# Remarks on Toland's duality, convexity constraint and optimal transport

G. Carlier *

October 13, 2007

*This article is dedicated to Michel Théra on the occasion of his 60th birthday.*

**Abstract**

We show that minimizing the difference of squared Wasserstein distances to two reference probability measures in a suitable set of probability measures is equivalent to a linear programming problem posed on set of convex functions (problem which has its own interest and motivations). This is naturally related to Toland's duality for the minimization of the difference of convex (DC for short) functions. We therefore end the paper by some remarks on DC problems with a convex (or concave) dual in the sense of Toland.

**Keywords :** Toland's duality, convexity constraint, optimal transportation, DC minimization.

It is a great pleasure and a real honor to contribute to this series of papers dedicated to Michel on the occasion of his 60th birthday. Michel's numerous contributions to nonlinear and variational analysis are of course very influential, the energy and enthusiasm Michel devoted to promote the french applied mathematics community deserve sincere and warm gratitude. Last but not least, I have always been impressed by (and have often benefited from) Michel's kindness to young researchers at the beginning of their career, by his wise advises and his help to facilitate interactions between mathematicians of different horizons. For all this: "merci Michel et bon anniversaire"!

* Université Paris Dauphine, CEREMADE, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE carlier@ceremade.dauphine.fr.

# 1 Introduction

Toland's duality (see [7], [8]) concerns minimization problems of the form:

$$(\mathcal{Q}) \inf_{x \in V} \{F(x) - G(x)\}$$

where $V$ is a real normed vector space and $F$ and $G$ are two convex lower semi-continuous proper (i.e. not identically $+\infty$) functions on $V$ with values in $\mathbb{R} \cup \{+\infty\}$. Such problems where the objective is the difference of two convex functions (henceforth DC) has received a lot of attention in the literature and Toland's duality states that the value of the DC problem $(\mathcal{Q})$ coincides with that of its dual:

$$(\mathcal{Q}^*) \inf_{p^* \in V^*} \{G^*(p^*) - F^*(p^*)\}$$

where $V^*$ stands for the topological dual of $V$, $G^*$ and $F^*$ denote the Fenchel-Legendre transform of $F$ and $G$:

$$F^*(p) := \sup_{x \in V}\{p(x) - F(x)\}, \; G^*(p) := \sup_{x \in V}\{p(x) - G(x)\}, \; \forall p \in V^*.$$

Toland's duality also enables one, under appropriate subdifferentiability conditions, to construct solutions of $(\mathcal{Q})$ from solutions of $(\mathcal{Q}^*)$ and vice versa. This duality for DC minimization, as convex duality, is particularly useful when $(\mathcal{Q}^*)$ is simpler to study than $(\mathcal{Q})$. It is not surprising that Toland's duality as well as other forms of nonconvex duality have proved to be efficient tools for a variety of applications, in particular in mechanics (the initial work of Toland was motivated by the motion of a heavy rotating chain). It is not our purpose here to give a complete list of references on nonconvex duality, we rather refer the interested reader to [4], [5], [7], [8] and the references therein. Let us remark that, at least formally, the Euler-Lagrange equation of $(\mathcal{Q})$ reads as

$$0 \in \partial F(x) - \partial G(x)$$

and that of $(\mathcal{Q}^*)$ reads as

$$0 \in \partial G^*(p) - \partial F^*(p).$$

In [1], Michel and Hedy Attouch developed a general duality principle for equations of the form $0 \in Ax - Bx$ for general operators $A$ and $B$ which enabled them to recover as a particular case Toland's duality.

The present article is concerned with special DC problems for which the dual (or predual) is convex (or equivalent to a convex problem). Although this is a very particular situation, we give various applications through the paper (related to optimal transport, convexity constraint, balayages of measures, Moreau-Yosida approximation). In particular, we study in details a particular case of such a situation in section 3 in which we relate precisely the problem of minimizing the difference of squared Wasserstein distances to a simple linear programming problem posed on set of convex functions (motivation for this linear programming problem being given in section 2). In section 4, we make some remarks on DC problems with a convex dual.

