HAL open science

Fenchel-Young inequality with a remainder and applications to convex duality and optimal transport

Guillaume Carlier

▶ To cite this version:

Guillaume Carlier. Fenchel-Young inequality with a remainder and applications to convex duality and optimal transport. 2022. hal-03614052v2

HAL Id: hal-03614052
https://hal.science/hal-03614052v2

Preprint submitted on 16 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fenchel-Young inequality with a remainder and applications to convex duality and optimal transport

Guillaume Carlier*

April 16, 2022

Abstract

This short note is devoted to some applications of a simple quantitative form of the Fenchel-Young inequality in Hilbert spaces both for convex functions and for Fitzpatrick functions of maximal monotone operators. Our initial motivation comes from a stability question in optimal transport. We derive from the quantitative form of the Fenchel-Young inequality a simple and constructive proof of the Brøndsted-Rockafellar theorem and a perturbed primal-dual attainment result in Hilbert spaces.

Keywords: Fenchel-Young inequality in quantitative form, Brøndsted-Rockafellar theorem, tilted convex duality, stability of optimal transport, Fitzpatrick function.

1 A quantitative Fenchel-Young inequality

In what follows, $(E, .)$ is a Hilbert space (identified with its dual), and $\Gamma_0(E)$ denotes the set of convex, lsc and proper (i.e. not identically $+\infty$) functions from E to $\mathbb{R} \cup \{+\infty\}$. Let $u \in \Gamma_0(E)$, u^* be its Legendre transform:

$$u^*(p) := \sup_{x \in E} \{p \cdot x - u(x)\}, \ \forall p \in E.$$

*CEREMADE, UMR CNRS 7534, Université Paris Dauphine, PSL, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE and INRIA-Paris, MOKAPLAN, carlier@ceremade.dauphine.fr

We then denote by $G_u : E \times E \to \mathbb{R} \cup \{+\infty\}$, the *gap* function:

$$G_u(x,p) := u(x) + u^*(p) - p \cdot x, \ \forall (x,p) \in E \times E.$$

By the very definition of u^*, G_u is nonnegative on $E \times E$ (Fenchel-Young inequality) and vanishes exactly on the graph of ∂u i.e. when $p \in \partial u(x)$ with

$$\partial u(x) := \{p \in E \ : \ u(y) \geq u(x) + p \cdot (y - x), \ \forall y \in E\}.$$

An interesting refined Fenchel-Young inequality involving the Fitzpatrick function [10] can be found in [1], we will discuss an extension to maximal monotone operators using the Fitzpatrick function in section 2. Let us also emphasize that, In [15], Santambrogio has developed an elegant and powerful duality argument for the regularity of solutions of convex variational problems in which quantitative versions of the Fenchel-Young inequality play a key role. Below, we prove a quantitative form involving the resolvent (or proximal operator). Let $(x,p) \in E \times E$ and $(x',p') \in E \times E$, Fenchel-Young inequality gives

$$u(x) \geq p' \cdot x - u^*(p'), \ u^*(p) \geq p \cdot x' - u(x')$$

summing these inequalities and rearranging the scalar product terms immediately gives

$$G_u(x,p) \geq -G_u(x',p') + (p' - p) \cdot (x - x'). \tag{1.1}$$

From this basic inequality and using a celebrated trick due to Minty [14], we obtain a quantitative form of the Fenchel-Young inequality:

Lemma 1.1 (Young's inequality with a remainder)**.** *Let $u \in \Gamma_0(E)$. For every $(x,p) \in E \times E$ and $\lambda > 0$, one has*

$$G_u(x,p) \geq \frac{1}{\lambda} \|x - (\mathrm{id} + \lambda \partial u)^{-1}(x + \lambda p)\|^2. \tag{1.2}$$

Proof. Define the resolvent $J_\lambda := (\mathrm{id} + \lambda \partial u)^{-1}$ and note that $p' \in \partial u(x')$ is equivalent to

$$x' = J_\lambda(x' + \lambda p').$$

Fix x and p in E. Define then $X := x + \lambda p$, $h := x - J_\lambda(X)$, and x' and p' by

$$x' = J_\lambda(X) = x - h, \ p' := \frac{X - x'}{\lambda} = p + \frac{h}{\lambda}$$

and observe that by construction $x' = J_\lambda(x' + \lambda p')$ so that $G_u(x', p') = 0$. Using (1.1), we then have

$$\begin{aligned} G_u(x,p) &\geq (p' - p) \cdot (x - x') \\ &= \frac{1}{\lambda}\|h\|^2 = \frac{1}{\lambda}\|x - (\mathrm{id} + \lambda \partial u)^{-1}(x + \lambda p)\|^2. \end{aligned}$$

