A General Nonconvex Multiduality Principle

Francesca Bonenti ·

Juan Enrique Martínez-Legaz ·

Rossana Riccardi

Communicated by Radu Ioan Boţ

Abstract We introduce a (possibly infinite) collection of mutually dual non-

convex optimization problems, which share a common optimal value, and give a

Francesca Bonenti

Open Capital Partners SGR, Milano, Italy

francesca.bonenti@opencapital.it

Juan Enrique Martínez-Legaz, Corresponding author

Departament d'Economia i d'Història Econòmica

Universitat Autònoma de Barcelona, Spain

and

BGSMath, Barcelona, Spain

JuanEnrique.Martinez.Legaz@uab.cat

Rossana Riccardi

Dipartimento di Economia e Management

Università degli Studi di Brescia, Italy

rossana.riccardi@unibs.it

characterization of their global optimal solutions. As immediate consequences of our general multiduality principle, we obtain Toland - Singer duality theorem as well as an analogous result involving generalized perspective functions. Based on our duality theory, we propose an extension of an existing algorithm for the minimization of d.c. functions, which exploits Toland - Singer duality, to a more general class of nonconvex optimization problems.

Keywords Nonconvex optimization · Multiduality · Toland-Singer duality

Mathematics Subject Classification (2000) 90C26 · 49M29

1 Introduction

One of the main tools in optimization is duality theory, which associates to a given (primal) problem $(\mathcal{P})$ another (dual) problem $(\mathcal{D})$, in such a way that the relation between the two problems provides useful information about $(\mathcal{P})$. In the case of convex optimization problems, under suitable assumptions, the optimal values of $(\mathcal{P})$ and $(\mathcal{D})$ coincide, and the primal optimal solutions can be recovered from the dual optimal solutions; this is particularly useful when $(\mathcal{D})$ happens to be easier to solve than $(\mathcal{P})$. The essential tools of convex duality theory are convex conjugation and the notion of subgradient; we refer to the classical book by Rockafellar [1] for a detailed treatment of classical convex duality. In the last decades, there has also been a very active research in nonconvex programming, motivated by the fact that many real life optimization problems are nonconvex. In this paper, we propose a generalization of one of

the most classical nonconvex duality theories, namely, Toland - Singer duality for d.c. (difference of convex) functions [2–4], consisting in a multiduality principle involving a (possibly infinite) collection of mutually dual problems. Our simple multiduality result does not require any convexity assumption, and one of its consequences is a new duality theorem involving just two mutually dual problems, expressed in terms of classical Fenchel conjugates. This theorem generalizes the well-known Toland - Singer duality theorem, and we use it to characterize (approximate) global optimal solutions, thus generalizing a well known necessary and sufficient global optimality condition due to Hiriart-Urruty [5]. Another application of the new theorem yields a duality result involving the generalized perspective functions introduced by Maréchal in [6].

There are some numerical algorithms based on Toland - Singer duality, such as the DC algorithm [7] and branch-and-bound/cutting-plane type algorithms [8,9]. Following the approach by Tao & El Bernoussi in [7], we propose an algorithm for searching local optimal solutions of nonconvex problems having the format considered in our duality theorem.

The paper consists of three more sections. In Section 2, we state the new multiduality principle, which is naturally formulated in the framework of generalized conjugation theory. Section 3 considers the special case when the collection of mutually dual problems consists of just two problems; for simplicity, this section is presented in the setting of classical convex conjugation.

In Section 4 we propose an extension of the DC algorithm to the broader class of nonconvex problems considered in Section 3.

2 Multiduality

Let $(X_i)_{i \in I}$ be a family of nonempty sets, indexed by $I \neq \emptyset$. We denote $X := \prod_{i \in I} X_i$ and $X_{-i} := \prod_{j \in I \setminus \{i\}} X_j$ for $i \in I$. For $y \in X_{-i}$ and $z \in X_i$, we define $x_{(-i,y),(i,z)} \in X$ by

$$\left(x_{(-i,y),(i,z)}\right)_i := z, \ \left(x_{(-i,y),(i,z)}\right)_j := y_j \text{ for } j \in I \setminus \{i\}. \tag{1}$$

Given a function $c : X \to \overline{\mathbb{R}}$, for each $i \in I$ we consider the coupling function $c_i : X_{-i} \times X_i \to \overline{\mathbb{R}}$ defined by

$$c_i(y, z) := c(x_{(-i,y),(i,z)}).$$

Notice that, for $x \in X$, one has $c_i(x_{-i}, x_i) = c(x)$; here and throughout we denote by x_{-i} the projection of $x \in X$ onto X_{-i}, that is,

$$(x_{-i})_j := x_j \quad \text{for} \quad j \in I \setminus \{i\}.$$

Following the generalized conjugation scheme of [10], we define a new coupling function $c'_i : X_i \times X_{-i} \to \overline{\mathbb{R}}$ by

$$c'_i(z, y) := c_i(y, z)$$

and consider the conjugation operators associated with c_i and c'_i, namely the c_i-conjugate of $g : X_{-i} \to \overline{\mathbb{R}}$ is the function $g^{c_i} : X_i \to \overline{\mathbb{R}}$ defined by

$$g^{c_i}(z) := \sup_{y \in X_{-i}} \{c_i(y, z) - g(y)\},$$

and, analogously, the c_i'-conjugate of $h : X_i \to \overline{\mathbb{R}}$ is $h^{c_i'} : X_{-i} \to \overline{\mathbb{R}}$, defined by

$$h^{c_i'}(y) := \sup_{z \in X_i} \{c_i'(z, y) - h(z)\};$$

here and in the sequel we adopt the conventions

$$+\infty + (-\infty) = -\infty + (+\infty) = +\infty - (+\infty) = -\infty - (-\infty) := -\infty.$$

