On unconstrained optimization problems solved using CDT and triality theory

Constantin Zălinescu
Octav Mayer Institute of Mathematics, Iași, and
"Al. I. Cuza" University, Iași
Faculty of Mathematics
Iași, Romania

Cluj-Napoca, GDO2019

Aim

In the preface of the book

[CDT] DY Gao, V Latorre, N Ruan (eds.), Canonical Duality Theory. Unified Methodology for Multidisciplinary Study. Advances in Mechanics and Mathematics (DY Gao, T. Ratiu eds.), Vol. 37. Cham: Springer (2017) [16/18 articles have DY Gao as (co)author]

it is said that

Canonical duality theory is a **breakthrough methodological theory** *that can be used not only for modeling complex systems within a unified framework, but also for solving a large class of challenging problems in multidisciplinary fields of engineering, mathematics, and sciences. ...*

This theory is composed mainly of

(1) a canonical dual transformation, which can be used to formulate perfect dual problems without duality gap;

(2) a complementary-dual principle, which solved the open problem in finite elasticity and provides a unified analytical solution form for general nonconvex/nonsmooth/discrete problems;

(3) a triality theory, which can be used to identify both global and local optimality conditions and to develop powerful algorithms for solving challenging problems in complex systems. ...

The original motivation of this book was a colloquium talk presented by David Yang Gao at UC Berkeley in 2013. ...

The research projects on the canonical duality theory have been continuously supported by US National Science Foundation and US Air Force Office of Scientific Research (AFOSR).

In the abstract of the first (survey) article from [CDT],

[GRL17] DY Gao, N Ruan, V Latorre: Canonical Duality-Triality Theory: Bridge Between Nonconvex Analysis/Mechanics and Global Optimization in Complex System, [CDT], pp. 1–47,

it is mentioned that *Breakthrough from recent challenges and conceptual mistakes by M. Voisei, C. Zalinescu and his coworker are addressed.*

Related to the global optimization problem

$$\min f(x), \quad \text{s.t.} \quad h_i(x) = 0, \ \ g_j(x) \leq 0 \ \ \forall i \in I_m, \ j \in I_p, \qquad (1).$$

it is said : *Without detailed information on these arbitrarily given functions, it is impossible to have a general theory for finding global extrema of the general nonconvex problem (1). ...*

This could be the reason why there was no ***breakthrough*** *in nonlinear programming during the past 60 years.*

A reviewer's opinion for WCGO2019, Metz, about arXiv:1811.04469 (On constrained optimization problems solved using CDT):
Notation Grid : 0 - Strong reject
In my opinion the **Canonical Duality Theory**, as developed either by the author or by Gao and co-authors **is of no practical value.**

However, **DY Gao had Invited Lectures and Colloquium Talks at important Institutions** on CDT:

Analytical Solutions and Canonical Dual Finite Element Method for a Class of Challenging/NP-Hard Problems in Nonconvex Mechanics and Complex Systems, Department of Mechanical Engineering, **University of California, Berkeley**, January 24, 2013

4. Canonical Duality and Triality: Unified Understanding and Analytical Solutions for Nonconvex, Nonsmooth and Discrete Problems in Complex Systems, Colloquium Lecture at the Department of Math, **University of Melbourne**, Sept. 12, 2011.

12. Canonical duality theory and applications in global optimization, Dept of Industrial and Systems Engineering, **University of Florida**, April 22, 2008

13. Canonical duality theory for solving some challenging problems in mechanics and global optimization, Dept Mechanical Science and Engineering, **University of Illinois at Urbana-Champaign**, Jan. 29, 2008.

14. Canonical Duality Approaches for Solving a Class of NP-hard Problems in Global Optimization and Nonconvex Systems, Department of Electric Engineering and Computer Science, **MIT**, May 11, 2007

20. Beauty and Unity in Optimization and System Science: Canonical Duality Theory, Department of Industrial and System Engineering, **Virginia Tech**, Oct. 12, 2007.

21. Unified Canonical Duality Theory for Solving a Class of Nonconvex Problems with Applications in Integer Programming, Department of Mathematics, **Simon Fraser University**, Nov. 21, 2007.

27. Canonical duality and triality, a potentially powerful method for solving nonlinear variational/optimization problems, Department of Mathematics, **University of Oakland**, Nov. 9, 2006.

29. Canonical duality theory in global optimization and application, department of electrical engineering, **Princeton University**, August 4, 2006.

35. Canonical duality theory and method for solving nonconvex variational- optimization problems with applications, Institute for Scientific Computing and Applied Mathematics, **Indiana University**, Bloomington, IN. Sep. 21, 2005.

36. Primal-dual methods and algorithm in large-scale nonconvex optimization and application, Department of Math, University of Wisconsin, Milwaukee, September 5, 2005.

37. Duality, triality and unity in arts, science, and religion, Institute of Information Science, **Konan University**, Japan, June 8, 2005.

45. Duality and Triality in Mathematics and Scientific Computations, Department of Mathematics, University of Auckland, New Zealand, January 22, 2002.

47. Duality and Triality: Unifying Mathematics and Natural Sciences, Colloquium talk at Department of Math., **University of Glasgow**, January 18, 2001

DY Gao solely or together with some of his collaborators applied his Canonical duality theory (CDT) for solving a class of **unconstrained optimization problems**, getting the so-called **triality theorems**.

Unfortunately, the **double-min duality** from these results published before 2010 revealed to be false, even if in 2003 DY Gao announced that *certain additional conditions* are needed for getting it.

After 2010 DY Gao together with some of his collaborators published several papers in which they added additional conditions for getting **double-min** and **double-max** dualities in the triality theorems.

Our aim is to treat rigorously this kind of problems and to discuss several results concerning the **triality theory** obtained up to now.

