J Glob Optim (2011) 49:173–183
DOI 10.1007/s10898-010-9592-y

Counterexamples to some triality and tri-duality results

M. D. Voisei · C. Zălinescu

Received: 23 October 2009 / Accepted: 12 August 2010 / Published online: 2 September 2010

Abstract The aim of this paper is to show that the results on triality and tri-duality in Gao (J Glob Optim 17:127–160, 2000; J Glob Optim 29:377–399, 2004; J Glob Optim 35:131–143, 2006; Encyclopedia of optimization, 2nd edn. Springer, New York, pp 822–828, 2009) and Gao et al. (J Glob Optim 45:473–497, 2009) are false. To prove this we provide simple counterexamples.

Keywords Counterexample · Triality · Tri-duality

1 Introduction

As D.Y. Gao says on his web-page http://www.math.vt.edu/people/gao/triality/triality.html, "the bi-minmax statements (tri-2, tri-3) were discovered in 1996 in a post-buckling analysis of large deformed beam model". The terms bi-duality, triality and tri-duality are intensively used in the book [1] and several other papers of D.Y. Gao.

This paper is organized as follows. In Sect. 2 the focus is on the triality and tri-duality results in [2, Th. 5, Th. 6] while Sect. 3 deals with several counterexamples to the results in [4] concerning a quadratic programming problem. The counterexample in Sect. 4 refers to a statement in [5]. This paper concludes with a counterexample to a triality theorem in [6].

M. D. Voisei
Department of Mathematics, Towson University, Towson, MD, USA
e-mail: mvoisei@towson.edu

C. Zălinescu (✉)
Faculty of Mathematics,
University "Al. I. Cuza" Iaşi, Iaşi, Romania
e-mail: zalinesc@uaic.ro

C. Zălinescu
Institute of Mathematics Octav Mayer, Iaşi, Romania

In the sequel we make several quotations from the respective papers of D.Y. Gao and his collaborators in order to make the reading and the understanding of the framework more easy a process.

2 Counterexamples to some results in [2,3]

In [2] one considers $\mathcal{X}$, $\mathcal{Y}$ as being locally convex spaces. "The so-called *geometrical operator* $\Lambda : \mathcal{X} \to \mathcal{Y}$ is a continuous, Gâteaux differentiable operator such that for any given $x \in \mathcal{X}_a \subset \mathcal{X}$, there exists a $y \in \mathcal{Y}_a \subset \mathcal{Y}$ satisfying the geometrical equation $y = \Lambda x$".

The so called extended Lagrangian is introduced as

$$\text{“}L(x, y^*) = \left\langle \Lambda(x); y^* \right\rangle - W^*(y^*) - \langle x, c\rangle \quad (5.65)\text{”},$$

where $W \in \Gamma(\mathcal{Y})$ (meaning that either W is convex, denoted by $W \in \check{\Gamma}(\mathcal{Y})$, or W is concave, denoted by $W \in \hat{\Gamma}(\mathcal{Y})$), W^* is its corresponding Fenchel conjugate (see [2, p. 131]), and $c \in \mathcal{X}^*$.

On [2, p. 148] one says "Let $\mathcal{Z}_a = \mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{Z}$" $= \mathcal{X} \times \mathcal{Y}^*$ "be the effective domain of L," that is $\mathcal{Z}_a = \{(x, y^*) \in \mathcal{X} \times \mathcal{Y}^* \mid |L(x, y^*)| < +\infty\}$ (see [2, p. 127]), "and let $\mathcal{L}_c \subset \mathcal{Z}_a$ be a critical point set of L, i.e.

$$\mathcal{L}_c = \{(\overline{x}, \overline{y}) \in \mathcal{X}_a \times \mathcal{Y}_a^* \mid \delta L(\overline{x}, \overline{y}^*; x, y^*) = 0 \quad \forall (x, y^*) \in \mathcal{X}_a \times \mathcal{Y}_a^*\}.$$

For any given critical point $(\overline{x}, \overline{y}^*) \in \mathcal{L}_c$, we let $\mathcal{X}_r \times \mathcal{Y}_r^*$ be its neighborhood such that on $\mathcal{X}_r \times \mathcal{Y}_r^*$, the pair $(\overline{x}, \overline{y}^*)$ is the only critical point of L. The following result is of fundamental importance in global optimization.

Theorem 5 (Triality theorem) *Let $(\overline{x}, \overline{y}^*) \in \mathcal{L}_c$ be a critical point of L and $\mathcal{X}_r \times \mathcal{Y}_r^*$ a neighborhood of $(\overline{x}, \overline{y}^*)$.*
I. Suppose that $W \in \check{\Gamma}(\mathcal{Y}_a)$ is convex. If $\left\langle \Lambda(x); \overline{y}^* \right\rangle$ is convex on $\mathcal{X}_r$, then

$$\min_{x \in \mathcal{X}_r} \max_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\overline{x}, \overline{y}^*) = \max_{y^* \in \mathcal{Y}_r^*} \min_{x \in \mathcal{X}_r} L(x, y^*). \quad (5.66)$$

However, if $\left\langle \Lambda(x); \overline{y}^* \right\rangle$ is concave on $\mathcal{X}_r$, then either

$$\min_{x \in \mathcal{X}_r} \max_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\overline{x}, \overline{y}^*) = \min_{y^* \in \mathcal{Y}_r^*} \max_{x \in \mathcal{X}_r} L(x, y^*), \quad (5.67)$$

or

$$\max_{x \in \mathcal{X}_r} \max_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\overline{x}, \overline{y}^*) = \max_{y^* \in \mathcal{Y}_r^*} \max_{x \in \mathcal{X}_r} L(x, y^*). \quad (5.68)\text{”}$$

Part II of [2, Th.5] is obtained from part I by changing L in $-L$. Note that the statement of [3, Th.3] is the same as that of [2, Th.5].

In order to make the writing simpler, we replace y^* by ς in Examples 1 and 2 below.