For $h : X_i \to \overline{\mathbb{R}}$, $\overline{z} \in h^{-1}(\mathbb{R})$ and $\epsilon \geq 0$, we set

$$\partial_\epsilon^{c_i'} h(\overline{z}) := \{y \in X_{-i} \; : \; h(z) - h(\overline{z}) \geq c_i'(z, y) - c_i'(\overline{z}, y) - \epsilon \quad \text{for every } z \in X_i\}.$$

One can easily check that

$$\partial_\epsilon^{c_i'} h(\overline{z}) = \left\{y \in X_{-i} \; : \; -(h^{c_i'}(y) - c_i'(\overline{z}, y)) \geq h(\overline{z}) - \epsilon\right\}.$$

Given $f : X \to \overline{\mathbb{R}}$, for $i \in I$ and $z \in X_i$ we define $f_{i,z} : X_{-i} \to \overline{\mathbb{R}}$ by

$$f_{i,z}(y) := f(x_{(-i,y),(i,z)}).$$

Notice that, for $x \in X$, one has $f_{i,x_i}(x_{-i}) = f(x)$.

Theorem 2.1 *Let $f : X \to \overline{\mathbb{R}}$. For every $i \in I$ one has*

$$\sup_{z \in X_i} f_{i,z}^{c_i}(z) = \sup_{x \in X} \{c(x) - f(x)\}.$$

Hence, the optimal value of problem

$$(\mathcal{P}_i) \qquad \sup_{z \in X_i} f_{i,z}^{c_i}(z)$$

does not depend on i.

Proof We have

$$\sup_{z \in X_i} f_{i,z}^{c_i}(z) = \sup_{z \in X_i} \sup_{y \in X_{-i}} \{c_i(y,z) - f_{i,z}(y)\}$$

$$= \sup_{x \in X} \{c_i(x_{-i}, x_i) - f_{i,x_i}(x_{-i})\}$$

$$= \sup_{x \in X} \{c(x) - f(x)\}. \quad \square$$

Corollary 2.1 *Let $G \subseteq X$ and, for $i \in I$, denote by $C_i : X_i \rightrightarrows X_{-i}$ the set-valued mapping defined by*

$$C_i(z) := \{y \in X_{-i} \ : \ x_{(-i,y),(i,z)} \in G\}.$$

Then

$$\sup_{z \in X_i} \delta_{C_i(z)}^{c_i}(z) = \sup_{x \in G} c(x).$$

Proof Apply Theorem 2.1 with $f := \delta_G$. $\hfill\square$

Corollary 2.2 *Let $G_i \subseteq X_i$ for every $i \in I$. Then*

$$\sup_{z \in G_i} \delta_{G_{-i}}^{c_i}(z) = \sup_{x \in G} c(x).$$

Proof Apply Corollary 2.1 with $G := \prod_{i \in I} X_i$. $\hfill\square$

Corollary 2.3 *Let $f_i : X_i \to \mathbb{R} \cup \{+\infty\}(i = 1, ..., n)$, and define*

$$f : \prod_{i=1}^{n} X_i \to \mathbb{R} \cup \{+\infty\} \quad by \quad f(x_1, ..., x_n) := \sum_{i=1}^{n} f_i(x_i).$$

Then, for every $i \in I$, one has

$$\sup_{z \in X_i} \left\{ \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} f_j \right)^{c_i} (z) - f_i(z) \right\} = \sup_{x \in X} \{c(x) - f(x)\}.$$

Theorem 2.2 *Assume that the optimal value of $(\mathcal{P}_i)$ is finite, and let $\epsilon, \eta \geq 0$.*

If $\bar{z}$ is an ϵ-optimal solution to $(\mathcal{P}_i)$ such that $f_{i,\bar{z}}^{c_i c_i'} = f_{i,\bar{z}}$, then, for every

$\bar{y} \in \partial_\eta f_{i,\bar{z}}^{c_i}(\bar{z})$ and $j \in I \setminus \{i\}$, the point $\bar{y}_j$ is an $(\epsilon + \eta)$-optimal solution to

$(\mathcal{P}_j)$.