Notations

$\mathfrak{S}_n$ denotes the class of symmetric matrices from $\mathfrak{M}_n := \mathbb{R}^{n\times n}$

$q_k : \mathbb{R}^n \to \mathbb{R}$, $q_k(x) := \frac{1}{2}\langle x, A_k x\rangle - \langle b_k, x\rangle + c_k \quad \left(k \in \overline{0,m}\right)$

with $A_k \in \mathfrak{S}_n$, $b_k \in \mathbb{R}^n$ (seen as column vectors), $c_k \in \mathbb{R}$, $\langle \cdot, \cdot\rangle$ the usual inner product on $\mathbb{R}^n$.

The fact that $A \in \mathfrak{S}_n$ is positive (semi) definite is denoted by $A \succ 0$ ($A \succeq 0$) and we set

$\mathfrak{S}_n^+ := \{A \in \mathfrak{S}_n \mid A \succeq 0\}$, $\mathfrak{S}_n^{++} := \{A \in \mathfrak{S}_n \mid A \succ 0\}$;

similarly for $\mathfrak{S}_n^-$, $\mathfrak{S}_n^{--}$. It is well known that $\mathfrak{S}_n^{++} = \operatorname{int} \mathfrak{S}_n^+$.

For $g : \mathbb{R}^m \to \overline{\mathbb{R}}$, $\operatorname{dom} g := \{y \in \mathbb{R}^m \mid g(y) < \infty\}$

g is proper when $\operatorname{dom} g \neq \emptyset$ and $g(y) \neq -\infty$ for $y \in \mathbb{R}^m$.

$\Gamma := \Gamma(\mathbb{R}^m)$ is the class of proper convex lsc functions $g : \mathbb{R}^m \to \overline{\mathbb{R}}$.

The Fenchel conjugate $g^* : \mathbb{R}^m \to \overline{\mathbb{R}}$ of the proper function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is defined by

$$g^*(\sigma) := \sup\{\langle y, \sigma\rangle - g(y) \mid y \in \mathbb{R}^m\} = \sup\{\langle y, \sigma\rangle - g(y) \mid y \in \operatorname{dom} g\}$$

while its subdifferential at $y \in \operatorname{dom} g$ is

$$\partial g(y) := \left\{\sigma \in \mathbb{R}^m \mid \left\langle y' - y, \sigma\right\rangle \le g(y') - g(y)\ \forall y' \in \mathbb{R}^m\right\},$$

and $\partial g(y) := \emptyset$ if $y \notin \operatorname{dom} g$; clearly, for $(y, \sigma) \in \mathbb{R}^m \times \mathbb{R}^m$

$$g(y) + g^*(\sigma) \ge \langle y, \sigma\rangle \ \wedge\ [\sigma \in \partial g(y) \Longleftrightarrow g(y) + g^*(\sigma) = \langle y, \sigma\rangle].$$

$\Gamma_{sc} := \Gamma_{sc}(\mathbb{R}^m)$ is the class of those $g \in \Gamma(\mathbb{R}^m)$ which are essentially strictly convex and essentially smooth convex functions, that is the class of proper lsc convex functions of Legendre type as defined in chapter 26 in Rockafellar's book Convex Analysis (1970).

Problems

We are concerned by the following unconstrained global minimization problem

$$(P) \quad \min f(x) \ \text{ s.t. } x \in \mathbb{R}^n$$

as well as by local minimum or/and maximal points of f, where $f := q_0 + V \circ q$ with $q(x) := (q_1(x), ..., q_m(x))^T$, q_i ($i \in \overline{0, m}$) being quadratic functions defined on $\mathbb{R}^n$, and $V \in \Gamma$.

To (P) one associates the so called "total complementary function" in
[GW17] DY Gao, C Wu, *Triality theory for general unconstrained global optimization problems*, CDT-book,

or "Gao–Strang complementary function" in
[GRP12] DY Gao, N Ruan, PM Pardalos, *Canonical dual solutions to sum of fourth-order polynomials minimization problems with applications to sensor network localization*, in Sensors: Theory, Algorithms, and Applications. Springer Optimization and Its Applications 61, Springer (2012),

or "extended Lagrangian" in
[G00]DY Gao, *Duality principles in nonconvex systems: theory, methods and applications,* Kluwer, Dordrecht (2000), and
[G03] DY Gao, *Perfect duality theory and complete solutions to a class of global optimization problems*, Optimization 52 (2003).

This is the function $\Xi : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ defined by

$$\Xi(x, \sigma) = q_0(x) + \langle q(x), \sigma \rangle - V^*(\sigma) = L(x, \sigma) - V^*(\sigma),$$

where L is the (usual) Lagrangian associated to $(q_k)_{k \in \overline{0,m}}$, that is

$$L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \quad L(x, \sigma) := q_0(x) + \langle q(x), \sigma \rangle .$$

It follows that

$$\Xi(x, \sigma) = \tfrac{1}{2} \langle x, A(\sigma)x \rangle - \langle b(\sigma), x \rangle + c(\sigma) - V^*(\sigma),$$

where, for $\sigma_0 := 1$ and $\sigma := (\sigma_1, ..., \sigma_m)^T \in \mathbb{R}^m$,

$$A(\sigma) := \sum\nolimits_{k=0}^{m} \sigma_k A_k, \quad b(\sigma) := \sum\nolimits_{k=0}^{m} \sigma_k b_k, \quad c(\sigma) := \sum\nolimits_{k=0}^{m} \sigma_k c_k;$$

clearly, $A(\cdot)$, $b(\cdot)$, $c(\cdot)$ are affine functions. Hence, $\Xi(\cdot, \sigma)$ is quadratic for each $\sigma \in \operatorname{dom} V^*$ and $\Xi(x, \cdot)$ is concave for each $x \in \mathbb{R}^n$.