## 2 Convexity constraint

Variational problems subjet to a convexity constraint arise in various areas (Newton's problem of least resistance [3], economics [9]...). In general, such problems are difficult to study, we will therefore restrict here to the case of a cost functional that "does not depend on $\nabla u$". Let us consider then as model problem

$$\inf_{u\in\mathcal{A}} \int_{\Omega} F(x, u(x))dx \tag{1}$$

where $\Omega$ is some bounded open convex subset of $\mathbb{R}^d$ and given $K$, some compact subset of $\mathbb{R}^d$, $\mathcal{A}$ is the set of convex functions defined by:

$$\mathcal{A} := \{u \;:\; \Omega \to \mathbb{R}, \;\; u \text{ convex and } \nabla u \in K \text{ a.e. in } \Omega\}. \tag{2}$$

Firstly, let us remark that $u \in \mathcal{A}$ if and only if it can be written in the form $u = v_K^*$ for some $v \in C^0(K, \mathbb{R})$ with

$$v_K^*(x) := \sup_{y\in K}\{y \cdot x - v(y)\}. \tag{3}$$

Secondly, let us note that as soon as $F$ is continuous (we do not look for minimal assumptions here) and satisfies a coercivity assumption of the form

$$F(x, t) \geq f(t) \text{ with } f(t) \to \infty \text{ as } |t| \to \infty$$

then (1) possesses at least a solution (this follows at once from the compactness of $K$, Ascoli's theorem and the fact that $\mathcal{A}$ is closed in the $C^0$ topology).

Finally, if we further assume that $F$ is differentiable with respect to its second argument and that $\partial_u F$ is continuous (say), then any solution $\overline{u}$ of (1) satisfies the variational inequality

$$\int_{\Omega} \partial_u F(x, \overline{u}(x))(v(x) - \overline{u}(x))dx \geq 0, \; \forall v \in \mathcal{A}. \tag{4}$$

Taking $v = \overline{u} \pm 1$ in the previous we immediately find that $\partial_u F(., u(.))$ has integral 0, we then define

$$\mu_+ := (\partial_u F(., \overline{u}(.)))_+ = \max(0, \partial_u F(., \overline{u}(.))),$$
$$\mu_- := (\partial_u F(., \overline{u}(.)))_- = \max(0, -\partial_u F(., \overline{u}(.))$$

so that $\mu_+$ and $\mu_-$ are nonnegative functions (continuous here) with same total mass. Now the variational inequality (4) can be expressed as

$$\int_\Omega \overline{u}(\mu_+ - \mu_-) = \inf_{u \in \mathcal{A}} \int_\Omega u(\mu_+ - \mu_-). \tag{5}$$

# 3 Linear programming with a convexity constraint and optimal transport

The aim of this section is to study the linear programming problem arising from the variational inequality (5)

$$\inf_{u \in \mathcal{A}} \int_\Omega u \; d(\mu_+ - \mu_-). \tag{6}$$

where $\mu_+$ and $\mu_-$ are nonnegative measures on $\Omega$ of same total mass and which we assume absolutely continuous with respect to the Lebesgue measure. Using the representation of elements of $\mathcal{A}$ in the form (3), we may rewrite in an equivalent way (6) as the (nonlinear and not even convex but unconstrained) problem:

$$(\mathcal{P}) \inf_{v \in C^0(K, \mathbb{R})} \int_\Omega v_K^* \; d(\mu_+ - \mu_-).$$

Before, we go further, we need the following definition:

**Definition 1** *Let $\mu$ be a nonnegative finite Borel measure on $\Omega$ and $\sigma$ be a Borel map $\Omega \to \mathbb{R}^d$, the push forward of $\mu$ through $\sigma$ is the finite Borel measure on $\mathbb{R}^d$ denoted $\sigma \# \mu$ defined by*

$$(\sigma \# \mu)(A) = \mu(\sigma^{-1}(A)), \quad \textit{for every } A \textit{ Borel subset of } \mathbb{R}^d.$$