Example 1 Let X be a Hilbert space with $\dim X \geq 2$, $Y = \mathbb{R}$, $\Lambda(x) = -\frac{1}{2}\|x\|^2$, $W(y) = \frac{1}{2}y^2 + 3y$ (hence W is convex and $W^*(\varsigma) = \frac{1}{2}(\varsigma - 3)^2$), $c \in X$ with $\|c\| = 2$. Hence

$$L(x, \varsigma) = -\frac{1}{2}\varsigma \|x\|^2 - \frac{1}{2}(\varsigma - 3)^2 - \langle c, x\rangle. \quad (1)$$

Then $\nabla_x L(x, \varsigma) = -c - \varsigma x$, $\nabla_\varsigma L(x, \varsigma) = -\frac{1}{2}\|x\|^2 - \varsigma + 3$. If (x, ς) is a critical point for L then $\varsigma \neq 0$ because $c \neq 0$ and so $x = -\varsigma^{-1}c$; hence $\|x\|^2 = 4/\varsigma^2 = 6 - 2\varsigma$. It follows that $\varsigma \in \{1, 1+\sqrt{3}, 1-\sqrt{3}\}$. Hence the critical points of L are $(-c, 1)$, $\left(\frac{1+\sqrt{3}}{2}c, 1-\sqrt{3}\right)$, and $\left(\frac{1-\sqrt{3}}{2}c, 1+\sqrt{3}\right)$. Clearly, all the critical points of L are isolated.

Consider $(\overline{x}, \overline{\varsigma}) := (-c, 1)$. We fix the neighborhood $U_0 := \{x \in X \mid \|x-\overline{x}\| \leq 1\}$ of $\overline{x}$ and the neighborhood $V_0 := (0, \infty)$ of $\overline{\varsigma}$. We have that $(\overline{x}, \overline{\varsigma})$ is the only critical point of L in $U_0 \times V_0$ since $\|\frac{-1\pm\sqrt{3}}{2}\overline{x} - \overline{x}\| = 3 \mp \sqrt{3} > 1$.

Because $\langle \Lambda(x); \overline{\varsigma}\rangle = -\frac{1}{2}\|x\|^2$, the hypotheses of the second portion of part I in [2, Th.5] hold. For $t > 0$, $x_t := (1-t)\overline{x}$, $\varsigma_t := 1+t$ we have

$$L(x_t, \varsigma_t) = -2(1+t)(1-t)^2 - \frac{1}{2}(t-2)^2 + 4(1-t) = t^2\left(\frac{3}{2} - 2t\right).$$

For any $U \times V$ a neighborhood of $(\overline{x}, \overline{\varsigma})$ there exists $\varepsilon \in (0, 3/4)$ such that $(x_t, \varsigma_t) \in U \times V$, for every $t \in (0, \varepsilon)$, and so $L(x_t, \varsigma_t) > 0 = L(\overline{x}, \overline{\varsigma})$. Hence $\sup_{x\in U}\sup_{\varsigma\in V} L(x, \varsigma) > L(\overline{x}, \overline{\varsigma})$, whence [2, (5.68)] does not hold on any neighborhood of $(\overline{x}, \overline{\varsigma})$.

Take again $U \times V$ an arbitrary neighborhood of $(\overline{x}, \overline{\varsigma})$ and $x \in U$.

Since $\nabla_\varsigma L(x, 3 - \frac{1}{2}\|x\|^2) = 0$ and $L(x, \cdot)$ is concave we have that

$$\sup_{\varsigma\in V} L(x,\varsigma) \leq \sup_{\varsigma\in\mathbb{R}} L(x,\varsigma) = L\left(x, 3 - \frac{1}{2}\|x\|^2\right) = \frac{1}{8}\|x\|^4 - \frac{3}{2}\|x\|^2 + \langle \overline{x}, x\rangle =: Q(x), \tag{2}$$

due to the fact that every concave function reaches its global maximum value at its critical points.

For every $u \in X$ with $\langle \overline{x}, u\rangle = 0$ we have

$$\begin{aligned} Q(\overline{x}+u) &= \frac{1}{8}(\|\overline{x}\|^2 + \|u\|^2)^2 - \frac{3}{2}(\|\overline{x}\|^2 + \|u\|^2) + \|\overline{x}\|^2 \\ &= \frac{1}{8}(4 + \|u\|^2)^2 - \frac{3}{2}(4 + \|u\|^2) + 4 = -\frac{1}{2}\|u\|^2\left(1 - \frac{1}{4}\|u\|^2\right). \end{aligned}$$

Since $\dim X \geq 2$, there exists $u \in X$ such that $\overline{x}+u \in U$, $\langle \overline{x}, u\rangle = 0$ and $0 < \|u\| < 2$; hence $Q(\overline{x}+u) < 0$. It follows that

$$\inf_{x\in U}\sup_{\varsigma\in V} L(x,\varsigma) \leq \inf_{x\in U} Q(x) \leq Q(\overline{x}+u) < 0 = L(\overline{x}, \overline{\varsigma}) = Q(\overline{x}). \tag{3}$$

Hence [2, (5.68)] also does not hold on any neighborhood of $(\overline{x}, \overline{\varsigma})$. Therefore [2, Th.5] is false.

On [2, p. 149] one says that the next result "is a special case of the triality theorem".

"**Theorem 6** (Tri-duality theorem) *Suppose that $W \in \check{\Gamma}(\mathcal{Y}_a)$, $(\overline{x}, \overline{y}^*) \in \mathcal{L}_c$ is a critical point of L and $\mathcal{X}_r \times \mathcal{Y}_r^*$ is a neighborhood of $(\overline{x}, \overline{y}^*)$. If $\langle \Lambda(x); \overline{y}^*\rangle$ is convex on $\mathcal{X}_r$, then*

$$P(\overline{x}) = \min_{x\in\mathcal{X}_r} P(x) \text{ if and only if } P^d(\overline{y}^*) = \max_{y^*\in\mathcal{Y}_r^*} P^d(y^*). \quad (5.77)$$

However, if $\langle \Lambda(x); \overline{y}^*\rangle$ is concave on $\mathcal{X}_r$, then

$$P(\overline{x}) = \min_{x\in\mathcal{X}_r} P(x) \text{ if and only if } P^d(\overline{y}^*) = \min_{y^*\in\mathcal{Y}_r^*} P^d(y^*), \quad (5.78)$$

$$P(\overline{x}) = \max_{x\in\mathcal{X}_r} P(x) \text{ if and only if } P^d(\overline{y}^*) = \max_{y^*\in\mathcal{Y}_r^*} P^d(y^*). \quad (5.79)\text{"}$$

In the above statement

"$P(x) = W(\Lambda(x)) - \langle x; c \rangle, c \in \mathcal{X}^*$ (5.64)" *and*
"$P^d(y^*) = \text{sta}\{L(x, y^*) \mid x \in \mathcal{X}\} = -G^d(y^*) - W^*(y^*) \quad \forall y^* \in \mathcal{Y}_s^*$, (5.74)

where $G^d : \mathcal{Y}^* \to \mathbb{R}$ is the so-called pure complementary gap function" and "sta" stands for finding critical values.