Since $V^{**} := (V^*)^* = V$, we obtain that

$$f(x) = \sup_{\sigma \in \text{dom } V^*} \Xi(x,\sigma) = \sup_{\sigma \in \text{ri}(\text{dom } V^*)} \Xi(x,\sigma) \quad \forall x \in \mathbb{R}^n,$$

Moreover,

$$\nabla_x \Xi(x,\sigma) = A(\sigma)x - b(\sigma), \quad \nabla^2_{xx} \Xi(x,\sigma) = A(\sigma),$$
$$\partial\left(-\Xi(x,\cdot)\right)(\sigma) = \partial V^*(\sigma) - q(x),$$

for all $(x,\sigma) \in \mathbb{R}^n \times \text{dom } V^*$. Hence, for $(x,\sigma) \in \mathbb{R}^n \times \text{dom } V^*$ one has

$$\nabla_x \Xi(x,\sigma) = 0 \iff A(\sigma)x = b(\sigma),$$
$$0 \in \partial\left(-\Xi(x,\cdot)\right)(\sigma) \iff q(x) \in \partial V^*(\sigma) \iff \sigma \in \partial V\left((q(x)\right).$$

Consider the following sets in which σ is taken from $\mathbb{R}^m$ if not specified otherwise:

$$Y_0 := \{\sigma \mid \det A(\sigma) \neq 0\}, \quad Y_{\text{col}} := \{\sigma \mid b(\sigma) \in \operatorname{Im} A(\sigma)\} \ \ (\supset Y_0),$$
$$Y^+ := \{\sigma \mid A(\sigma) \succ 0\}, \quad Y^- := \{\sigma \mid A(\sigma) \prec 0\},$$
$$Y^+_{\text{col}} := \{\sigma \in Y_{\text{col}} \mid A(\sigma) \succeq 0\}, \quad Y^-_{\text{col}} := \{\sigma \in Y_{\text{col}} \mid A(\sigma) \preceq 0\},$$
$$S_0 := Y_0 \cap \operatorname{dom} V^*, \quad S^+ := Y^+ \cap \operatorname{dom} V^*, \quad S^- := Y^- \cap \operatorname{dom} V^*,$$
$$S_{\text{col}} := Y_{\text{col}} \cap \operatorname{dom} V^*, \quad S^+_{\text{col}} := Y^+_{\text{col}} \cap \operatorname{dom} V^*, \quad S^-_{\text{col}} := Y^-_{\text{col}} \cap \operatorname{dom} V^*.$$

In

[Z18a] Z. On quadratic optimization problems and canonical duality, arXiv:1809.09032,

we considered a dual function associated to the family $(q_k)_{k\in\overline{0,m}}$, which is denoted by D_L here. More precisely,

$$D_L : Y_{\text{col}} \to \mathbb{R}, \quad D_L(\sigma) := L(x,\sigma) \text{ with } A(\sigma)x = b(\sigma).$$

In a similar way, we consider the (dual objective) function D associated to $(q_k)_{k\in\overline{0,m}}$ and V defined by

$$D : S_{\text{col}} \to \mathbb{R}, \quad D(\sigma) := \Xi(x,\sigma) \text{ with } A(\sigma)x = b(\sigma);$$

hence

$$D(\sigma) = D_L(\sigma) - V^*(\sigma) \quad \forall \sigma \in S_{\text{col}}.$$

Setting

$$x(\sigma) := A(\sigma)^{-1} b(\sigma) := [A(\sigma)]^{-1} \cdot b(\sigma)$$

for $\sigma \in Y_0$, we obtain for $\sigma \in S_0$ that

$$D(\sigma) = \Xi\left(x(\sigma),\sigma\right) = -\tfrac{1}{2}\left\langle b(\sigma), A(\sigma)^{-1} b(\sigma)\right\rangle + c(\sigma) - V^*(\sigma).$$

We have that D_L is concave and usc on Y^+_{col}, and convex and lsc on Y^-_{col}, and [Z18a, Eq. (9)]holds; moreover $D_L(\sigma)$ is attained at any $x \in \mathbb{R}^n$ such that $A(\sigma)x = b(\sigma)$ whenever $\lambda \in Y^+_{\text{col}} \cup Y^-_{\text{col}}$, being attained uniquely at $x := x(\sigma)$ for $\sigma \in Y^+ \cup Y^-$.

It follows that

$$D(\sigma) = \begin{cases} \min_{x \in \mathbb{R}^n} \Xi(x, \sigma) & \text{if} \quad \sigma \in S^+_{\text{col}}, \\ \max_{x \in \mathbb{R}^n} \Xi(x, \sigma) & \text{if} \quad \sigma \in S^-_{\text{col}}, \end{cases}$$

the value of $D(\sigma)$ being attained uniquely at $x := x(\sigma)$ when $\sigma \in S^+ \cup S^-$ $(\subset S_0)$; moreover, we have that D is concave and usc on S^+_{col} as the sum of two concave and usc functions, while D is a d.c. function (difference of convex functions) on S^-_{col}. In general, D is neither convex nor concave on (the convex set) S^-_{col}.

As in [Z18a] (or by direct calculations), we have that

$$\begin{aligned} \frac{\partial D}{\partial \sigma_i}(\sigma) &= \tfrac{1}{2} \langle x(\sigma), A_i x(\sigma) \rangle - \langle b_i, x(\sigma) \rangle + c_i - \frac{\partial V^*}{\partial \sigma_i}(\sigma) \\ &= q_i(x(\sigma)) - \frac{\partial V^*}{\partial \sigma_i}(\sigma) \end{aligned} \tag{1*}$$

for those $\sigma \in \operatorname{int} S_0$ and $i \in \overline{1, m}$ for which $\frac{\partial V^*}{\partial \sigma_i}(\sigma)$ exists.

When V is sublinear one has that $V^* = \iota_{\partial V(0)}$. So, we get the following simple result.

Proposition 1

Assume that $V \in \Gamma(\mathbb{R}^m)$ is sublinear. Then $D|_{S_{\text{col}}^-}$ is convex; moreover, $\nabla D(\sigma) = q(x(\sigma))$ for every $\sigma \in S_0 \cap \operatorname{int}(\operatorname{dom} V^*)$.