Proof We have

$$
\begin{aligned}
f_{j,\bar{y}_j}^{c_j}(\bar{y}_j) &\geq c_j\left(x_{(i,\bar{z}),(I\setminus\{i,j\},\bar{y}_{-j})}, \bar{y}_j\right) - f_{j,\bar{y}_j}\left(x_{(i,\bar{z}),(I\setminus\{i,j\},\bar{y}_{-j})}\right) \\
&= c(x_{(-i,\bar{y}),(i,\bar{z})}) - f(x_{(i,\bar{z}),(-i,\bar{y})}) \\
&= c_i(\bar{y},\bar{z}) - f_{i,\bar{z}}(\bar{y}) = c_i'(\bar{z},\bar{y}) - f_{i,\bar{z}}(\bar{y}) \\
&= c_i'(\bar{z},\bar{y}) - f_{i,\bar{z}}^{c_i c_i'}(\bar{y}) \geq f_{i,\bar{z}}^{c_i}(\bar{z}) - \eta \\
&\geq \sup_{z \in X_i} f_{i,z}^{c_i}(z) - \epsilon - \eta = \sup_{v \in X_j} f_{j,v}^{c_j}(v) - (\epsilon + \eta),
\end{aligned}
$$

the last equality being an immediate consequence of Theorem 2.1. $\qquad\square$

3 Nonconvex Duality

3.1 A Generalization of Toland - Singer Duality

Let $(X_1, X_2, \langle\cdot,\cdot\rangle)$ be a dual pair of locally convex spaces and $f : X_1 \times X_2 \to \overline{\mathbb{R}}$.

In this section we will apply the general results of Section 2 to the special case

when $I := \{1, 2\}$ and c is the duality pairing $\langle\cdot,\cdot\rangle$.

We recall that the Fenchel conjugates of $g : X_2 \to \overline{\mathbb{R}}$ and $h : X_1 \to \overline{\mathbb{R}}$ are

the functions $g^* : X_1 \to \overline{\mathbb{R}}$ and $h^* : X_2 \to \overline{\mathbb{R}}$, respectively, defined by

$$
g^*(z) := \sup_{y \in X_2} \{\langle z, y \rangle - g(y)\}, \qquad h^*(y) := \sup_{z \in X_1} \{\langle z, y \rangle - h(z)\}.
$$

It is easy to see that, for the coupling function c considered in this section, one has $g^{c_1} = g^*$ and $h^{c_2} = h^*$. Thus, Theorem 2.1 and Corollaries 2.1 and 2.2 yield:

Corollary 3.1 *Let* $f : X_1 \times X_2 \to \overline{\mathbb{R}}$. *Then*

$$\sup_{z \in X_1} f(z, \cdot)^*(z) = \sup_{(z,y) \in X_1 \times X_2} \{\langle z, y \rangle - f(z, y)\} = \sup_{y \in X_2} f(\cdot, y)^*(y).$$

Corollary 3.2 *Let* $T : X_1 \rightrightarrows X_2$. *Then*

$$\sup_{z \in X_1} \delta^*_{T(z)}(z) = \sup_{(z,y) \in Graph\ T} \langle z, y \rangle = \sup_{y \in X_2} \delta^*_{T^{-1}(y)}(y).$$

Proof Apply Corollary 2.1 with $G := Graph\ T$.   □

Corollary 3.3 *Let* $Z \subseteq X_1$ *and* $Y \subseteq X_2$. *Then*

$$\sup_{z \in Z} \delta^*_Y(z) = \sup_{x \in Z \times Y} \langle z, y \rangle = \sup_{y \in Y} \delta^*_Z(y).$$

Similarly, from Corollary 2.3 or, alternatively, Corollary 3.1 (by setting $f(z, y) := h(z) + g(y)$), one obtains the following version of the classical Toland - Singer duality theorem [2,4]:

Corollary 3.4 *Let* $g : X_2 \to \overline{\mathbb{R}}$ *and* $h : X_1 \to \overline{\mathbb{R}}$. *Then one has*

$$\sup_{z \in X_1} \{g^*(z) - h(z)\} = \sup_{(z,y) \in X_1 \times X_2} \{\langle z, y \rangle - h(z) - g(y)\} = \sup_{y \in X_2} \{h^*(y) - g(y)\}.$$

From the preceding corollary, by taking $g := k^*$, with $k : X_1 \to \overline{\mathbb{R}}$ such that $k^{**} = k$, one immediately obtains the standard Toland - Singer formula.

We recall that, for $\eta \geq 0$, the η-subdifferential of $h : X_1 \to \overline{\mathbb{R}}$ at $\bar{z} \in h^{-1}(\mathbb{R})$ is the set

$$\partial_\eta h(\bar{z}) := \{y \in X_2 :\ h(z) \geq h(\bar{z}) + \langle z - \bar{z}, y \rangle - \eta \quad \text{for every } z \in X_1\}.$$

From Theorem 2.2, one gets the following result on approximate optimal solutions of the pair of dual problems

$$(\mathcal{P}) \qquad\qquad \text{maximize } f(z, \cdot)^*(z)$$

and

$$(\mathcal{D}) \qquad\qquad \text{maximize } f(\cdot, y)^*(y)$$

of Corollary 3.1. A related result, showing how to obtain optimal dual solutions from optimal primal solutions of nonconvex problems, can be found in [3, Theorem 2.4].

Corollary 3.5 *Assume that the optimal value of* $(\mathcal{P})$ *is finite, and let* $\epsilon, \eta \geq 0$. *If* $\bar{z}$ *is an* ϵ-*optimal solution to* $(\mathcal{P})$ *such that* $f(\bar{z}, \cdot)$ *is convex and l.s.c. then, for every* $\bar{y} \in \partial_\eta f(\bar{z}, \cdot)^*(\bar{z})$, *the point* $\bar{y}$ *is an* $(\epsilon + \eta)$-*optimal solution to* $(\mathcal{D})$.

The following characterization of approximate global optimal solutions generalizes a well known necessary and sufficient global optimality condition due to Hiriart-Urruty [5].

