Commun. Optim. Theory 2024 (2024) 3 https://doi.org/10.23952/cot.2024.3

Communications in Optimization Theory
Available online at http://cot.mathres.org

CONSTANT DUALITY GAP AND APPLICATIONS

JEAN-BERNARD BAILLON[1,*], ELADIO OCAÑA[2]

[1]SAMM, Université Paris 1 Panthéon-Sorbonne, Paris, France
[2]IMCA, Universidad Nacional de Ingeniería, Calle Los Biólogos 245 Urb. San Cesar, La Molina, Lima 12, Perú

Dedicated to Francis Clarke on the occasion of his 75th birthday

Abstract. Inspired by the role played by zero duality gap in optimization problems, especially in the stopping strategy of algorithms, we design in this work a similar scheme but addressing non-convex quadratic optimization problems subject to linear equality constraints having possibly nonzero duality gap. In fact, we get a formula for determining it at least approximately.
Keywords. Convex Analysis; Duality ; Gap; Optimization; Quadratic.
2020 Mathematics Subject Classification. 90C26, 90C46.

1. Introduction

All optimization problem associates a dual problem that, under certain regularity conditions, provides relevant informations about existence, uniqueness, and also strategies to solve it algorithmically.

The duality gap between the primal and its respective dual problems, which is defined as the difference of their optimal values, has been widely studied by many authors by appealing certain conditions (called constraint qualification) ensuring zero duality gap, which basically implies convexity in the problem, see for example [3]. In fact, from the algorithmic point of view, one of the most important effects of the zero duality gap is the stopping criterium of the algorithms in order to solve the primal (and also its dual) problem.

In the general setting, i.e., without assuming any constraint qualification, the duality gap maybe nonzero and determining it, even approximately, could be very difficult.

Corresponding the optimization problem

$$\overline{\alpha} := \min_x [f(x) : g(x) = 0], \qquad (Opt)$$

*Corresponding author.
E-mail address: baillon@univ-paris1.fr (J.B. Baillon), eocana@imca.edu.pe (E. Ocaña).
Received: January 29, 2022; Accepted: May 12, 2023.

where $f:\mathbb{R}^n\to\mathbb{R}$ and $g:\mathbb{R}^n\to\mathbb{R}^q$ are two given functions, the marginal (or value) function is the function $h:\mathbb{R}^q\to\overline{\mathbb{R}}$ defined as

$$h(v)=\inf_x[f(x)+\Phi(x):g(x)+v=0], \qquad (P_v)$$

where $\Phi:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a given function, whose main role in the problem is to penalize or approximate it, making possibly easier than the original problem. For exemple, taking Φ appropriately, the original problem can be approximated by a (maybe not convex) quadratic optimization problem.

Corresponding to problem (P_v), we get in this work an explicit expression of the so called gap function of h, which is the difference of h and its corresponding Fenchel bi-conjugate function h^{**}, provided f is quadratic and g linear. In fact, by considering $f(x)=\langle Ax,x\rangle+\langle a,x\rangle$, $g(x)=Bx-b$, with A, B, a and b, matrices and vectors of appropriated orders, and $\Phi(x)=\varphi(Bx-b)$ (for a given function φ), we shall prove that for any $v\in\mathbb{R}^q$,

$$h(v)-h^{**}(v)=\varphi(v)-[\langle A\cdot,\cdot\rangle+\varphi(v+B\cdot)]^{**}(0),$$

which, as can be seen, is independent on b. Then, assuming φ to be even satisfying $\varphi(0)=0$, the next explicit expression of the duality gap is deduced:

$$h(0)-h^{**}(0)=-\inf_x[\langle Ax,x\rangle+\varphi(Bx)].$$

In another context, J.-P Aubin and I. Ekeland [1] introduced a measure of lack convexity in order to estimate the duality gap in nonconvex optimization. In fact, for a given real-valued function f defined on a convex set X, its lack of convexity is defined as

$$\rho(f)=\sup\left\{f\left(\sum\alpha_i x_i\right)-\sum\alpha_i f(x_i)\right\}$$

over all finite families $\alpha_i\in]0,1[$, $x_i\in X$ with $\sum\alpha_i=1$. Clearly, $0\le\rho(f)\le\infty$, and $\rho(f)=0$ if and only if f is convex.

The plan of the paper is as follows. In the next section some relationships on the convex hull and the Fenchel bi-conjugate for a function are discussed. In the third section some properties on affine–quadratic functions are also discussed highlighting the characterization of the affine-quadratic property by means of the parallelogram law. The duality gap of the marginal function corresponding to a general optimization problem with linear equality constraint is discussed in the fourth section, an explicit expression of the duality gap function is also deduced. When the objective function in the problem is the sum of an affine-quadratic function and an arbitrary function, where the second one playing the role of penalizing or regularizing the problem, is dealt in the fifth section. We also get in this part an explicit expression of the duality gap corresponding to the marginal function. As a consequence of this expression we deduce that zero duality gap implies convexity of the problem. The case where the admissible set is a linear inequality constraint, is briefly discusses in a corollary. Finally, in section six, we provide a general primal-dual algorithm where the duality gap (assuming it known) is used as stoping criterium.