□

Remark 1.2 (Equality case). Inequality (1.2) is an equality exactly when

$$J_\lambda(x + \lambda p) \in \partial u^*(p) \text{ and } (x + \lambda p) - J_\lambda(x + \lambda p) \in \lambda \partial u(x)$$

i.e. when there exists $q \in \partial u(x)$ such that both p and q belong to $\partial u(x + \lambda(p - q))$.

2 A variant for Fitzpatrick functions of maximal monotone operators

The goal of this paragraph is to present a variant of (1.2) for Fitzpatrick functions associated with a maximal monotone operator [10]. Recall that a set-valued operator A from the Hilbert space E to 2^E is monotone if $(p - p') \cdot (x - x') \geq 0$ whenever $p \in A(x)$ and $p' \in A(x')$; it is maximal monotone if, in addition, it has no strict extension that is still monotone. In this case, for every $\lambda > 0$, $\mathrm{id} + \lambda A$ is onto and the resolvent $(\mathrm{id} + \lambda A)^{-1}$ is a single-valued one-Lipschitz monotone map, see [4]. Slightly abusing notations, we shall identify A with its graph i.e. the subset of $E \times E$, $A = \{(x,p) \in E \times E \ : \ p \in A(x)\}$ and only consider the case where A has a nonempty domain i.e. there exists $x \in E$ such that $A(x) \neq \emptyset$. Given A, a maximal monotone operator with nonempty domain, the Fitzpatrick function of A, $F_A : E \times E \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$F_A(x,p) := p \cdot x - \inf_{(x',p') \in A} \{(p - p') \cdot (x - x')\}, \ \forall (x,p) \in E \times E. \tag{2.1}$$

Fitzpatrick proved the Fenchel-Young like inequality

$$F_A(x,p) \geq p \cdot x, \ \forall (x,p) \in E \times E, \tag{2.2}$$

and characterized A by

$$F_A(x,p) = p \cdot x \iff (x,p) \in A \iff (\forall \lambda > 0, x = (\mathrm{id} + \lambda A)^{-1}(x + \lambda p)).$$

Given $(x,p) \in E$ and $\lambda > 0$, let us mimick the proof of (1.2) by considering

$$x' := (\mathrm{id} + \lambda A)^{-1}(x + \lambda p),\ p' := \frac{x - x'}{\lambda} + p$$

then since $(x', p') \in A$, (2.1) gives

$$F_A(x,p) - p \cdot x \geq (p - p') \cdot (x' - x) = \frac{1}{\lambda}\|x - (\mathrm{id} + \lambda A)^{-1}(x + \lambda p)\|^2 \quad (2.3)$$

which is the announced quantitative version of (2.2).

3 Stability of optimal transport

Let us denote by $\mathscr{P}(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$ and $\mathscr{P}_2(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$ with finite second moment. Given μ and ν in $\mathscr{P}_2(\mathbb{R}^d)$, the squared Wasserstein distance between μ and ν is by definition

$$W_2^2(\mu,\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \mathrm{d}\gamma(x,y) \quad (3.1)$$

where $\Pi(\mu,\nu)$ is the set of transport plans between μ and ν i.e. the set of Borel probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having μ and ν as marginals. Thanks to the seminal results of Brenier [3] and McCann [12], we know that there exists a convex function u with the property that $\gamma \in \Pi(\mu,\nu)$ is optimal for (3.1) if and only if its support is included in the graph of ∂u, the subdifferential of u. A natural stability question is whether an almost optimal plan is (in a sense to be made precise) close to the graph of ∂u. This question, which is of partical importance for numerical and discretization purposes, has been addressed recently by Berman [2], Li and Nochetto [11], Delalande and Mérigot [9], also see [8] for convergence of entropic optimal transport. Under some conditions on the marginals μ and ν and their supports, Caffarelli's regularity theory for Monge-Ampère equations [6, 7] implies the regularity of u and in particular that Brenier's optimal transport map ∇u is Lipschitz. Under this assumption, Li and Nochetto proved the following.