Theorem 3.1 *Assume that the optimal value of* $(\mathcal{P})$ *is finite, and let* $\bar{z} \in X_1$ *and* $\epsilon \geq 0$. *Then* $\bar{z}$ *is an* ϵ-*optimal solution to* $(\mathcal{P})$ *if and only if for every* $\eta \geq 0$ *and every* $y \in f(\bar{z}, \cdot)^{-1}(\mathbb{R})$ *such that* $\bar{z} \in \partial_\eta f(\bar{z}, \cdot)(y)$ *one has*

$$y \in \partial_{\epsilon + \eta} f(\cdot, y)(\bar{z}). \qquad\qquad (2)$$

Proof Let $\bar{z}$ be an ϵ-optimal solution to $(\mathcal{P})$, and $\eta \geq 0$ and $y \in f(\bar{z}, \cdot)^{-1}(\mathbb{R})$ be such that $\bar{z} \in \partial_\eta f(\bar{z}, \cdot)(y)$. Using Corollary 3.1, we obtain

$$f(\cdot, y)(\bar{z}) = f(\bar{z}, \cdot)(y) \leq \langle \bar{z}, y \rangle + \eta - f(\bar{z}, \cdot)^*(\bar{z})$$

$$\leq \langle \bar{z}, y \rangle + \eta - \sup_{z \in X_1} f(z, \cdot)^* + \epsilon \leq \langle \bar{z}, y \rangle + \eta - f(\cdot, y)^*(y) + \epsilon,$$

which proves (2).

Conversely, assume that the condition stated after "if and only if" holds,
let $y \in f(\bar{z}, \cdot)^{-1}(\mathbb{R})$, and define $\eta := f(\bar{z}, y) + f(\bar{z}, \cdot)^*(\bar{z}) - \langle \bar{z}, y \rangle$. Then $\eta \geq 0$
and $\bar{z} \in \partial_\eta f(\bar{z}, \cdot)(y)$, and therefore (2) holds. Hence

$$f(\bar{z}, \cdot)^*(\bar{z}) = \langle \bar{z}, y \rangle + \eta - f(\bar{z}, y) \geq f(\cdot, y)^*(y) - \epsilon,$$

which shows that $\bar{z}$ is an ϵ-optimal solution to $(\mathcal{P})$.   □

To apply the duality theory presented in this section to a given optimization
problem, one has to be able to recognize whether the objective function of the
problem under consideration can be written in the form $f(z, \cdot)^*(z)$ for some
function $f : X_1 \times X_2 \to \overline{\mathbb{R}}$. The following theorem provides a useful criterion
to make this recognition possible.

Theorem 3.2 *For every $f : X_1 \times X_2 \to \overline{\mathbb{R}}$, the function $\varphi : X_1^2 \to \overline{\mathbb{R}}$ defined
by $\varphi(z, z') := f(z, \cdot)^*(z')$ is convex, proper (accepting as proper the constant
functions $+\infty$ and $-\infty$) and l.s.c. in its second argument. Conversely, for
every function $\varphi : X_1^2 \to \overline{\mathbb{R}}$ with these properties there exists $f : X_1 \times X_2 \to \overline{\mathbb{R}}$
such that $f(z, \cdot)^*(z) = \varphi(z, z)$ for every $z \in X_1$.*

Proof The first part of the statement is an immediate consequence of well
known properties of conjugate functions. To prove the converse, define $f :
X_1 \times X_2 \to \overline{\mathbb{R}}$ by $f(z, y) := \varphi(z, \cdot)^*(y)$; then, since $\varphi(z, \cdot)$ is convex, proper
and l.s.c, from the equality $f(z, \cdot) := \varphi(z, \cdot)^*$ it follows that $\varphi(z, \cdot) := f(z, \cdot)^*$,
and therefore $\varphi(z, z) := f(z, \cdot)^*(z)$.   □

According to the preceding theorem, the class of problems

$(\mathcal{P})$ maximize $f(z, \cdot)^{*}(z)$

to which our duality theory applies coincides with that consisting of problems
with format

$(\mathcal{P})$ maximize $\varphi(z, z)$,   (3)

the function $\varphi : X_1^2 \to \overline{\mathbb{R}}$ being as in the statement. Since, in view of the
proof, the objective function of these two problems are linked by the relation
$f(z, \cdot) := \varphi(z, \cdot)^{*}$, a straightforward computation yields the dual objective
function in terms of φ :

$$f(\cdot, y)^{*}(y) = \sup_{z \in X_1} \{\langle z, y \rangle - f(z, y)\} = \sup_{z \in X_1} \{\langle z, y \rangle - \varphi(z, \cdot)^{*}(y)\}. \quad (4)$$

Hence, if the primal problem is stated as (3), the formulation of the dual
problem is

$(\mathcal{D})$ $\sup_{z \in X_1} \{\langle z, y \rangle - \varphi(z, \cdot)^{*}(y)\}.$

To illustrate that solving the dual problem may be advantageous, we present
the following academic example, in which the primal problem consists in maxi-
mizing a non concave function and, on the contrary, the dual objective function
is concave.