2. Convex hull and Fenchel bi-conjugate

Given a function $f:\mathbb{R}^n\to\overline{\mathbb{R}}:=[-\infty,\infty]$, its domain is the set $\operatorname{dom}(f):=\{x\in\mathbb{R}^n: f(x)<\infty\}$. When $\operatorname{dom}(f)\neq\emptyset$ and $f>-\infty$ everywhere, f is said to be *proper*. The function f is said to be *even* if $f(-x) = f(x)$ for all $x \in \mathbb{R}^n$; *convex* if its epigraph $\operatorname{epi}(f) := \{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \lambda\}$ is convex; and *lower semicontinuos* (lsc in short) if $\operatorname{epi}(f)$ is closed.

The Fenchel conjugate of f is the lsc convex function $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by $f^*(x^*) := \sup[\langle x^*, x\rangle - f(x) : x \in \mathbb{R}^n]$. We denote by $f^{**} := (f^*)^*$ the Fenchel bi-conjugate of f. The difference $f - f^{**}$ is termed the duality gap function corresponding to f. In general,

$$-\infty \leq f^{**}(x) = \sup_{x^*} \inf_{y} [f(y) + \langle x^*, x - y\rangle] \leq f(x) \text{ for all } x \in \mathbb{R}^n.$$

Additional relationships between f and its conjugate/bi-conjugate are listed below, their proofs can be easily deduced from the definitions.

Proposition 2.1. *We get the following properties*

i) *f^* is even if f is so;*
ii) *If f is convex, then $f(0) = \inf f(x) = -f^*(0)$ whenever f is even.*
iii) *$f^{**}(0) = \inf f(x) = -f^*(0)$ whenever f is even.*

The *convex hull function* of f, denoted by $\operatorname{co} f$, is defined as the greatest convex function majorized by f. One deduce that $\operatorname{co} f \leq f$ and due the Carathéodory's theorem [6, Proposition 2.3.1],

$$\operatorname{co} f(x) = \inf \left\{ \sum_{i=0}^{n} \lambda_i f(x_i) : \sum_{i=0}^{n} \lambda_i x_i = x,\, \lambda_i \geq 0,\, \sum_{i=0}^{n} \lambda_i = 1 \right\}.$$

To complete the properties given previously in Proposition 2.1, one get

iv) *If f is even then $\operatorname{co} f(0) = f^{**}(0)$.*

The proof of the next proposition can also be deduced directly by definitions.

Proposition 2.2. *We get the following properties*

i) $\operatorname{dom}(\operatorname{co} f) = \operatorname{co}(\operatorname{dom}(f))$*;*
ii) $\operatorname{co} f$ *is lsc on* $\operatorname{ri}(\operatorname{co}(\operatorname{dom}(f)))$*;*
iii) *If $\operatorname{co} f(\bar{x}) = -\infty$ at some $\bar{x} \in \mathbb{R}^n$, then $f^{**} \equiv -\infty$ on $\mathbb{R}^n$ and coincides with $\operatorname{co} f$ on* $\operatorname{ri}(\operatorname{co}(\operatorname{dom}(f)))$*;*
iv) *$\operatorname{co} f(x) > -\infty$ for all $x \in \mathbb{R}^n$ if and only if $\operatorname{dom}(f^*) \neq \emptyset$ if and only if $f^{**}(x) > -\infty$ for all $x \in \mathbb{R}^n$ if and only if there is an affine function minorizing f on $\mathbb{R}^n$;*
v) *If $\operatorname{co} f(x) > -\infty$ for all $x \in \mathbb{R}^n$, then f^{**} coincides with the greatest lsc convex function, denoted by $\overline{\operatorname{co}} f$, majorized by f (or equivalently by $\operatorname{co} f$). In particular, f^{**} and $\operatorname{co} f$ coincide on $\mathbb{R}^n$ except maybe on the boundary of* $\operatorname{co}(\operatorname{dom}(f))$.

The next example shows that both operations biconjugate and convex hull are different.

Example 2.3. Let $g : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ be defined as

$$g(x_1, x_2) = \begin{cases} |x_2| e^{-|x_1 x_2|} & \text{if } 0 \leq x_1 \leq 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Simple calculations get

$$g^{**}(x_1, x_2) = \begin{cases} 0 & \text{if } 0 \leq x_1 \leq 1, \\ +\infty & \text{otherwise} \end{cases}$$

and

$$\operatorname{co} g(x_1,x_2) = \begin{cases} 0 & \text{if } 0 < x_1 \leq 1, \\ |x_2| & \text{if } x_1 = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Of course, biconjugate and convex hull coincide over all lsc convex functions, it is the same over all quadratic functions $f(x) = \langle Ax, x\rangle$ corresponding to $n \times n$ symmetric matrices A, both values being $-\infty$ if A is not positive semidefinite.

Unfortunately, equality between both operations, biconjugate and convex hull, is not stable under addition. More specifically, $f_1^{**} = \operatorname{co} f_1$ and $f_2^{**} = \operatorname{co} f_2$ does not implies $(f_1 + f_2)^{**} = \operatorname{co}(f_1 + f_2)$, as shown the next example where one of the functions is convex lsc and the other quadratic.

