Unified Interior Point Methodology for Canonical Duality in Global Optimization

Vittorio Latorre

Abstract We propose an interior point method to solve instances of the nonconvex optimization problems reformulated with canonical duality theory. To this aim we propose an interior point potential reduction algorithm based on the solution of the primal–dual total complementarity function. We establish the global convergence result for the algorithm under mild assumptions. Our methodology is quite general and can be applied to several problems which dual has been formulated with canonical duality theory and shows the possibility of devising efficient interior points methods for nonconvex duality.

1 Introduction

We want to introduce a framework to solve the following saddle point problem:

$$\min_{x\in\mathbb{R}^n}\max_{\sigma\in\mathbb{R}^m} \Xi(x,\sigma) = \frac{1}{2}x^T G(\sigma)x - F(\sigma)^T x - V^*(\sigma), \quad s.t. \quad G(\sigma)\succeq 0, \tag{1}$$

where $\succeq$ indicates that G is positive semidefinite, $G(\sigma)$ is a $n\times n$ symmetric matrix such that the map $G(\sigma):\mathbb{R}^m\to\mathbb{R}^{n\times n}$ is positive semidefinite convex, that is,

$$G(t\sigma_1+(1-t)\sigma_2)\succeq tG(\sigma_1)+(1-t)G(\sigma_2), \quad \forall\sigma_1,\sigma_2\in\mathbb{R}^m, \forall t\in(0,1).$$

$V^*(\sigma)$ is a convex and two times continuously differentiable function in σ. it is easy to notice that Problem (1) is convex in x for every σ such that $G(\sigma)\succeq 0$ and it is concave for every σ.

Such problem arises from the reformulation of nonconvex optimization problems in Canonical Duality Theory. Canonical duality is a methodology to formulate the dual of nonconvex optimization problems without any duality gap between the

V. Latorre (✉)
Department of Computer, Control and Management Engineering,
“Sapienza”, University of Rome, Via Ariosto 25, 00185 Rome, Italy
e-mail: latorre@dis.uniroma1.it

D.Y. Gao et al. (eds.), *Canonical Duality Theory*, Advances in Mechanics
and Mathematics 37, DOI 10.1007/978-3-319-58017-3_12

stationary points of the primal problem and the stationary points of the dual problem. The interest in canonical duality is not only due to the absence of duality gap, but also for the possibility to define global optimality conditions for many of such nonconvex optimization problems. In the recent years, canonical duality theory has been applied in biology, engineering, sciences [6, 16], and recently in network communications [7, 15], radial basis neural networks [10] and constrained optimization [9].

In spite of its theoretical prowess and range of applications, there are few results regarding the numerical solution of problems formulated with canonical duality theory. In [16] several mid-sized instances of the maximum cut problem are solved, to a maximum of 500 variables, with good performances in terms of speed; however, no convergence result is given. A convergence result is given in [17]; however, the assumptions on the convergence are rather strong. In a more recent work on the application of canonical duality theory to Quasi-Variational Inequalities [11], the authors reformulate problem (1) as a monotone Variational Inequality (VI) and are able to solve high-dimensional problems with several thousand of variables, without giving any convergence result, but suggesting that the methodology could have some interesting proprieties.

In this paper we partially resume the approach presented in [11]. We consider the Karush–Kunt–Tucker conditions of the monotone variational inequality associated with (1), reformulate the problem as a system of constrained equations and then prove the convergence of a potential reduction interior point method to the desired solution under mild assumptions.

The approach we consider is a potential reduction algorithm based on the damped Newton method reported in [3, 13]. The framework of this algorithm rests on six main assumptions on the operator, the feasible set, and the potential reduction merit function. The convergence result easily follows once it is proved that the proposed methodology satisfies these assumptions. The same framework has been applied to Generalized Nash Equilibrium Problems [1] and more recently to Quasi-Variational Inequalities [2], providing in both cases new important benchmarks to solve these problems.