Example 3.1 Let us consider the primal problem

$(\mathcal{P})$ maximize $\alpha(z)$,

where α is a non concave function defined on a normed space X_1. Let us define
$\varphi : X_1^2 \to \overline{\mathbb{R}}$ by

$$\varphi(z, z') := \alpha(z) - ||z||^2 + \langle z', z \rangle + \delta_{NB}(z' - z);$$

here B is the closed unit ball in X_1 and $N > 0$. Clearly, $\varphi(z, z) = \alpha(z)$, and

therefore $(\mathcal{P})$ can be rewritten as (3). Since

$$
\begin{aligned}
\varphi\left(z, \cdot\right)^{*}(y) &= \sup_{z' \in X_1} \left\{ \langle z', y \rangle - \varphi\left(z, z'\right) \right\} \\
&= \sup_{z' \in X_1} \left\{ \langle z', y - z \rangle + \delta_{NB}(z' - z) \right\} - \alpha(z) + ||z||^2 \\
&= \sup_{v \in X_1} \left\{ \langle z + v, y - z \rangle - \delta_{NB}(v) \right\} - \alpha(z) + ||z||^2 \\
&= \sup_{v \in NB} \langle v, y - z \rangle + \langle z, y \rangle - \alpha(z) \\
&= N||y - z|| + \langle z, y \rangle - \alpha(z),
\end{aligned}
$$

according to (4) we get the following expression for the dual objective function:

$$
f\left(\cdot, y\right)^{*}(y) = \sup_{z \in X_1} \left\{ \langle z, y \rangle - \varphi\left(z, \cdot\right)^{*}(y) \right\} = \sup_{z \in X_1} \left\{ \alpha(z) - N||y - z|| \right\},
$$

which, in view of [11, p. 200], shows that it is the smallest N-Lipschitz majorant

of α. This N-Lipschitz envelope is concave in some cases; for instance, let us

consider the case when $X_1 := \mathbb{R}$,

$$
\alpha(z) := -2z^6 + 15z^4 - 36z^2
$$

and $N := 10$. Then one can easily prove that

$$
f\left(\cdot, y\right)^{*}(y) = \begin{cases}
10y + r + \alpha(-r), & \text{if } y \leq -r, \\
\alpha(y), & \text{if } y \in [-r, r], \\
-10y + r + \alpha(r), & \text{if } y \geq -r.
\end{cases}
$$

with r being the smallest positive real number satisfying $\alpha'(-r) = 10$.

Figure 1 depicts the graphs of both α and $y \to f\left(\cdot, y\right)^{*}(y)$ and shows that the
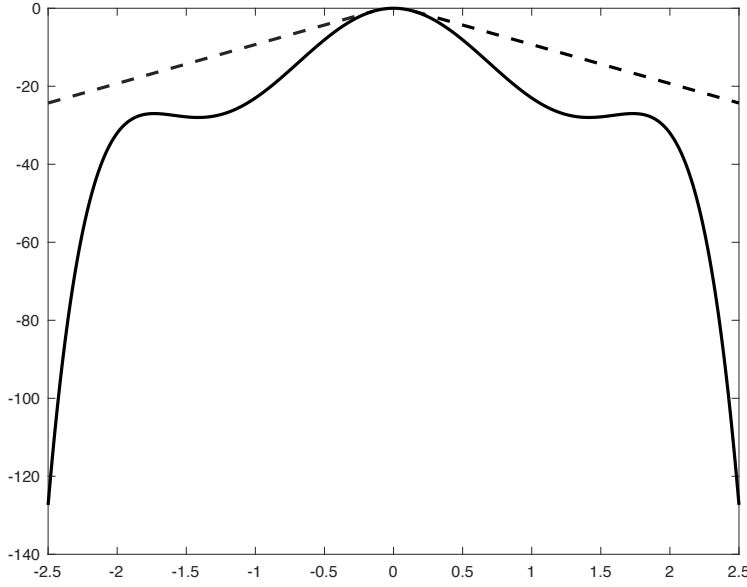
latter function is concave.

Fig. 1 Graph of $\alpha(y)$ (black line) and $y \to f(\cdot, y)^*(y)$ (dash line)

3.2 Example: Generalized Perspective Functions

Let $g : X_2 \to]0, +\infty[$ and $h : X_1 \to]0, +\infty[$, and define $f : X_1 \times X_2 \to]0, +\infty[$ by $f(z, y) := h(z)g(y)$. For this particular function f, the objective functions of problems $(\mathcal{P})$ and $(\mathcal{D})$ can be expressed by means of the operation Δ introduced in [6], which constitutes a generalization of the so called perspective function of Convex Analysis. Indeed, for $z \in X_1$ one has $f(z, \cdot) = h(z)g$, and hence

$$f(z, \cdot)^*(z) = (h(z)g)^*(z) = h(z)g^*\left(\frac{z}{h(z)}\right) = (g^* \Delta h)(z, z).$$

Similarly, for $y \in X_2$ one has

$$f(\cdot, y)^*(y) = (g(y)h)^*(y) = g(y)h^*\left(\frac{y}{g(y)}\right) = (h^* \Delta g)(y, y).$$

Thus, Theorem 3.1 yields:

Corollary 3.6 *Let $g : X_2 \to \,]0, +\infty[$ and $h : X_1 \to \,]0, +\infty[$. Then*

$$\sup_{z \in X_1} (g^* \Delta h)(z, z) = \sup_{y \in X_2} (h^* \Delta g)(y, y).$$