Example 2.4. Let g be as the previous example and let f be the function defined on $\mathbb{R}^2$ as $f(x_1,x_2) = g(x_1,x_2) - \alpha(x_1^2 + x_2^2)$ for some negative real parameter α. Clearly, f is lsc and for $|\alpha|$ sufficiently large, is also convex.

The next proposition [5, Lemma 1.5.3] (see also [6, Corollary 3.47]), shows that under a coerciveness property both operations biconjugate and convex hull coincide.

Proposition 2.5. *If f is lsc minorized by an affine function and satisfying the following coercivity property*

$$\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = \infty,$$

then $f^{**} = \operatorname{co} f$ everywhere.

3. Affine–quadratic functions

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be affine–quadratic if it is of the form $f(x) = \langle Ax, x\rangle + \langle a, x\rangle + \alpha$ on some affine linear subspace; and $+\infty$ or $-\infty$ otherwise, where A is an $n \times n$ symmetric matrix, $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

It is clear that the sum of two affine–quadratic functions is affine–quadratic; and if f is affine–quadratic, then for any $y \in \mathbb{R}^n$ the function f_y defined as $f_y(x) = f(x+y)$ for all $x \in \mathbb{R}^n$, is affine–quadratic.

This property is also stable under other operations as listed below.

Lemma 3.1. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a linear–quadratic function. Then for a linear subspace $E \subset \mathbb{R}^n$, the function f_E defined on $\mathbb{R}^n$ as*

$$f_E(x) := \min_{y \in E} f(y + x) \ \text{ for all } x \in \mathbb{R}^n,$$

is affine–quadratic. In particular,

i) if B is an $m \times n$ matrix, then the function f_B defined on $\mathbb{R}^m$ by

$$f_B(x) := \min_{By = x} f(y) \ \forall x \in \mathbb{R}^m,$$

is affine–quadratic.

ii) if f_1 and f_2 are linear–quadratics defined on $\mathbb{R}^n$, then the inf–convolution $f_1 \sharp f_2$ defined on $\mathbb{R}^n$ as

$$(f_1 \sharp f_2)(x) := \inf_{x_1+x_2=x} [f_1(x_1) + f_2(x_2)] \ \forall x \in \mathbb{R}^n,$$

is affine–quadratic.

Proof. By considering $E = \operatorname{img}(C)$ for some matrix C, the function f_E can be set as

$$f_E(x) = \langle Ax, x\rangle + \langle a, x\rangle + \alpha + g(x),$$

where

$$g(x) = \min_z [\langle ACz, Cz\rangle + \langle 2Ax + a, Cz\rangle].$$

So, $g(x)$ (or equivalently $f_E(x)$) is finite if and only if the matrix $D := C^t AC$ is positive semidefinite and there exists z such that $2Dz + C^t(2Ax + a) = 0$. One deduce that $g(x) = 0$ if and only if $Dz = 0$, which, without loss of generality, we can assume that D is invertible. It follows that $z = -\frac{1}{2} D^{-1} C^t (2Ax + a)$ and hence

$$g(x) = -\langle ACD^{-1}C^t Ax, x\rangle - \langle ACD^{-1}C^t a, x\rangle - \frac{1}{4}\langle CD^{-1}C^t a, a\rangle.$$

The result follows. □

The quadratic property can be characterized by means the parallelogram law as shows the following proposition.

Proposition 3.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function satisfying*

$$f(x+y) + f(x-y) = 2[f(x) + f(y)] \quad \forall x, y \in \mathbb{R}^n. \tag{3.1}$$

Then there exists an $n \times n$ symmetric matrix A such that

$$f(x) = \langle Ax, x\rangle \quad \forall x \in \mathbb{R}^n.$$

Proof. Taking $x = y = 0$ in (3.1) one get $f(0) = 0$ and hence

$$f(2x) = 4f(x) \quad \forall x \in \mathbb{R}^n. \tag{3.2}$$

Let $\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be defined as

$$\varphi(x, y) = \frac{1}{2}[f(x+y) - f(x) - f(y)] \quad \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{3.3}$$

It is symmetric and from (3.2),

$$\varphi(x, x) = f(x) \quad \forall x \in \mathbb{R}^n. \tag{3.4}$$

So, in order to prove the result it is enough to show that φ is bilinear. We first note that for any $x, y \in \mathbb{R}^n$,

$$\varphi(2x, y) = \frac{1}{2}[f(2x+y) - f(2x) - f(y)]$$

and hence by using (3.2), one has

$$\varphi(2x, y) = f(x+y) - f(x) - f(y) = 2\varphi(x, y). \tag{3.5}$$

For any x_1, x_2, y in $\mathbb{R}^n$, one has

$$\varphi(x_1 + x_2, y) = \frac{1}{2}\left[f(x_1 + \frac{1}{2}y + x_2 + \frac{1}{2}y) - f(x_1 + x_2) - f(y)\right]$$

which is equivalent to

$$\varphi(x_1+x_2,y) = f\left(x_1+\frac{y}{2}\right) + f\left(x_2+\frac{y}{2}\right) - \frac{1}{2}[f(x_1+x_2)+f(x_1-x_2)+f(y)]$$

or again to

$$\varphi(x_1+x_2,y) = f\left(x_1+\frac{y}{2}\right) + f\left(x_2+\frac{y}{2}\right) - f(x_1) - f(x_2) - 2f\left(\frac{y}{2}\right).$$

