vol. 3   no. 2   2015

Mathematics and Mechanics
of
Complex Systems

Daniel Morales-Silva and David Y. Gao

CANONICAL DUALITY THEORY AND TRIALITY
FOR SOLVING GENERAL GLOBAL OPTIMIZATION PROBLEMS IN
COMPLEX SYSTEMS

CANONICAL DUALITY THEORY AND TRIALITY FOR SOLVING GENERAL GLOBAL OPTIMIZATION PROBLEMS IN COMPLEX SYSTEMS

DANIEL MORALES-SILVA AND DAVID Y. GAO

General nonconvex optimization problems are studied by using the canonical duality-triality theory. The triality theory is proved for sums of exponentials and quartic polynomials, which solved an open problem left in 2003. This theory can be used to find the global minimum and local extrema, which bridges a gap between global optimization and nonconvex mechanics. Detailed applications are illustrated by several examples.

1. Introduction and motivation

This paper intends to solve the following nonconvex optimization problem (($\mathcal{P}$) in short):

$$(\mathcal{P}) : \text{ext}\big\{\Pi(\mathbf{x}) = W(\mathbf{x}) + \tfrac{1}{2}\mathbf{x}^t \mathbf{A}\mathbf{x} - \mathbf{f}^t\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\big\}, \tag{1}$$

where ext$\{\cdot\}$ denotes finding extremum points of a function given in $\{\cdot\}$, $\mathbf{f} \in \mathbb{R}^n$ is a given (input) vector, $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a given symmetric matrix, and $W : \mathbb{R}^n \to \mathbb{R}$ is a combination of fourth-order polynomials (double-well functions) and quadratic-exponential functions, namely

$$W(\mathbf{x}) := \sum_{i \in I_m} \exp\big(\tfrac{1}{2}\mathbf{x}^t \mathbf{B}_i \mathbf{x} - \alpha_i\big) + \sum_{j \in I_p} \tfrac{1}{2}b_j\big(\tfrac{1}{2}\mathbf{x}^t \mathbf{C}_j \mathbf{x} - \theta_j\big)^2,$$

where $I_m = \{1, \ldots, m\}$ and $I_p = \{1, \ldots, p\}$ are two integer sets with m and p that are fixed integers; all the coefficients b_j with $j \in I_p$ are positive constants, and $\alpha_i, \theta_j \in \mathbb{R}$ for all $i \in I_m$ and $j \in I_p$ are given parameters; the matrices $\{\mathbf{B}_i\}_{i \in I_m}$ and $\{\mathbf{C}_j\}_{j \in I_p}$ are assumed to be symmetric, positive semidefinite such that the cone generated by them contains a positive-definite matrix.

The nonconvex optimization problem ($\mathcal{P}$) arises naturally in complex systems with a wide range of applications, including chaotic dynamical systems [Gao 2003a; Gao and Ogden 2008a; Gao and Ruan 2008], computational biology [Zhang et al.

2011], chemical-database analysis [Xie and Schlick 2000], large-deformation computational mechanics [Gao 1996; Santos and Gao 2012], population growing [Ruan and Gao 2014a], location/allocation, network communication [Gao et al. 2012], and transitions of solids [Gao and Ogden 2008a; 2008b; Gao and Yu 2008], etc.

For example, the popular sensor-network location problem is to solve the following system of nonlinear equations (see [Aspnes et al. 2004; Moré and Wu 1997]):

$$\|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2^2 = d_{ij}^2 \quad \forall (i, j) \in \mathscr{I}_p, \qquad \boldsymbol{u}_k = \boldsymbol{a}_k \quad \forall k \in \mathscr{I}_b, \tag{2}$$

where the vectors $\boldsymbol{u}_i = \{u_i^\alpha\} \in \mathbb{R}^d$ ($i = 1, \ldots, p$) represent the locations of the unknown sensors, $\mathscr{I}_p = \{(i, j) : i < j, \ d_{ij}$ is specified$\}$ and $\mathscr{I}_b = \{k : \boldsymbol{u}_k = \boldsymbol{a}_k$ is specified$\}$ are two given index sets, d_{ij} are given distances for $(i, j) \in \mathscr{I}_p$, and the given vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_q \in \mathbb{R}^d$ are the so-called anchors. The notation $\|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2$ denotes the Euclidean distance between $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$; i.e.,

$$\|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2 = \sqrt{\sum_{\alpha=1}^{d} (u_i^\alpha - u_j^\alpha)^2}.$$

By using the least-squares method, the quadratic equations (2) of the sensor-localization problem can be reformulated as an optimization problem:

$$\min\left\{ P(\boldsymbol{u}) = \sum_{(i, j) \in \mathscr{I}_p} \tfrac{1}{2} \big(\|\boldsymbol{u}_i - \boldsymbol{u}_j\|_2^2 - d_{ij}^2 \big)^2 : \boldsymbol{u}_i \in \mathscr{U}_a \right\}, \tag{3}$$

where $\mathscr{U}_a = \{\boldsymbol{u} \in \mathbb{R}^{d \times p} : \boldsymbol{u}_k = \boldsymbol{a}_k \ \forall k \in \mathscr{I}_b\}$ is a feasible space. Let $\boldsymbol{x} = \{\{u_1^1, \ldots, u_1^d\}, \ldots, \{u_p^1, \ldots, u_p^d\}\} \in \mathbb{R}^n$ ($n = d \times p$) denote an extended vector. By using the Lagrange-multiplier method to relax the boundary conditions in $\mathscr{U}_a$, the least-squares method for the sensor-localization problem (3) can be written in the problem (1) for certain properly defined matrices $\{\boldsymbol{C}_j\}$, which are the so-called deformation matrices in structural mechanics. The sensor-network-localization-type problems also appear in computational biology, Euclidean ball packing, molecular confirmation, recently wireless network communication, etc. [Ruan and Gao 2014b; Zhang et al. 2011]. Due to the nonconvexity, the sensor-network localization problem is considered to be NP-hard even for the simplest case $d = 1$ [Moré and Wu 1997; Saxe 1979]. A recent result of Aspnes et al. [2004] shows that the problem of computing a realization of the sensors on the plane is NP-complete in general.

Mathematics and mechanics have been complementary partners since Newton. Many fundamental ideas, concepts, and mathematical methods extensively used in calculus of variations and optimization originated from mechanics. For example, the Lagrange-multiplier method was first proposed by Lagrange from the classical analytic mechanics while the concepts of superpotential and subdifferential in modern convex analysis were introduced in [Moreau 1968; Moreau et al. 1988] from

frictional mechanics. From the point of view of computational large-deformation mechanics, both the fourth-order polynomial-minimization problem ($\mathscr{P}$) and the sensor-localization problem (3) are actually two special cases of discretized finite-deformation problems [Gao 1996]. It is known that, in continuum mechanics and differential geometry, the deformation $\boldsymbol{u}(\boldsymbol{x}) : \Omega \to \mathbb{R}^r$ is a vector field over an open domain $\Omega \subset \mathbb{R}^r$, and the minimal-potential variational problem is defined by

$$\min\left\{ P(\boldsymbol{u}) = \int_{\Omega} [W(\nabla \boldsymbol{u}) - \boldsymbol{u}^T \boldsymbol{f}] \, d\Omega : \boldsymbol{u} \in \mathcal{U}_a \right\}, \tag{4}$$

where $W(\boldsymbol{F})$ is the so-called *stored strain energy*, which is usually a nonconvex function of the deformation gradient $\boldsymbol{F} = \nabla \boldsymbol{u}$, the feasible set $\mathcal{U}_a$ in this nonconvex variational problem is called the *kinematically admissible space*, where certain boundary conditions are prescribed. According to the hyperelasticity law (see [Gao 2000b, Chapter 6.1.2] or [Marsden and Hughes 1983]), the stored strain energy should be an *objective function* of the deformation gradient $\boldsymbol{F}$; i.e., there exists an objective strain measure $\boldsymbol{E}(\boldsymbol{F})$ and a convex function $V(\boldsymbol{E})$ such that

$$W(\nabla \boldsymbol{u}) = V(\boldsymbol{E}(\nabla \boldsymbol{u})). \tag{5}$$

One of the most simple objective strain measures is the well known *Green–Saint-Venant strain tensor* $\boldsymbol{E} = \frac{1}{2}[\boldsymbol{F}^T \boldsymbol{F} - \boldsymbol{I}]$. Clearly, this strain measure satisfies the objectivity condition; i.e., $\boldsymbol{E}(\boldsymbol{Q}\boldsymbol{F}) = \boldsymbol{E}(\boldsymbol{F})$ for any given orthonormal (rotation) matrix $\boldsymbol{Q}$. For the most simple *Saint-Venant–Kirchhoff material*, $V(\boldsymbol{E})$ is a quadratic function of $\boldsymbol{E}$; i.e.,

$$V(\boldsymbol{E}) = \tfrac{1}{2}\lambda(\operatorname{tr} \boldsymbol{E})^2 + \mu \operatorname{tr}(\boldsymbol{E})^2, \tag{6}$$

where $\lambda, \mu > 0$ are the classical Lamé constants and $\operatorname{tr} \boldsymbol{E}$ represents the trace of $\boldsymbol{E}$. Therefore, the stored energy $W(\boldsymbol{F})$ is a fourth-order polynomial tensor function of $\boldsymbol{F} = \nabla \boldsymbol{u}$ while for biomaterials the stored energy could be the combination of the polynomial and exponential functions of the Cauchy–Green strain tensor. By using a finite-difference method (FDM), the deformation gradient $\nabla \boldsymbol{u}$ can be directly approximated by the difference $\boldsymbol{D}\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}_i) - \boldsymbol{u}(\boldsymbol{x}_j) = \boldsymbol{u}_i - \boldsymbol{u}_j$ while, in a finite-element method (FEM), the domain $\Omega = \bigcup_e^m \Omega^e$ is discretized by a finite number of elements $\Omega^e \subset \Omega$ and, in each element, the deformation field $\boldsymbol{u}(\boldsymbol{x}) = \sum_i N_i(\boldsymbol{x})\boldsymbol{u}_i$ is numerically represented by the nodal vectors $\boldsymbol{u}_i$ via piecewise interpolation (polynomial) function $N_i(\boldsymbol{x})$ [Gao 1996]. Therefore, by either FDM or FEM, the minimal potential variational problem (4) can be eventually reduced to a very complicated large-scale fourth-order polynomial/exponential minimization problem with the problems ($\mathscr{P}$) as its the most simple case. In the contact mechanics and elastoplastic design of large deformed structures, the nonconvex problems are usually subjected to inequality constraints. In these cases, the global

optimal solution could be local minima (see [Cai et al. 2014]), and to solve such problems is fundamentally difficult by using traditional direct methods.