The statement of [3, Th. 4] is similar to the statement of the previous theorem.

Example 2 Consider the framework of Example 1. Then (using also (2))

$$P(x) = W(\frac{1}{2}\|x\|^2) + \langle x, \overline{x} \rangle = \frac{1}{8}\|x\|^4 - \frac{3}{2}\|x\|^2 + \langle \overline{x}, x \rangle = \sup_{\varsigma \in \mathbb{R}} L(x, \varsigma).$$

Also, using (1), we find

$$P^d(\varsigma) = \text{sta}\{L(x, \varsigma) \mid x \in X\} = L(\varsigma^{-1}\overline{x}, \varsigma) = 2\varsigma^{-1} - \frac{1}{2}(\varsigma - 3)^2 = \frac{1}{2}(\varsigma - 1)^2(4 - \varsigma)/\varsigma$$

for $\varsigma \neq 0$. It is easily seen that $\overline{\varsigma} = 1$ is a local minimum point of P^d. However, as seen in Example 1 (note that $Q = P$ and recall (3)), $\overline{x}$ is not a minimum point of P on any neighborhood of $\overline{x}$ if $\dim X \geq 2$. Hence [2, Th.6] (more precisely, (5.78)) does not hold.

3 Counterexamples to some results in [4]

In [4] one considers the problem $(\mathcal{P})$ min $P(x) := \frac{1}{2}x^T Ax - x^T f$ subject to $x \in \mathcal{X}_f$, where $\mathcal{X}_f := \{x \in \mathbb{R}^n \mid Bx \leq b\}$, with A a non-singular symmetric real $n \times n$ matrix, B a real $m \times n$ matrix, $f \in \mathbb{R}^n, b \in \mathbb{R}^m$, and $\nu_1 \leq \nu_2$ means that $\nu_2 - \nu_1 \in \mathbb{R}_+^m := \{z \in \mathbb{R}^m \mid z_i \geq 0 \ \forall i \in \overline{1, m}\}$. One considers the Lagrangian function $L : \mathbb{R}^n \times \mathbb{R}_+^m \to \mathbb{R}$ defined by

$$L(x, \nu) := \frac{1}{2}x^T Ax - x^T f + \nu^T(Bx - b).$$

The element $(\overline{x}, \overline{\nu})$ is a KKT point for $(\mathcal{P})$ if

$$A\overline{x} + B^T\overline{\nu} = f, \quad \overline{\nu} \geq 0, \quad B\overline{x} \leq b, \quad \overline{\nu}^T(B\overline{x} - b) = 0. \tag{4}$$

Note that $\nabla_x L(x, \nu) = Ax - f + B^T\nu, \nabla_\nu L(x, \nu) = Bx - b$; hence $\nabla_x L(\overline{x}, \overline{\nu}) = 0$ whenever $(\overline{x}, \overline{\nu})$ is KKT point for $(\mathcal{P})$. On [4, p. 394], one introduces

$$\begin{aligned} P^d(\nu) &:= \text{ext}\{L(x, \nu) \mid x \in \mathbb{R}^n\} = L(A^{-1}(f - B^T\nu), \nu) \\ &= -\frac{1}{2}(f - B^T\nu)^T A^{-1}(f - B^T\nu) - \nu^T b, \end{aligned}$$

where "ext" stands for finding all the extremum values. The value ext $L(\cdot, \nu)$ is unique when A is either positive or negative definite (and implicitly invertible) and is obtained at critical points of $L(\cdot, \nu)$, i.e., $\nabla_x L(x, \nu) = Ax - f + B^T\nu = 0$ or equivalently at $x = A^{-1}(f - B^T\nu)$. More precisely, ext $L(\cdot, \nu)$ is a global maximum value, that is, $P^d(\nu) = \sup_{x \in \mathbb{R}^n} L(x, \nu)$ for $\nu \in \mathbb{R}^m$, when A is negative definite since in this case $L(\cdot, \nu)$ is concave and it is a global minimum value, i.e., $P^d(\nu) = \inf_{x \in \mathbb{R}^n} L(x, \nu)$ for $\nu \in \mathbb{R}^m$, when A is positive definite since in this case $L(\cdot, \nu)$ is convex.

On [4, page 395] one says: "Then by the general triality theory developed in [11], we have

Lemma 1 (Triality for canonical quadratic programming) *Suppose that the vector* $(\overline{x}, \overline{\nu})$ *is a KKT point of the quadratic programming problem* $(\mathcal{P})$.

If A *is positive definite, then the saddle minmax theorem in the form*

$$\min_{x\in\mathcal{X}_f}\max_{\nu\geq 0} L(x,\nu) = L(\overline{x},\overline{\nu}) = \max_{\nu\geq 0}\min_{x\in\mathbb{R}^n} L(x,\nu) \quad (56)$$

holds.

If A *is negative definite, then either the super-maximum theorem in the form*

$$\max_{x\in\mathcal{X}_f}\max_{\nu\geq 0} L(x,\nu) = L(\overline{x},\overline{\nu} = \max_{\nu\geq 0}\max_{x\in\mathbb{R}^n} L(x,\nu) \quad (57)$$

holds, or the super-minimum theorem in the form

$$\min_{x\in\mathcal{X}_f}\max_{\nu\geq 0} L(x,\nu) = L(\overline{x},\overline{\nu}) = \min_{\nu\geq 0}\max_{x\in\mathbb{R}^n} L(x,\nu) \quad (58)$$

holds."

The reference [11] mentioned above is our reference [1].

Assume that A is negative definite. On one hand

$$\begin{aligned}\sup_{\nu\geq 0}\sup_{x\in\mathbb{R}^n} L(x,\nu) &= \sup_{x\in\mathbb{R}^n}\sup_{\nu\geq 0} L(x,\nu) = \sup_{x\in\mathbb{R}^n}\sup_{\nu\geq 0}\left[\frac{1}{2}x^TAx - x^Tf + \nu^T(Bx-b)\right]\\ &= \sup_{x\in\mathbb{R}^n}\left[\frac{1}{2}x^TAx - x^Tf + \sup_{\nu\geq 0}\nu^T(Bx-b)\right] = \infty,\end{aligned}$$

whenever there exists $x_0 \in \mathbb{R}^n$ such that $b - Bx_0 \notin \mathbb{R}^m_+$. Note that $b - Bx \in \mathbb{R}^m_+$ for every $x \in \mathbb{R}^n$ iff $B = 0$ and $b \in \mathbb{R}^m_+$. So, assuming that $B \neq 0$ or $b \notin \mathbb{R}^m_+$, we have that (57) in [4, Lem.1] does not hold.