So, by using (3.5), one gets

$$\varphi(x_1+x_2,y) = \varphi(x_1,y)+\varphi(x_2,y).$$

φ being continuous and symmetric, one deduces that it is bilinear. □

Proposition 3.3. *Let $f:\mathbb{R}^n \to \mathbb{R}$ be a continuous function. It is affine–quadratic if and only if*

$$f(y+x)+f(y-x)+2f(0) = 2f(y)+f(x)+f(-x) \quad \forall x,y \in \mathbb{R}^n \tag{3.6}$$

Proof. It is clear that f verifies (3.6) if it is affine–quadratic. To prove the convere, let f_s be defined as

$$f_s(x) = \frac{1}{2}(f(x)+f(-x)) - f(0).$$

Then,

$$f_s(y+x)+f_s(y-x) = 2f_s(y)+2f_s(x),$$

and hence, in view of Proposition 3.2, there exists an $n\times n$ symmetric matrix A such that $f_s(x) = \langle Ax,x\rangle$. Now, let f_i be the function defined by

$$f_i(x) = \frac{1}{2}(f(x)-f(-x)).$$

It follows that $f = f_s+f_i+f(0)$ and from (3.6),

$$f_i(y+x)+f_i(y-x) = 2f_i(y) \quad \forall x,y\in\mathbb{R}^n,$$

which implies that

$$f_i(y_1+y_2) = f_i(y_1)+f_i(y_2) \quad \forall y_1,y_2\in\mathbb{R}^n.$$

f_i beings continuous, it is linear and hence there exists $a\in\mathbb{R}^n$ such that $f_i(x) = \langle a,x\rangle$ for all $x\in\mathbb{R}^n$. One deduce that

$$f(x) = \langle Ax,x\rangle + \langle a,x\rangle + f(0) \quad \forall x\in\mathbb{R}^n.$$

The proof follows. □

4. The duality gap function

For a given proper function $r:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, let h be the function defined on $\mathbb{R}^m$ by

$$h(u) := \inf[r(x) : Bx-b=u],$$

where B is an $m\times n$ matrix of maximal rank and $b\in\mathbb{R}^m$. Then, for each $u^*\in\mathbb{R}^m$ one has

$$h^*(u^*) = -\langle b,u^*\rangle + r^*(B^t u^*), \tag{4.1}$$

and hence for each $u \in \mathbb{R}^m$,

$$\begin{aligned} h^{**}(u) &= \sup_{u^*}[\langle u+b, u^*\rangle - r^*(B^t u^*)] \\ &= \sup_{u^*}\inf_x[r^{**}(x) + \langle u^*, u+b-Bx\rangle] \\ &\leq \inf_x \sup_{u^*}[r^{**}(x) + \langle u^*, u+b-Bx\rangle] \\ &= \inf_x[r^{**}(x) : Bx-b=u]. \end{aligned}$$

4.1. **Relationship between h^{**} and $\operatorname{co} h$.**

We denote by $\widehat{h}$ the convex function defined on $\mathbb{R}^m$ by

$$\widehat{h}(u) = \inf_x[r^{**}(x) : Bx-b=u]. \tag{4.2}$$

Clearly, $h^{**} \leq \widehat{h}$ and also $\widehat{h} \leq h$ because $r^{**} \leq r$. So,

$$h^{**}(u) \leq \widehat{h}(u) \leq \operatorname{co} h(u) \quad \forall u \in \mathbb{R}^m,$$

and hence, assuming $h^{**}(u) = \operatorname{co} h(u)$ (see the discussions given en Section 2), one gets

$$h^{**}(u) = \inf_x[r^{**}(x) : Bx-b=u].$$

In general the following identity is deduced.

Proposition 4.1. *For each $u \in \mathbb{R}^m$, it holds*

$$\operatorname{co} h(u) = \inf[\operatorname{co} r(x) : Bx-b=u].$$

Proof. Take $\varepsilon > 0$. There exist finitely many $u_j \in \mathbb{R}^m$ and $t_j \geq 0$ with $\sum t_j = 1$ such that $\sum t_j u_j = u$ and

$$\sum t_j h(u_j) < \operatorname{co} h(u) + \varepsilon.$$

For each j, there exists x_j satisfying $Bx_j - b = u_j$ such that

$$r(x_j) < h(u_j) + \varepsilon.$$

Set $x = \sum t_j x_j$. One has $Bx - b = u$ and

$$\operatorname{co} r(x) \leq \sum t_j r(x_j) \leq \sum t_j h(u_j) + \varepsilon \leq \operatorname{co} h(u) + 2\varepsilon$$

and hence

$$\inf[\operatorname{co} r(x) : Bx-b=u] \leq \operatorname{co} h(u).$$

To show the reverse inequality, take $u \in \mathbb{R}^m$ and let $x \in \mathbb{R}^n$ be such that $Bx - b = u$ and let finitely many of $x_j \in \mathbb{R}^n$ and $t_j \geq 0$ such that $\sum t_j = 1$ and $\sum t_j x_j = x$. Denoting $u_j := Bx_j - b$ one has $\sum t_j u_j = u$ and hence