Note that $\sigma \# \mu$ can also be defined by

$$\int_{\mathbb{R}^d} h(y) d(\sigma \# \mu)(y) = \int_\Omega h(\sigma(x)) d\mu(x), \forall \; h \text{ continuous and bounded.}$$

A first optimality condition for (6) is given by

**Proposition 1** *If $\overline{u}$ is a solution of (6) then $\nabla\overline{u}\#\mu_+ = \nabla\overline{u}\#\mu_-$.*

**Proof.** Let us write $\overline{u}$ in the form $\overline{u} = v_K^*$ for some $v \in C^0(K, \mathbb{R})$. Let $h \in C^0(K, \mathbb{R})$ for $\varepsilon > 0$, define $u_\varepsilon := (v + \varepsilon h)_K^*$. By construction, $u_\varepsilon \in \mathcal{A}$, and a result of Gangbo [6] implies that for every point of differentiability $x$ of $\overline{u}$ one has

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\Big(u_\varepsilon(x) - \overline{u}(x)\Big) = -h(\nabla\overline{u}(x)). \tag{7}$$

Since $\mu_\pm$ are absolutely continuous with respect to the Lebesgue measure and $\overline{u}$ is convex, $\overline{u}$ is differentiable $\mu_\pm$ a.e. so that (7) holds $\mu_\pm$ a.e.. Now, since $\overline{u}$ solves (6), we have for every $\varepsilon > 0$

$$\frac{1}{\varepsilon}\int_\Omega (u_\varepsilon - \overline{u})(\mu_+ - \mu_-) \geq 0.$$

Using Lebesgue's dominated convergence theorem, (7), and passing to the limit in the previous inequality, we obtain

$$\int_\Omega h(\nabla\overline{u})\, d(\mu_+ - \mu_-) \leq 0.$$

Since $h \in C^0(K, \mathbb{R})$ is arbitrary, the previous inequality is in fact an equality, which yields $\nabla\overline{u}\#\mu_+ = \nabla\overline{u}\#\mu_-$.

$\square$

The condition $\nabla u\#\mu_+ = \nabla u\#\mu_-$ is not very informative (it is satisfied for every affine $u$) and is certainly not a sufficient condition for (6) (it is also satisfied by the maximizers of $\int_\Omega u(\mu_+ - \mu_-)$ over $\mathcal{A}$ or more generally by any $u = v_K^*$ with $v$ a critical point of $v \mapsto \int_\Omega v_K^*(\mu_+ - \mu_-)$). At this point, it is natural to expect a variational characterization of the measure $\overline{\nu} := \nabla\overline{u}\#\mu_+ = \nabla\overline{u}\#\mu_-$. To get this variational characterization, we need to recall some results on optimal transportation.

Given two nonnegative finite Borel measures on $\mathbb{R}^d$, $\mu$ and $\nu$ having finite second moments and the same total mass, the squared Wasserstein distance between $\mu$ and $\nu$ is by definition:

$$W_2^2(\mu, \nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) \tag{8}$$

where $\Pi(\mu, \nu)$ is the set of transport plans between $\mu$ and $\nu$ i.e. the set of nonnegative finite Borel measures on $\mathbb{R}^d \times \mathbb{R}^d$ having $\mu$ and $\nu$ as marginals.

In other words $\gamma \in \Pi(\mu, \nu)$ if for every continuous and bounded test functions $g$ and $h$ one has:

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \Big(g(x) + h(y)\Big) d\gamma(x, y) = \int_{\mathbb{R}^d} g(x) d\mu(x) + \int_{\mathbb{R}^d} h(y) d\nu(y).$$