Assume that $g \in \Gamma_{sc}$. Then: $g^* \in \Gamma_{sc}$, $\operatorname{dom} \partial g = \operatorname{int}(\operatorname{dom} g)$, and g is differentiable on $\operatorname{int}(\operatorname{dom} g)$; moreover,
$\nabla g : \operatorname{int}(\operatorname{dom} g) \to \operatorname{int}(\operatorname{dom} g^*)$ is bijective and continuous with $(\nabla g)^{-1} = \nabla g^*$.
In the rest of this section we assume $V \in \Gamma_{sc}$; hence $V^* \in \Gamma_{sc}$.

We set

$$X_0 := \{x \in \mathbb{R}^n \mid q(x) \in \operatorname{int}(\operatorname{dom} V)\} \subset \operatorname{dom} f.$$

Because V is differentiable on $\operatorname{int}(\operatorname{dom} V)$ and V^* is differentiable on $\operatorname{int}(\operatorname{dom} V^*)$, clearly

$$\nabla f(x) = A_0 x - b_0 + \sum\nolimits_{i=1}^{m} \frac{\partial V}{\partial y_i}(q(x)) \cdot (A_i x - b_i) \;\; \forall x \in X_0, \quad (2^*)$$

$$\nabla_\sigma \Xi(x,\sigma) = q(x) - \nabla V^*(\sigma) \quad \forall (x,\sigma) \in \mathbb{R}^n \times \operatorname{int}(\operatorname{dom} V^*). \quad (3^*)$$

It follows that (1*) holds for $\sigma \in \operatorname{int} S_0$ $[= S_0 \cap \operatorname{int}(\operatorname{dom} V^*)]$ and $i \in \overline{1,m}$, whence for $\sigma' \in S_0 \cap \operatorname{int}(\operatorname{dom} V^*)$ one has

$$\nabla D(\sigma') = q(x(\sigma')) - \nabla V^*(\sigma') = \nabla_\sigma \Xi(x(\sigma'),\sigma'). \qquad (4^*)$$

From (3*) and (1*) we get

$$\nabla_\sigma \Xi(x,\sigma) = 0 \iff \big[\sigma \in \operatorname{int} S_0 \ \wedge \ q(x) = \nabla V^*(\sigma)\big] \\ \iff \big[x \in X_0 \ \wedge \ \sigma = \nabla V\left(q(x)\right)\big]. \qquad (5^*)$$

From the concavity of $\Xi(x,\cdot)$ for $x \in \mathbb{R}^n$ and (4*) we obtain

$$f(x) = \sup_{\sigma \in \operatorname{dom} V^*} \Xi(x,\sigma) = \sup_{\sigma \in \operatorname{int}(\operatorname{dom} V^*)} \Xi(x,\sigma) = \Xi\big(x, \nabla V\left(q(x)\right)\big)$$

for all $x \in X_0$; moreover, using (2*) and (3*) we obtain that

$$[x \in X_0 \wedge \sigma = \nabla V\left(q(x)\right)] \Rightarrow [\nabla f(x) = \nabla_x \Xi(x,\sigma) \wedge f(x) = \Xi(x,\sigma)].$$

Furthermore, using the expressions of $\nabla \Xi$, $\nabla^2 \Xi$ and (3*), for $(x,\sigma) \in \mathbb{R}^n \times \operatorname{int}(\operatorname{dom} V^*)$ we have that

$$\nabla \Xi(x,\sigma) = 0 \Leftrightarrow [x \in X_0 \wedge \sigma = \nabla V\left(q(x)\right) \wedge A(\sigma)x = b(\sigma)].$$

The preceding considerations yield directly the next result.

Proposition 2

Let $V \in \Gamma(\mathbb{R}^m)$ and $(\overline{x}, \overline{\sigma}) \in \mathbb{R}^n \times \operatorname{dom} V^*$.
(i) Assume that $\nabla_x \Xi(\overline{x}, \overline{\sigma}) = 0$ and $q(\overline{x}) \in \partial V^*(\overline{\sigma})$. Then $(\overline{x}, \overline{\sigma}) \in \operatorname{dom} f \times S_{\text{col}}$, $\overline{\sigma} \in \partial V\left(q(\overline{x})\right)$, and

$$f(\overline{x}) = \Xi(\overline{x}, \overline{\sigma}) = D(\overline{\sigma}). \tag{PDF}$$

(ii) Moreover, assume that $A(\overline{\sigma}) \succeq 0$. Then $\overline{\sigma} \in S^{+}_{\text{col}}$ and

$$f(\overline{x}) = \inf_{x \in \operatorname{dom} f} f(x) = \Xi(\overline{x}, \overline{\sigma}) = \sup_{\sigma \in S^{+}_{\text{col}}} D(\sigma) = D(\overline{\sigma}); \tag{MMD}$$

furthermore, if $\overline{\sigma} \in S^{+}$, then $\overline{x}$ is the unique global solution of problem (P).

Proposition3

Let $V \in \Gamma_{sc}$ and $(\overline{x}, \overline{\sigma}) \in \mathbb{R}^n \times \text{int}(\text{dom } V^*)$.

(i) Assume that $(\overline{x}, \overline{\sigma})$ is a critical point of Ξ. Then $(\overline{x}, \overline{\sigma}) \in X_0 \times S_{\text{col}}$, $\overline{x}$ is a critical point of f, and (PDF) holds; moreover, if $\overline{\sigma} \in S_0$ then $\overline{\sigma}$ is a critical point of D.

(ii) Assume that $(\overline{x}, \overline{\sigma})$ is a critical point of Ξ such that $A(\overline{\sigma}) \succeq 0$. Then $\overline{\sigma} \in S^+_{\text{col}}$ and (MMD) holds; moreover, if $A(\overline{\sigma}) \succ 0$ then $\overline{x}$ is the unique global solution of problem (P).