$$\operatorname{co} h(u) \leq \sum t_j h(u_j).$$

On the other hand, for each j one has $h(u_j) \leq r(x_j)$ and, from the previous inequality,

$$\operatorname{co} h(u) \leq \sum u_j r(x_j).$$

One deduce that $\operatorname{co} h(u) \leq \operatorname{co} r(x)$ for all $x \in \mathbb{R}^n$ such that $Bx - b = u$, and hence

$$\operatorname{co} h(u) \leq \inf[\operatorname{co} r(x) : Bx-b=u].$$

The result is stablished. $\square$

The linear–quadratic case: We now assume that $r(x) = \langle Ax, x\rangle + \langle a, x\rangle + \varphi(Bx - b)$ for some symmetric matrix A, vector a and function φ, then

$$h(u) = \mathscr{Q}(u) + \varphi(u), \tag{4.3}$$

where

$$\mathscr{Q}(u) = \inf[\, \langle Ax, x\rangle + \langle a, x\rangle : Bx - b = u\,].$$

Due the Frank–Wolfe theorem [4], $\mathscr{Q}(u) > -\infty$ if and only if there exists $\bar{x}$ satisfying $B\bar{x} - b = u$ such that

$$\mathscr{Q}(u) = \langle A\bar{x}, \bar{x}\rangle + \langle a, \bar{x}\rangle,$$

which is also equivalent to

$$\langle Ay, y\rangle + \langle 2A\bar{x} + a, y\rangle \geq 0 \ \text{ for all } \ y \in \ker(B).$$

One deduce the following finiteness characterization. It also appears ($u = 0$) in [2], Corollary 4.3.

Proposition 4.2. $\mathscr{Q}(u) > -\infty$ *if and only if*

- A *is positive semidefinite on* $\ker(B)$*, and*
- $a \in \operatorname{img}(A) + \operatorname{img}(B^t)$.

Notice that this characterization does not depend on u.

Under this finiteness condition, $\mathscr{Q}$ is of real value because B is of maximal rank.

Since φ is more or less chosen, we can take it appropriately in order to get $\operatorname{co} h = h^{**}$ (see again the discussions given in Section 2), for example we can choose in (4.3), $\varphi(u) = \|u\|_3^3$, see Proposition 2.5.

The next two bounds are deduced for any function r.

Proposition 4.3. *It hold that*

$$\inf[\, r(x) - r^{**}(x) : Bx - b = u\,] \leq h(u) - h^{**}(u)$$

and

$$h(u) - \widehat{h}(u) \leq \sup[\, r(x) - r^{**}(x) : Bx - b = u\,].$$

Proof. Take $u \in \mathbb{R}^m$. One get $h^{**}(u) \leq r^{**}(x)$ for any x such that $Bx - b = u$ and hence $r(x) - r^{**}(x) \leq r(x) - h^{**}(u)$, which implies

$$\inf[\, r(x) - r^{**}(x) : Bx - b = u\,] \leq h(u) - h^{**}(u).$$

The definition of $\widehat{h}$ in (4.2) and the elementary inequality,

$$\inf f - \inf g \leq \sup(f - g),$$

for any functions f and g, the second bound also holds. $\square$

Remark 4.1. *If* $\operatorname{co} h = h^{**}$*, the two bounds in the previous proposition can be set as*

$$\begin{aligned}\inf[r(x) - r^{**}(x) : Bx - b = u] &\leq h(u) - h^{**}(u) \\ &\leq \sup[r(x) - r^{**}(x) : Bx - b = u].\end{aligned}$$

In general, in view of Proposition 4.1, one get

$$\begin{aligned}\inf[r(x) - \operatorname{co} r(x) : Bx - b = u] &\leq h(u) - \operatorname{co} h(u) \\ &\leq \sup[r(x) - \operatorname{co} r(x) : Bx - b = u].\end{aligned}$$

4.2. **Explicit expression of the duality gap function.**

For $f, \Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $y \in \mathbb{R}^n$ given, let $r_y : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be the function defined by

$$r_y(x) := f(x+y) + \Phi(x)$$

and its corresponding gap function

$$g_y := r_y - (r_y)^{**}.$$

Theorem 4.4. *Assume that* f *is affine–quadratic corresponding to a symmetric matrix* A*. If for some (and hence for any)* $y \in \mathbb{R}^n$*, the assumptions of Proposition 2.5 are fulfilled for* r_y*, then*

$$g_y(x) = \Phi(x) - [\langle A\cdot, \cdot\rangle + \Phi(x + \cdot)]^{**}(0) \text{ for all } x \in \mathbb{R}^n.$$

Proof. Take $\varepsilon > 0$. There exist finite families $x_j \in \mathbb{R}^n$ and $t_j \geq 0$ with $\sum t_j = 1$ such that $\sum t_j x_j = x$ and

$$\sum t_j [f(x_j + y) + \Phi(x_j)] = \sum t_j r_y(x_j) < r_y^{**}(x) + \varepsilon.$$