The optimal transportation problem defining the Wasserstein distance in (8) was solved by Brenier (see [2]) who proved that provided $\mu$ is absolutely continuous with respect to the Lebesgue measure, the infimum in (8) is uniquely attained by a transport plan of the form $\gamma = (\mathrm{id}, \nabla u)\#\mu$ with $u$ convex. Conversely, if $\nu = \nabla u \# \mu$ with $u$ convex then $(\mathrm{id}, \nabla u)\#\mu$ is the solution of (8). In particular, if, $\overline{u}$ solves (6), by proposition 1, we have $\nabla \overline{u} \# \mu_+ = \nabla \overline{u} \# \mu_- =: \overline{\nu}$, then we get

$$\begin{aligned}
&\frac{1}{2} W_2^2(\mu_\pm, \overline{\nu}) = \frac{1}{2} \int_\Omega |x - \nabla \overline{u}(x)|^2 \, d\mu_\pm(x) \\
&= \frac{1}{2} \int_\Omega |x|^2 \, d\mu_\pm(x) + \frac{1}{2} \int_\Omega |\nabla \overline{u}(x)|^2 \, d\mu_\pm(x) - \int_\Omega x \cdot \nabla \overline{u}(x) d\mu_\pm(x) \\
&= \frac{1}{2} \int_\Omega |x|^2 \, d\mu_\pm(x) + \frac{1}{2} \int_K |y|^2 \, d\overline{\nu}(y) - \int_\Omega \Big(\overline{u}(x) + \overline{u}^*(\nabla \overline{u}(x))\Big) d\mu_\pm(x)
\end{aligned}$$

($\overline{u}^*$ denotes the Legendre-Fenchel transform of $\overline{u}$ and we have used above the identity $x \cdot \nabla \overline{u}(x) = \overline{u}(x) + \overline{u}^*(\nabla \overline{u}(x))$ $\mu_\pm$-a.e.). We then get

$$\frac{1}{2} W_2^2(\mu_\pm, \overline{\nu}) = \frac{1}{2} \int_\Omega |x|^2 \, d\mu_\pm + \frac{1}{2} \int_K |y|^2 \, d\overline{\nu} - \int_\Omega \overline{u} \, d\mu_\pm - \int_K \overline{u}^* \, d\overline{\nu}(y). \quad (9)$$

This yields:

$$\int_\Omega \overline{u} \, d(\mu_+ - \mu_-) = \frac{1}{2} W_2^2(\mu_-, \overline{\nu}) - \frac{1}{2} W_2^2(\mu_+, \overline{\nu}) + \frac{1}{2} \int_\Omega |x|^2 \, d(\mu_+ - \mu_-)(x). \quad (10)$$

Denoting $\mathcal{M}_+(K)$ the set of nonnegative Borel measures on $K$ with total mass $\mu_+(\Omega) = \mu_-(\Omega)$, let us consider the variational problem

$$(\mathcal{P}^*) \inf_{\nu \in \mathcal{M}_+(K)} \left\{ \frac{1}{2} W_2^2(\mu_-, \nu) - \frac{1}{2} W_2^2(\mu_+, \nu) \right\}.$$

We thus deduce from (10)

$$\inf(\mathcal{P}) = \int_\Omega \overline{u} \, d(\mu_+ - \mu_-) \geq \inf(\mathcal{P}^*) + \frac{1}{2} \int_\Omega |x|^2 \, d(\mu_+ - \mu_-)(x). \quad (11)$$

We are now in position to prove the main result of this section.

**Theorem 1**

$$\inf(\mathcal{P}) = \inf(\mathcal{P}^*) + \frac{1}{2}\int_\Omega |x|^2 \, d(\mu_+ - \mu_-)(x). \tag{12}$$

*Moreover,*

1. *$\overline{u} \in \mathcal{A}$ solves (6) if and only if $\nabla\overline{u}\#\mu_+ = \nabla\overline{u}\#\mu_- = \overline{\nu}$ where $\overline{\nu} \in \mathcal{M}_+(K)$ solves $(\mathcal{P}^*)$,*

2. *$\overline{\nu} \in \mathcal{M}_+(K)$ solves $(\mathcal{P}^*)$ if and only if $\overline{\nu} = \nabla\overline{u}\#\mu_+ = \nabla\overline{u}\#\mu_-$ where $\overline{u} \in \mathcal{A}$ solves (6).*