Canonical duality theory was developed originally by Gao and Strang [1989] for solving general variational problem (4) in finite-deformation theory, where the stored energy $W(\boldsymbol{F})$ is nonconvex and even nonsmooth. By introducing a so-called complementary gap function, they recovered the complementary energy principle in large deformation (geometrically nonlinear) systems. They proved that the non-negative gap function can be used to identify the global minimizer of the nonconvex potential variational problems. Seven years later, it was discovered that the nega-tive gap function can be used to identify the largest local minimum and maximum. Therefore, a so-called triality theory was first proposed in nonconvex mechanics [Gao 1997] and then generalized to global optimization [Gao 2000a]. This triality theory is composed of a canonical min-max duality and two pairs of double-min, double-max dualities, which reveals an intrinsic duality pattern in complex systems and has been used successfully for solving a wide class of challenging problems in complex systems [Gao 1998; 1999; 2009; Gao and Sherali 2009]. However, it was realized [Gao 2003a; 2003b] that the double-min duality holds under certain additional conditions. Recently, this problem is partly solved for a class of fourth-order polynomial optimization problems [Gao and Wu 2012]. Based on these results, this paper intends to solve the more challenging problem $(\mathcal{P})$. We will show that, by the canonical dual transformation, all critical solutions of $(\mathcal{P})$ can be analytically presented in terms of the canonical dual solutions. The extremality of these solutions can be identified by the triality theory. Several solved examples are listed in the last section.

2. Canonical dual problem and analytical solutions

Following the standard procedure of the canonical dual transformation (e.g., [Gao 2003b]), first we need to choose a geometric operator $\Lambda = (\Lambda_1(\boldsymbol{x}), \Lambda_2(\boldsymbol{x})) : \mathbb{R}^n \to \mathbb{R}^{m+p}$, where

$$\Lambda_1(\boldsymbol{x}) = \left\{ \tfrac{1}{2}\boldsymbol{x}^t \boldsymbol{B}_i \boldsymbol{x} - \alpha_i \right\} : \mathbb{R}^n \to \mathbb{R}^m,$$
$$\Lambda_2(\boldsymbol{x}) = \left\{ \tfrac{1}{2}\boldsymbol{x}^t \boldsymbol{C}_j \boldsymbol{x} - \theta_j \right\} : \mathbb{R}^n \to \mathbb{R}^p.$$

Therefore, the nonconvex function $W(\boldsymbol{x})$ can be written in the canonical form

$$W(\boldsymbol{x}) = V(\Lambda(\boldsymbol{x})) = V_1(\Lambda_1(\boldsymbol{x})) + V_2(\Lambda_2(\boldsymbol{x})) \tag{7}$$

with

$$V_1(\boldsymbol{\epsilon}) = \sum_{i \in I_m} \exp(\epsilon_i) \quad \text{and} \quad V_2(\boldsymbol{\gamma}) = \sum_{j \in I_p} \tfrac{1}{2} b_j \gamma_j^2. \tag{8}$$

Clearly, the canonical function $V(\boldsymbol{\varepsilon})$ is convex on

$$\mathcal{V}_a = \{\boldsymbol{\varepsilon} = (\boldsymbol{\epsilon}, \boldsymbol{\gamma}) \in \mathbb{R}^{m+p} : \epsilon_i \in [-\alpha_i, +\infty), \ \gamma_j \in [-\theta_j, +\infty) \ \forall i \in I_m, \ j \in I_p\} \quad (9)$$

such that the canonical dual variable $\boldsymbol{\varsigma} = (\boldsymbol{\tau}, \boldsymbol{\sigma})$ of $\boldsymbol{\varepsilon} = (\boldsymbol{\epsilon}, \boldsymbol{\gamma})$ can be uniquely defined by

$$\boldsymbol{\varsigma} = \nabla V(\boldsymbol{\varepsilon}) \implies \boldsymbol{\tau} = \nabla V_1(\boldsymbol{\epsilon}) = \{\exp(\epsilon_i)\}, \ \boldsymbol{\sigma} = \nabla V_2(\boldsymbol{\gamma}) = \{b_j \gamma_j\}, \quad (10)$$

and on the canonical dual space

$$\mathcal{V}_a^* = \{\boldsymbol{\varsigma} = (\boldsymbol{\tau}, \boldsymbol{\sigma}) \in \mathbb{R}^{m+p} : \tau_i \in [\exp(-\alpha_i), \infty), \ \sigma_j \in [-b_j \theta_j, \infty), \ \forall i \in I_m, \ j \in I_p\},$$

the Legendre conjugate of $V(\boldsymbol{\varepsilon})$ can be defined by

$$V^c(\boldsymbol{\varsigma}) = \text{sta}\{\boldsymbol{\varepsilon}^t \boldsymbol{\varsigma} - V(\boldsymbol{\varepsilon}) : \boldsymbol{\varepsilon} \in \mathcal{V}_a\} = V_1^c(\boldsymbol{\tau}) + V_2^c(\boldsymbol{\sigma}), \quad (11)$$

where sta$\{*\}$ denotes finding stationary points of the function given in $\{*\}$ and

$$V_1^c(\boldsymbol{\tau}) = \sum_{i \in I_m} (\tau_i \ln \tau_i - \tau_i) \quad \text{and} \quad V_2^c(\boldsymbol{\sigma}) = \sum_{j \in I_p} \frac{1}{2b_j} \sigma_j^2. \quad (12)$$

By using the canonical dual transformation $W(\boldsymbol{x}) = V(\Lambda(\boldsymbol{x})) = \Lambda(\boldsymbol{x})^T \boldsymbol{\varsigma} - V^c(\boldsymbol{\varsigma})$, the Gao–Strang *total complementary function* $\Xi : \mathbb{R}^n \times \mathcal{V}_a^* \to \mathbb{R}$ associated with the problem $(\mathcal{P})$ can be given by

$$\begin{aligned}
\Xi(\boldsymbol{x}, \boldsymbol{\varsigma}) &= \langle \Lambda(\boldsymbol{x}), \boldsymbol{\varsigma} \rangle - V^c(\boldsymbol{\varsigma}) + \tfrac{1}{2} \boldsymbol{x}^t \boldsymbol{A} \boldsymbol{x} - \boldsymbol{f}^t \boldsymbol{x} \\
&= \tfrac{1}{2} \boldsymbol{x}^t \boldsymbol{G}(\boldsymbol{\varsigma}) \boldsymbol{x} - \boldsymbol{\alpha}^t \boldsymbol{\tau} - \boldsymbol{\theta}^t \boldsymbol{\sigma} - V_1^c(\boldsymbol{\tau}) - V_2^c(\boldsymbol{\sigma}) - \boldsymbol{f}^t \boldsymbol{x}, \quad (13)
\end{aligned}$$

where

$$\boldsymbol{G}(\boldsymbol{\varsigma}) = \boldsymbol{A} + \sum_{i \in I_m} \tau_i \boldsymbol{B}_i + \sum_{j \in I_p} \sigma_j \boldsymbol{C}_j. \quad (14)$$

Via this $\Xi(\boldsymbol{x}, \boldsymbol{\varsigma})$, the canonical dual function $\Pi^d : \mathcal{V}_a^* \to \mathbb{R}$ can be defined by

$$\Pi^d(\boldsymbol{\varsigma}) := \text{sta}\{\Xi(\boldsymbol{x}, \boldsymbol{\varsigma}) : \boldsymbol{x} \in \mathbb{R}^n\} = \{\Xi(\boldsymbol{x}(\boldsymbol{\varsigma}), \boldsymbol{\varsigma}) : \nabla_{\boldsymbol{x}} \Xi(\boldsymbol{x}(\boldsymbol{\varsigma}), \boldsymbol{\varsigma}) = 0\}.$$

Notice that $\nabla_{\boldsymbol{x}} \Xi(\boldsymbol{x}, \boldsymbol{\varsigma}) = \boldsymbol{G}(\boldsymbol{\varsigma}) \boldsymbol{x} - \boldsymbol{f} = 0$ if and only if

$$\boldsymbol{G}(\boldsymbol{\varsigma}) \boldsymbol{x} = \boldsymbol{f}. \quad (15)$$

Let $\mathcal{C}_{ol}(\boldsymbol{G}(\boldsymbol{\varsigma}))$ be the space generated by the columns of the matrix $\boldsymbol{G}(\boldsymbol{\varsigma})$. Then on the dual feasible space

$$\mathcal{S}_a = \{\boldsymbol{\varsigma} \in \mathcal{V}_a^* : \boldsymbol{f} \in \mathcal{C}_{ol}(\boldsymbol{G}(\boldsymbol{\varsigma}))\},$$

the primal solution $\boldsymbol{x} = (\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1} \boldsymbol{f}$ is well defined (if $\boldsymbol{G}(\boldsymbol{\varsigma})$ is singular, $(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}$ denotes its pseudoinverse; see [Desoer and Whalen 1963; Peters and Wilkinson

1970] and references therein), and we have $\Pi^d : \mathscr{S}_a \to \mathbb{R}$, where

$$\Pi^d(\varsigma) = -\tfrac{1}{2} f^t (G(\varsigma))^{-1} f - V_1^c(\tau) - V_2^c(\sigma) - \alpha^t \tau - \theta^t \sigma. \qquad (16)$$

Therefore, the canonical dual problem is proposed in the form

$$(\mathscr{P}^d) : \text{ext}\{\Pi^d(\varsigma) : \varsigma \in \mathscr{S}_a\}. \qquad (17)$$

By the canonical duality theory, it is not difficult to show that

$$\Pi(x) = \text{sta}\{\Xi(x, \varsigma) : \varsigma \in \mathscr{S}_a\} = \Xi(x, \varsigma(x)), \qquad (18)$$

where $\varsigma(x) = (\tau(x), \sigma(x))$ and

$$(\tau(x))_i = \exp((\Lambda_1(x))_i), \quad i \in I_m,$$
$$(\sigma(x))_j = b_j (\Lambda_2(x))_j, \qquad j \in I_p.$$

According to the general theory presented in [Gao 2003b], we have:

Theorem 1 (analytical solutions). *Suppose that for a given $f \in \mathbb{R}^n$ the canonical dual space $\mathscr{S}_a$ is not empty. If $\bar{\varsigma} \in \mathscr{S}_a$ is a stationary point of Π^d, then*

$$\bar{x} = (G(\bar{\varsigma}))^{-1} f \qquad (19)$$

is a stationary point of Π and

$$\Pi(\bar{x}) = \Pi^d(\bar{\varsigma}). \qquad (20)$$

Proof. Let us calculate $\nabla \Pi^d(\varsigma)$ and $\nabla^2 \Pi^d(\varsigma)$. We know that

$$\nabla \Pi^d(\varsigma) = \begin{bmatrix} \nabla_\tau \Pi^d(\varsigma) \\ \nabla_\sigma \Pi^d(\varsigma) \end{bmatrix} \in \mathbb{R}^{m+p};$$

then

$$(\nabla_\tau \Pi^d(\varsigma))_i = \tfrac{1}{2} f^t (G(\varsigma))^{-1} B_i (G(\varsigma))^{-1} f - \ln \tau_i - \alpha_i, \quad i \in I_m, \qquad (21)$$
$$(\nabla_\sigma \Pi^d(\varsigma))_j = \tfrac{1}{2} f^t (G(\varsigma))^{-1} C_j (G(\varsigma))^{-1} f - \frac{\sigma_j}{b_j} - \theta_j, \quad j \in I_p. \qquad (22)$$

On the other hand,

$$\nabla^2 \Pi^d(\varsigma) = \begin{bmatrix} \nabla^2_{\tau\tau} \Pi^d(\varsigma) & \nabla^2_{\tau\sigma} \Pi^d(\varsigma) \\ \nabla^2_{\sigma\tau} \Pi^d(\varsigma) & \nabla^2_{\sigma\sigma} \Pi^d(\varsigma) \end{bmatrix} \in \mathbb{R}^{m+p} \times \mathbb{R}^{m+p},$$

where $\nabla_{\tau\sigma}^2 \Pi^d(\boldsymbol{\varsigma}) := (\nabla_\tau(\nabla_\sigma \Pi^d(\boldsymbol{\varsigma}))^t)$. Let δ_{ij} be the Kronecker delta. Then

$$(\nabla_{\tau\tau}^2 \Pi^d(\boldsymbol{\varsigma}))_{ij} = -\boldsymbol{f}^t(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{B}_i(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{B}_j(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{f} - \frac{\delta_{ij}}{\tau_j}, \quad i, j \in I_m,$$

$$(\nabla_{\tau\sigma}^2 \Pi^d(\boldsymbol{\varsigma}))_{ij} = -\boldsymbol{f}^t(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{B}_i(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{C}_j(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{f}, \quad i \in I_m, \ j \in I_p,$$

$$(\nabla_{\sigma\tau}^2 \Pi^d(\boldsymbol{\varsigma}))_{ij} = -\boldsymbol{f}^t(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{C}_i(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{B}_j(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{f}, \quad i \in I_m, \ j \in I_p,$$

$$(\nabla_{\sigma\sigma}^2 \Pi^d(\boldsymbol{\varsigma}))_{ij} = -\boldsymbol{f}^t(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{C}_i(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{C}_j(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{f} - \frac{\delta_{ij}}{b_j}, \quad i, j \in I_p.$$

By making $\boldsymbol{x} = (\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{f}$ and $\boldsymbol{F}(\boldsymbol{x}) \in \mathbb{R}^{n \times (m+p)}$ be $\boldsymbol{F}(\boldsymbol{x}) = [\boldsymbol{B}_1\boldsymbol{x}, \dots, \boldsymbol{B}_m\boldsymbol{x}, \boldsymbol{C}_1\boldsymbol{x}, \dots, \boldsymbol{C}_p\boldsymbol{x}]$, we have

$$\nabla^2 \Pi^d(\boldsymbol{\varsigma}) = -\boldsymbol{F}(\boldsymbol{x})^t(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{F}(\boldsymbol{x}) - \mathrm{Diag}\left(\frac{1}{\tau_1}, \dots, \frac{1}{\tau_m}, \frac{1}{b_1}, \dots, \frac{1}{b_p}\right). \quad (23)$$

Let $\boldsymbol{D} = \mathrm{Diag}(\tau_1, \dots, \tau_m, b_1, \dots, b_p)$; then $\nabla^2\Pi^d(\boldsymbol{\varsigma})$ can be written as

$$\nabla^2 \Pi^d(\boldsymbol{\varsigma}) = -\boldsymbol{F}(\boldsymbol{x})^t(\boldsymbol{G}(\boldsymbol{\varsigma}))^{-1}\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{D}^{-1}. \quad (24)$$

Calculating $\nabla\Pi(\boldsymbol{x})$ and $\nabla^2\Pi(\boldsymbol{x})$, we have respectively

$$\nabla\Pi(\boldsymbol{x}) = \sum_{i\in I_m} \exp\left(\tfrac{1}{2}\boldsymbol{x}^t\boldsymbol{B}_i\boldsymbol{x} - \alpha_i\right)\boldsymbol{B}_i\boldsymbol{x} + \sum_{j\in I_p} b_j\left(\tfrac{1}{2}\boldsymbol{x}^t\boldsymbol{C}_j\boldsymbol{x} - \theta_j\right)\boldsymbol{C}_j\boldsymbol{x} + \boldsymbol{A}\boldsymbol{x} - \boldsymbol{f}, \quad (25)$$

$$\nabla^2\Pi(\boldsymbol{x}) = \boldsymbol{A} + \sum_{i\in I_m} \exp\left(\tfrac{1}{2}\boldsymbol{x}^t\boldsymbol{B}_i\boldsymbol{x} - \alpha_i\right)(\boldsymbol{B}_i\boldsymbol{x}(\boldsymbol{B}_i\boldsymbol{x})^t + \boldsymbol{B}_i)$$
$$+ \sum_{j\in I_p} b_j\left(\boldsymbol{C}_j\boldsymbol{x}(\boldsymbol{C}_j\boldsymbol{x})^t + \left(\tfrac{1}{2}\boldsymbol{x}^t\boldsymbol{C}_j\boldsymbol{x} - \theta_j\right)\boldsymbol{C}_j\right). \quad (26)$$

Since $\bar{\boldsymbol{\varsigma}} = (\bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\sigma}})$ is a stationary point of Π^d, then by (21) and (22),

$$(\Lambda_1(\bar{\boldsymbol{x}}))_i = \ln\bar{\tau}_i, \quad i \in I_m, \quad (27)$$

$$(\Lambda_2(\bar{\boldsymbol{x}}))_j = \frac{\bar{\sigma}_j}{b_j}, \quad j \in I_p. \quad (28)$$

Using (27) and (28) in (25), we obtain

$$\nabla\Pi(\bar{\boldsymbol{x}}) = \boldsymbol{G}(\bar{\boldsymbol{\varsigma}})\bar{\boldsymbol{x}} - \boldsymbol{f} = \boldsymbol{G}(\bar{\boldsymbol{\varsigma}})(\boldsymbol{G}(\bar{\boldsymbol{\varsigma}}))^{-1}\boldsymbol{f} - \boldsymbol{f} = 0.$$

Notice that (27) and (28) together with (16) and (18) imply that

$$\Pi(\bar{\boldsymbol{x}}) = \Xi(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\varsigma}}) = \Xi((\boldsymbol{G}(\bar{\boldsymbol{\varsigma}}))^{-1}\boldsymbol{f}, \bar{\boldsymbol{\varsigma}}) = \Pi^d(\bar{\boldsymbol{\varsigma}}). \quad (29)$$

And this finishes the proof. $\qquad\square$

Remark 1. This theorem shows that the problem $(\mathcal{P}^d)$ is canonical dual to the nonconvex primal problem $(\mathcal{P})$ in the sense that $\Pi(\bar{\boldsymbol{x}}) = \Pi^d(\bar{\boldsymbol{\varsigma}})$ at each critical point

of $\Xi(x, \varsigma)$. By the criticality condition (15), we know that, if $G(\varsigma)$ is singular at $\bar{\varsigma}$, the canonical equilibrium equation (15) may have an infinite number of solutions: $\bar{x} = G(\bar{\varsigma})^{\dagger} f + N x^o$, where $G^{\dagger}$ represents the Moore–Penrose generalized inverse, N is a basis matrix of the null space of $G(\bar{\varsigma})$, and x^o is a free vector. In this case, Theorem 1 still holds, but the canonical dual function Π^d will have the additional parametric vector x^o. In order to avoid this case, a quadratic perturbation method is introduced in [Ruan and Gao 2014b]; i.e., in the case that $G(\bar{\varsigma})$ is singular, replace it by the perturbed form

$$G_{\alpha}(\bar{\varsigma}) = G(\bar{\varsigma}) + \alpha D, \tag{30}$$

where $\alpha > 0$ is a perturbation parameter and D is a given positive-definite matrix. Very often, $D = I$. A detailed study on this quadratic perturbation method is given in [Ruan and Gao 2014b].

In the next section, we will show that the extremality of some of these solutions can be identified by a refined triality theory.

3. Triality theory

Before presenting the refined triality theory, we need the sets

$$\mathcal{S}_a^+ := \{\varsigma \in \mathcal{S}_a : G(\varsigma) \succeq 0\} \quad \text{and} \quad \mathcal{S}_a^- := \{\varsigma \in \mathcal{S}_a : G(\varsigma) \prec 0\}.$$

Lemma 1. *Suppose that $m + p < n$, $\bar{\varsigma} \in \mathcal{S}_a^-$ is a stationary point and a local minimizer of Π^d, and $\bar{x} = (G(\bar{\varsigma}))^{-1} f$. Then there exists a matrix $L \in \mathbb{R}^{n \times (m+p)}$ with $\mathrm{Rank}(L) = m + p$ such that*

$$L^t \nabla^2 \Pi(\bar{x}) L \succeq 0. \tag{31}$$

Proof. Since $\bar{\varsigma} \in \mathcal{S}_a^-$ is a local minimizer of Π^d, we have that $\nabla^2 \Pi^d(\bar{\varsigma}) \succeq 0$. It follows from (24) that

$$-F(\bar{x})^t (G(\bar{\varsigma}))^{-1} F(\bar{x}) \succeq D^{-1} \succ 0.$$

Thus, $\mathrm{Rank}(F(\bar{x})) = m + p$. Since $\bar{\varsigma} \in \mathcal{S}_a^-$ and $F(\bar{x}) D F(\bar{x})^t \succeq 0$, there exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that

$$T^t G(\bar{\varsigma}) T = \mathrm{Diag}(-\lambda_1, \dots, -\lambda_n), \tag{32}$$

$$T^t F(\bar{x}) D F(\bar{x})^t T = \mathrm{Diag}(a_1, \dots, a_{m_1+m_2}, 0, \dots, 0), \tag{33}$$

where $\lambda_i > 0$ for every $i = 1, \dots, n$ and $a_j > 0$ for every $j = 1, \dots, m + p$ (see [Feng et al. 2012; Horn and Johnson 1985] and references therein). According to Lemma A1 in the Appendix, we know that there exist orthogonal matrices $U \in \mathbb{R}^{n \times n}$ and $E \in \mathbb{R}^{(m+p) \times (m+p)}$ such that

$$T^t F(\bar{x}) D^{1/2} = U R E, \tag{34}$$

where $\boldsymbol{R} \in \mathbb{R}^{n \times (m+p)}$ and

$$
\boldsymbol{R}_{ij} = \begin{cases} \sqrt{a_i} & \text{if } i = j \text{ and } i = 1, \ldots, m+p, \\ 0 & \text{otherwise.} \end{cases}
$$