Proposition 3.1 (Li and Nochetto [11])**.** *If Brenier's optimal transport map ∇u is M-Lipschitz, then for every $\gamma \in \Pi(\mu,\nu)$, one has*

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x,y) \geq W_2^2(\mu,\nu) + \frac{1}{M} \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - \nabla u(x)|^2 d\gamma(x,y). \quad (3.2)$$

In the general case where u is nonsmooth, we can easily deduce from (1.2) a surrogate based for the stability inequality (3.2) which holds for arbitrary marginals. The relevance of Minty's trick for optimal transport was first observed by McCann, Pass and Warren [13].

Proposition 3.2. *for every $\gamma \in \Pi(\mu,\nu)$, one has*

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\gamma(x,y) \geq W_2^2(\mu,\nu) + 2\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-(\mathrm{id}+\partial u)^{-1}(x+y)|^2 d\gamma(x,y).$$

Proof. Let u^* denote the Legendre transform of u and $\overline{\gamma} \in \Pi(\mu,\nu)$ be an optimal plan between μ and ν. Since γ and $\overline{\gamma}$ share the same marginals and since $u(x)+u^*(y)-x\cdot y = 0$ on the support of $\overline{\gamma}$, we have

$$\begin{aligned}\frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \mathrm{d}\gamma(x,y) - \frac{1}{2}W_2^2(\mu,\nu) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} [u(x)+u^*(y)-x\cdot y]\mathrm{d}\gamma(x,y)\\ &\geq \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-(\mathrm{id}+\partial u)^{-1}(x+y)|^2 \mathrm{d}\gamma(x,y)\end{aligned}$$

where the last line follows from Lemma 1.1. □

Remark 3.3. In case μ is absolutely continuous with respect to the Lebesgue measure, u is differentiable μ-a.e. and ∇u solves the Monge formulation of (3.1):

$$\inf\left\{\int_{\mathbb{R}^d} |x-T(x)|^2 \mathrm{d}\mu(x) \ : \ T_{\#}\mu = \nu\right\} \tag{3.3}$$

where $T_{\#}\mu$ is the pushforward of μ through T. In this setting, it is instructive to compare (3.2) which states that whenever $T_{\#}\mu = \nu$, one has

$$\|\,\mathrm{id}-T\|^2_{L^2(\mu)} - \|\,\mathrm{id}-\nabla u\|^2_{L^2(\mu)} \geq \frac{1}{M}\|T-\nabla u\|^2_{L^2(\mu)}$$

where M is the Lipschitz constant of Brenier's optimal transport map ∇u and the inequality from proposition 3.2 which reads as

$$\|\,\mathrm{id}-T\|^2_{L^2(\mu)} - \|\,\mathrm{id}-\nabla u\|^2_{L^2(\mu)} \geq 2\|\,\mathrm{id}-(\mathrm{id}+\partial u)^{-1}\circ(\mathrm{id}+T)\|^2_{L^2(\mu)}.$$

4 Connection with the Brøndsted-Rockafellar theorem

One can deduce from inequality (1.2) a short (and constructive but restricted to the Hilbertian case) proof of the Brøndsted-Rockafellar theorem [5]. Let

$u \in \Gamma_0(E)$, $x \in E$ and $\varepsilon > 0$, recall that the ε-subdifferential of u at x, $\partial_\varepsilon u(x)$ is by definition

$$\partial_\varepsilon u(x) := \{p \in E \ : \ G_u(x,p) \leq \varepsilon\}$$

which is a non empty closed and convex set as soon as x is in the domain of u.

Theorem 4.1. *Let $u \in \Gamma_0(E)$, $\varepsilon > 0$ and $(x,p) \in E \times E$ such that $p \in \partial_\varepsilon u(x)$ and $\lambda > 0$. Define*

$$x' := (\mathrm{id} + \lambda \partial u)^{-1}(x + \lambda p), \ p' := p + \frac{x - x'}{\lambda}$$

then x' and p' satisfy

$$\|x - x'\| \leq \sqrt{\lambda \varepsilon}, \ \|p - p'\| \leq \sqrt{\frac{\varepsilon}{\lambda}}, \ p' \in \partial u(x') \tag{4.1}$$

and

$$u(x') + \frac{\lambda}{2}\|p'\|^2 \leq u(x) + \frac{\lambda}{2}\|p\|^2. \tag{4.2}$$