On the other hand,

$$\inf_{x\in\mathcal{X}_f}\sup_{\nu\geq 0} L(x,\nu) = \inf_{x\in\mathcal{X}_f}\left[\frac{1}{2}x^TAx - x^Tf\right] = -\infty$$

whenever $\mathcal{X}_f \neq \emptyset$ and there exists $x_0 \in \mathbb{R}^n \setminus \{0\}$ such that $Bx_0 \geq 0$ (such a B is easily found; take $m = n$ and $B = I_m$). Indeed, in this case $B(\overline{x} - tx_0) \leq b$, for every $t > 0$, that is, $\overline{x} - tx_0 \in \mathcal{X}_f$ for every $t > 0$; it follows that

$$\inf_{x\in\mathcal{X}_f}\left[\frac{1}{2}x^TAx - x^Tf\right] \leq \inf_{t>0}\left[\frac{1}{2}(\overline{x}-tx_0)^TA(\overline{x}-tx_0) - (\overline{x}-tx_0)^Tf\right] = -\infty$$

because the expression inside the infimum is a quadratic polynomial in t with negative leading coefficient $\frac{1}{2}x_0^TAx_0$. Hence (58) in [4, Lem.1] does not hold.

Therefore, when A is negative definite [4, Lem.1] is false.

On [4, p. 382] one considers also a parametrization of problem $(\mathcal{P})$:

$$\text{“}(\mathcal{P}_\mu): \min_{x\in\mathbb{R}^n} P(x) = \frac{1}{2}x^TAx - f^Tx, \quad (12)$$

$$\text{s.t.} Bx \leq b,\ |x|^2 \leq 2\mu. \quad (13)$$

Since for a given $\mu > 0$, the feasible space

$$\mathcal{X}_\mu = \left\{x \in \mathbb{R}^n \mid Bx \leq b,\ \frac{1}{2}|x|^2 \leq \mu\right\} \quad (14)$$

is a closed convex subset of $\mathbb{R}^n$, the parametric optimization problem $(\mathcal{P}_\mu)$ has at least one global minimizer $\overline{x}_\mu$."

Actually, $(\mathcal{P}_\mu)$ has at least one global minimizer because $\mathcal{X}_\mu$ is compact (that is, bounded and closed) in $\mathbb{R}^n$.

A dual problem is also introduced as follows: "On the dual feasible space defined by

$$\begin{aligned}\mathcal{Y}^*_\mu &= \{(\epsilon^*, \rho^*) \in \mathcal{Y}^*_a \mid \det(A + \rho^* I) \neq 0\},\\ &= \{(\epsilon^*, \rho^*) \in \mathbb{R}^m \times \mathbb{R} \mid \epsilon^* \geq 0,\ \rho^* \geq 0,\ \det(A + \rho^* I) \neq 0\},\end{aligned}$$

the canonical dual function $P^d(y^*) = \overline{F}^\Lambda(y^*) - \overline{W}^\natural(y^*)$ takes the following form

$$P^d(\in^*, \rho^*) = -\frac{1}{2}(f - B^T\epsilon^*)^T(A + \rho^* I)^{-1}(f - B^T\epsilon^*) - \mu\rho^* - b^T\epsilon^* \quad (18)$$

Thus, the canonical dual problem ($(\mathcal{P}^d_\mu)$ in short) associated with the parametric problem can be eventually formulated as the following (See Gao 2003)

$$\begin{aligned}&(\mathcal{P}^d_\mu) : \text{ext } P^d(\epsilon^*, \rho^*) \quad (19)\\ &\text{s.t.} \quad \epsilon^* \geq 0,\ \rho^* \geq 0,\ \det(A + \rho^* I) \neq 0, \quad (20)\end{aligned}$$

where ext $P(x)$ stands for finding all the extremum values of $P(x)$."

Note that in [4] there is no reference Gao 2003, but 2003a and 2003b.

It is worth observing that $\mathcal{Y}^*_\mu = \mathbb{R}^m_+ \times \{\rho \in \mathbb{R} \mid \rho \geq 0, \det(A + \rho I) \neq 0\}$.

On [4, p. 384] one finds the following definition.

"**Definition 1** (Index of the matrix A) Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. The index i_d of A is defined to be the total number of distinct negative eigenvalues of A.

By this definition, the quadratic function $P(x)$ is nonconvex if and only if the index $i_d > 0$. Suppose that the vector $(\bar{\epsilon}^*_i, \bar{\rho}^*_i)$ is a KKT point of $(\mathcal{P}^d_\mu)$, and the matrix A with index i_d has $p \leq n$ distinct eigenvalues $\{a_i\}$, $i = 1, \ldots, p \leq n$ in the order of

$$a_1 < a_2 < \cdots < a_{i_d} < 0 \leq a_{i_d+1} < \cdots < a_p.$$

Then, if $\bar{\rho}^*_i > -a_1$, the matrix $A + \bar{\rho}^*_i I$ is positive definite. However, if $a_{i_d+1} = \cdots = a_p = 0$ and the KKT point $\bar{\rho}^*_i < -a_{i_d}$, the matrix $A + \bar{\rho}^*_i I$ will be negative definite. Let

$$\begin{aligned}\mathcal{Y}^*_{\mu+} &= \{(\epsilon^*, \rho^*) \in \mathcal{Y}^*_\mu \mid A + \rho^* I \text{ is positive definite}\} \quad (27)\\ \mathcal{Y}^*_{\mu-} &= \{(\epsilon^*, \rho^*) \in \mathcal{Y}^*_\mu \mid A + \rho^* I \text{ is negative definite}\} \quad (28)\\ \mathcal{Y}^*_{\mu i} &= \{(\epsilon^*, \rho^*) \in \mathcal{Y}^*_{\mu-} \mid \rho^* > 0)\}, \quad (29)\end{aligned}$$

and

$$\mathcal{X}_{\mu s} = \left\{x \in \mathcal{X}_\mu \mid \frac{1}{2}|x|^2 = \mu\right\}. \quad (30)\text{''}$$

Observe that $A + \rho I$ is positive definite iff $\rho > -a_1$ and $A + \rho I$ is negative definite iff $\rho < -a_p$. Hence if $i_d \neq 0$ and $p = i_d$ then

$$\mathcal{Y}^*_{\mu+} = \mathbb{R}^m_+ \times (-a_1, \infty), \quad \mathcal{Y}^*_{\mu-} = \mathbb{R}^m_+ \times [0, -a_{i_d}); \tag{5}$$

moreover, if $p > i_d$ then $\mathcal{Y}^*_{\mu-} = \emptyset$ (contrary to what is said above, namely that $\bar{\rho}^*_i \in \mathcal{Y}^*_{\mu-}$).