(iii) Assume that $\overline{\sigma} \in S_0$ and $\overline{\sigma}$ is a critical point of D. Then $(\overline{x}, \overline{\sigma})$ is a critical point of Ξ, where $\overline{x} := A(\overline{\sigma})^{-1} b(\overline{\sigma})$; therefore, (i) and (ii) apply.

In many papers by DY Gao and his collaborators one speaks about **"triality theorems"** in which, besides the minimax result established for the case $A(\overline{\sigma}) \succeq 0$ (see Proposition 2), one obtains also **"bi-duality"** results (**"double-min duality"** and **"double-max duality"**) established for $A(\overline{\sigma}) \prec 0$, that is $\overline{x}$ and $\overline{\sigma}$ are simultaneously local minimizers (maximizers) for f on $\operatorname{dom} f$ and for D on S^{-}, respectively.

In [GRL17] it is said: *"the triality was proposed originally from post-buckling analysis [42] in "either-or" format since the double-max duality is always true but the double-min duality was proved only in one-dimensional nonconvex analysis [49]"*.

The next example shows that such triality results are not valid for general $V \in \Gamma(\mathbb{R}^m)$, even for $n = m = 1$. We concentrate on the case $\overline{\sigma} \in S^{-}$ of Proposition 2 (i), that is $(\overline{x}, \overline{\sigma}) \in \mathbb{R}^n \times \mathbb{R}^m$ is such that $A(\overline{\sigma})\overline{x} = b(\overline{\sigma})$ and $\overline{\sigma} \in S^{-} \cap \partial V(q(\overline{x}))$, and so $\overline{x} \in q^{-1}(\operatorname{dom} V) = \operatorname{dom} f$.

Example 7

Consider $V := \iota_{\mathbb{R}_-}$, and $q_0(x) := -\frac{1}{2}x^2 + x$, $q(x) =:= \frac{1}{2}\left(x^2 - 1\right)$ for $x \in \mathbb{R}$. Then $f := f_\emptyset = q_0 + \iota_{[-1,1]}$, $A(\sigma) = \sigma - 1$, $b(\sigma) = -1$, $c(\sigma) = -\frac{1}{2}\sigma$, whence $L(x,\sigma) = \frac{1}{2}(\sigma - 1)x^2 + x - \frac{1}{2}\sigma$, $Y_{\text{col}} = Y_0 = \mathbb{R} \setminus \{1\}$, $x(\sigma) = 1/(1-\sigma)$, and so $D_L(\sigma) = \frac{1}{2}\left(\frac{1}{1-\sigma} - \sigma\right)$, for $\sigma \in Y_0$; moreover, $D = D_L$ on $S_{\text{col}} = [0,1) \cup (1,\infty)$. For $\overline{\sigma} = 0$ $[\in S^-_{\text{col}} = S^- = [0,1)]$ we get $\overline{x} := x(0) = 1$. Clearly, $0 \in \partial V(q(1)) = \partial V(0)$ $(= \mathbb{R}_+)$. Hence the pair $(1,0)$ verifies the hypothesis of Proposition 2 (i), even more, $(1,0)$ is a critical point of L. However, by direct verification, we obtain that $\overline{x} = 1$ is the unique global maximizer of f on $\operatorname{dom} f = [-1,1]$, while applying [Z18a, Prop. 4] we obtain that $\overline{\sigma} = 0$ is the unique global minimizer of D_L on Y^-_{col} $[= (-\infty, 1)]$, whence 0 is the unique global minimizer of D on S^-. These facts show that "double-min duality" and "double-max duality" are not verified in the present case.

In DY Gao's works published after 2011 the "triality theorems" are established for V a twice differentiable strictly convex function.

Our aim in the sequel is to study the problems of "double-min duality" and "double-max duality" for the special class Γ^2_{sc} of those functions $V \in \Gamma_{sc}$ which are twice differentiable on $\operatorname{int}(\operatorname{dom} V)$ with $\nabla^2 V(y) \succ 0$ for $y \in \operatorname{int}(\operatorname{dom} V)$.

First, in the next section, we establish a result on positive semidefinite operators in Euclidean spaces needed for getting our "bi-duality" results.

An auxiliary result

In order to study the case when $\overline{\sigma} \in S^-$, we need the following result which is probably known, but we have not a reference for it.

Proposition 8 (to be continued)

Let X, Y be nontrivial Euclidean spaces and $H : Y \to X$ be a linear operator with $H^* : X \to Y$ its adjoint. Consider $Q := HH^* := H \circ H^*$, $R := H^*H$, and

$$\varphi : X \to \mathbb{R}, \quad \varphi(x) := \|H^*x\|^2, \qquad \psi : Y \to \mathbb{R}, \quad \psi(y) := \|Hy\|^2.$$

Then the following assertions hold:

(a) Q and R are self-adjoint positive semi-definite operators, $\ker Q = \ker H^*$, $\operatorname{Im} Q = \operatorname{Im} H$, $\ker R = \ker H$, $\operatorname{Im} R = \operatorname{Im} H^*$; consequently, $H = 0 \Leftrightarrow Q = 0 \Leftrightarrow R = 0$.

Proposition 8 (continued)

(b) Setting $S_X := \{x \in X \mid \|x\| = 1\}$, one has $\alpha = \beta$, where

$$\alpha := \max_{x \in S_X} \varphi(x) = \max\{\lambda \in \mathbb{R} \mid \exists x \in X \setminus \{0\} : Qx = \lambda x\},$$
$$\beta := \max_{y \in S_Y} \psi(y) = \max\{\lambda \in \mathbb{R} \mid \exists y \in Y \setminus \{0\} : Ry = \lambda y\}.$$

(c) If $H \neq 0$, then $\operatorname{Im} Q \neq \{0\}$, $\operatorname{Im} R \neq \{0\}$, and $\gamma = \delta > 0$, where

$$\gamma := \min_{x \in S_X \cap \operatorname{Im} Q} \varphi(x) = \min\{\lambda > 0 \mid \exists x \in X \setminus \{0\} : Qx = \lambda x\},$$
$$\delta := \min_{x \in S_Y \cap \operatorname{Im} R} \psi(y) = \min\{\lambda > 0 \mid \exists y \in Y \setminus \{0\} : Ry = \lambda y\}.$$