**Proof.** Let $\nu \in \mathcal{M}_+(K)$, by Brenier's results recalled above, there exists $u_+ \in \mathcal{A}$ such that $\nu = \nabla u_+\#\mu_+$ (and $(\mathrm{id}, \nabla u_+)\#\mu_+)$ is an optimal plan between $\mu_+$ and $\nu$), by the same computation as in (9), we have

$$\frac{1}{2}W_2^2(\mu_+, \nu) = \frac{1}{2}\int_\Omega |x|^2 \, d\mu_+ + \frac{1}{2}\int_K |y|^2 \, d\nu - \int_\Omega u_+ \, d\mu_+ - \int_K u_+^* \, d\nu(y). \tag{13}$$

Now Young's inequality yields

$$\frac{1}{2}|x|^2 + \frac{1}{2}|y|^2 - u_+(x) - u_+^*(y) \leq \frac{1}{2}|x-y|^2 \ \forall (x,y) \in \Omega \times K,$$

so that for every $\gamma \in \Pi(\mu_-, \nu)$

$$\begin{aligned}\int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}|x-y|^2 d\gamma(x,y) \geq \frac{1}{2}\int_\Omega |x|^2 \, d\mu_- + \frac{1}{2}\int_K |y|^2 \, d\nu \\ - \int_\Omega u_+ \, d\mu_- - \int_K u_+^* \, d\nu(y)\end{aligned}$$

which yields

$$\frac{1}{2}W_2^2(\mu_-, \nu) \geq \frac{1}{2}\int_\Omega |x|^2 \, d\mu_- + \frac{1}{2}\int_K |y|^2 \, d\nu - \int_\Omega u_+ \, d\mu_- - \int_K u_+^* \, d\nu(y). \tag{14}$$

With (13), we then have:

$$\begin{aligned}\forall \nu \in \mathcal{M}_+(K),\ \frac{1}{2}W_2^2(\mu_-, \nu) - \frac{1}{2}W_2^2(\mu_+, \nu) + \frac{1}{2}\int_\Omega |x|^2 \, d(\mu_+ - \mu_-) \\ \geq \int_\Omega u_+ d(\mu_+ - \mu_-) \geq \inf(\mathcal{P})\end{aligned}$$

we then deduce (12) from the previous inequality and (11). Assertion 1. and the "if part" of assertion 2. then easily follow from proposition 1, (10) and

(12). Finally, if $\overline{\nu}$ solves $(\mathcal{P}^*)$, as previously, we write $\overline{\nu} = \nabla u_+ \# \mu_+$ with $u_+ \in \mathcal{A}$, with (12) and the same inequality as above, we get:

$$\inf(\mathcal{P}) = \frac{1}{2}W_2^2(\mu_-, \overline{\nu}) - \frac{1}{2}W_2^2(\mu_+, \overline{\nu}) + \frac{1}{2}\int_\Omega |x|^2 \, d(\mu_+ - \mu_-) \geq \int_\Omega u_+ d(\mu_+ - \mu_-)$$

so that $u_+$ solves (6) and then, by proposition 1, $\overline{\nu} = \nabla u_+ \# \mu_-$, which completes the proof.

$\square$

In proposition 1 and then in theorem 1, we have assumed that $\mu_+$ and $\mu_-$ are absolutely continuous with respect to the Lebesgue measure and proved, under this assumption, the duality relation

$$2 \inf_{u \in \mathcal{A}} \int_\Omega u d(\mu_+ - \mu_-) = \inf_{\nu \in \mathcal{M}_+(K)} \left\{ W_2^2(\mu_-, \nu) - W_2^2(\mu_+, \nu) \right\} + \int_\Omega |x|^2 (\mu_+ - \mu_-). \tag{15}$$