One continues with: "Based on the triality theory developed in [11] as well as the recent result (see Gao 2003), we have the following interest result.

Theorem 2 (Local and global extrema) *Suppose that the matrix A has no zero index $i_d > 0$, and for a given parameter $\mu > 0$, the vector $(\bar{\epsilon}_i^*, \bar{\rho}_i^*)$ is a KKT point of the duals problem $(\mathcal{P}_\mu^d)$, and let $\overline{x}_i = (A + \bar{\rho}_i^* I)^{-1}(f - B^T \bar{\epsilon}_i^*)$.*

If $\bar{\rho}_i^* > -a_1$, then the vector $(\bar{\epsilon}_i^*, \bar{\rho}_i^*)$ is a maximizer of P^d on $\mathcal{Y}_{\mu+}^*$ if and only if the vector $\overline{x}_i$ is a minimizer of P on $\mathcal{X}_{\mu s}$, and

$$P(\overline{x}_i) = \min_{x \in \mathcal{X}_{\mu s}} P(x) = \max_{(\epsilon^*, \rho^*) \in \mathcal{Y}_{\mu+}^*} P^d(\epsilon^*, \rho^*) = P^d(\bar{\epsilon}_i^*, \bar{\rho}_i^*). \quad (31)$$

If $0 \le \bar{\rho}_i^* < -a_{i_d}$, then $(\bar{\epsilon}_i^*, \bar{\rho}_i^*)$ is a maximizer of P^d on $\mathcal{Y}_{\mu-}^*$ if and only if $\overline{x}_i$ is a global maximizer of P on $\mathcal{X}_\mu$, and

$$P(\overline{x}_i) = \max_{x \in \mathcal{X}_\mu} P(x) = \max_{(\epsilon^*, \rho^*) \in \mathcal{Y}_{\mu-}^*} P^d(\epsilon^*, \rho^*) = P^d(\bar{\epsilon}_i^*, \bar{\rho}_i^*) \quad (32)$$

If $0 < \bar{\rho}_i^* < -a_{i_d}$, the KKT point $(\bar{\epsilon}_i^*, \bar{\rho}_i^*)$ is a minimizer of P^d on the open set $\mathcal{Y}_{\mu i}^*$ if and only if $\overline{x}_i$ is a minimizer of P on the set $\mathcal{X}_{\mu s}$, and

$$P(\overline{x}_i) = \min_{x \in \mathcal{X}_{\mu s}} P(x) = \min_{(\epsilon^*, \rho^*) \in \mathcal{Y}_{\mu i}^*} P^d(\epsilon^*, \rho^*) = P^d(\bar{\epsilon}_i^*, \bar{\rho}_i^*). \quad (33)\text{”}$$

First note that the index i in the above statement is not related to any eigenvalue a_i of A. Second, the set $\mathcal{Y}_{\mu i}^*$ is open in $\mathbb{R}^m \times \mathbb{R}$ only if $\mathcal{Y}_{\mu i}^*$ is empty. Third, as seen above, if $p > i_d$ then the sets $\mathcal{Y}_{\mu-}^*, \mathcal{Y}_{\mu i}^*$ are empty; so, if we find some KKT point $(\bar{\epsilon}_i^*, \bar{\rho}_i^*)$ of $(\mathcal{P}_\mu^d)$ such that $0 < \bar{\rho}_i^* < -a_{i_d}$ then $(\bar{\epsilon}_i^*, \bar{\rho}_i^*) \notin \mathcal{Y}_{\mu-}^*, \mathcal{Y}_{\mu i}^* = \emptyset$ and the last two parts of the previous theorem conclusion do not make sense as we will see more precisely in the next example. Hence, in order for the last two cases of [4, Th.2] to be meaningful one must have that $p = i_d$, in which case (5) holds.

Example 3 Let $A = \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}$, $f = (-2, 0)^T$, $B = (-1, 0)$, $b = -1$, $\mu = \frac{1}{2}$. Then

$$P(x_1, x_2) = -x_1^2 + \frac{1}{2}x_2^2 + 2x_1, \quad P^d(\varepsilon, \rho) = -\frac{1}{2}\frac{(\varepsilon - 2)^2}{\rho - 2} - \frac{1}{2}\rho + \varepsilon = 1 + \frac{(\varepsilon - \rho)^2}{2(2 - \rho)}. \quad (6)$$

We have that $(\bar{\varepsilon}, \bar{\rho}) = (1, 1)$ is a KKT point for $(\mathcal{P}_\mu^d)$ with $0 < \bar{\rho} < -a_{i_d} = 2$; more precisely $\delta P^d(\bar{\varepsilon}, \bar{\rho}) = 0$. Moreover, $\overline{x} = (1, 0)^T$, $\mathcal{X}_\mu = \mathcal{X}_{\mu s} = \{\overline{x}\}$, and $\mathcal{Y}_{\mu+}^* = \mathbb{R}_+ \times (2, \infty)$ from which

$$P(\overline{x}) = \min_{x \in \mathcal{X}_{\mu s}} P(x) = \max_{x \in \mathcal{X}_\mu} P(x) = 1 = \max_{(\varepsilon, \rho) \in \mathcal{Y}_{\mu+}^*} P^d(\varepsilon, \rho) = P^d(1, 1). \quad (7)$$

Clearly, $\mathcal{Y}_{\mu-}^* = \mathcal{Y}_{\mu i}^* = \emptyset$, $A + \bar{\rho} I$ is not negative definite, and $(\bar{\varepsilon}, \bar{\rho})$ cannot belong to $\mathcal{Y}_{\mu-}^*$ even though $0 < \bar{\rho} < -a_{i_d} = 2$. These facts show that the last two parts of the conclusion of [4, Th.2] do not hold because they do not make sense.