(d) The following implications hold:

$$\min_{x \in S_X} \varphi(x) = 0 \Leftrightarrow \ker Q \neq \{0\} \Leftrightarrow \operatorname{Im} Q \neq X \Leftrightarrow \operatorname{Im} H \neq X,$$
$$\min_{y \in S_Y} \psi(y) = 0 \Leftrightarrow \ker R \neq \{0\} \Leftrightarrow \operatorname{Im} R \neq Y \Leftrightarrow \ker H \neq \{0\}.$$

Throughout this section we assume that $V \in \Gamma^2_{sc}$. Observe that for $g \in \Gamma^2_{sc}$ one has $g^* \in \Gamma^2_{sc}$ and

$$\nabla^2 g^*(\sigma) = \left(\nabla^2 g\left((\nabla g)^{-1}(\sigma)\right)\right)^{-1} \quad \forall \sigma \in \operatorname{int}(\operatorname{dom} g^*).$$

It follows that for $x \in X_0$ and $u \in \mathbb{R}^n$,

$$\left\langle u, \nabla^2 f(x) u\right\rangle = \left\langle u, \left[A_0 + \sum\nolimits_{i=1}^m \frac{\partial V}{\partial y_i}(q(x)) \cdot A_i\right] u\right\rangle + \left\langle v_u, \nabla^2 V(q(x)) v_u\right\rangle,$$

where $v_u := \left(\langle u, A_1 x - b_1\rangle, ..., \langle u, A_m x - b_m\rangle\right)^T$, and

$$\frac{\partial^2 D}{\partial\sigma_i \partial\sigma_k}(\sigma) = -\left\langle A_i x(\sigma) - b_i, A(\sigma)^{-1}\left(A_k x(\sigma) - b_k\right)\right\rangle - \frac{\partial^2 V^*}{\partial\sigma_i \partial\sigma_k}(\sigma)$$

for all $\sigma \in \operatorname{int} S_0$ and $i, k \in \overline{1, m}$. It follows that

$$\langle v, \nabla^2 D(\sigma) v\rangle = -\langle A_v x(\sigma) - b_v, A(\sigma)^{-1}(A_v x(\sigma) - b_v)\rangle - \left\langle v, \nabla^2 V^*(\sigma) v\right\rangle$$

for all $v \in \mathbb{R}^m$ and $\sigma \in S_0$, where

$$A_v := \sum\nolimits_{i=1}^m v_j A_j, \quad b_v := \sum\nolimits_{j=1}^m v_j b_j \quad (v \in \mathbb{R}^m).$$

Assume that $(\overline{x}, \overline{\sigma}) \in X_0 \times S^-$ is a critical point of Ξ; then $\overline{\sigma} = \nabla V(q(\overline{x}))$. Because $A(\overline{\sigma}) \prec 0$ and $\nabla^2 V(q(\overline{x})) \succ 0$ there exist non-singular matrices $E \in \mathfrak{M}_n$ and $F \in \mathfrak{M}_m$ such that $-A(\overline{\sigma}) = E^* E$ and $\nabla^2 V(q(\overline{x})) = F^* F$, where E^* and F^* are the transposed matrices of E and F, respectively; hence $A(\overline{\sigma})^{-1} = -E^{-1}(E^{-1})^*$ and $\nabla^2 V^*(\overline{\sigma}) = F^{-1}(F^{-1})^*$. Let us set

$$d_i := (E^{-1})^*(A_i \overline{x} - b_i) \in \mathbb{R}^n \quad (i \in \overline{1, m}).$$

Because for A_v defined above one has

$$A_v \overline{x} - b_v = \sum\nolimits_{i=1}^m v_i (A_i \overline{x} - b_i) = \sum\nolimits_{i=1}^m v_i E^* d_i = E^* \sum\nolimits_{i=1}^m v_i d_i,$$

from the expression of $\left\langle v, \nabla^2 D(\overline{\sigma}) v \right\rangle$ we obtain that

$$\left\langle v, \nabla^2 D(\overline{\sigma}) v \right\rangle = \left\| \sum\nolimits_{i=1}^m v_i d_i \right\|^2 - \left\| (F^{-1})^* v \right\|^2 \quad \forall v \in \mathbb{R}^m.$$

Taking into account the expression of $\left\langle u, \nabla^2 f(\overline{x}) v \right\rangle$, we have that

$$\left\langle u, \nabla^2 f(\overline{x}) u \right\rangle = \langle u, A(\overline{\sigma}) u \rangle + \left\langle v_u, \nabla^2 V\left(q(\overline{x})\right) v_u \right\rangle = \|F v_u\|^2 - \|Eu\|^2,$$

for all $u \in \mathbb{R}^n$, where

$$v_u = \left(\langle u, A_i \overline{x} - b_i \rangle\right)_{i \in \overline{1,m}} = \left(\langle Eu, d_i \rangle\right)_{i \in \overline{1,m}},$$

Let us set

$$J : \mathbb{R}^m \to \mathbb{R}^n, \quad Jv := \sum\nolimits_{i=1}^{m} v_i d_i \quad (v \in \mathbb{R}^m);$$

then the adjoint $J^* : \mathbb{R}^n \to \mathbb{R}^m$ is given by

$$J^* u = \left(\langle u, d_1 \rangle, \ldots, \langle u, d_m \rangle\right)^T =: \left(\langle u, d_i \rangle \right)_{i \in \overline{1,m}} \quad (u \in \mathbb{R}^n).$$

Take $H : \mathbb{R}^m \to \mathbb{R}^n$ defined by $H := J \circ F^*$. Then $H^* = F \circ J^* : \mathbb{R}^n \to \mathbb{R}^m$.