We now claim that (15) in fact holds for general nonnegative Borel measures on $\overline{\Omega}$ with same finite total mass $\mu_+$ and $\mu_-$. To see this, it is enough to proceed by approximation (in the weak-$\star$ topology of $\mathcal{M}(\overline{\Omega})$) and to show that both sides of (15) are continuous functions of $(\mu_+, \mu_-)$ in the the weak-$\star$ topology. For the right hand side, it easily follows from the triangle inequality and the fact that $W_2$ is a metric that metrizes the weak $\star$ topology on nonnegative measures with a fixed total mass on $\overline{\Omega}$ (see [10]). As for the left-hand side, it is easy to show that it is $2\text{diam}(K)$-Lipschitz in the product Wasserstein $W_1$ metric (which also metrizes the weak-$\star$ topology on nonnegative measures with a fixed total mass on $\overline{\Omega}$, [10]) defined by:

$$W_1(\mu, \mu') = \sup \left\{ \int_{\overline{\Omega}} u d(\mu - \mu'), \quad u \text{ 1-Lipschitz on } \overline{\Omega} \right\}.$$

We end this section by an application of the previous duality to the notion of balayage that we now recall.

**Definition 2** *Let $\mu_+$ and $\mu_-$ be two nonnegative finite Borel measures on (the convex compact set) $\overline{\Omega}$ with same total mass, $\mu_+$ is said to be a balayée of $\mu_-$ if*

$$\int_{\overline{\Omega}} u d\mu_+ \geq \int_{\overline{\Omega}} u d\mu_-$$

*for every convex function $u$ continuous on $\overline{\Omega}$.*

By standard approximation arguments, it is easy to see that in the previous definition one can also require the convex test-functions $u$ to be Lipschitz. So that another way to express that $\mu_+$ is a balayée of $\mu_-$ is that the value of the linear programming problem (6) is 0 for every compact set $K$ containing the origin (closed ball centered at the origin with an arbitrarily large radius, say). We deduce from this remark and the duality relation (15) the following characterization of balayages.

**Theorem 2** *Let $\mu_+$ and $\mu_-$ be two nonnegative finite Borel measures on $\overline{\Omega}$ with same total mass $\alpha$, then $\mu_+$ is a balayée of $\mu_-$ if and only if:*

$$W_2^2(\mu_+, \nu) \leq W_2^2(\mu_-, \nu) + \int_{\overline{\Omega}} |x|^2 d(\mu_+ - \mu_-)$$

*for every compactly supported Borel measure $\nu$ with total mass $\alpha$.*

We end this section by remarking that the duality of theorem 1 is nothing but an application of the Toland's duality for DC (difference of convex functions) problems. Indeed, defining

$$F_\pm(v) := \int_\Omega v_K^* d\mu_\pm, \ \forall v \in C^0(K, \mathbb{R})$$

it is easy to check that $F_+$ and $F_-$ are convex continuous functionals on $C^0(K, \mathbb{R})$. Let us denote by $\mathcal{M}(K)$ the dual of $C^0(K, \mathbb{R})$ and recall that $\mathcal{M}_+(K)$ is the subset of $\mathcal{M}(K)$ consisting of nonnegative measures having total mass $\mu_\pm(\Omega)$. Firstly, it is easy to check that if $\nu \in \mathcal{M}(K)$ and $\nu \notin \mathcal{M}_+(K)$ then

$$F_\pm^*(-\nu) = - \inf_{v \in C^0(K,\mathbb{R})} \left\{ \int_K v d\nu + \int_\Omega v_K^* d\mu_\pm \right\} = +\infty.$$

Secondly, if $\nu \in \mathcal{M}_+(K)$, the Kantorovich duality formula (see for instance [10] for details) can be written as

$$F_\pm^*(-\nu) = \frac{1}{2} W_2^2(\mu_\pm, \nu) - \frac{1}{2} \int_\Omega |x|^2 d\mu_\pm(x) - \frac{1}{2} \int_K |y|^2 d\nu(y).$$

So that finally, using Toland's duality ([7], [8]), we recover (12)

$$\begin{aligned} \inf(\mathcal{P}) &= \inf_{v \in C^0(K,\mathbb{R})} \{F_+(v) - F_-(v)\} = \inf_{\nu \in \mathcal{M}_+(K)} \left\{ F_-^*(-\nu) - F_+^*(-\nu) \right\} \\ &= \inf(\mathcal{P}^*) + \frac{1}{2} \int_\Omega |x|^2 \, d(\mu_+ - \mu_-)(x). \end{aligned}$$

# 4 DC problems with a convex dual

In the example of the previous section, we saw that the nonconvex problem $(\mathcal{P}^*)$ is dual in the sense of Toland to $(\mathcal{P})$ which, in turn (making the change of variable $u = v_K^*$), is equivalent to the linear programming (6). This enabled us to characterize all minimizers of $(\mathcal{P}^*)$ in terms of solutions of (6). In this final and independent section, we make some remarks on DC problems for which the dual problem has special convexity properties (hence is, in principle, easier to solve) and discuss possible applications of this (very particular) situation.