For each j, set $x'_j = x_j - x$. Then $\sum t_j x'_j = 0$ and so

$$\sum t_j [f(x_j + y) + \Phi(x_j)] = f(x+y) + \Phi(x) - \Phi(x) + \sum t_j [\langle A x'_j, x'_j\rangle + \Phi(x + x'_j)].$$

Hence,

$$r_y(x) - \Phi(x) + [\langle A\cdot, \cdot\rangle + \Phi(x + \cdot)]^{**}(0) \leq r_y^{**}(x).$$

To show the reverse inequality, let x'_j and $t_j \geq 0$ be finite families such that $\sum t_j = 1$ and $\sum t_j x'_j = 0$. Take $x \in \mathbb{R}^n$ and, for each j, $x_j = x'_j + x$. One get,

$$\begin{aligned} r_y^{**}(x) &\leq \sum t_j r_y(x_j) \\ &= \sum t_j [f(x'_j + x + y) + \Phi(x'_j + x)] \\ &= r_y(x) - \Phi(x) + \sum t_j [\langle A x'_j, x'_j\rangle + \Phi(x'_j + x)] \end{aligned}$$

One deduce that

$$r_y^{**}(x) \leq r_y(x) - \Phi(x) + [\langle A\cdot, \cdot\rangle + \Phi(x + \cdot)]^{**}(0).$$

The result is stablished. □

So, in view of Proposition 2.1, the next explicit expression is deduced.

Corollary 4.5. *If* Φ *is even, then for any* $y \in \mathbb{R}^n$*,*

$$g_y(0) = \Phi(0) - \inf_x [\langle Ax, x\rangle + \Phi(x)]. \tag{4.4}$$

Example 4.6. Let us consider the case where Φ is the indicator function of the unit ball $C = \{x \in \mathbb{R}^n : \|x\|_\infty \le 1\}$ and $A = (a_{ij})$ symmetric whose entries are nonnegative. Then

$$\inf_x [\langle Ax, x\rangle + \Phi(x)] = -\sum_{i,j=1}^n a_{ij},$$

and hence $g_y(0) = \sum_{i,j} a_{ij}$ for any $y \in \mathbb{R}^n$.

Proposition 4.7. *If for any even function Φ satisfying $\Phi(0) = 0$ one has*

$$r_y(0) - \operatorname{co} r_y(0) = c_\Phi \ \textit{ for all } y \in \mathbb{R}^n,$$

where c_Φ is a constant only depending on Φ, then f is affine–quadratic.

Proof. Take $a \in \mathbb{R}^n$ and for $\xi \in \mathbb{R}$ consider Φ defined as

$$\Phi(x) = \begin{cases} 0 & \text{if } x = 0, \\ \xi & \text{if } x = \pm a, \\ +\infty & \text{otherwise.} \end{cases}$$

One has

$$\operatorname{co}(f(y+\cdot) + \Phi(\cdot))(0) = \min[f(y), \frac{1}{2}(f(y+a) + f(y-a)) + \xi]$$

and hence

$$\min[f(y), \frac{1}{2}(f(y+a) + f(y-a)) + \xi] = f(y) + c_\Phi,$$

which implies, at $y = 0$,

$$\min[f(0), \frac{1}{2}(f(a) + f(-a)) + \xi] = f(0) + c_\Phi.$$

So, by taking ξ such that $\frac{1}{2}(f(a) + f(-a)) + \xi < f(0)$, one get $c_\Phi < 0$ and hence

$$\begin{aligned} \frac{1}{2}(f(y+a) + f(y-a)) &= f(y) + c_\Phi - \xi, \\ \frac{1}{2}(f(a) + f(-a)) &= f(0) + c_\Phi - \xi. \end{aligned}$$

One deduce that

$$f(y+a) + f(y-a) = 2f(y) + f(a) + f(-a) - 2f(0) \ \text{ for all } y, a \in \mathbb{R}^n,$$

which means, in view of Proposition 3.3, that f is affine–quadratic. □

4.3. **Boundedness of the duality gap function.**

For a family $\{r_j\}_{j\in J}$ of functions defined on $\mathbb{R}^n$ with value in $\mathbb{R} \cup \{+\infty\}$, let $\underline{r}$ and $\overline{r}$ be the functions defined on $\mathbb{R}^n$ as

$$\underline{r}(x) = \inf_{j\in J} r_j(x) \quad \text{and} \quad \overline{r}(x) = \sup_{j\in J} r_j(x),$$

and then for a matrix B and vector b, let $\underline{h}$ and $\overline{h}$ be the functions defined on $\mathbb{R}^m$ as

$$\underline{h}(u) = \inf_x [\underline{r}(x) : Bx - b = u] \quad \text{and} \quad \overline{h}(u) = \inf_x [\overline{r}(x) : Bx - b = u].$$

Also, for each j, let h_j be the function defined on $\mathbb{R}^m$ as

$$h_j(u) = \inf_x [r_j(x) : Bx - b = u].$$

The following upper bounds are deduced

Proposition 4.8. *It holds that*

$$\inf_{j\in J}(h_j - \operatorname{co} h_j) \le \underline{h} - \operatorname{co}\underline{h} \quad \textit{and} \quad \inf_{j\in J}(h_j - h_j^{**}) \le \underline{h} - \underline{h}^{**}.$$

In the particular case when r_j is defined as

$$r_j(x) = \langle Ax, x\rangle + \langle a_j, x\rangle + \varphi(Bx-b),$$

for some matrix A, vectors a_j and function φ, the following proposition gets a lower bound.