Denoting $u' := Eu$ for $u \in \mathbb{R}^n$ and $v' := (F^{-1})^* v$ for $v \in \mathbb{R}^m$, we obtain that

$$\left\langle u, \nabla^2 f(\overline{x})u \right\rangle = \left\| H^* u' \right\|^2 - \left\| u' \right\|^2, \quad \left\langle v, \nabla^2 D(\overline{\sigma})v \right\rangle = \left\| Hv' \right\|^2 - \left\| v' \right\|^2.$$

Because E and F are non-singular, for $\rho \in \{>, \geq, <, \leq\}$ and $\rho' \in \{\succ, \succeq, \prec, \preceq\}$ with the natural correspondence, we have

$$\nabla^2 f(\overline{x}) \ \rho' \ 0 \Longleftrightarrow \left[\left\| H^* u' \right\|^2 \ \rho \ 1 \ \ \forall u' \in S_n \right] \Longleftrightarrow \left[\varphi(u) \ \rho \ 1 \ \ \forall u \in S_n \right],$$
$$\nabla^2 D(\overline{\sigma}) \ \rho' \ 0 \Longleftrightarrow \left[\left\| Hv' \right\|^2 \ \rho \ 1 \ \ \forall v' \in S_m \right] \Longleftrightarrow \left[\psi(v) \ \rho \ 1 \ \ \forall v \in S_m \right],$$

where $S_p := S_{\mathbb{R}^p}$, and φ, ψ are defined in Proposition 8 with

$$H := J \circ F^* : \mathbb{R}^m \to \mathbb{R}^n, \quad H^* = F \circ J^* : \mathbb{R}^n \to \mathbb{R}^m.$$

Recall that $E \in \mathfrak{M}_n$ and $F \in \mathfrak{M}_m$ are such that $-A(\overline{\sigma}) = E^* E$ and $\nabla^2 V(q(\overline{x})) = F^* F$, $d_i := (E^{-1})^* (A_i \overline{x} - b_i)$ $(i \in \overline{1, m})$.

In the next result we shall use Proposition 8 for the operator $H := J \circ F^{*}$. Setting $\dim\{0\} := 0$, the following assertions hold:

- $\dim(\operatorname{Im} H) = \dim(\operatorname{Im} H^{*}) \leq \min\{n, m\}$,
- $\dim(\ker H) + \dim(\operatorname{Im} H) = m$, $\dim(\ker H^{*}) + \dim(\operatorname{Im} H^{*}) = n$,
- $\dim(\ker H^{*})$ $[= \dim(\ker Q)]$ is equal to the multiplicity of the eigenvalue 0 of $Q := H \circ H^{*}$, while $\dim(\ker H)$ is equal to the multiplicity of the eigenvalue 0 of $R := H^{*} \circ H$.

Proposition 9 (to be continued)

Let $(\overline{x}, \overline{\sigma}) \in X_0 \times S^{-}$ be a critical point of Ξ.

(i) If $\overline{x}$ (resp. $\overline{\sigma}$) is a local maximizer of f (resp. D), then $\|Hv\| \leq 1$ for all $v \in S_m$, or, equivalently, $(\alpha =)$ $\beta \leq 1$. Conversely, if $\|Hv\| < 1$ for all $v \in S_m$, then $\overline{x}$ (resp. $\overline{\sigma}$) is a local strict maximizer of f (resp. D). In particular, if $A_i\overline{x} = b_i$ (or equivalently $d_i = 0$) for all $i \in \overline{1, m}$, then $\overline{x}$ and $\overline{\sigma}$ are local strict maximizers of f and D, respectively.

Proposition 9 (continued)

(ii) If $\overline{x}$ is a local minimizer of f, then $\|H^*u\| \geq 1$ for all $u \in S_n$; in particular H is surjective, $m \geq n$, and every positive eigenvalue of $H^* \circ H$ is greater than or equal to 1. Conversely, if $\|H^*u\| > 1$ for all $u \in S_n$, then $\overline{x}$ is a local strict minimizer of f; moreover, if $m > n$ then $\overline{\sigma}$ is not a local extremum for D.

(iii) If $\overline{\sigma}$ is a local minimizer of D, then $\|Hv\| \geq 1$ for all $v \in S_m$; in particular H is injective, $m \leq n$, and every positive eigenvalue of $H \circ H^*$ is greater than or equal to 1. Moreover, if $m < n$ then $\overline{x}$ is not a local extremum for f. Conversely, if $\|Hv\| > 1$ for all $v \in S_m$, then $\overline{\sigma}$ is a local strict minimizer of D.

(iv) Assume that $m = n$ and $\{A_i\overline{x} - b_i \mid i \in \overline{1, m}\}$ is a basis of $\mathbb{R}^m$. If $\|Hv\| > 1$ for all $v \in S_m$, then $\overline{x}$ and $\overline{\sigma}$ are local strict minimizers of f and D, respectively.

Gao & Wu (in [GW17]) use the assumption
"(A3) The critical points of problem $(\mathcal{P})$ are non-singular, i.e., if $\nabla\Pi(\overline{x}) = 0$, then $\det \nabla^2\Pi(\overline{x}) \neq 0$"
Under such a condition we have the following result.

Corollary 10

Let $(\overline{x}, \overline{\sigma}) \in X_0 \times S^{-}$ be a critical point of Ξ such that $\det \nabla^2 f(\overline{x}) \neq 0$ [that is 0 is not an eigenvalue of $\nabla^2 f(\overline{x})$]. The following assertions hold:

(a) $\overline{x}$ is a local maximizer of f if and only if $\|Hv\| < 1$ for all $v \in S_m$, if and only if $\overline{\sigma}$ is a local maximizer of D.

(b) Assume that $m = n$. Then $\overline{x}$ is a local minimizer of f if and only if $\|Hv\| > 1$ for all $v \in S_m$, if and only if $\overline{\sigma}$ is a local minimizer of D.

Relations with previous results

Let us compare first our results with those from the most recently published paper on this topic for general V, that is Gao and Wu's paper [GW17].