Proof. Since $\varepsilon \geq G_u(x,p)$, (1.2) directly implies $\|x - x'\| \leq \sqrt{\lambda\varepsilon}$ and then

$$\|p - p'\| = \lambda^{-1}\|x - x'\| \leq \sqrt{\frac{\varepsilon}{\lambda}}.$$

By construction, we also have

$$x + \lambda p \in x' + \partial u(x') \text{ i.e. } p' = \frac{x - x'}{\lambda} + p \in \partial u(x')$$

and x' minimizes

$$y \in E \mapsto \lambda u(y) + \frac{1}{2}\|y - (x + \lambda p)\|^2 = \lambda u(y) + \frac{1}{2}\|y - (x' + \lambda p')\|^2$$

from which (4.2) follows. □

Observe that in the previous result, by construction, we have $x + \lambda p = x' + \lambda p'$.

5 Primal and dual attainment for tilted dual convex problems

Let E and F be two Hilbert spaces, $f \in \Gamma_0(E)$, $g \in \Gamma_0(F)$ and A be a bounded linear operator between E and F. Consider the convex minimization:

$$\inf_{x\in E}\Big\{f(x)+g(Ax)\Big\} \tag{5.1}$$

and its Fenchel-Rockafellar dual

$$\sup_{q\in F}\Big\{-f^*(A^*q)-g^*(-q)\Big\} \tag{5.2}$$

where A^* is the adjoint of A. Note that the duality gap

$$\overline{\delta} := \inf\,(5.1) - \sup\,(5.2) \in \mathbb{R}_+ \cup \{+\infty\}$$

between these two problems can also be written as

$$\overline{\delta} = \inf_{(x,q)\in E\times F}\Big\{G_f(x,A^*q)+G_g(Ax,-q)\Big\}$$

The duality gap $\overline{\delta}$ can be positive (and even infinite); even if $\overline{\delta}=0$, one cannot take for granted that (5.1) or (5.2) have solutions, primal or dual attainments require further assumptions in general. However, as we shall see below, when one *tilts* the data with linear perturbations of the order of $\sqrt{\overline{\delta}}$, the corresponding *tilted* primal and dual problems have solutions (and of course, no gap).

Theorem 5.1. *For every $\delta > \overline{\delta}$, there exists $(h,k) \in E\times F$ such that*

$$\|h\|^2+\|k\|^2 \le \delta \tag{5.3}$$

and, the tilted functions

$$f_{h,k}(x) := f(x-h)-(A^*k+h)\cdot x,\ \forall x\in E,\ g_k(y) := g(y-k),\ \forall y\in F,$$

satisfy

$$\min_{x\in E}\Big\{f_{h,k}(x)+g_k(Ax)\Big\} = \max_{q\in F}\Big\{-f^*_{h,k}(A^*q)-g^*_k(-q)\Big\} \tag{5.4}$$

(where we have written min and max on purpose to emphasize the fact that both are achieved).

Proof. If $\overline{\delta} = +\infty$, there is nothing to prove so we assume $\overline{\delta} \in \mathbb{R}_+$. Thanks to the definition of $\overline{\delta}$ and (1.2), we get

$$\delta > \inf_{(x,q)\in E\times F} \left\{ \|x - (\mathrm{id} + \partial f)^{-1}(x + A^*q)\|^2 + \|Ax - (\mathrm{id} + \partial g)^{-1}(Ax - q)\|^2 \right\}$$

so that there exists $(x, q) \in E \times F$ such that defining $h \in E$ and $k \in F$ by

$$h := x - (\mathrm{id} + \partial f)^{-1}(x + A^*q),\ k := Ax - (\mathrm{id} + \partial g)^{-1}(Ax - q), \tag{5.5}$$

the pair (h, k) satisfies (5.3). By the very definition of h and k we have

$$A^*(q-k) \in -h - A^*k + \partial f(x-h) = \partial f_{h,k}(x),\ -(q-k) \in \partial g(Ax-k) = \partial g_k(Ax)$$

which readily implies that:

- x minimizes $f_{h,k} + g_h \circ A$ over E,
- $q - k$ minimizes $f^*_{h,k} \circ A^* + g^*_k(-.)$ over F,
- $0 = f_{h,k}(x) + g_k(Ax) + f^*_{h,k}(A^*(q - k)) + g^*_k(-q + k)$.