Given $V$, a normed vector space, $V^*$ its topological dual, $\Gamma_0(V)$ denotes the set of convex lower semi continuous proper functions on $V$ with values in $\mathbb{R}\cup\{+\infty\}$. For $F$ and $G$ in to $\Gamma_0(V)$, $F^*$ and $G^*$ denote the Fenchel-Legendre transforms of $F$ and $G$ respectively, that is $F^*$ and $G^*$ are the elements of $\Gamma_0(V^*)$ defined by:

$$F^*(p) := \sup_{x\in V}\{p(x) - F(x)\},\ G^*(p) := \sup_{x\in V}\{p(x) - G(x)\},\ \forall p \in V^*.$$

Toland's duality ([7], [8]) relates the primal DC problem:

$$(\mathcal{Q})\ \inf_{x\in V}\{F(x) - G(x)\}$$

to its dual defined by

$$(\mathcal{Q}^*)\ \inf_{p\in V^*}\{G^*(p) - F^*(p)\}.$$

One has to be cautious in the previous problems to handle the case where both $F$ and $G$ (or $G^*$ and $F^*$) take the value $+\infty$ at the same point, and the natural way to cope with this case is to adopt the convention $+\infty - (+\infty) = +\infty$. With this convention in mind, $(\mathcal{Q})$ and $(\mathcal{Q}^*)$ are then understood as

$$(\mathcal{Q})\ \inf_{\mathrm{Dom}(F)}\{F - G\} \quad \text{and } (\mathcal{Q}^*)\ \inf_{\mathrm{Dom}(G^*)}\{G^* - F^*\}.$$

Where as usual, Dom denotes the domain of the function i.e. the set where it is not $+\infty$. The basic duality results concerning $(\mathcal{Q})$ and $(\mathcal{Q}^*)$ can be summarized as follows:

- $\inf(\mathcal{Q}) = \inf(\mathcal{Q}^*)$,
- if $\overline{x}$ solves $(\mathcal{Q})$ and if $G$ is subdifferentiable at $\overline{x}$ then any $\overline{p} \in \partial G(\overline{x})$ solves $(\mathcal{Q}^*)$,

- if $\overline{p}$ solves $(\mathcal{Q}^*)$ and if $F^*$ is subdifferentiable at $\overline{p}$ then any $\overline{x} \in \partial F^*(\overline{p})$ solves $(\mathcal{Q})$.

Now, we are interested in the particular case where the objective function in the dual problem is itself convex i.e.:

$$G^* - F^* = H^*, \text{ for some } H \in \Gamma_0(V)$$

this yields, since $G = G^{**}$,

$$G = (F^* + H^*)^*$$

and it is well known that $(F^* + H^*) = (F \,\Box\, H)^*$ where $F \,\Box\, H$ is the infimal convolution of $F$ and $H$ defined by

$$F \,\Box\, H(x) = \inf_{y \in V} \{F(x-y) + H(y)\}.$$

We thus deduce from the general duality relation $\inf(\mathcal{Q}) = \inf(\mathcal{Q}^*)$ recalled above, the following relation (for arbitrary $F$ and $H$ in $\Gamma_0(V)$)

$$\inf_{\mathrm{Dom}(F)} \{F - (F \,\Box\, H)^{**}\} = \inf_{\mathrm{Dom}(F^*) \cap \mathrm{Dom} H^*} H^* = \inf_{\mathrm{Dom}(F^*)} H^*. \tag{16}$$

If, in addition, $\mathrm{Dom}(F) = V$, $\mathrm{Dom}(F^*) = V^*$, $\overline{p}$ minimizes $H^*$ (i.e. $\overline{p} \in \partial H(0)$) and $\overline{x} \in \partial F^*(\overline{p})$ then $\overline{x}$ minimizes $F - (F \,\Box\, H)^{**}$ over $V$.