A natural question is whether [4, Th.2] is true when $p = i_d$, that is, A is negative definite. To this end we have the following example.

Example 4 Let $A = \mathrm{Diag}(-2, -4)$, $f = (-2, 0)^T$, $B = (-1, 0)$, $b = -1$, $\mu = \frac{1}{2}$. Then $P(x_1, x_2) = -x_1^2 - 2x_2^2 + 2x_1$ and P^d is as in (6). As in Example 3 we have that $(\bar{\varepsilon}, \bar{\rho}) = (1, 1)$ is a KKT point for $(\mathcal{P}_\mu^d)$ with $0 < \bar{\rho} < -a_{i_d} = 2$, $\overline{x} = (1, 0)^T$, $\mathcal{X}_\mu = \mathcal{X}_{\mu s} = \{\overline{x}\}$, and (7) holds (with $\mathcal{Y}_{\mu+}^* = \mathbb{R}_+ \times (4, \infty)$). Moreover, $(\bar{\varepsilon}, \bar{\rho}) \in \mathcal{Y}_{\mu i}^* = \mathbb{R}_+ \times (0, 2) \subset \mathcal{Y}_{\mu-}^* =$

$\mathbb{R}_+ \times [0, 2)$. From the expression of P^d in (6) it is clear that relation (32) in [4, Th.2] does not hold since $\max_{(\varepsilon,\rho)\in\mathcal{Y}^*_{\mu-}} P^d(\varepsilon, \rho) = +\infty$. Hence [4, Th.2] is false.

4 Counterexample to a statement in [5]

On [5, p. 135] one says: "The primal problem $(\mathcal{P})$ for $p = 1$ is to find all critical points of the nonconvex function

$$P(x) = \frac{1}{2}\alpha_1 \left(\frac{1}{2}|x|^2 - \lambda_1\right)^2 - x^T f.$$

The canonical dual function for this simple case is

$$P^d(\varsigma) = -\frac{|f|^2}{2\varsigma} - \frac{1}{2}\alpha_1^{-1}\varsigma^2 - \varsigma\lambda_1."$$

On p. 136 of [5] one considers

$$\text{“}L(x, \varsigma) = \frac{1}{2}|x|^2\varsigma - \frac{1}{2}\alpha_1^{-1}\varsigma^2 - \varsigma\lambda_1 - x^T f \quad (12)$$

which is actually the generalized complementary energy studied by Gao and Strang in nonconvex/nonsmooth variational problem [12]".

The reference [12] above is our reference [7]. Here $\alpha_1, \lambda_1 > 0$.

The statement in [5] we focus on is

"If $(\overline{x}, \overline{\varsigma})$ is a critical point of $L(x, \varsigma)$, and $\overline{\varsigma} < 0$, then in the neighborhood of $(\overline{x}, \overline{\varsigma})$, we have either

$$P(\overline{x}) = \min_{x\in\mathbb{R}^n} \max_{\varsigma<0} L(x, \varsigma) = \min_{\varsigma<0} \max_{x\in\mathbb{R}^n} L(x, \varsigma) = P^d(\overline{\varsigma}), \quad (14)$$

or

$$P(\overline{x}) = \max_{x\in\mathbb{R}^n} \max_{\varsigma<0} L(x, \varsigma) = \max_{\varsigma<0} \max_{x\in\mathbb{R}^n} L(x, \varsigma) = P^d(\overline{\varsigma}). \quad (15)\text{”}$$

From their definitions it follows that $P(x) = \max_{\varsigma\in\mathbb{R}} L(x, \varsigma)$, for every $x \in \mathbb{R}^n$ while from a direct convex conjugate computation it follows that $P^d(\varsigma) = \max_{x\in\mathbb{R}^n} L(x, \varsigma)$, for every $\varsigma < 0$.

Note that [5, (14)] implies that $\overline{\varsigma}$ is a global minimum point of P^d on $(-\infty, 0)$, while [5, (15)] implies that $\overline{\varsigma}$ is a global maximum of P^d on $(-\infty, 0)$. However, in the previous statement one also specifically declares that all happens "in the neighborhood of $(\overline{x}, \overline{\varsigma})$"; making us wonder about the meaning of this part of the statement.

To construct a counterexample for the above statement we use a version of Example 1 with ς replaced by $-\varsigma$.

Example 5 Let $X = \mathbb{R}^n$ with $n \geq 2$, $f \in X$, $|f| = 2$, $\alpha_1 = 1$, and $\lambda_1 = 3$. Hence

$$L(x, \varsigma) = \frac{1}{2}\varsigma \|x\|^2 - \frac{1}{2}\varsigma^2 - 3\varsigma - \langle f, x\rangle,$$

where $\|\cdot\|$, $\langle\cdot,\cdot\rangle$ stand for the norm and inner product of $\mathbb{R}^n$, respectively.

Then $\nabla_x L(x, \varsigma) = \varsigma x - f$, $\nabla_\varsigma L(x, \varsigma) = \frac{1}{2}\|x\|^2 - \varsigma - 3$, and $(\overline{x}, \overline{\varsigma}) := (-f, -1)$ is a critical point of L. It follows that

$$P(x) = \sup_{\varsigma\in\mathbb{R}} L(x, \varsigma) = L(x, \frac{1}{2}\|x\|^2 - 3) = \frac{1}{2}\left(\frac{1}{2}\|x\|^2 - 3\right)^2 - \langle f, x\rangle,$$

due to the fact that every concave function (in this case $L(x, \cdot)$) attains its global maximum value at its critical points ($\varsigma = \frac{1}{2}\|x\|^2 - 3$). Similarly, for $\varsigma < 0$ and the concave function $L(\cdot, \varsigma)$ we have

$$P^d(\varsigma) = \sup_{x\in\mathbb{R}^n} L(x, \varsigma) = L(\varsigma^{-1} f, \varsigma) = -\frac{1}{2}\varsigma^2 - 3\varsigma - \frac{2}{\varsigma}.$$

Note that $P(\overline{x}) = 9/2 = P^d(\overline{\varsigma})$ and $\max_{\varsigma<0} L(x, \varsigma) = L(x, 0) = -\langle f, x\rangle$, for $\|x\| \geq \sqrt{6}$; from which

$$\inf_{x\in\mathbb{R}^n} \max_{\varsigma<0} L(x, \varsigma) = -\infty = \inf_{\varsigma<0} P^d(\varsigma) = \inf_{\varsigma<0} \max_{x\in\mathbb{R}^n} L(x, \varsigma),$$

that is, [5, (14)] does not hold, and

$$\sup_{x\in\mathbb{R}^n} \max_{\varsigma<0} L(x, \varsigma) = \infty = \sup_{\varsigma<0} P^d(\varsigma) = \sup_{\varsigma<0} \max_{x\in\mathbb{R}^n} L(x, \varsigma),$$

that is, [5, (15)] does not hold, too.