According to the singular value decomposition theory, we know that $\boldsymbol{U}$ is the identity matrix. Then

$$
\begin{aligned}
\nabla^2 \Pi^d(\bar{\varsigma}) &= -\boldsymbol{F}(\bar{x})^t (\boldsymbol{G}(\bar{\varsigma}))^{-1} \boldsymbol{F}(\bar{x}) - \boldsymbol{D}^{-1} \\
&= -(\boldsymbol{F}(\bar{x})^t \boldsymbol{T})[\boldsymbol{T}^t \boldsymbol{G}(\bar{\varsigma}) \boldsymbol{T}]^{-1} (\boldsymbol{T}^t \boldsymbol{F}(\bar{x})) - \boldsymbol{D}^{-1} \\
&= -\boldsymbol{D}^{-1/2} \boldsymbol{E}^t \boldsymbol{R}^t \operatorname{Diag}\left(-\frac{1}{\lambda_1}, \ldots, -\frac{1}{\lambda_n}\right) \boldsymbol{R} \boldsymbol{E} \boldsymbol{D}^{-1/2} - \boldsymbol{D}^{-1} \succeq 0.
\end{aligned}
$$

Multiplying by $\boldsymbol{D}^{1/2}$ from the left and the right,

$$
\boldsymbol{D}^{1/2} \nabla^2 \Pi^d(\bar{\varsigma}) \boldsymbol{D}^{1/2} = -\boldsymbol{E}^t \boldsymbol{R}^t \operatorname{Diag}\left(-\frac{1}{\lambda_1}, \ldots, -\frac{1}{\lambda_n}\right) \boldsymbol{R} \boldsymbol{E} - \boldsymbol{I}_{(m+p) \times (m+p)} \succeq 0.
$$

If we multiply the right side of the last equation by $\boldsymbol{E}$ from the left and $\boldsymbol{E}^t$ from the right, we have

$$
\begin{aligned}
0 \preceq &-\boldsymbol{R}^t \operatorname{Diag}\left(-\frac{1}{\lambda_1}, \ldots, -\frac{1}{\lambda_n}\right) \boldsymbol{R} - \boldsymbol{I}_{(m+p) \times (m+p)} \\
&\preceq \operatorname{Diag}\left(\frac{a_1}{\lambda_1} - 1, \ldots, \frac{a_{m+p}}{\lambda_{m+p}} - 1\right);
\end{aligned} \tag{35}
$$

thus, $a_i \geq \lambda_i$ for every $i = 1, \ldots, m+p$. On the other hand,

$$
\begin{aligned}
\boldsymbol{T}^t \nabla^2 \Pi(\bar{x}) \boldsymbol{T} &= \boldsymbol{T}^t \boldsymbol{G}(\bar{\varsigma}) \boldsymbol{T} + \boldsymbol{T}^t \boldsymbol{F}(\bar{x}) \boldsymbol{D} \boldsymbol{F}(\bar{x})^t \boldsymbol{T} \\
&= \operatorname{Diag}(-\lambda_1, \ldots, -\lambda_n) + \operatorname{Diag}(a_1, \ldots, a_{m+p}, 0, \ldots, 0) \\
&= \operatorname{Diag}(a_1 - \lambda_1, \ldots, a_{m+p} - \lambda_{m+p}, -\lambda_{m+p+1}, \ldots, -\lambda_n).
\end{aligned}
$$

Let $\boldsymbol{J} \in \mathbb{R}^{n \times n}$ be defined by

$$
J_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } i = 1, \ldots, m+p, \\ 0 & \text{otherwise.} \end{cases}
$$

Then we have

$$
\boldsymbol{J}^t \boldsymbol{T}^t \nabla^2 \Pi(\bar{x}) \boldsymbol{T} \boldsymbol{J} = \operatorname{Diag}(a_1 - \lambda_1, \ldots, a_{m+p} - \lambda_{m+p}) \succeq 0. \tag{36}
$$

Let $\boldsymbol{L} = \boldsymbol{T} \boldsymbol{J}$; clearly $\operatorname{Rank}(\boldsymbol{L}) = m + p$ and $\boldsymbol{L}^t \nabla^2 \Pi(\bar{x}) \boldsymbol{L} \succeq 0$. □

In a similar way, we can prove the following lemma:

Lemma 2. *Suppose that $m + p > n$, $\bar{\varsigma} \in \mathcal{S}_a^-$ is a stationary point Π^d, and $\bar{x} = (\boldsymbol{G}(\bar{\varsigma}))^{-1} \boldsymbol{f}$ is a local minimizer of Π. Then there exists a matrix $\boldsymbol{Q} \in \mathbb{R}^{(m+p) \times n}$ with $\operatorname{Rank}(\boldsymbol{Q}) = n$ such that*

$$
\boldsymbol{Q}^t \nabla^2 \Pi^d(\bar{\varsigma}) \boldsymbol{Q} \succeq 0. \tag{37}
$$

Let the $m + p$ column vectors of $\boldsymbol{L}$ be $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_{m+p}$, respectively, and the n column vectors of $\boldsymbol{Q}$ be $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$, respectively. Clearly, $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_{m+p}$ are $m + p$ independent vectors and $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are n independent vectors. Now the subspaces $\mathscr{X}_b$ and $\mathscr{S}_b$ are defined as

$$\mathscr{X}_b = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = \bar{\boldsymbol{x}} + \sum_{i=1}^{m+p} \upsilon_i \boldsymbol{l}_i, \; \{\upsilon_i\}_{i=1}^{m+p} \subset \mathbb{R} \right\}, \tag{38}$$

$$\mathscr{S}_b = \left\{ \boldsymbol{\varsigma} \in \mathbb{R}^{m+p} : \boldsymbol{\varsigma} = \bar{\boldsymbol{\varsigma}} + \sum_{j=1}^{n} \vartheta_j \boldsymbol{q}_j, \; \{\vartheta_j\}_{j=1}^{n} \subset \mathbb{R} \right\}. \tag{39}$$

Now we are ready to present the refined triality theory.

Theorem 2 (triality theory). *Let $\bar{\boldsymbol{\varsigma}}$ be a stationary point of Π^d and $\bar{\boldsymbol{x}} = (\boldsymbol{G}(\bar{\boldsymbol{\varsigma}}))^{-1} \boldsymbol{f}$. Assume that $\det(\nabla^2 \Pi(\bar{\boldsymbol{x}})) \neq 0$.*

(i) *If $\bar{\boldsymbol{\varsigma}} \in \mathscr{S}_a^+$, then $\bar{\boldsymbol{\varsigma}}$ is the only global maximizer of Π^d in $\mathscr{S}_a^+$ and $\bar{\boldsymbol{x}}$ is the only global minimizer of Π.*

(ii) *If $\bar{\boldsymbol{\varsigma}} \in \mathscr{S}_a^-$, then $\bar{\boldsymbol{\varsigma}}$ is a local maximizer of Π^d in $\mathscr{S}_a^-$ if and only if $\bar{\boldsymbol{x}}$ is a local maximizer of Π.*

(iii) *If $\bar{\boldsymbol{\varsigma}} \in \mathscr{S}_a^-$ and:*

   (a) *If $n = m + p$, then $\bar{\boldsymbol{\varsigma}}$ is a local minimizer of Π^d if and only if $\bar{\boldsymbol{x}}$ is a local minimizer of Π; i.e., there exist neighborhoods $\mathscr{X}, \mathscr{S} \subset \mathbb{R}^n$ of $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{\varsigma}}$, respectively, such that*

$$\Pi(\bar{\boldsymbol{x}}) = \min_{\boldsymbol{x} \in \mathscr{X}} \Pi(\boldsymbol{x}) = \min_{\boldsymbol{\varsigma} \in \mathscr{S}} \Pi^d(\boldsymbol{\varsigma}) = \Pi^d(\bar{\boldsymbol{\varsigma}}). \tag{40}$$

   (b) *If $m + p < n$ and $\bar{\boldsymbol{\varsigma}}$ is a local minimizer of Π^d, then $\bar{\boldsymbol{x}}$ is a saddle point of Π and there exist neighborhoods $\mathscr{X}, \mathscr{S} \subset \mathbb{R}^n$ of $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{\varsigma}}$, respectively, such that*

$$\Pi(\bar{\boldsymbol{x}}) = \min_{\boldsymbol{x} \in \mathscr{X} \cap \mathscr{X}_b} \Pi(\boldsymbol{x}) = \min_{\boldsymbol{\varsigma} \in \mathscr{S}} \Pi^d(\boldsymbol{\varsigma}) = \Pi^d(\bar{\boldsymbol{\varsigma}}). \tag{41}$$

   (c) *If $n < m + p$ and $\bar{\boldsymbol{x}}$ is a local minimizer of Π, then $\bar{\boldsymbol{\varsigma}}$ is a saddle point of Π^d and there exist neighborhoods $\mathscr{X}, \mathscr{S} \subset \mathbb{R}^n$ of $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{\varsigma}}$, respectively, such that*

$$\Pi(\bar{\boldsymbol{x}}) = \min_{\boldsymbol{x} \in \mathscr{X}} \Pi(\boldsymbol{x}) = \min_{\boldsymbol{\varsigma} \in \mathscr{S} \cap \mathscr{S}_b} \Pi^d(\boldsymbol{\varsigma}) = \Pi^d(\bar{\boldsymbol{\varsigma}}). \tag{42}$$

Proof. (i) Since $\bar{\boldsymbol{\varsigma}} \in \mathscr{S}_a^+$, from (24), it is not difficult to show that Π^d is strictly concave in $\mathscr{S}_a^+$ and $\Xi(\,\cdot\,, \bar{\boldsymbol{\varsigma}})$ is strictly convex in $\mathbb{R}^n$ and therefore $\bar{\boldsymbol{\varsigma}}$ must be the only global maximizer of Π^d in $\mathscr{S}_a^+$ and $\bar{\boldsymbol{x}}$ is the only global minimizer of $\Xi(\,\cdot\,, \bar{\boldsymbol{\varsigma}})$.