Another particular case is the one where the objective function in the dual problem is concave i.e.:

$$F^* - G^* = H^*, \text{ for some } H \in \Gamma_0(V),$$

as before, this yields $F = (G \,\Box\, H)^{**}$. In this case, by Toland's duality, we obtain

$$\inf_{\mathrm{Dom}(F)} \{(G \,\Box\, H)^{**} - G\} = \inf_{\mathrm{Dom}(G^*)} -H^* \tag{17}$$

or, equivalently

$$\sup_{\mathrm{Dom}(G \,\Box\, H)^{**}} \{G - (G \,\Box\, H)^{**}\} = \sup_{\mathrm{Dom}(G^*)} H^*. \tag{18}$$

We thus have proved the following

**Proposition 2** *Let $F$ and $H$ be in $\Gamma_0(V)$, one has*

$$\inf_{\mathrm{Dom}(F)} \{F - (F \,\Box\, H)^{**}\} = \inf_{\mathrm{Dom}(F^*)} H^*$$

*and*

$$\sup_{\mathrm{Dom}(F \,\Box\, H)^{**}} \{F - (F \,\Box\, H)^{**}\} = \sup_{\mathrm{Dom}(F^*)} H^*.$$

Let us illustrate the previous result in the particular case where $V$ is a Hilbert space (identified with its dual), $F \in \Gamma_0(V)$ and $H = H_\lambda = \frac{1}{2\lambda}\|.\|^2$ for some $\lambda > 0$. In this case, $F_\lambda := F \,\square\, H_\lambda$ is the Moreau-Yosida approximation of $F$. Since $\mathrm{Dom}(F_\lambda) = V$ and $H_\lambda^* = \frac{\lambda}{2}\|.\|^2$, the previous result gives

$$\inf_V \{F - F_\lambda\} = \frac{\lambda}{2} \inf_{p \in \mathrm{Dom}(F^*)} \|p\|^2$$

and

$$\sup_V \{F - F_\lambda\} = \frac{\lambda}{2} \sup_{p \in \mathrm{Dom}(F^*)} \|p\|^2.$$

In particular, $F - F_\lambda$ is bounded from above if and only if $\mathrm{Dom}(F^*)$ is bounded i.e. $F$ is Lipschitz.

# References


[1] H. Attouch and M. Théra, A general duality principle for the sum of two operators, *J. Convex Anal.,* vol. 3, (1996).

[2] Y. Brenier, *Polar Factorization and monotone rearrangements of vector valued functions*, Communications in Pure and Applied Mathematics, vol. 44 (1991).

[3] G. Buttazzo, V. Ferone and B. Kawohl, *Minimum problems over sets of concave functions and related questions*, Math. Nachr. , vol. 173 (1995).

[4] I. Ekeland, *Duality in non-convex optimization and calculus of variations*, SIAM J. Opt. and Control, vol. 15 (1977).

[5] I. Ekeland, *Dual variational methods in non-convex optimization and differential equations*, in Mathematical Theories of Optimization, Cecconi and Zolezzi ed., Springer Lecture Notes in Mathematics 979 (1983).

[6] W. Gangbo, *An elementary proof of the polar factorization of vector-valued functions*, Arch. Rational Mech. Anal., vol. 128 (1994).

[7] J. F. Toland, *Duality in Nonconvex Optimization*, J. Math. Anal. Appl., vol. 66, (1978).

[8] J. F. Toland, *A duality principle for non-convex optimisation and the calculus of variations*, Arch. Rational Mech. Anal., vol. 71 (1979).

[9] J.-C. Rochet and P. Choné, *Ironing, sweeping and multidimensional screening*, Econometrica, vol. 66 (1998).

[10] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58, American Mathematical Society, Providence, RI, (2003).