The paper is organized as follows. In the next section we briefly show how problem (1) is obtained from general nonconvex optimization problem. In Sect. 3 we reformulate problem (1) as a system of equations, while in Sect. 4 we briefly report the key assumptions of the framework introduced in [13] and present the interior point method together with its convergence proprieties and the boundedness of the generated sequence. In Sect. 5 we report the conclusions.

Notation. For a given subset of S of $\mathbb{R}^n$ we let int S, cl S, and bd S denote, respectively, the interior, the closure, and the boundary of S; Given a set $\mathscr{A}$ we indicate with $|\mathscr{A}|$ the number of elements in $\mathscr{A}$. If the mapping $H : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable in a point x in its domain, the Jacobian matrix of H at x is denoted $JH(x)$.

The set of real matrices with n rows and m columns is defined as $\mathbb{R}^{n\times m}$; the set of $n - dimensional$ squared and symmetric matrices is denoted as $\mathscr{S}^n$; given a matrix A, we denote with a_{ij} its element on the i^{th} row and j^{th} column. The inner product defined on the set $\mathbb{R}^{n\times n}$ of squared matrices is given by

$$X \bullet Y = tr(X^T Y), \quad (X; Y) \in \mathbb{R}^{n\times n},$$

where “tr” denotes the trace of a matrix. This inner product induces the Frobenius norm for matrices given by

$$\|X\|_F = \sqrt{tr(X^T X)}, \quad X \in \mathbb{R}^{n\times n}.$$

Given a mapping $F(x, Y) : \mathbb{R}^n \times \mathscr{S}^n \to \mathbb{R}^n \times \mathscr{S}^n$ defined as

$$F(x, Y) = \begin{pmatrix} g(x, Y) \\ h(x, Y) \end{pmatrix},$$

with $g(x, Y) : \mathbb{R}^n \times \mathscr{S}^n \to \mathbb{R}^n$ and $h(x, Y) : \mathbb{R}^n \times \mathscr{S}^n \to \mathscr{S}^n$, a vector $\bar{x} \in \mathbb{R}^n$ and a matrix $\bar{Y} \in \mathscr{S}^n$, with a small abuse of notation we define the product between the mapping and the elements of $\mathbb{R}^n \times \mathscr{S}^n$ as:

$$F(x, Y) \bullet (\bar{x}, \bar{Y}) = g(x, Y)^T \bar{x} + h(x, Y) \bullet \bar{Y}.$$

The subsets of $\mathscr{S}^n$ consisting of the positive semidefinite and positive definite matrices are denoted by $\mathscr{S}^n_+$ and $\mathscr{S}^n_{++}$, respectively. For two matrices A and B in $\mathscr{S}^n$, we write $A \succeq B$ if $A - B \in \mathscr{S}^n_+$; similarly, $A \succ B$ means $A - B \in \mathscr{S}^n_{++}$; furthermore we define $\preceq$ and $\prec$ such that $A \preceq B$ if $-A \succeq -B$ and $A \prec B$ if $-A \succ -B$. $\mathbb{R}^n_+ \subset \mathbb{R}^n$ denotes the set of nonnegative numbers in $\mathbb{R}^n$; $\mathbb{R}^n_{++} \subset \mathbb{R}^n$ denotes the set of positive numbers in $\mathbb{R}^n$; sta$\{f(x) : x \in \mathscr{X}\}$ denotes the set of stationary points of function f in $\mathscr{X}$; diag(a) denotes the (square) diagonal matrix whose diagonal entries are the elements of the vector a; $vect\{A\}$ denotes the vector $\in \mathbb{R}^{n^2}$ such that the first n elements are the elements in the first column of A, the elements from $n + 1$ to $2n$ are the elements in the second column of A and so on till the last n elements that correspond to the elements in the n^{th} column of A; $\circ$ denotes the Hadamard (component-wise) product operator; and $\mathbf{0}_n$ denotes the origin in $\mathbb{R}^n$, likewise $\mathbf{0}_{n\times m}$ denotes the origin in $\mathbb{R}^{n\times m}$. If no index is indicated, the dimension of $\mathbf{0}$ is deduced from the context; $\mathbf{1}_n$ denotes the vectors of all ones in $\mathbb{R}^n$; I_n denotes the identity matrix in $\mathbb{R}^{n\times n}$.