This shows primal and dual attainment and the absence of duality gap for (5.4), the tilted version of (5.1)-(5.2). □

Now, let us consider the more general situation where $\Phi \in \Gamma_0(E \times F)$ (E and F are again Hilbert spaces), and we consider the gap between

$$\inf_{x\in E} \Phi(x, 0) \tag{5.6}$$

and its dual

$$\sup_{q\in F} -\Phi^*(0, q). \tag{5.7}$$

This gap is

$$\overline{\delta} = \inf_{(x,q)\in E\times F} G_\Phi((x, 0), (0, q)).$$

Of course, if $(x, q) \in E \times F$ are such that $(0, q) \in \partial\Phi(x, 0)$ (equivalently $(x, 0) \in \partial\Phi^*(q, 0)$) then x solves (5.6), q solves (5.7) and $\overline{\delta} = 0$. In the general case, using (1.2) again we deduce that, for any $\delta > \overline{\delta}$, there exists $(x, q) \in E \times F$ such that

$$\delta \geq \|(x, 0) - (\mathrm{id} + \partial\Phi)^{-1}(x, q)\|^2$$

so that defining

$$(h, k) := (x, 0) - (\mathrm{id} + \partial\Phi)^{-1}(x, q)$$

one has

$$\|(h,k)\| \leq \sqrt{\delta} \tag{5.8}$$

and

$$(0,q) \in (-h,-k) + \partial\Phi(x-h,-k).$$

Defining the tilted function

$$\Phi_{h,k}(u,v) := \Phi(u-h,v-k) - h\cdot u - k\cdot v,\ \forall (u,v) \in E\times F$$

we thus have

$$(0,q) \in \partial\Phi_{h,k}(x,0)$$

so that

$$\Phi_{h,k}(x,0) + \Phi^*_{h,k}(0,q) = 0,$$

and that the tilted versions of (5.6) and (5.7) where Φ is replaced by $\Phi_{h,k}$ both admit solutions.

Acknowledgments: G.C. acknowledges the support of the Lagrange Mathematics and Computing Research Center and is grateful to Heinz Bauschke for his kind suggestions and comments.

References

[1] Heinz H. Bauschke, D. Alexander McLaren, and Hristo S. Sendov. Fitzpatrick functions: inequalities, examples, and remarks on a problem by S. Fitzpatrick. *J. Convex Anal.*, 13(3-4):499–523, 2006.

[2] Robert J. Berman. Convergence rates for discretized Monge-Ampère equations and quantitative stability of optimal transport. *Found. Comput. Math.*, 21(4):1099–1140, 2021.

[3] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[4] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* North-Holland Mathematics Studies, No. 5. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1973.

[5] A. Brøndsted and R. T. Rockafellar. On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.*, 16:605–611, 1965.

[6] Luis A. Caffarelli. Boundary regularity of maps with convex potentials. II. *Ann. of Math. (2)*, 144(3):453–496, 1996.

[7] Luis A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):547–563, 2000.

[8] G Carlier, P Pegon, and L. Tamanini. Convergence rate of general entropic optimal transport costs. in preparation, 2022.

[9] A. Delalande and Q. Mérigot. Quantitative stability of optimal transport maps under variations of the target measure. arXiv:2103.05934, 2021.

[10] Simon Fitzpatrick. Representing monotone operators by convex functions. In *Workshop/Miniconference on Functional Analysis and Optimization (Canberra, 1988)*, volume 20 of *Proc. Centre Math. Anal. Austral. Nat. Univ.*, pages 59–65. Austral. Nat. Univ., Canberra, 1988.

[11] Wenbo Li and Ricardo H. Nochetto. Quantitative stability and error estimates for optimal transport plans. *IMA J. Numer. Anal.*, 41(3):1941–1965, 2021.

[12] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

[13] Robert J. McCann, Brendan Pass, and Micah Warren. Rectifiability of optimal transportation plans. *Canad. J. Math.*, 64(4):924–934, 2012.

[14] George J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, 29:341–346, 1962.

[15] Filippo Santambrogio. Regularity via duality in calculus of variations and degenerate elliptic pdes. *Journal of Mathematical Analysis and Applications*, 457(2):1649–1674, 2018. Special Issue on Convex Analysis and Optimization: New Trends in Theory and Applications.