Putting together Assumptions (A1) and (A2) of [GW17] (see also its arxiv (2012) version, say [18]), the function V considered there is real-valued, strictly convex, and twice continuously differentiable on $\operatorname{Im} q$. Hence V from [GW17] is more general than being in Γ^2_{sc} when $\operatorname{dom} V = \mathbb{R}^m$. Of course, the strict convexity of V implies $\nabla^2 V(y) \succeq 0$ for $y \in \mathbb{R}^m$, but this property does not imply $(\nabla^2 V)(q(\overline{x})) \succ 0$, which is used for example in [GW17, Eq. (36)].

In "Theorem 2 (Tri-duality Theorem)" (the case $n = m$) and "Theorem 3. (Triality Theorem)" (the case $n \neq m$), $\overline{\sigma} \in S_{\text{col}}$ is a "critical point of the canonical problem $(\mathcal{P}^d)$" and $\overline{x} := [A(\overline{\sigma})]^{-1} b(\overline{\sigma})$, and Assumption (A3) holds, that is $[\nabla f(x) = 0 \Rightarrow \det \nabla^2 f(x) \neq 0]$.

Our result in the case $A(\overline{\sigma}) \succeq 0$ is more general than those in [GW17, Ths. 2, 3] not only because the hypothesis on V in Proposition 2 is weaker and Assumption (A3) is not present, but also because the conclusion in [GW17, Ths. 2, 3] is weaker, more precisely $f(\overline{x}) = \inf_{x\in\mathbb{R}^n} f(x) \Leftrightarrow \sup_{\sigma\in S^+_{\mathrm{col}}} D(\sigma) = D(\overline{\sigma})$.

In what concerns the case $A(\overline{\sigma}) \prec 0$ and $V \in \Gamma^2_{sc}$, Corollary 10 is much more precise than the corresponding results in [GW17, Ths. 2, 3] because it is mentioned when $\overline{x}$ and $\overline{\sigma}$ are local minimizers (maximizers). Moreover, our proofs are very different from those of [GW17], and follow the lines of the proof of Prop. 1 in Voisei-Z. (2013).

Gao and Wu in (2011a, b and 2012) (which are essentially the same) prove [GW17, Ths. 2, 3] for $V(y) := \frac{1}{2}\sum_{k=1}^{m} \beta_k y_k^2$ with $\beta_k > 0$ and $b_k := 0$ ($k \in \overline{1,m}$) [under Assumption (A3)], using similar arguments. Note that $\overline{\sigma}$ is taken to be a critical point of D in "Theorem 4.3 (Refined Triality Theorem)"(the case $n \neq m$) instead of being a "critical point of Problem $(\mathcal{P}^d)$", as in "Theorem 3.1 (Tri-Duality Theorem)".

Morales-Silva and Gao (2011) discuss the problem from Gao and Wu in (2011a) with $A_0 := 0$ and $m := 1$.

Morales-Silva and Gao (2012, 2015) consider $V(y) := \sum_{k=1}^{p} \exp(y_k) + \frac{1}{2}\sum_{k=p+1}^{m} \beta_k y_k^2$ for $0 \leq p \leq m$ (setting $\sum_{k=i}^{j} \gamma_k := 0$ when $j < i$) with $\beta_k > 0$ for $k \in \overline{p+1,m}$; moreover, $b_k := 0$ and $A_k \succeq 0$ for $k \in \overline{1,m}$ are such that there exists $(\alpha_k)_{k\in\overline{1,m}} \subset \mathbb{R}^m_+$ with $\sum_{k=1}^{m} \alpha_k A_k \succ 0$. Under Assumption (A3) and using similar arguments to those in Th. 2 of Gao and Wu in (2012), they prove [GW17, Ths. 2, 3] for $\overline{\sigma}$ "a stationary point of" D.

A special place among DY Gao's papers published after 2010 is occupied by [GRP12] and [RG14] Ruan and Gao (2014).

In [GRP12] one takes the same V as in Gao and Wu (2012) but Assumption (A3) is not considered. Putting together Theorems 2 and 3 from Gao and Wu (2012) for "$\overline{\varsigma}$ a critical point of the canonical dual function $P^d(\overline{\varsigma})$," with the mention "If $n \neq m$, the double-min duality (25) holds conditionally", one gets "Theorem 2 (Triality Theorem)" of [GRP12] .

A detailed proof is provided in the case $\overline{\varsigma} \in \mathcal{S}_a^+$ $(= S^+_{\text{col}})$.

The proof for the case $\overline{\varsigma} \in \mathcal{S}_a^-$ $(= S^-)$ is the following:

"If $\overline{\varsigma} \in \mathcal{S}_a^-$, the matrix $G(\overline{\varsigma})$ is a negative definite. In this case, the Gao–Strang complementary function $\Xi(\overline{x}, \overline{\varsigma})$ is a so-called super-Lagrangian [14], i.e., it is locally concave in both $x \in \mathcal{X}_0 \subset \mathcal{X}_a$ and $\varsigma \in \mathcal{S}_0 \subset \mathcal{S}_a^-$.

By the fact that

$\max_{x\in\mathcal{X}_0} \max_{\varsigma\in\mathcal{S}_0} \Xi(x, \varsigma) = \max_{\varsigma\in\mathcal{S}_0} \max_{x\in\mathbb{R}^n} \Xi(x, \varsigma)$ (26)

holds on the neighborhood $\mathcal{X}_0 \times \mathcal{S}_0$ of $(\overline{x}, \overline{\varsigma})$, we have the double-max duality statement (24). If $n = m$, we have [33]:

$\min_{x\in\mathcal{X}_0} \max_{\varsigma\in\mathcal{S}_0} \Xi(x, \varsigma) = \min_{\varsigma\in\mathcal{S}_0} \max_{x\in\mathbb{R}^n} \Xi(x, \varsigma)$ (27)

which leads to the double-min duality statement (25). This proves the theorem."

AND SO ON.

Thank you for your attention!