

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. Rockafellar in the mid-twentieth century, concerns 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 $f: X \rightarrow]-\infty, +\infty]$. We shall assume that f is *proper*, that is, $\text{dom } f = \{x \in X \mid f(x) \in \mathbb{R}\} \neq \emptyset$. Then the *Fenchel conjugate* f^* at $u \in X$ is

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

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

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

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

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

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 color, we are able to display an entire family of conjugates (see Figures 1 and 2).

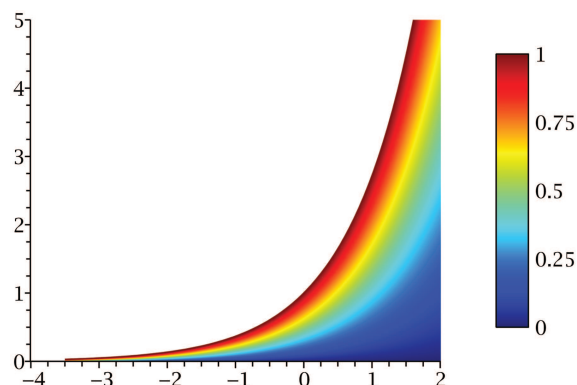


Figure 1. The family $(\alpha \exp)_{\alpha \in]0, 1]}$.

Heinz H. Bauschke is professor of mathematics at the University of British Columbia in Kelowna. His email address is heinz.bauschke@ubc.ca.

Yves Lucet is associate professor of computer science at the University of British Columbia in Kelowna. His email address is yves.lucet@ubc.ca.

DOI: <http://dx.doi.org/10.1090/noti788>.

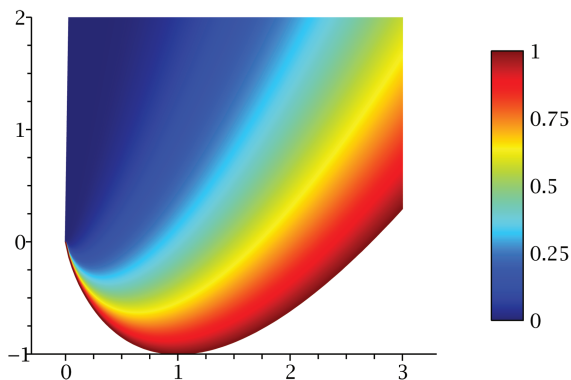


Figure 2. The associated family of conjugates.

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

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

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* $f \square g$, defined by

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

Under appropriate hypotheses, one has

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

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 , that is, it maps from X to the power set of X , and it is defined by

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

Now equality in the Fenchel-Young inequality characterizes *subgradients*, that is, 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 $\text{dom } f = X = \mathbb{R}$, then the left (f'_-) and right (f'_+) derivatives exist at every x and

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

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, $A: X \rightarrow Y$ is continuous and linear, and $g \in \Gamma_Y$. The most important theorem concerns *Fenchel-Rockafellar duality*, which involves the *primal problem*

$$(P) \quad \text{minimize}_{x \in X} f(x) + g(Ax),$$

and the associated *dual problem*

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

Set $\mu = \inf \{f(x) + g(Ax) \mid x \in X\}$ and $\mu^* = \inf \{f^*(-A^*y) + g^*(y) \mid y \in Y\}$. Then $\mu \geq -\mu^*$. The key result asserts that in the presence of a so-called (primal) *constraint qualification* such as $0 \in \text{int}(\text{dom } g - A \text{dom } f)$, 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

$$\inf \{\langle c, x \rangle \mid x \geq 0, Ax = b\},$$

and

$$\sup \{\langle b, y \rangle \mid y \in \mathbb{R}^m, A^*y \leq c\},$$

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

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

and, for all x and y in X ,

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

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

Turning to further applications, let

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

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

Moreau envelope and Moreau decomposition: The beautiful identity

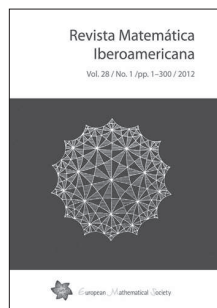
$$(f \square q) + (f^* \square q) = q$$

becomes

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



New journal from the
European Mathematical Society



New in 2012

Revista Matemática Iberoamericana

A scientific journal of the Real Sociedad Matemática Española
Published by the EMS Publishing House as of 2012

Revista Matemática Iberoamericana publishes original research articles on all areas of mathematics. Its distinguished Editorial Board selects papers according to the highest standards. Founded in 1985, *Revista* is the scientific journal of Real Sociedad Matemática Española.

Editors:

Antonio Córdoba, Universidad Autónoma de Madrid, Spain
José Luis Fernández, Universidad Autónoma de Madrid, Spain
Luis Vega, Universidad del País Vasco, Bilbao, Spain

Editorial Board:

Antonio Ambrosetti, SISSA, Trieste, Italy
Kari Astala, University of Helsinki, Finland
Rodrigo Bañuelos, Purdue University, West Lafayette, USA
Luis Barreira, Instituto Superior Técnico, Lisboa, Portugal
Pilar Bayer, Universitat de Barcelona, Spain
Joaquim Bruna, Universitat Autònoma de Barcelona, Spain
Luis A. Caffarelli, University of Texas at Austin, USA
Fernando Chamizo, Universidad Autónoma de Madrid, Spain
Sun-Yung Alice Chang, Princeton University, USA
Guy David, Université Paris-Sud, Orsay, France
Charles Fefferman, Princeton University, USA
Pierre-Louis Lions, Université de Paris IX-Dauphine, France
Rafael de la Llave, University of Texas at Austin, USA
Terry Lyons, University of Oxford, UK
Antonio Ros, Universidad de Granada, Spain
Elias M. Stein, Princeton University, USA
Gunther Uhlmann, University of Washington, Seattle, USA

Subscription information

ISSN print 0213-2230
ISSN online 2235-0616
2012. Vol. 28. 4 issues
Approx. 1200 pages. 17.0 x 24.0 cm
Price of subscription:
328 € online only / 388 € print+online

European Mathematical Society Publishing House
Seminar for Applied Mathematics, ETH-Zentrum FLI C4
CH-8092 Zürich, Switzerland
subscriptions@ems-ph.org
www.ems-ph.org

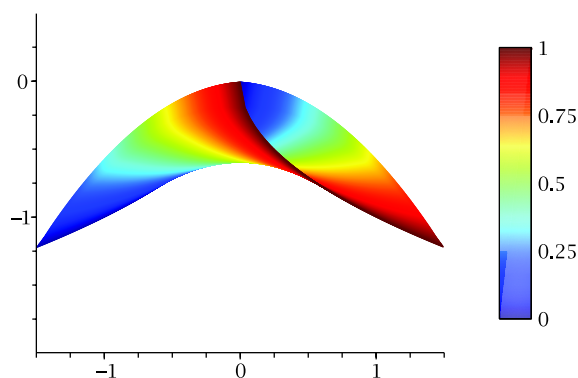


Figure 3.

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

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

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

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

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

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

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

The proximal average provides a homotopy between f_0 and f_1 , even when $\text{dom } f_0 \cap \text{dom } f_1 = \emptyset$, and it is useful for the construction of antiderivatives and maximally monotone operators. By associating each $\lambda \in [0, 1]$ with a color, we are able to display the full family of proximal averages; see Figure 3 for the 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 $f_1 = f_0 \circ (-\text{Id})$. The following reading list is a starting point from which 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, 1970.