Proposition 4.9. *For each $u \in \mathbb{R}^m$, it holds*

$$(\overline{h} - \operatorname{co}\overline{h})(u) \le \sup_{j\in J}(h_j - \operatorname{co} h_j)(u) = \varphi(u) - \operatorname{co}[\langle A\cdot,\cdot\rangle + \varphi(u+B\cdot)](0).$$

Proof. Fix $j \in J$. Following the proof of Theorem 4.4 one has for each x,

$$\operatorname{co}\overline{r}(x) \ge \operatorname{co} r_j(x) = r_j(x) - [\varphi(Bx-b) - \operatorname{co}[\langle A\cdot,\cdot\rangle + \varphi(Bx-b+B\cdot)](0)]$$

and hence for each $x\in\mathbb{R}^n$ such that $Bx-b=u$ one has $\operatorname{co}\overline{r}(x) \ge \overline{r}(x) - [\varphi(u) - \operatorname{co}[\langle A\cdot,\cdot\rangle + \varphi(u+B\cdot)](0)]$ and hence by taking the infimum on both side one get the desired result in view of Proposition 4.1. □

5. The optimization problem

We now discuss the duality gap corresponding to the optimization problem

$$\overline{\alpha} := \min_x[f(x) : g(x) = 0], \tag{Opt}$$

where $f:\mathbb{R}^n \to \mathbb{R}$ and $g:\mathbb{R}^n\to\mathbb{R}^q$ are two given functions. Its corresponding Lagrangian function l defined on $\mathbb{R}^n\times\mathbb{R}^q$ is

$$l(x,v^*) = f(x) + \langle g(x), v^*\rangle, \tag{lp}$$

and then the associated primal and dual problems are

$$\overline{\theta} := \min_x[\,\alpha(x) := \max_{v^*} l(x,v^*)\,] \tag{P}$$

and

$$\underline{\theta} := \max_{v^*}[\,\beta(v^*) := \min_x l(x,v^*)\,]. \tag{D}$$

Clearly, $\overline{\theta} \ge \underline{\theta}$. The difference $\overline{\theta} - \underline{\theta}$ is called the *duality gap* between problems (P) and (D).

The corresponding marginal function h defined on R^q is

$$h(v) = \inf_x[f(x) : g(x) + v = 0]. \tag{P_v}$$

Of course, $h(0) = \overline{\alpha}$.

Both primal and dual optimal values can also be put in terms of h.

Lemma 5.1. *The following relationships hold*

$$\overline{\theta} = h(0) \quad \textit{and} \quad \underline{\theta} = h^{**}(0).$$

Proof. The first one follows because $\max_{v^*} l(x,v^*)$ coincides with $f(x)$ whenever $g(x) = 0$, and is $+\infty$ otherwise. The second one follows from the fact that $h^*(v^*) = -\inf_x[f(x) + \langle g(x),v^*\rangle]$ and hence

$$h^{**}(0) = \sup_{v^*} -h^*(v^*) = \sup_{v^*}\inf_x[f(x) + \langle g(x),v^*\rangle] = \underline{\theta}.$$

The aforementioned relationships follow. □

5.1. **The quadratic case.**

Let us consider now the case $f(x) = \langle Ax,x\rangle + \langle a,x\rangle + \Phi(x)$ and $g(x) = Bx-b$, where A is an $n\times n$ symmetric matrix, B an $m\times n$ matrix of maximal rank, $a\in\mathbb{R}^n$, $b\in\mathbb{R}^m$, and Φ a given function. Then, the associated optimization problem is

$$\overline{\alpha} := \min_x[r(x) := \langle Ax,x\rangle + \langle a,x\rangle + \Phi(x) : Bx-b = 0] \tag{q_E}$$

and its corresponding marginal function

$$h(u) := \min_x[r(x) : Bx-b = u]. \tag{q_{E_u}}$$

Theorem 5.2. *If $\Phi(x) = \varphi(Bx-b)$ for all $x\in\mathbb{R}^n$, for some function $\varphi : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$, then*

$$h(u) - \mathrm{co}\,h(u) = \varphi(u) - \mathrm{co}\,[\langle A\cdot,\cdot\rangle + \varphi(u+B\cdot)](0) \;\; \textit{for all } u\in\mathbb{R}^m.$$

Proof. Following the same idea from the proof of Theorem 4.4, one has

$$r(x) - \mathrm{co}\,r(x) = \varphi(Bx-b) - \mathrm{co}\,[\langle A\cdot,\cdot\rangle + \varphi(Bx-b+B\cdot)](0)$$

and hence, for each $u\in\mathbb{R}^m$,

$$\begin{aligned}\inf_x[r(x) - \mathrm{co}\,r(x) : Bx-b = u] &= \varphi(u) - \mathrm{co}\,[\langle A\cdot,\cdot\rangle + \varphi(u+B\cdot)](0)\\ &= \sup_x[r(x) - \mathrm{co}\,r(x) : Bx-b = u],\end{aligned}$$

which coincides with $h(u) - \mathrm{co}\,h(u)$ in view of Remark 4.1. The result follows. □