4 An Algorithm

In the setting of the preceding section, in this one we will assume that X_1 and X_2 are the Euclidean space $\mathbb{R}^n$ and $\langle \cdot, \cdot \rangle$ is the standard Euclidean product. Let $f : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$, and consider the sets

$$\mathcal{P}_l = \{\bar{z} \in \mathbb{R}^n : y \in \mathbb{R}^n \text{ and } \bar{z} \in \partial f(\bar{z}, \cdot)(y) \text{ imply } y \in \partial f(\cdot, y)(\bar{z})\}$$

and

$$\mathcal{D}_l = \{\bar{y} \in \mathbb{R}^n : z \in \mathbb{R}^n \text{ and } \bar{y} \in \partial f(\cdot, \bar{y})(z) \text{ imply } z \in \partial f(z, \cdot)(\bar{y})\}.$$

According to Theorem 3.1, the set $\mathcal{P}_l$ contains all the optimal solutions to $(\mathcal{P})$ and the set $\mathcal{D}_l$ contains all the optimal solutions to $(\mathcal{D})$.

For $\bar{z} \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^n$, we consider the auxiliary problems

$$S(\bar{z}) \qquad \text{maximize } f(\cdot, y)^*(y) \text{ subject to } y \in \partial f(\bar{z}, \cdot)^*(\bar{z})$$

and

$$T(\bar{y}) \qquad \text{maximize } f(z, \cdot)^*(z) \text{ subject to } z \in \partial f(\cdot, \bar{y})^*(\bar{y}).$$

We will denote by $\mathcal{O}_{S(\bar{z})}$ and $\mathcal{O}_{T(\bar{y})}$ the sets of optimal solutions to problems $S(\bar{z})$ and $T(\bar{y})$, respectively.

Proposition 4.1 *(1) If $\bar{z} \in \mathcal{P}_l$ and $f(\bar{z}, \cdot)$ is convex, proper and l.s.c., then*

the function $y \mapsto f(\cdot, y)^*(y)$ is constant on $\partial f(\bar{z}, \cdot)^*(\bar{z})$.

(2) If $\bar{y} \in \mathcal{D}_l$ and $f(\cdot, \bar{y})$ is convex, proper and l.s.c., then the function

$z \mapsto f(z, \cdot)^*(z)$ is constant on $\partial f(\cdot, \bar{y})^*(\bar{y})$.

Proof By symmetry, we only need to prove *(1)*. Let $\bar{z} \in \mathcal{P}_l$. If $y \in \partial f(\bar{z}, \cdot)^*(\bar{z})$,

then

$$f(\bar{z}, \cdot)^*(\bar{z}) = \langle \bar{z}, y \rangle - f(\bar{z}, \cdot)^{**}(y) = \langle \bar{z}, y \rangle - f(\bar{z}, y) \tag{5}$$

and, since $y \in \partial f(\cdot, y)(\bar{z})$ (as $\bar{z} \in \mathcal{P}_l$),

$$f(\cdot, y)^*(y) = \langle \bar{z}, y \rangle - f(\bar{z}, y). \tag{6}$$

From (6) and (5) it immediately follows that

$$f(\cdot, y)^*(y) = f(\bar{z}, \cdot)^*(\bar{z}). \tag{7}$$

$\square$

Starting with an initial point $z^0 \in \mathbb{R}^n$, we construct two sequences z^k and

y^k as follows:

$$
\begin{aligned}
z^0 &\mapsto y^0 \in \mathcal{O}_{S(z^0)} \\
z^1 \in \mathcal{O}_{T(y^0)} &\mapsto y^1 \in \mathcal{O}_{S(z^1)} \\
\vdots &\qquad \vdots \\
z^{k+1} \in \mathcal{O}_{T(y^k)} &\mapsto y^{k+1} \in \mathcal{O}_{S(z^{k+1})}.
\end{aligned}
\tag{8}
$$

This algorithm requires the auxiliary problems occurring at each iteration to

be solvable, and from now on we will assume that this is the case. The follow-

ing proposition states sufficient conditions for solvability of these problems.

We recall that a function is called co-finite if its conjugate is finite valued everywhere.

Proposition 4.2 *Let $\bar{z} \in \mathbb{R}^n$ and $\bar{y} \in \mathbb{R}^n$, and assume that f is continuous. If $f(\bar{z}, \cdot)$ is co-finite and the mapping $y \longmapsto dom f(\cdot, y)$ is compact-valued and continuous, then $\mathcal{O}_{S(\bar{z})} \neq \emptyset$. Analogously, if $f(\cdot, \bar{y})$ is co-finite and the mapping $z \longmapsto dom f(z, \cdot)$ is compact-valued and continuous, then $\mathcal{O}_{T(\bar{y})} \neq \emptyset$.*

Proof The co-finiteness of $f(\bar{z}, \cdot)$ guarantees that the feasible set $\partial f(\bar{z}, \cdot)^*(\bar{z})$ of $S(\bar{z})$ is nonempty and compact. On the other hand, continuity of f together with compact-valuedness and continuity of the mapping $y \longmapsto dom f(\cdot, y)$ implies, by Berge' maximum theorem, continuity of the objective function $y \longmapsto f(\cdot, y)^*(y)$. It thus suffices to apply Weierstrass' extreme value theorem to conclude that $\mathcal{O}_{S(\bar{z})} \neq \emptyset$. The proof of the nonemptiness of $\mathcal{O}_{T(\bar{y})}$ is the same, mutatis mutandis.   □