Moreover, as seen in Example 1, relations [5, (14)] and [5, (15)] do not hold, even if we replace $\mathbb{R}^n$ and $(-\infty, 0)$ by small neighborhoods of $\overline{x}$ and $\overline{\varsigma}$, respectively.

5 Counterexample to a result in [6]

In [6] one considers quite a complicated problem. In order to make the statement of [6, Th.3] clear we introduce the data of the problem; which is

$$\text{“}(\mathcal{P}) \;:\; \min\left\{P(x) = \frac{1}{2}x^T Ax - x^T f : x \in \mathcal{X}_c\right\}, \quad (1)$$

where $A = \{A_{ij}\} \in \mathbb{R}^{n\times n}$ is a symmetric matrix, $f \in \mathbb{R}^n$ is a given vector, the feasible space $\mathcal{X}_k \subset \mathbb{R}^n$ is defined as

$$\mathcal{X}_c = \{x \in \mathcal{X}_a \mid g(x) \leq d \in \mathbb{R}^m\}, \quad (2)$$

where $g(x) = \{g_\alpha(x)\} : \mathcal{X}_a \to \mathbb{R}^m$ is a given vector-valued differentiable (not necessary convex) function, $\mathcal{X}_a$ is a convex open set in $\mathbb{R}^n$, and $d \in \mathbb{R}^m$ is a given vector”. Let “$U(x) \equiv -P(x) = x^T f - \frac{1}{2}x^T Ax$”.“Since $g(x)$ is a nonconvex function, following the standard procedure of the canonical dual transformation (see [10]), we assume that there exists a Gâteaux differentiable *geometrical operator*

$$\xi = \{\xi^\alpha_\beta\} = \Lambda(x) : \mathcal{X}_a \subset \mathbb{R}^n \to \mathcal{E}_a \subset \mathbb{R}^{m\times p_\alpha}, \quad (11)$$

and a *canonical function* $V : \mathcal{E}_a \to \mathbb{R}^m$ such that the nonconvex constraint $g(x)$ can be written in the canonical form:

$$g(x) = V(\Lambda(x)). \quad (12)\text{”}$$

As usual “the Legendre conjugate $V^* : \mathcal{E}^*_a \to \mathbb{R}^m$ of V can be defined by

$$V^*(\varsigma) = \text{sta}\{\langle\xi; \varsigma\rangle - V(\xi) : \xi \in \mathcal{E}_a\},$$

where the notation sta$\{*\}$ denotes computing the stationary points of $\{*\}$.”

For the present case, "the geometrical mapping $\Lambda(x) : \mathcal{X}_a \to \mathcal{E}_a$ is usually a quadratic operator

$$\Lambda(x) = \left\{ \frac{1}{2} x^T B_\beta^\alpha x + x^T C_\beta^\alpha \right\} : \mathbb{R}^n \to \mathcal{E}_a \subset \mathbb{R}^{m \times p_\alpha} \quad (27)$$

where $B_\beta^\alpha = \left\{ B_{ij\beta}^\alpha \right\} = \left\{ B_{ji\beta}^\alpha \right\} \in \mathbb{R}^{n \times n}$, $C_\beta^\alpha = \left\{ C_{i\beta}^\alpha \right\} \in \mathbb{R}^n$, and the range $\mathcal{E}_a$ depends on both B_β^α and C_β^α. In this case, the generalized complementary function has the form:

$$\Xi(x, \sigma, \varsigma) = \frac{1}{2} x^T G_a(\sigma, \varsigma) x - \sigma^T \left(V^*(\varsigma) + d \right) - x^T F(\sigma, \varsigma), \quad (28)$$

where

$$G_a(\sigma, \varsigma) = A + \sum_{\alpha=1}^{m} \sum_{\beta \in I_\alpha} \sigma_\alpha B_\beta^\alpha \varsigma_\alpha^\beta \quad (29)$$

is the Hessian matrix of $\Xi(x, \sigma, \varsigma)$, which does not depend on x, and

$$F(\sigma, \varsigma) = f - \sum_{\alpha=1}^{m} \sum_{\beta \in I_\alpha} \sigma_\alpha C_\beta^\alpha \varsigma_\alpha^\beta. \quad (30)\text{''}$$

"The canonical dual feasible space $\mathcal{S}_c$ can be defined as

$$\mathcal{S}_c = \{ (\sigma, \varsigma) \in \mathbb{R}_+^m \times \mathcal{E}_a^* \mid F(\sigma, \varsigma) \in \mathcal{C}_{ol}(G_a) \}, \quad (33)$$

and the canonical dual function P^d can be formulated as

$$P^d(\sigma, \varsigma) = -\frac{1}{2} F(\sigma, \varsigma)^T G_a^+(\sigma, \varsigma) F(\sigma, \varsigma) - \sigma^T \left(V^*(\varsigma) + d \right) \quad (34)\text{''}$$

"where G_a^+ is the Moore–Penrose generalized inverse of G_a". "Let

$$\mathcal{S}_c^+ = \{ (\sigma, \varsigma) \in \mathcal{S}_c \mid G_a(\sigma, \varsigma) \succeq 0 \}, \quad (37)$$
$$\mathcal{S}_c^- = \{ (\sigma, \varsigma) \in \mathcal{S}_c \mid G_a(\sigma, \varsigma) \prec 0 \}. \quad (38)\text{''}$$

Here for the matrix $G \in \mathbb{R}^{n \times n}$, $G \prec (\succeq) 0$ means that G is negative (semi-positive) definite and $\mathcal{C}_{ol}(G)$ stands for the subspace of $\mathbb{R}^n$ spanned by the columns of G.