2 Problem Description

Canonical duality theory is applied to the following general nonconvex problem:

$$(\mathscr{P}) : \min_{x\in\mathbb{R}^n} \left\{ \Pi(x) = W(x) + \frac{1}{2} x^T A x - c^T x \right\},$$

where $W(x)$ is a nonconvex term in the objective function, $A \in \mathscr{S}^n$ and $c \in \mathbb{R}^n$. The canonical dual transformation can be applied if the following assumption is satisfied:

Assumption 1 *There exists a nonlinear operator*

$$\xi = \Lambda(x) : \mathbb{R}^n \to \mathscr{E}_a \subseteq \mathbb{R}^m$$

function of x, such that the nonconvex functional $W(x)$ can be rewritten as

$$W(x) = V(\Lambda(x)) = V(\xi) : \mathscr{E}_a \to \mathbb{R}, \tag{2}$$

where V is a convex and differentiable function in ξ.

If Assumption 1 is satisfied, the primal problem can be rewritten in the following form:

$$\min_{x \in \mathbb{R}^n} \left\{ \Pi(x) = V(\Lambda(x)) + \frac{1}{2} x^T A x - c^T x \right\}.$$

As $V(\xi)$ is convex and differentiable, it is possible to apply the Legendre transformation, and write the total complementarily function in the primal variable x and dual variable $\sigma \in \mathscr{S}_a \subseteq \mathbb{R}^m$:

$$\Xi(x, \sigma) = \Lambda(x)^T \sigma - V^*(\sigma) + \frac{1}{2} x^T A x - c^T x,$$

where $V^*(\sigma)$ is the Fenchel conjugate of $V(\xi)$.

In many real-world applications, the geometrically nonlinear operator $\Lambda(x)$ is usually a quadratic function, say

$$\Lambda(x) = \left\{ \frac{1}{2} x^T C_k x - x^T b_k \right\}^m : \mathbb{R}^n \to \mathscr{E}_a \subset \mathbb{R}^m. \tag{3}$$

In the following we focus on the transformation for a general quadratic operator. With operator (3) the total complementarity function can be reformulated as

$$\begin{aligned} \Xi(x, \sigma) &= \frac{1}{2} x^T G(\sigma) x - F(\sigma)^T x - V^*(\sigma), \\ G(\sigma) &= A + \sum_{k=1}^m C_k \sigma_k, \quad F(\sigma) = c + \sum_{k=1}^m \sigma_k b_k. \end{aligned} \tag{4}$$

The dual is obtained by exploiting the stationarity conditions of (4) in the primal variable:

$$\nabla_x \Xi(x, \sigma) = \mathbf{0}_n \Rightarrow x = G(\sigma)^{-1} F(\sigma),$$

and substituting the newfound value in the total complementarity function:

$$\Pi^d(\sigma) = -\frac{1}{2}F(\sigma)^T G(\sigma)^{-1} F(\sigma) - V^*(\sigma). \tag{5}$$

Note that the feasible set $\mathscr{S}_a$ is not convex; then, in order to identify the global optimality conditions, we need to introduce the following subset of $\mathscr{S}_a$:

$$\mathscr{S}_a^+ = \{\sigma \in \mathscr{S}_a | \ G(\sigma) \succeq 0\}.$$

Theorem 1. (**Global Optimality** *[5]*) *Given a critical point* $(\bar{x}, \bar{\sigma})$ *of* $\Xi(x, \sigma)$, $\bar{x}$ *is the unique global minimizer of* $\Pi(x)$ *if* $\bar{\sigma} \in \mathscr{S}_a^+$ *is the global maximizer of* $\Pi^d(\sigma)$ *on* $\mathscr{S}_a^+$, *and there is no duality gap between the primal, dual, and total complementarity functions, i.e.,*

$$\min_{x \in \mathbb{R}^n} \Pi(x) = P(\bar{x}) = \Xi(\bar{x}, \bar{\sigma}) = \Pi^d(\bar{\sigma}) = \max_{\sigma \in \mathscr{S}_a^+} \Pi^d(\sigma). \tag{6}$$