By the definition of Ξ given in (13) and the convexity of V, the Fenchel inequality leads to

$$\Xi(x, \varsigma) \leq \Pi(x) \quad \forall (x, \varsigma) \in \mathbb{R}^n \times \mathscr{S}_a.$$

Let us assume now that there exists a vector $x' \in \mathbb{R}^n \setminus \{\bar{x}\}$ such that $\Pi(x') \leq \Pi(\bar{x})$; then

$$\Pi(\bar{x}) \geq \Pi(x') \geq \Xi(x', \bar{\varsigma}) > \Xi(\bar{x}, \bar{\varsigma}) = \Pi(\bar{x}),$$

where the last equality comes from (29). This contradiction proves that $\bar{x}$ must be the only global minimizer of Π.

(ii) Notice first that using (27) and (28) in (26) we have

$$\nabla^2 \Pi(\bar{x}) = G(\bar{\varsigma}) + F(\bar{x}) D F(\bar{x})^t, \tag{43}$$

where $F(x)$ and D are defined in (24). If $\bar{\varsigma}$ is a local maximizer of Π^d in $\mathscr{S}_a^-$, we must have that $\nabla^2 \Pi^d(\bar{\varsigma}) \preceq 0$ from (24), which is equivalent to

$$D^{-1} + F(\bar{x})^t (G(\bar{\varsigma}))^{-1} F(\bar{x}) \succeq 0. \tag{44}$$

• If $m + p = n$ and F is invertible, multiplying (44) by $(F(\bar{x})^t)^{-1}$ from the left and $(F(\bar{x}))^{-1}$ from the right, we have

$$(F(\bar{x})^t)^{-1} D^{-1} (F(\bar{x}))^{-1} + (G(\bar{\varsigma}))^{-1} \succeq 0. \tag{45}$$

This is equivalent to

$$(F(\bar{x})^t)^{-1} D^{-1} (F(\bar{x}))^{-1} \succeq -(G(\bar{\varsigma}))^{-1} \succ 0,$$

which in turn is equivalent to (Lemma A2 in the Appendix)

$$-G(\bar{\varsigma}) \succeq F(\bar{x}) D F(\bar{x})^t \iff \nabla^2 \Pi(\bar{x}) \preceq 0.$$

By assumption, $\det(\nabla^2 \Pi(\bar{x})) \neq 0$; then $\bar{x}$ is a local maximum of Π.

• If $m + p \neq n$ or F is not invertible, then by Lemma A1, there exist orthogonal matrices $E \in \mathbb{R}^{n \times n}$ and $K \in \mathbb{R}^{(m+p) \times (m+p)}$ and a matrix $R \in \mathbb{R}^{n \times (m+p)}$ such that

$$R_{ij} = \begin{cases} s_i & \text{if } i = j \text{ and } i = 1, \dots, r, \\ 0 & \text{otherwise,} \end{cases}$$

where $s_i > 0$ for every i, $r = \text{Rank}(F(\bar{x}))$, and

$$F(\bar{x}) D^{1/2} = E R K. \tag{46}$$

Using (46), (44) can be rewritten as

$$D^{-1} + D^{-1/2} K^t R^t E^t (G(\bar{\varsigma}))^{-1} E R K D^{-1/2} \succeq 0.$$

After multiplying this equation by $KD^{1/2}$ from the left and $D^{1/2}K^t$ from the right, we have

$$I_{(m+p)\times(m+p)} + R^t(E^tG(\bar{\varsigma})E)^{-1}R \succeq 0.$$

This equation is equivalent to

$$-I_{(m+p)\times(m+p)} - R^t(E^tG(\bar{\varsigma})E)^{-1}R \preceq 0.$$

By [Lemma A3](#) in the [Appendix](#), the last equation is equivalent to

$$0 \succeq E^tG(\bar{\varsigma})E + RR^t = E^tG(\bar{\varsigma})E + R(KD^{-1/2}DD^{-1/2}K^t)R^t.$$

Multiplying by E from the left and E^t from the right, we can obtain that

$$0 \succeq G(\bar{\varsigma}) + (ERKD^{-1/2})D(D^{-1/2}K^tR^tE^t) = G(\bar{\varsigma}) + F(\bar{x})DF(\bar{x})^t = \nabla^2\Pi(\bar{x}).$$

By the assumption $\det(\nabla^2\Pi(\bar{x})) \neq 0$, $\bar{x}$ is a local maximum of Π.

Notice that every step of the proof is equivalent, so if $\bar{x}$ is a local maximum of Π, then $\bar{\varsigma}$ must be a local maximum of Π^d.

[(iii)](#) Let us consider the three cases:

[(a)](#) Assume $n = m + p$. If $\bar{\varsigma}$ is a local minimizer of Π^d, then

$$\nabla^2\Pi^d(\bar{\varsigma}) = -F(\bar{x})^t(G(\bar{\varsigma}))^{-1}F(\bar{x}) - D^{-1} \succeq 0 \iff -F(\bar{x})^t(G(\bar{\varsigma}))^{-1}F(\bar{x}) \succeq D^{-1}.$$

This implies that $\mathrm{Rank}(F(\bar{x})) = n$. By multiplying the last inequality by $(F(\bar{x})^t)^{-1}$ from the left and by $(F(\bar{x}))^{-1}$ from the right, we have

$$-(G(\bar{\varsigma}))^{-1} \succeq (F(\bar{x})^t)^{-1}D^{-1}(F(\bar{x}))^{-1}.$$

By [Lemma A2](#), this is equivalent to

$$-G(\bar{\varsigma}) \preceq F(\bar{x})DF(\bar{x})^t \iff \nabla^2\Pi(\bar{x}) \succeq 0.$$

And since $\det(\nabla^2\Pi(\bar{x})) \neq 0$, $\bar{x}$ is a local minimizer of Π. In a similar way, we can prove the converse.

[(b)](#) From [(24)](#), we know that

$$-F(\bar{x})^t(G(\bar{\varsigma}))^{-1}F(\bar{x}) \succeq D^{-1}.$$

Then $-F(\bar{x})^t(G(\bar{\varsigma}))^{-1}F(\bar{x})$ is a nonsingular matrix, and $\mathrm{Rank}(F(\bar{x})) = m+p < n$. We claim now that $\bar{x}$ is not a local minimizer of Π. This is because, if $\bar{x}$ were also a local minimizer, we would have

$$\nabla^2\Pi(\bar{x}) = G(\bar{\varsigma}) + F(\bar{x})DF(\bar{x})^t \succeq 0.$$

Thus,

$$F(\bar{x})DF(\bar{x})^t \succeq -G(\bar{\varsigma}).$$

This implies that

$$n = \text{Rank}(-\boldsymbol{G}(\bar{\boldsymbol{\varsigma}})) = \text{Rank}(\boldsymbol{F}(\bar{\boldsymbol{x}})\boldsymbol{DF}(\bar{\boldsymbol{x}})^t) = m + p,$$

which is a contradiction. Therefore, $\bar{\boldsymbol{x}}$ is a saddle point of Π.

To prove (41), we let $\boldsymbol{L}$ be the matrix as given in Lemma 1 and $\{\boldsymbol{l}_i\}_{i=1}^{m+p}$ be the column vectors of $\boldsymbol{L}$. Define

$$\varphi(t_1, \ldots, t_{m+p}) := \Pi(\bar{\boldsymbol{x}} + t_1\boldsymbol{l}_1 + \cdots + t_{m+p}\boldsymbol{l}_{m+p}).$$

We need to show that $(0, \ldots, 0) \in \mathbb{R}^{m+p}$ is a local minimizer of the function φ. Notice that

$$\nabla\varphi(0, \ldots, 0) = \boldsymbol{L}^t\nabla\Pi(\bar{\boldsymbol{x}}) = 0$$

and

$$\nabla^2\varphi(0, \ldots, 0) = \boldsymbol{L}^t\nabla^2\Pi(\bar{\boldsymbol{x}})\boldsymbol{L} \succeq 0,$$

which is a consequence of Lemma 1. Furthermore, from (36) we have that

$$\nabla^2\varphi(0, \ldots, 0) = \text{Diag}(a_1 - \lambda_1, \ldots, a_{m+p} - \lambda_{m+p}),$$

and since $\det(\nabla^2\Pi(\bar{\boldsymbol{x}})) \neq 0$, it can be proven that $a_i > \lambda_i$ for every i.

(c) The proof is similar to that of part (b).   □

Remark 2. Theorem 2 shows that, in order to solve the problem $(\mathscr{P})$ by means of the canonical duality theory, a necessary condition is that the problem $(\mathscr{P})$ should have a unique solution. It was indicated in [Ruan and Gao 2014b] that, if the nonconvex minimization problem has more than one global minimizer, it could be NP-hard. In order to solve this type of problems, the perturbation methods should be used.

Remark 3. The triality theory states precisely that, if $\boldsymbol{\varsigma}$ is a global maximizer of Π^d on a certain set, then $\boldsymbol{x}$ is a global minimizer for Π. This is known from the general result by Gao and Strang [1989]. If $\boldsymbol{\varsigma}$ is a local maximizer for Π^d, then $\boldsymbol{x}$ is also a local maximizer for Π. This is the so-called double-max duality statement. If $\boldsymbol{\varsigma}$ is a local minimizer for Π^d, then $\boldsymbol{x}$ is also a local minimizer for Π in certain directions. This is the so-called double-min duality in the standard triality form proposed in [Gao 2000b]. The triality theory was first discovered in nonconvex mechanics [Gao 1997]. Gao [2003a; 2003b] realized that the double-min duality holds under certain additional condition, which was left as an open problem. Recently, this open problem was solved for the quartic polynomial optimization problem [Gao and Wu 2012]. This result is now generalized to the general nonconvex problem $(\mathscr{P})$. Part (iii) of Theorem 2 shows that, if $m + p = n$, then $\boldsymbol{\varsigma}$ is a local minimizer if and only if $\boldsymbol{x}$ is also a local minimizer. In other cases, either $\boldsymbol{x}$ is a saddle point of Π or $\boldsymbol{\varsigma}$ is a saddle point of Π^d.