Theorem 4.1 *Let $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and let the sequences z^k and y^k be constructed according to (8). Then*

$$f(z^{k+1}, \cdot)^*(z^{k+1}) \geq f(\cdot, y^k)^*(y^k) \geq f(z^k, \cdot)^*(z^k). \tag{9}$$

The first inequality holds with the equal sign if and only if $z^{k+1} \in \partial f(z^{k+1}, \cdot)(y^k)$, in which case, assuming that the function $f(\cdot, y^k)$ is convex, proper and l.s.c., we have $y^k \in \mathcal{D}_l$. The second inequality holds with the equal sign if and only if $y^k \in \partial f(\cdot, y^k)(z^k)$, in which case, assuming that the function $f(z^k, \cdot)$ is convex, proper and l.s.c., we have $z^k \in \mathcal{P}_l$.

Proof Because of Fenchel inequality and the relations $z^{k+1} \in \partial f(\cdot, y^k)^*(y^k)$ and $y^k \in \partial f(z^k, \cdot)^*(z^k)$, we obtain

$$
\begin{aligned}
f(z^{k+1}, \cdot)^*(z^{k+1}) &\geq \langle z^{k+1}, y^k \rangle - f(z^{k+1}, y^k) \\
&= f(\cdot, y^k)^*(y^k) \\
&\geq \langle z^k, y^k \rangle - f(z^k, y^k) \\
&= f(z^k, \cdot)^*(z^k).
\end{aligned}
$$

The 'if and only if' assertions follow from the above chain of inequalities, combined with the well-known characterization of subgradients as those elements that satisfy the Fenchel inequality with the equal sign.

Let us assume that the first inequality holds with the equal sign, and let $z \in \mathbb{R}^n$ be such that $y^k \in \partial f(\cdot, y^k)(z)$. Since we have $z^{k+1} \in \mathcal{O}_{T(y^k)}$ and $z^{k+1} \in \partial f(z^{k+1}, \cdot)(y^k)$, using that $f(\cdot, y^k)$ is convex, proper and l.s.c., we obtain

$$
\begin{aligned}
f(z, \cdot)^*(z) &\leq f(z^{k+1}, \cdot)^*(z^{k+1}) = \langle z^{k+1}, y^k \rangle - f(z^{k+1}, y^k) \\
&= \langle z^{k+1}, y^k \rangle - f(\cdot, y^k)^{**} (z^{k+1}) \leq \langle z, y^k \rangle - f(\cdot, y^k)^{**} (z) \\
&= \langle z, y^k \rangle - f(z, y^k) \leq f(z, \cdot)^*(z);
\end{aligned}
$$

therefore $f(z, \cdot)^*(z) = \langle z, y^k \rangle - f(z, y^k)$, that is, $z \in \partial f(z, \cdot)(y^k)$. This shows that $y^k \in \mathcal{D}_l$.

Let us assume that the second inequality holds with the equal sign, and let $y \in \mathbb{R}^n$ be such that $z^k \in \partial f(z^k, \cdot)(y)$. Since we have $y^k \in \mathcal{O}_{S(z^k)}$ and $y^k \in \partial f(\cdot, y^k)(z^k)$, using that $f(z^k, \cdot)$ is convex, proper and l.s.c. and the fact

that $z^k \in \partial f(z^k, \cdot)^{**}(y)$, we obtain

$$f(\cdot, y)^*(y) \le f(\cdot, y^k)^*(y^k) = \langle z^k, y^k \rangle - f(z^k, y^k) = \langle z^k, y^k \rangle - f(z^k, \cdot)^{**}(y^k)$$

$$\le \langle z^k, y \rangle - f(z^k, \cdot)^{**}(y) = \langle z^k, y \rangle - f(z^k, y) \le f(\cdot, y)^*(y);$$

therefore $f(\cdot, y)^*(y) = \langle z^k, y \rangle - f(z^k, y)$, that is, $y \in \partial f(\cdot, y)(z^k)$. This shows that $z^k \in \mathcal{P}_l$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Corollary 4.1 *(1) Let f, z^k and y^k be as in Theorem 4.1. Then*

$$\lim_{k \to \infty} f(z^k, \cdot)^*(z^k) = \lim_{k \to \infty} f(\cdot, y^k)^*(y^k). \qquad (10)$$

(2) If f is continuous, $\overline{z}$ is a cluster point of the sequence z^k, and the sequence y^k is bounded, then

$$f(\overline{z}, \cdot)^*(\overline{z}) = \lim_{k \to \infty} f(z^k, \cdot)^*(z^k). \qquad (11)$$

(3) If f is continuous, $\overline{y}$ is a cluster point of the sequence y^k, and the sequence z^k is bounded, then

$$f(\cdot, \overline{y})^*(\overline{y}) = \lim_{k \to \infty} f(\cdot, y^k)^*(y^k).$$

Proof (1) The equality (10) is an immediate consequence of (9).