The result reported in equation (6) clearly shows the global optimality conditions. The original nonconvex primal problem is reduced to the maximization of the dual function $\Pi^d(\sigma)$ on the convex set $\mathscr{S}_a^+$. Furthermore it easy to notice from the (5) that the dual is concave on $\mathscr{S}_a^+$, and therefore the resulting problem is convex. Finally, we want to underline that there is no duality gap between the solution of the dual and the global minimum in the primal.

3 Reformulation of the Problem as a System of Constrained Equations

By the results of Theorem 1, it is possible to find the global solution of Problem ($\mathscr{P}$) by different approaches. One approach is to directly solve the dual formulation on $\mathscr{S}_a^+$, but this method has several faults:

- It is necessary to calculate the inverse of matrix $G(\sigma)$ every time the objective function is evaluated, and such operation could be necessary several times per iteration;
- The inverse matrix operation can become even more time expensive or generate errors in the case $G(\sigma)$ is ill-conditioned or it is not full rank;
- If the algorithm that solves the dual problem fails to converge to a good enough approximation of a stationary point, it is difficult to retrieve informations on the corresponding point in the primal problem.

For these reasons we propose a method that exploits the information available on both the primal and dual problems and search for a saddle point of the total complementarity function in $\mathscr{S}_a^+$, that is exactly the problem in the form of (1). As a matter of facts, it is easy to notice that finding the maximum of $\Pi^d(\sigma)$ in $\mathscr{S}_a^+$ is equivalent to solve the following canonical saddle point problem:

$$\min_{x\in\mathbb{R}^n}\max_{\sigma\in\mathbb{R}^m} \Xi(x,\sigma) = \frac{1}{2}x^T G(\sigma)x - F(\sigma)^T x - V^*(\sigma) \quad \text{s.t.,} \quad G(\sigma) \succeq 0, \tag{7}$$

that is the same problem presented in the introduction. The solution of (7) can be found by solving a monotone variational inequality on a convex set [3]:

$$\Gamma(x,\sigma) = 0, \quad G(\sigma) \succeq 0, \tag{8}$$

where $\Gamma : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ is defined as

$$\Gamma(x,\sigma) = \begin{pmatrix} \nabla_x \Xi(x,\sigma) \\ -\nabla_\sigma \Xi(x,\sigma) \end{pmatrix}.$$

The operator Γ is monotone because $\Xi(x,\sigma)$ is convex in the primal variables for $\sigma \in \mathscr{S}_a^+$ and it is concave for all $\sigma \in \mathscr{S}_a$ [14], while the set of positive definite matrices is a convex cone. We want to find a solution of (8) by solving the Karush–Kunt–Tucker (KKT) conditions associated with the problem, that is,

$$\begin{gathered} \Gamma_L(x,\sigma,L) = \begin{pmatrix} \nabla_x \Xi(x,\sigma) \\ -\nabla_\sigma \Xi(x,\sigma) - \nabla_\sigma (L \bullet G(\sigma)) \end{pmatrix} = \mathbf{0}_{n+m} \\ L \bullet G(\sigma) = 0, \quad L \succeq 0, \quad G(\sigma) \succeq 0, \end{gathered} \tag{9}$$

where $L \in \mathscr{S}_+^n$ is the matrix of the Lagrangian multipliers. The mapping $\Gamma_L(x,\sigma,L)$ is monotone as a result of Lemma 7 in [12]. Problems can arise when searching for the solution of (8) when there are KKT points located on the boundary of the feasible set. As a matter of facts, a point satisfying conditions (9) with $L \neq 0$ does not correspond to a saddle point of the total complementarity function $\Xi(x,\sigma)$ (in fact they generally correspond to stationary points of the primal problem). In other words we are interested in KKT points which matrix of multipliers L is equal to $\mathbf{0}_{n\times n}$.