Remark 4. The canonical duality-triality theory has been challenged recently by C. Zălinescu and his coworkers R. Strugariu and M. D. Voisei in several papers (see [Strugariu et al. 2011]). By listing some simple "counterexamples", they claimed that this theory is false. Unfortunately, most of these counterexamples are not new and were first discovered by Gao [2003a; 2003b], who was never cited. Some of their "counterexamples" are fundamentally wrong; i.e. they incorrectly choose linear functions as the stored energy and nonlinear functions as external energy (see [Voisei and Zălinescu 2011]). These conceptual mistakes show a big gap between mathematics and mechanics.

4. Numerical examples

In the following examples, $m = p = 1$ and $b_1 = 1$. The graphs provided and the numerical results were obtained using MAXIMA [2010].

4.1. *One stationary point in $\mathcal{S}_a^+$.* First, we consider the case that the primal function has a unique solution. We let $\alpha_1 = \theta_1 = 1$ and

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad f = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Clearly, the function $\Pi : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\Pi(x, y) = \exp\left(\tfrac{1}{2}(x^2 + 2y^2) - 1\right) + \tfrac{1}{2}\left(\tfrac{1}{2}(x^2 + y^2) - 1\right)^2 + \tfrac{1}{2}(x^2 - y^2) - x - y,$$

and the dual function has the form of

$$\Pi^d(\tau, \sigma) = -\frac{1}{2}\left(\frac{1}{1+\tau+\sigma} + \frac{1}{2\tau+\sigma-1}\right) - \tau \cdot \ln(\tau) - \tfrac{1}{2}\sigma^2 - \sigma.$$

It can be shown that Π^d has only one critical point in $\mathcal{S}_a^+$ and it is given (approximately) by

$$\bar{\varsigma} = (1.171057661103504, -0.34599084656216).$$
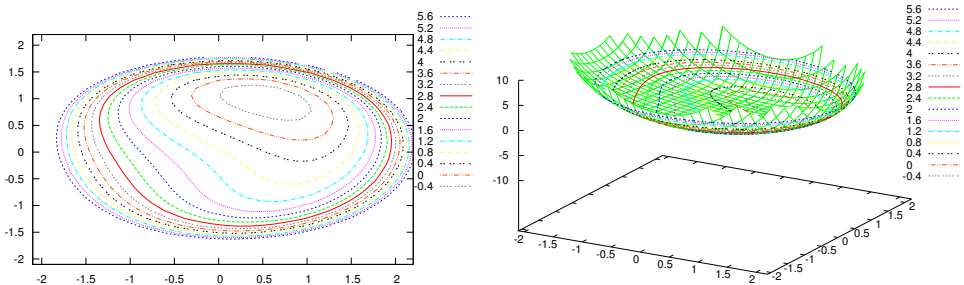
Figure 1. Contours and graph of the primal function Π of Section 4.1.
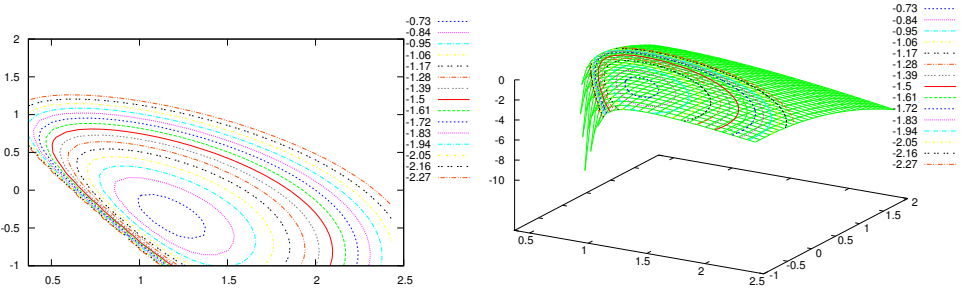
Figure 2. Contours and graph of the dual function Π^d of Section 4.1.

By the triality theory, the vector

$$\bar{x} = G(\bar{\varsigma})^{-1} f = (0.54792514555217, 1.003890602479819)$$

is the only global minimizer of the primal problem.

4.2. *One stationary point in $\mathscr{S}_a^+$ and one in $\mathscr{S}_a^-$.* Let $\alpha_1 = 1$, $\theta_1 = 50$, and

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -16 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \text{and} \quad f = \begin{bmatrix} -25 \\ 9 \end{bmatrix}.$$

The primal function $\Pi : \mathbb{R}^2 \to \mathbb{R}$ is then given by

$$\Pi(x, y) = \exp\left(\tfrac{1}{2}(x^2 + y^2) - 1\right) + \tfrac{1}{2}\left(\tfrac{1}{2}(x^2 + 2y^2) - 50\right)^2 + \tfrac{1}{2}(x^2 - 16y^2) + 25x - 9y,$$

and its canonical dual is

$$\Pi^d(\tau, \sigma) = -\frac{1}{2}\left(\frac{81}{-16 + \tau + 2\sigma} + \frac{625}{1 + \tau + \sigma}\right) - \tau \cdot \ln(\tau) - \tfrac{1}{2}\sigma^2 - 50\sigma,$$

which has two critical points:

$$\bar{\varsigma}_1 = (96.61711963278241, -38.94928057661689) \in \mathscr{S}_a^+,$$
$$\bar{\varsigma}_2 = (\ 0.42157060067968, -49.86072154366873) \in \mathscr{S}_a^-.$$
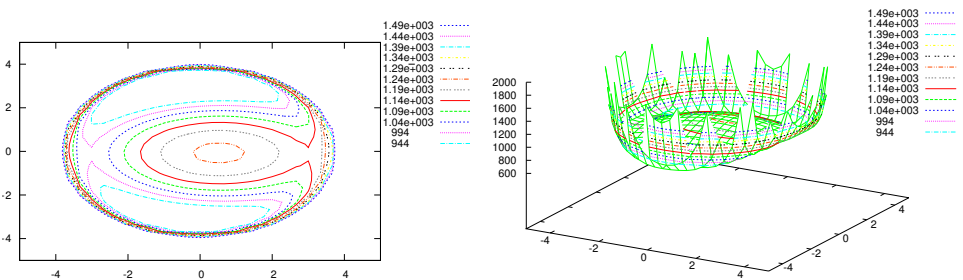
Figure 3. Contours and graph of the primal function Π of Section 4.2.
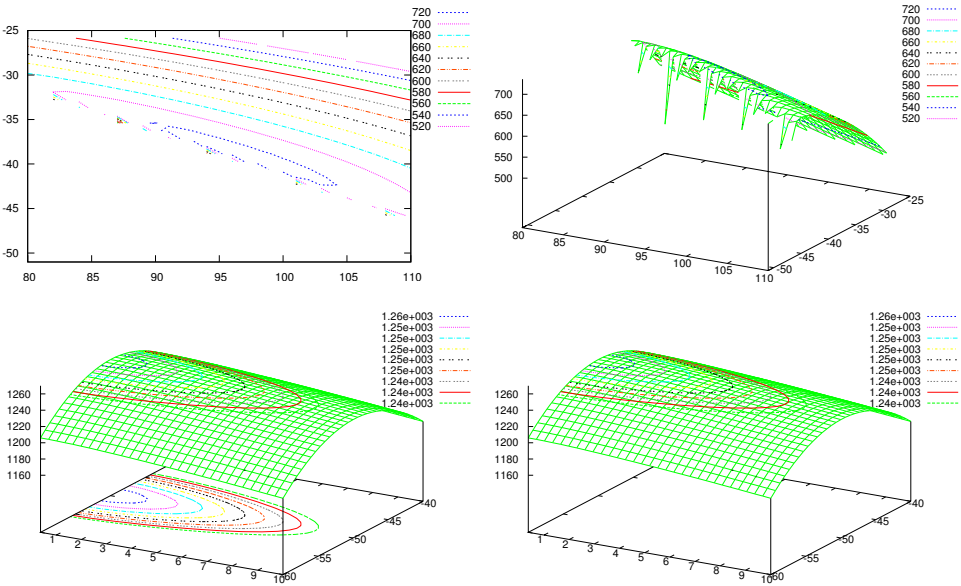
Figure 4. Contours and graph of the dual function Π^d of Section 4.2 in $\mathcal{S}_a^+$ (top) and in $\mathcal{S}_a^-$ (bottom).

Therefore, by the triality theory, the associated vector

$$\overline{x_1} = G(\overline{\varsigma_1})^{-1} f = (-0.42612784793499, 3.310578038951848)$$

is the only global minimizer of $\Pi(x)$, and

$$\overline{x_2} = (0.51611144112381, -0.078057328303129)$$

is a local maximizer (see Figure 3) since $\overline{\varsigma_2}$ is a local maximum of Π^d in $\mathcal{S}_a^-$ (see Figure 4, bottom).

4.3. _One stationary point in $\mathcal{S}_a^+$ and two in $\mathcal{S}_a^-$._ In order to illustrate the triality theory, we let $\alpha_1 = \theta_1 = 2$ and

$$A = \begin{bmatrix} -16 & 0 \\ 0 & -4 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad f = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

Accordingly, we have

$$\Pi(x, y) = \exp\left(\tfrac{1}{2}x^2 - 2\right) + \tfrac{1}{2}\left(\tfrac{1}{2}y^2 - 2\right)^2 + \tfrac{1}{2}(-16x^2 - 4y^2) - 2x - 2y,$$

$$\Pi^d(\tau, \sigma) = -\frac{1}{2}\left(\frac{4}{\sigma - 4} + \frac{4}{\tau - 16}\right) - \tau \cdot \ln(\tau) - \tau - \tfrac{1}{2}\sigma^2 - 2\sigma.$$

In this case, Π^d has in total six critical points but only one in $\mathcal{S}_a^+$,

$$\overline{\varsigma_1} = (16.64468576727409, 4.552474610531074) \in \mathcal{S}_a^+$$
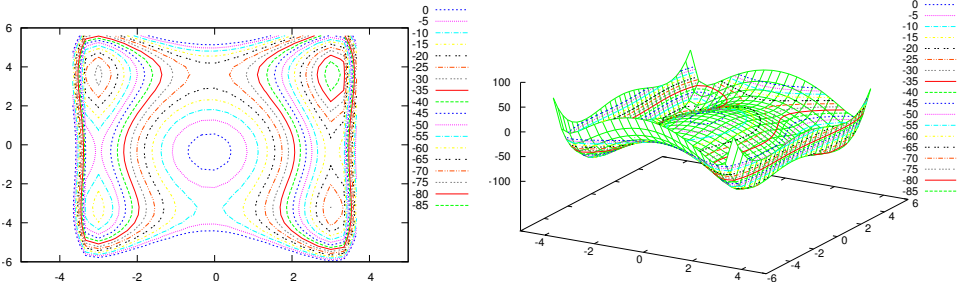
Figure 5. Contours and graph of the primal function Π of Section 4.3.