(2) If f is continuous, then the function $z \mapsto f(z, \cdot)^*(z)$ is l.s.c.. If $\overline{z}$ is a cluster point of z^k, then

$$\lim_{j \to \infty} z^{k_j} = \overline{z}$$

for some subsequence z^{k_j}. Since the sequence y^k is bounded, we can suppose (extracting again subsequences if necessary) that the sequence y^{k_j} converges to a point $\overline{y}$. Using the chain of inequalities at the beginning of the proof of

Theorem 4.1, the continuity of f, and the lower semicontinuity of the function $z \mapsto f(z, \cdot)^*(z)$, we obtain

$$
\begin{aligned}
\lim_{k \to \infty} f(z^k, \cdot)^*(z^k) = \lim_{j \to \infty} f(z^{k_j}, \cdot)^*(z^{k_j}) &= \lim_{j \to \infty} \left(\langle z^{k_j}, y^{k_j} \rangle - f(z^{k_j}, y^{k_j}) \right) \\
&= \langle \overline{z}, \overline{y} \rangle - f(\overline{z}, \overline{y}) \leq f(\overline{z}, \cdot)^*(\overline{z}) \leq \lim_{j \to \infty} f(z^{k_j}, \cdot)^*(z^{k_j}) \\
&= \lim_{k \to \infty} f(z^k, \cdot)^*(z^k),
\end{aligned}
$$

which proves (11).

The proof of *(3)* is similar to that of *(2)*. $\qquad\square$

Theorem 4.2 *Let f be continuous.*

(1) If f is convex and proper in its first argument, $\overline{z}$ is a cluster point of z^k, and the sequence y^k is bounded, then $\overline{z} \in \mathcal{P}_l$.

(2) If f is convex and proper in its second argument, $\overline{y}$ is a cluster point of y^k, and the sequence z^k is bounded, then $\overline{y} \in \mathcal{D}_l$.

Proof (1) Let $y \in \mathbb{R}^n$ be such that $\overline{z} \in \partial f(\overline{z}, \cdot)(y)$. Since $\overline{z}$ is a cluster point of z^k, we have

$$
\lim_{j \to \infty} z^{k_j} = \overline{z}
$$

for some subsequence z^{k_j}. We can suppose (extracting subsequences if necessary) that the sequence $y^{k_j - 1}$ converges to a point $\overline{y} \in \mathbb{R}^n$. Given that $z^{k_j} \in \mathcal{O}_{T(y^{k_j-1})}$ and $y^{k_j} \in \mathcal{O}_{S(z^{k_j})}$, we have $z^{k_j} \in \partial f(\cdot, y^{k_j-1})^*(y^{k_j-1})$ and $y^{k_j} \in \partial f(z^{k_j}, \cdot)^*(z^{k_j})$. Hence, using the properties of f and Corollary 4.1, we

get

$$f(\cdot,y)^*(y) \le \lim_{j\to\infty} f(\cdot,y^{k_j-1})^*(y^{k_j-1}) \tag{12}$$

$$= \lim_{j\to\infty} \left(\langle z^{k_j}, y^{k_j-1} \rangle - f(z^{k_j}, y^{k_j-1}) \right)$$

$$\le \lim_{j\to\infty} f(z^{k_j}, \cdot)^*(z^{k_j})$$

$$= f(\overline{z}, \cdot)^*(\overline{z})$$

$$= \langle \overline{z}, y \rangle - f(\overline{z}, y)$$

From (12) and Fenchel inequality, we obtain $y \in \partial f(\cdot,y)(\overline{z})$. This shows that $\overline{z} \in \mathcal{P}_l$.

The proof of *(2)* is similar. $\square$

References

1. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970).

2. Toland, J. F.: A duality principle for non-convex optimisation and the calculus of variations. Archive for Rational Mechanics and Analysis 71, pp. 41–61, (1979).

3. Toland, J. F.: Duality in nonconvex optimization. J. Math. Anal. Appl. 66, pp. 399-415, (1978).

4. Singer, I.: A Fenchel-Rockafellar type duality theorem for maximization. Bulletin of the Australian Mathematical Society 20, pp. 193–198, (1979).

5. Hiriart-Urruty, J.-B.: From convex optimization to nonconvex optimization: Necessary and sufficient conditions for global optimality. Nonsmooth Optimization and Related Topics, pp. 219–240, Plenum, New York/London, (1989).

6. Maréchal, P.: On a functional operation generating convex functions. I. Duality. J. Optim. Theory Appl. 126, no. 1, 175–189, (2005).

7. Tao, P.D., El Bernoussi, S.: Duality in D.C. (Difference of Convex functions) Optimization. Subgradient Methods. International Series of Numerical Mathematics, pp. 277–293, Birkhäuser Basel, (1988).

8. Horst, R., Thoai, N. V.: DC programming: overview. J. Optim. Theory Appl. 103, no. 1, 1–43, (1999).

9. Sun, W., Sampaio, R. J-B, Candido, M. A. B.: Proximal point algorithm for minimization of DC function. Journal of Computational Mathematics 21, no. 4, pp. 451–462, (2003).

10. Martínez-Legaz, J. E.: Generalized convex duality and its economic applications. Handbook of generalized convexity and generalized monotonicity, pp. 237–292, Nonconvex Optim. Appl., 76, Springer, New York, (2005).

11. Martínez-Legaz, J. E.: On lower subdifferentiable functions. Trends in mathematical optimization, pp. 197-232, Birkhäuser, Basel, (1988).