{dom} f = X = \mathbb{R}$, then the left ($f'_-$) and right ($f'_+$) derivatives exists at every x and

$$\partial f(x) = \big[f'_-(x), f'_+(x)\big].$$

Thus, the subdifferential operator is a powerful generalized derivative. It also has a property critical for Optimization:

$$0 \in \partial f(x) \;\Leftrightarrow\; x \text{ is a global minimizer of } f.$$

Suppose that Y is another real Hilbert space, $A\colon X \to Y$ is continuous and linear, and $g \in \Gamma_Y$, The most important theorem concerns *Fenchel-Rockafellar* duality, which involves the *primal problem*

$$(P) \qquad \underset{x \in X}{\text{minimize}} \;\; f(x) + g(Ax),$$

and the associated *dual problem*

$$(D) \qquad \underset{y \in Y}{\text{minimize}} \;\; f^*(-A^*y) + g^*(y).$$

Set $\mu = \inf\big\{f(x) + g(Ax) \;\big|\; x \in X\big\}$ and $\mu^* = \inf\big\{f^*(-A^*y) + g^*(y) \;\big|\; y \in Y\big\}$. Then $\mu \ge -\mu^*$. The key result asserts that in the presence of a so-called (primal) *constraint qualification* such as $0 \in \operatorname{int}\big(\operatorname{dom} g - A \operatorname{dom} f\big)$, one has $\mu = -\mu^*$ and the dual problem possesses at least one solution. Let y be an *arbitrary* dual solution. Then the *entire set* of primal solutions is obtained as

$$\partial f^*(-A^*y) \cap A^{-1}\partial g^*(y).$$

As an example, one may formally derive the well known *Linear Programming (LP) Duality*, which concerns

$$\inf\big\{\langle c, x\rangle \;\big|\; x \ge 0,\, Ax = b\big\},$$

and

$$\sup\big\{\langle b, y\rangle \;\big|\; y \in \mathbb{R}^m, A^*y \le c\big\},$$

where $c \in X = \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in Y = \mathbb{R}^m$, and vector inequalities are interpreted entry-wise, by setting $f = \langle \cdot, c\rangle + \iota_{\mathbb{R}^n_+}$ and $g = \iota_{\{b\}}$.

Let $f \in \Gamma_X$. Then the operator $\partial f + \operatorname{Id}$ is surjective; here Id denotes the identity mapping. The inverse operator $(\partial f + \operatorname{Id})^{-1}$ is actually single-valued and called the *proximal mapping* Prox_f. In view of

$$x = \operatorname{Prox}_f x \;\Leftrightarrow\; x \text{ is a global minimizer of } f$$

and, for all x and y in X,

$$\begin{aligned} \|\operatorname{Prox}_f x - \operatorname{Prox}_f y\|^2 &\le \|x-y\|^2 \\ &\quad - \|(\operatorname{Id} - \operatorname{Prox}_f)x - (\operatorname{Id} - \operatorname{Prox}_f)y\|^2, \end{aligned}$$

Prox_f is Lipschitz continuous with constant 1 and thus enables fixed-point algorithmic approaches to optimization problems.

Turning to further applications, let

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

Strict-smooth duality: When $X = \mathbb{R}^n$ and $\operatorname{dom} f = \operatorname{dom} f^* = X$, then f is strictly convex if and only if f^* is differentiable.

Moreau envelope and Moreau decomposition: The beautiful identity

$$(f \Box q) + (f^* \Box q) = q$$

becomes

$$\operatorname{Prox}_f + \operatorname{Prox}_{f^*} = \operatorname{Id}$$

after taking the derivative. Let P_Y denote the projector onto the closed subspace Y of X. Then this last decomposition turns into the well known orthogonal subspace decomposition

$$P_Y + P_{Y^\perp} = \operatorname{Id}$$

since $\iota_Y^* = \iota_{Y^\perp}$ and $\mathrm{Prox}_{\iota_Y} = P_Y$.

The material thus far has been classical, although significant refinements continue to be made. We conclude with a recent development.

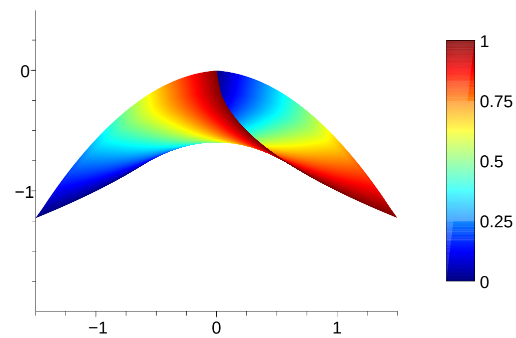
Proximal average: Let f_0 and f_1 be in Γ_X. Then the proximal average f_λ for $0 < \lambda < 1$ is defined by

$$\big((1-\lambda)(f_0+q)^* + \lambda(f_1+q)^*\big)^* - q.$$

We have $(f_\lambda)^* = (f^*)_\lambda$, i.e., taking the Fenchel conjugate and the proximal average commute, and

$$\mathrm{Prox}_{f_\lambda} = (1-\lambda)\,\mathrm{Prox}_{f_0} + \lambda\,\mathrm{Prox}_{f_1}.$$

The proximal average provides a homotopy between f_0 and f_1, even when $\mathrm{dom}\, f_0 \cap \mathrm{dom}\, f_1 = \varnothing$ and it is useful for the construction of antiderivatives and maximally monotone operators. By associating each $\lambda \in [0,1]$ with a colour, we are able to display the full family of proximal averages; here is a graph of the family of proximal averages $(f_\lambda)_{\lambda\in[0,1]}$ of $f_0(x) = -\sqrt{-x} + \iota_{[-3/2,0]}(x)$ and $f_1 = f_0 \circ (-\mathrm{Id})$.



The following reading list is a starting point to explore the theory, history, applications, and (symbolic and numerical) computation of Fenchel conjugates.

Further reading

(1) H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
(2) J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.
(3) Y. Lucet, What shape is your conjugate?, *SIAM Rev.* 52 (2010), 505–542.
(4) R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1996.