(see Figure 6, top), and two in $\mathscr{S}_a^-$:

$$\overline{\varsigma}_2 = (0.13641513779858, -1.943380912562619) \in \mathscr{S}_a^-,$$
$$\overline{\varsigma}_3 = (15.34981976568548, 3.390906302031545) \in \mathscr{S}_a^-.$$

From Figure 6, bottom, we can see that $\overline{\varsigma}_2$ is a local maximizer and $\overline{\varsigma}_3$ is a local minimizer of Π^d. Therefore, by the triality theory, we know that

$$\overline{x}_1 = G(\overline{\varsigma}_1)^{-1} f = (3.102286573591542, 3.620075858467906)$$
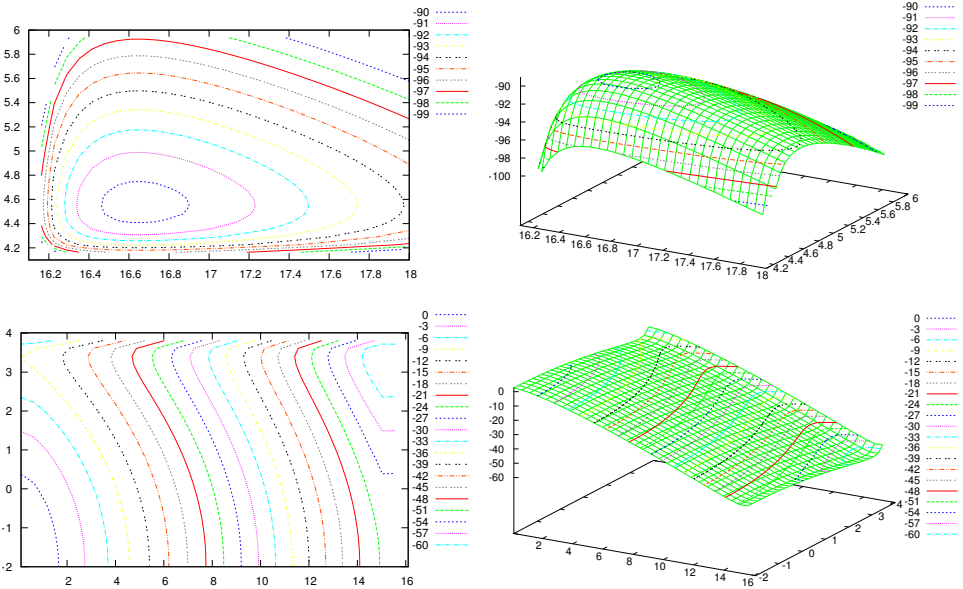
Figure 6. Contours and graph of the dual function Π^d of Section 4.3 in $\mathscr{S}_a^+$ (top) and in $\mathscr{S}_a^-$ (bottom).

is the only global minimizer,

$$\bar{x}_2 = (-0.12607490787063, -0.33650880356205)$$

is a local maximizer, and

$$\bar{x}_3 = (-3.076070133243102, -3.283567054905852)$$

is a local minimizer of $\Pi(x)$ (see Figure 5).

4.4. *Nonunique global minima.*
In the case that no stationary point can be found in $\mathcal{S}_a^+$, the primal problem could have more than one global minima. To see this, we let $f \equiv 0$, $\alpha_1 = \theta_1 = 2$, and

$$A \equiv 0, \quad B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad C_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

In this case, the primal function

$$\Pi(x, y) = \exp\left(\tfrac{1}{2}x^2 - 2\right) + \tfrac{1}{2}\left(\tfrac{1}{2}y^2 - 2\right)^2$$

has two global minimums at $(0, -2)$ and $(0, 2)$ and a local maximum at $(0, 0)$. While the dual function

$$\Pi^d(\tau, \sigma) = -\tau \ln \tau - \tau - \tfrac{1}{2}\sigma^2 - 2\sigma$$

does not have a stationary point in $\mathcal{S}_a^+$. There is however a critical point in the boundary of $\mathcal{S}_a^+$, namely $\bar{\varsigma} = (\exp(-2), 0)$. By defining $\bar{x} = G(\bar{\varsigma})^{-1}f$, we have that $\bar{x} = (0, 0)$.

In order to find a global minimum of Π, we need to introduce the perturbations

$$A_n = \begin{bmatrix} -\tfrac{16}{n} & 0 \\ 0 & -\tfrac{4}{n} \end{bmatrix} \quad \text{and} \quad f_n = \begin{bmatrix} \tfrac{2}{n} \\ \tfrac{2}{n} \end{bmatrix} \quad \text{for every } n \in \mathbb{N}.$$

Then the associated primal and dual functions are

$$\Pi_n(x, y) = \exp\left(\tfrac{1}{2}x^2 - 2\right) + \tfrac{1}{2}\left(\tfrac{1}{2}y^2 - 2\right)^2 + \tfrac{1}{2}\left(-\tfrac{16}{n}x^2 - \tfrac{4}{n}y^2\right) - \tfrac{2}{n}x - \tfrac{2}{n}y,$$

$$\Pi_n^d(\tau, \sigma) = -\frac{1}{2}\left(\frac{4}{n^2\left(\tau - \tfrac{16}{n}\right)} + \frac{4}{n^2\left(\sigma - \tfrac{4}{n}\right)}\right) - \tau \ln \tau + \tau - \tfrac{1}{2}\sigma^2 - 2\tau - 2\sigma.$$

Notice that if $n = 1$ we are in the case presented in Section 4.3. Let us show that for sufficiently large values of n we can find a stationary point for Π_n^d in $\mathcal{S}_a^+$, namely $\bar{\varsigma}_n$. Furthermore, by defining $\bar{x}_n = G(\bar{\varsigma}_n)^{-1} f_n$, we will have a convergent sequence.

Let us calculate the gradient of Π_n^d:

$$\nabla \Pi_n^d(\tau, \sigma) = \begin{bmatrix} -2 - \ln \tau + \frac{2}{(n\tau - 16)^2} \\ -\sigma - 2 + \frac{2}{(n\sigma - 4)^2} \end{bmatrix}.$$

Let $h(\tau) = -2 - \ln \tau + 2/(n\tau - 16)^2$ and $g(\sigma) = -\sigma - 2 + 2/(n\sigma - 4)^2$. It is not difficult to show that there exists a sufficiently large $N \in \mathbb{N}$ such that, if $n > N$,

- $n \cdot \exp\left(-2 + \frac{1}{n}\right) - 16$ and $n \cdot \exp(-2) - 16$ are positive numbers,

- $h\left(\exp\left(-2 + \frac{1}{n}\right)\right) = \dfrac{2}{\left(n \cdot \exp\left(-2 + \frac{1}{n}\right) - 16\right)^2} - \frac{1}{n} < 0 < h(\exp(-2))$

$$= \frac{2}{(n \cdot \exp(-2) - 16)^2},$$

- $g\left(\frac{5.1}{n}\right) \approx -\frac{5.1}{n} - 0.34710743801 < 0 < g\left(\frac{4.9}{n}\right) \approx 0.46913580247 - \frac{4.9}{n}$.

Based on these results, we know that, for every $n > N$, $\nabla \Pi_n^d$ has a stationary point $\bar{\varsigma}_n = (\bar{\tau}_n, \bar{\sigma}_n) \in \left[\exp(-2), \exp\left(-2 + \frac{1}{n}\right)\right] \times \left[\frac{4.9}{n}, \frac{5.1}{n}\right]$. Moreover, since $g(\bar{\sigma}_n) = 0$, it is easy to obtain $\lim_{n \to +\infty} n \cdot \bar{\sigma}_n = 5$.

Notice also that

$$G(\bar{\varsigma}_n) = \begin{bmatrix} \bar{\tau}_n - \frac{16}{n} & 0 \\ 0 & \bar{\sigma}_n - \frac{4}{n} \end{bmatrix}$$

is positive definite. Therefore, the perturbed solution can be obtained as

$$\bar{x}_n = G(\bar{\varsigma}_n)^{-1} f_n = \begin{bmatrix} 2/(n \cdot \bar{\tau}_n - 16) \\ 2/(n \cdot \bar{\sigma}_n - 4) \end{bmatrix}.$$

Since $\bar{\tau}_n \in \left[\exp(-2), \exp\left(-2 + \frac{1}{n}\right)\right]$, we have $\lim_{n \to +\infty} \tau_n = \exp(-2)$. From the fact that $\lim_{n \to +\infty} n \cdot \bar{\sigma}_n = 5$, we get

$$\lim_{n \to +\infty} \bar{x}_n = \begin{bmatrix} 0 \\ 2 \end{bmatrix},$$

which is a solution of Π.

Canonical perturbation method was originally introduced in [Ruan et al. 2010] for solving nonconvex polynomial minimization problems. This method has been used successfully in integer programming and network communication (see [Gao et al. 2012; Wang et al. 2012]).

5. Future research

Some open questions that will be studied in the future are the following:

- As stated in Remark 1, in order to use the canonical dual transformation, a necessary condition is that $(\mathcal{P})$ has a unique solution. Is this also a sufficient condition? In other words, given $(\mathcal{P})$ such that it has a unique solution, can we find a stationary point of Π^d in $\mathcal{S}_a^+$?

- Section 4.4 shows an interesting perturbation method that allows us to solve a problem when the necessary condition of Remark 1 is not satisfied. Can we generalize this method and develop an algorithm?

Appendix A: Some lemmas in matrix analysis

The following results are needed in the proofs of Section 2:

Lemma A1 (singular-value decomposition [Horn and Johnson 1985]). *For any given matrix $M \subset \mathbb{R}^{m \times n}$ with $\mathrm{Rank}(M) = r$, there exist $U \subset \mathbb{R}^{m \times m}$, $R \subset \mathbb{R}^{m \times n}$, and $E \subset \mathbb{R}^{n \times n}$ such that*

$$M = URE,$$

where U and E are orthogonal matrices, and

$$R_{ij} = \begin{cases} s_i & \text{if } i = j \text{ and } i = 1, \ldots, r, \\ 0 & \text{if } i \neq j, \end{cases}$$

where $s_i > 0$ for every $i = 1, \ldots, r$.