Remark 5.1. *Notice that the gap $h - \mathrm{co}\,h$ does not depend on b.*

Corollary 5.3. *If $h(u) - \mathrm{co}\,h(u) = 0$, then*

$$\langle Ax,x\rangle + \frac{1}{2}[\varphi(u+Bx) + \varphi(u-Bx)] \geq \varphi(u),\ \forall x\in\mathbb{R}^n,$$

and hence $q(x) := \langle Ax,x\rangle + \langle a,x\rangle$ is convex over $Bx = b$.

Corollary 5.4. *If φ is even, then*

$$h(0) - \mathrm{co}\,h(0) = \varphi(0) - \inf_x[\langle Ax,x\rangle + \varphi(Bx)].$$

We end this section by showing the boundedness of the duality gap function corresponding to problem (q_E) but considering it with inequality constraint. For that, let h_b be defined as

$$h_b(u) := \min_x[r(x) : Bx-b = u],$$

and then, for a given subset $C\subset\mathbb{R}^m$, the function h_C defined as

$$h_C(u) = \inf_{b\in C} h_b(u).$$

Corollary 5.5. *One get*

$$\operatorname{co} h_C(u) \le \inf_{b\in C} \operatorname{co} h_b(u) = h_C(u) - \varphi(u) + \operatorname{co}[\langle A\cdot,\cdot\rangle + \varphi(u+B\cdot)](0).$$

In particular, if $C = \bar{b} - \mathbb{R}^m_+$ *for some* $\bar{b} \in \mathbb{R}^m$*, then*

$$\operatorname{co} h(u) \le \inf_x [f(x) : Bx - u \le \bar{b}] + \operatorname{co}[\langle A\cdot,\cdot\rangle + \varphi(u+B\cdot)](0),$$

where

$$h(u) = \varphi(u) + \inf_x [f(x) : Bx - u \le \bar{b}].$$

Proof. For each $b \in C$, $h_C \le h_b$ and hence $\operatorname{co} h_C \le \operatorname{co} h_b$. The desired relationship is deduced from Theorem 5.2. □

6. Application: A primal-dual algorithm

Corresponding to problems (P) and (D) formulated through the lagrangian function defined in Section 5, the next result establishes an stopping criterium for global minimizers of problem (P). Its proof can be easily deduced from the definitions.

Proposition 6.1. *Let* $\bar{x}$ *and* $\bar{v}^*$ *be such that* $\alpha(\bar{x}) - \beta(\bar{v}^*) = \overline{\theta} - \underline{\theta} =: gap(P)$*. Then* $\bar{x}$ *is an optimal solution of* (P) *and* $\bar{v}^*$ *is an optimal solution of* (D)*. The converse is also true.*

The present algorithm provides a general scheme of a primal–dual algorithm in order to solve problem (P).

A primal–dual algorithm:

(1) **Initial step:** Take $v_0^* \in \mathbb{R}^q$ and do $k = 0$.
(2) **Step k:** Let x_k be an optimal solution of

$$\beta(v_k^*) = \min_x l(x, v_k^*) \tag{P_k}$$

   (a) If $\alpha(x_k) - \beta(v_k^*) = gap(P)$, STOP. x_k is an optimal solution of (P);
   (b) Otherwise, let $d_k \in \mathscr{A}(v_k^*, x_k, \cdots)$ be an ascending direction of β. Do $v_{k+1}^* = v_k^* + r_k d_k$ for an appropriate step size $r_k > 0$, and then $k = k+1$. Go to 2.

7. Conclusion

In this paper, we designed a general scheme of nonzero duality gap corresponding to a quadratic optimization problem subject to linear equality constraints, extending in some sense the same effect and properties generated by zero duality gap. In that sense, we introduced a general function Φ in the involved optimization problem whose effect is to change its structure, transforming the original problem, for example, into an unconstrained optimization problem. Another possible effect is to make the original optimization problem into another one with easier structure.

REFERENCES

[1] J.-P. Aubin, I. Ekeland, Estimates of the duality gap in nonconvex optimization, Math. Oper. Res. 1 (1976) 225–245.

[2] F. Flores–Bazán, G. Cárcamo, A geometric characterization of strong duality in nonconvex quadratic programming with linear and nonconvex quadratic constraints, Math. Program. 145 (2014) 263-290.

[3] F. Flores–Bazán, F. Thiele, On the lower semicontinuity of the value function and existence of solutions in quasiconvex optimization, J. Optim. Theory Appl. 195 (2022) 390–417.

[4] M. Frank, P. Wolfe, An algorithm for quadratic programming, Naval Research Logistics Quarterly 3 (1956) 95-110.

[5] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex analysis and minimization algorithms, II: Advanced Theory and Bundle Methods, Springer-Verlag, Berlin, 1993.

[6] R.T. Rockafellar, R.J.B. Wets, Variational Analysis, Springer-Verlag, Berlin, 1998.