"Then from Theorems 1, 2, and the triality theory developed in [10, 12, 14], we have the following result:

Theorem 3 (Triality Theory) *Suppose that $\Lambda(x)$ is a quadratic operator defined by* (27) *and the condition (ii) in* (24) *holds.*

If $(\overline{\sigma}, \overline{\varsigma}) \in \mathcal{S}_c$ is a critical point of $(\mathcal{P}^d)$, then $\overline{x} = G_a^+(\overline{\sigma}, \overline{\varsigma}) F(\overline{\sigma}, \overline{\varsigma})$ is a KKT point of $(\mathcal{P})$ and $P(\overline{x}) = P^d(\overline{\sigma}, \overline{\varsigma})$."

"If the critical point $(\overline{\sigma}, \overline{\varsigma}) \in \mathcal{S}_c^-$, then on the neighborhood $\mathcal{X}_o \times \mathcal{S}_o$ of $(\overline{x}, \overline{\sigma})$, we have either

$$P(\overline{x}) = \min_{x \in \mathcal{X}_o} P(x) = \min_{(\sigma, \varsigma) \in \mathcal{S}_o} P^d(\sigma, \varsigma) = P^d(\overline{\sigma}, \overline{\varsigma}) \quad (40)$$

or

$$P(\overline{x}) = \max_{x \in \mathcal{X}_o} P(x) = \max_{(\sigma, \varsigma) \in \mathcal{S}_o} P^d(\sigma, \varsigma) = P^d(\overline{\sigma}, \overline{\varsigma}). \quad (41)\text{''}$$

The references [10, 12, 14] are our references [1,2] and [3], respectively.

The "condition (ii) in (24)" referred above is "(ii) the canonical function $V : \mathcal{E}_a \to \mathbb{R}^m$ is convex."

Moreover, in the above statement probably $\mathcal{X}_o$ is a neighborhood of $\overline{x}$ and $\mathcal{S}_o$ is a neighborhood of $(\overline{\sigma}, \overline{\varsigma})$.

We adapt Example 1 to construct a counterexample for the above theorem.

Example 6 We take a very simple case in which $n \geq 2$, $A = -2I_n$, $\|f\| = 2$, $m = 1$, $p_1 = 1$(hence $\alpha = \beta = 1$), $I_1 = \{1\}$, $d = 0$, $B_1^1 = I_n$, $C_1^1 = 0_{n\times 1}$, $V(v) = \frac{1}{2}v^2 - v$ (whence $V^*(\varsigma) = \frac{1}{2}(\varsigma + 1)^2$); so $\Lambda(x) = \frac{1}{2}\|x\|^2$. It follows that $G_a(\sigma, \varsigma) = (\sigma\varsigma - 2)I_n$, $F(\sigma, \varsigma) = f$ and $\Xi(x, \sigma, \varsigma) = \frac{1}{2}(\sigma\varsigma - 2)\|x\|^2 - \frac{1}{2}\sigma(\varsigma + 1)^2 - x^T f$. Moreover,

$$P(x) = -\|x\|^2 - \langle f, x\rangle, \quad P^d(\sigma, \varsigma) = \frac{2}{2 - \sigma\varsigma} - \frac{\sigma}{2}(\varsigma + 1)^2, \quad P(-f) = 0 = P^d(1, 1),$$

$$\Xi(x, \sigma, \varsigma) = \frac{1}{2}(\sigma\varsigma - 2)\|x\|^2 - \frac{1}{2}\sigma(\varsigma + 1)^2 - \langle f, x\rangle .$$

Then $\nabla_x \Xi(x, \sigma, \varsigma) = (\sigma\varsigma - 2)x - f$, $\nabla_\sigma \Xi(x, \sigma, \varsigma) = \frac{1}{2}\varsigma\|x\|^2 - \frac{1}{2}(\varsigma + 1)^2$, $\nabla_\varsigma \Xi(x, \sigma, \varsigma) = \frac{1}{2}\sigma\|x\|^2 - \sigma(\varsigma + 1)$, and

$$\frac{\partial P^d}{\partial\sigma}(\sigma, \varsigma) = \frac{2\varsigma}{(2 - \sigma\varsigma)^2} - \frac{1}{2}(\varsigma + 1)^2, \quad \frac{\partial P^d}{\partial\varsigma}(\sigma, \varsigma) = \frac{2\sigma}{(2 - \sigma\varsigma)^2} - \sigma(\varsigma + 1).$$

It is easily checked that $(\overline{x}, \overline{\sigma}, \overline{\varsigma}) = (-f, 1, 1)$ is a critical point of Ξ and that $(1, 1)$ is a critical point of P^d.

We also have that $(\overline{\sigma}, \overline{\varsigma}) = (1, 1) \in \operatorname{int}\mathcal{S}_c^-$ because $\mathcal{S}_c^- = \{(\sigma, \varsigma) \in \mathbb{R}^2 \mid \sigma \geq 0, \sigma\varsigma < 2\}$.

A further calculus gives $\frac{\partial^2 P^d}{\partial\sigma^2}(1, 1) = 4$, $\frac{\partial^2 P^d}{\partial\sigma\partial\varsigma}(1, 1) = 4$, $\frac{\partial^2 P^d}{\partial\varsigma^2}(1, 1) = 3$. Hence $D^2 P^d$ $(1, 1)$ is not semidefinite (positive or negative), and so $(1, 1)$ is not a local extremum for P^d. This proves that [6, Th.3] is not true.

References

1. Gao, D.Y.: Duality Principles in Nonconvex Systems. Theory, Methods and Applications. Kluwer, Dordrecht (2000)
2. Gao, D.Y.: Canonical dual transformation method and generalized triality theory in nonsmooth global optimization. J. Glob. Optim. **17**, 127–160 (2000)
3. Gao, D.Y.: Duality theory: triduality in global optimization. In: Floudas, C.A., Pardalos, P.M. (eds.) Encyclopedia of Optimization, 2nd edn, pp. 822–828. Springer, New York (2009)
4. Gao, D.Y.: Canonical duality theory and solutions to constrained nonconvex quadratic programming. J. Glob. Optim. **29**, 377–399 (2004)
5. Gao, D.Y.: Complete solutions and extremality criteria to polynomial optimization problems. J. Glob. Optim. **35**, 131–143 (2006)
6. Gao, D.Y., Ruan, N., Sherali, H.D.: Solutions and optimality criteria for nonconvex constrained global optimization problems with connections between canonical and Lagrangian duality. J. Glob. Optim. **45**, 473–497 (2009)
7. Gao, D.Y., Strang, G.: Geometric nonlinearity: potential energy, complementary energy, and the gap function. Q. Appl. Math. **47**, 487–504 (1989)