Lemma A2 [Horn and Johnson 1985]. *If G and U are positive-definite matrices in $\mathbb{R}^{n \times n}$, then $G \succeq U$ if and only if $U^{-1} \succeq G^{-1}$.*

Lemma A3 [Gao and Wu 2012]. *Suppose P, U, and D are three matrices in $\mathbb{R}^{n \times n}$ such that*

$$D = \begin{bmatrix} D_{11} & 0_{m \times (n-m)} \\ 0_{(n-m) \times n} & 0_{(n-m) \times (n-m)} \end{bmatrix},$$

where $D_{11} \in \mathbb{R}^{m \times m}$ is nonsingular and

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \prec 0 \quad and \quad U = \begin{bmatrix} U_{11} & 0_{m \times (n-m)} \\ 0_{(n-m) \times m} & U_{22} \end{bmatrix} \succ 0,$$

where P_{ij} and U_{ii} are appropriate-dimensional matrices for $i, j = 1, 2$. Then

$$P + DUD^t \preceq 0 \iff -D^t P^{-1} D - U^{-1} \preceq 0.$$

Acknowledgements

References

[Aspnes et al. 2004] J. Aspnes, D. Goldberg, and Y. R. Yang, "On the computational complexity of sensor network localization", pp. 32–44 in *Algorithmic aspects of wireless sensor networks* (Turku, 2004), edited by S. Nikoletseas and J. D. P. Rolim, Lecture Notes in Computer Science **3121**, Springer, Berlin, 2004.

[Cai et al. 2014] K. Cai, D. Y. Gao, and Q. H. Qin, "Post-buckling solutions of hyper-elastic beam by canonical dual finite element method", *Math. Mech. Solids* **19**:6 (2014), 659–671. arXiv 1302.4136

[Desoer and Whalen 1963] C. A. Desoer and B. H. Whalen, "A note on pseudoinverses", *J. Soc. Indust. Appl. Math.* **11** (1963), 442–447.

[Feng et al. 2012] J.-M. Feng, G.-X. Lin, R.-L. Sheu, and Y. Xia, "Duality and solutions for quadratic programming over single non-homogeneous quadratic constraint", *J. Glob. Optim.* **54**:2 (2012), 275–293.

[Gao 1996] D. Y. Gao, "Complementary finite-element method for finite deformation nonsmooth mechanics", *J. Eng. Math.* **30**:3 (1996), 339–353.

[Gao 1997] D. Y. Gao, "Dual extremum principles in finite deformation theory with applications to post-buckling analysis of extended nonlinear beam theory", *Appl. Mech. Rev.* (*ASME*) **50**:11S (1997), S64–S71.

[Gao 1998] D. Y. Gao, "Duality, triality and complementary extremum principles in non-convex parametric variational problems with applications", *IMA J. Appl. Math.* **61**:3 (1998), 199–235.

[Gao 1999] D. Y. Gao, "General analytic solutions and complementary variational principles for large deformation nonsmooth mechanics", *Meccanica* (*Milano*) **34**:3 (1999), 169–198.

[Gao 2000a] D. Y. Gao, "Canonical dual transformation method and generalized triality theory in nonsmooth global optimizaton", *J. Glob. Optim.* **17**:1-4 (2000), 127–160.

[Gao 2000b] D. Y. Gao, *Duality principles in nonconvex systems: theory, methods and applications*, Nonconvex Optimization and its Applications **39**, Kluwer, Dordrecht, 2000.

[Gao 2003a] D. Y. Gao, "Nonconvex semi-linear problems and canonical duality solutions", pp. 261–312 in *Advances in mechanics and mathematics*, vol. 2, edited by D. Y. Gao and R. W. Ogden, Adv. Mech. Math. **4**, Kluwer, Boston, 2003.

[Gao 2003b] D. Y. Gao, "Perfect duality theory and complete solutions to a class of global optimization problems", *Optimization* **52**:4-5 (2003), 467–493.

[Gao 2009] D. Y. Gao, "Canonical duality theory: theory, method, and applications in global optimization", *Comput. Chem. Eng.* **33**:12 (2009), 1964–1972.

[Gao and Ogden 2008a] D. Y. Gao and R. W. Ogden, "Multiple solutions to non-convex variational problems with implications for phase transitions and numerical computation", *Quart. J. Mech. Appl. Math.* **61**:4 (2008), 496–522.

[Gao and Ogden 2008b] D. Y. Gao and R. W. Ogden, "Closed-form solutions, extremality and non-smoothness criteria in a large deformation elasticity problem", *Z. Angew. Math. Phys.* **59**:3 (2008), 498–517.

[Gao and Ruan 2008] D. Y. Gao and N. Ruan, "Solutions and optimality criteria for nonconvex quadratic-exponential minimization problem", *Math. Methods Oper. Res.* **67**:3 (2008), 479–491.

[Gao and Sherali 2009] D. Y. Gao and H. D. Sherali, "Canonical duality theory: connections between nonconvex mechanics and global optimization", pp. 257–326 in *Advances in applied mathematics and global optimization*, edited by D. Y. Gao and H. D. Sherali, Adv. Mech. Math. **17**, Springer, New York, 2009.

[Gao and Strang 1989] D. Y. Gao and G. Strang, "Geometric nonlinearity: potential energy, complementary energy, and the gap function", *Quart. Appl. Math.* **47**:3 (1989), 487–504.

[Gao and Wu 2012] D. Y. Gao and C. Wu, "On the triality theory for a quartic polynomial optimization problem", *J. Ind. Manag. Optim.* **8**:1 (2012), 229–242.

[Gao and Yu 2008] D. Y. Gao and H. Yu, "Multi-scale modelling and canonical dual finite element method in phase transitions of solids", *Int. J. Solids Struct.* **45**:13 (2008), 3660–3673.

[Gao et al. 2012] D. Y. Gao, N. Ruan, and P. M. Pardalos, "Canonical dual solutions to sum of fourth-order polynomials minimization problems with applications to sensor network localization",

pp. 37–54 in *Sensors: theory, algorithms, and applications*, edited by V. L. Boginski et al., Springer Optim. Appl. **61**, Springer, New York, 2012.

[Horn and Johnson 1985] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, 1985.

[Marsden and Hughes 1983] J. E. Marsden and T. J. R. Hughes, *Mathematical foundations of elasticity*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[MAXIMA 2010] MAXIMA, "Maxima, a computer algebra system", 2010, Available at http:// maxima.sourceforge.net. Version 5.22.1.

[Moré and Wu 1997] J. J. Moré and Z. Wu, "Global continuation for distance geometry problems", *SIAM J. Optim.* **7**:3 (1997), 814–836.

[Moreau 1968] J.-J. Moreau, "La notion de sur-potentiel et les liaisons unilatérales en élastostatique", *C. R. Acad. Sci. Paris Sér. A-B* **267A** (1968), A954–A957.

[Moreau et al. 1988] J.-J. Moreau, P. D. Panagiotopoulos, and G. Strang (editors), *Topics in nonsmooth mechanics*, Birkhäuser, Basel, 1988.

[Peters and Wilkinson 1970] G. Peters and J. H. Wilkinson, "The least squares problem and pseudo-inverses", *Comput. J.* **13**:3 (1970), 309–316.

[Ruan and Gao 2014a] N. Ruan and D. Y. Gao, "Canonical duality approach for non-linear dynamical systems", *IMA J. Appl. Math.* **79**:2 (2014), 313–325.

[Ruan and Gao 2014b] N. Ruan and D. Y. Gao, "Global optimal solutions to a general sensor network localization problem", *Perform. Eval.* **75–76** (2014), 1–16.

[Ruan et al. 2010] N. Ruan, D. Y. Gao, and Y. Jiao, "Canonical dual least square method for solving general nonlinear systems of quadratic equations", *Comput. Optim. Appl.* **47**:2 (2010), 335–347.

[Santos and Gao 2012] H. A. F. A. Santos and D. Y. Gao, "Canonical dual finite element method for solving post-buckling problems of a large deformation elastic beam", *Int. J. Nonlinear Mech.* **47**:2 (2012), 240–247.

[Saxe 1979] J. B. Saxe, "Embeddability of weighted graphs in k-space is strongly NP-hard", pp. 480–489 in *Proceedings of the 17th Allerton Conference on Communications, Control, and Computing* (Monticello, IL, 1979), University of Illinois at Urbana-Champaign, Urbana, IL, 1979. Also in *Two papers on graph embedding problems*, preprint, Carnegie Mellon University, Pittsburgh, PA, 1980.

[Strugariu et al. 2011] R. Strugariu, M. D. Voisei, and C. Zălinescu, "Counter-examples in bi-duality, triality and tri-duality", *Discrete Contin. Dyn. Syst.* **31**:4 (2011), 1453–1468.

[Voisei and Zălinescu 2011] M. D. Voisei and C. Zălinescu, "Some remarks concerning Gao–Strang's complementary gap function", *Appl. Anal.* **90**:6 (2011), 1111–1121.

[Wang et al. 2012] Z. Wang, S.-C. Fang, D. Y. Gao, and W. Xing, "Canonical dual approach to solving the maximum cut problem", *J. Glob. Optim.* **54**:2 (2012), 341–351.

[Xie and Schlick 2000] D. Xie and T. Schlick, "Visualization of chemical databases using the singular value decomposition and truncated-Newton minimization", pp. 267–286 in *Optimization in computational chemistry and molecular biology: local and global approaches*, edited by C. A. Floudas and P. M. Pardalos, Optimization in Computational Chemistry and Molecular Biology **40**, Springer, Dordrecht, 2000.

[Zhang et al. 2011] J. Zhang, D. Y. Gao, and J. Yearwood, "A novel canonical dual computational approach for prion AGAAAAGA amyloid fibril molecular modeling", *J. Theoret. Biol.* **284** (2011), 149–157.

DANIEL MORALES-SILVA: d.moralessilva@federation.edu.au
School of Science, Information Technology and Engineering, Federation University,
Mt. Helen VIC 3353, Australia

DAVID Y. GAO: d.gao@federation.edu.au
david.gao@anu.edu.au
School of Science, Information Technology and Engineering, Federation University,
Mt. Helen, VIC 3353, Australia

and

Research School of Engineering, Australian National University, Canberra, ACT 0200, Australia

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS

msp.org/memocs

Cover image: "Tangle" by © John Horigan; produced using the *Context Free* program (contextfreeart.org).