To this aim, we reformulate the conditions (9) as a system of Constrained Equations (CE) and propose an interior point method specifically designed to solve this system of Constrained Equations and send the matrix of Lagrange multipliers to zero. We introduce the matrix $W \in \mathscr{S}_+^n$ of slack variables and consider the $CE(H,\Omega)$ system:

$$H(z) = \mathbf{0}, \quad z = (x,\sigma,L,W) \in \Omega, \tag{10}$$

where $H : \Omega \to S$ with $\Omega = \mathbb{R}^{n+m} \times \mathscr{S}_+^n \times \mathscr{S}_+^n$ and $S = \mathbb{R}^{n+m} \times \mathscr{S}_+^n \times \mathscr{S}_+^n \times \mathscr{S}_+^n$, is defined as

$$H(x, \sigma, L, W) = \begin{pmatrix} \Gamma_L(x, \sigma, L) \\ \Phi(\sigma, L, W) \\ L \end{pmatrix} \tag{11}$$

with $\Phi(\sigma, L, W)$ defined as

$$\Phi(\sigma, L, W) = \begin{pmatrix} W - G(\sigma) \\ (LW + WL)/2 \end{pmatrix}.$$

The last set of equations in (11) forces the matrix of Lagrange multipliers to go to zero when the algorithm reaches convergence, assuring that the solution of $CE(\Omega, H)$ is a saddle point of (7).

4 Key Assumptions and Convergence Result

In this section we present the conditions which the operator H and the feasible set Ω must satisfy together with a suitable potential reduction function in order to assure the convergence to a solution of the (10). The framework we use is the same as the one presented in [3] and [13]. This framework is based on six main assumptions that we report here for convenience.

Given the set Ω, operator H, and a potential function $p : \text{int } S \to \mathbb{R}$, the following assumptions must be satisfied by a potential reduction method in order to assure convergence to a solution of the $CE(\Omega, H)$.

(A1) the closed set Ω has a nonempty interior.

(A2) there exists a closed set $S \subseteq \mathbb{R}^{n+m} \times \mathscr{S}^n_+ \times \mathscr{S}^n_+ \times \mathscr{S}^n_+$ such that

1. $\mathbf{0} \in S$;
2. the open set $\Omega_I = H^{-1}(\text{int } S) \cap \text{int } \Omega$ is nonempty;
3. the set $H^{-1}(\text{int } S) \cap \text{bd } \Omega$ is empty.

(A3) H is continuously differentiable on Ω_I, and $JH(x)$ is full rank for all $x \in \Omega_I$

(A4) for every sequence $\{u^k\} \subset \text{int } S$ such that

$$\text{either } \lim_{k\to\infty} \|u^k\| = \infty \text{ or } \lim_{k\to\infty} u^k = \bar{u} \in \text{bd } S \backslash \{0\}$$

we have

$$\lim_{k\to\infty} p(u^k) = \infty.$$

(A5) p is continuously differentiable in its domain and $u \bullet \nabla p(u) > 0$ for all nonzero $u \in \text{int } S$.

(A6) there exists a nonzero vector $o \in S$ and a scalar $\bar{\beta} \in (0, 1]$ such that

$$u \bullet \nabla p(u) \geq \bar{\beta} \frac{(o \bullet u)(o \bullet \nabla p(u))}{\|o\|^2}, \quad \forall u \in \text{int } S.$$

In the following theorems we show that operator H and the feasible set Ω satisfy the aforementioned assumptions with the choice of a suitable potential reduction function.

Theorem 2. *Suppose that $\Xi(x, \lambda)$ is twice differentiable in x and σ, then the set Ω and the operator H in (11) satisfy conditions (A1)–(A3).*

Proof. Condition $(A1)$ is trivially satisfied, also condition $(A2).1$ holds. The point $(\mathbf{0}_{n+m}, I_n, I_n)$ belongs to both Ω_I and int Ω, therefore condition $(A2).2$ holds. From condition

$$(LW + WL)/2,$$

we can define the following set:

$$\mathscr{U} = \{(L, W) \in \mathscr{S}^n_{++} \times \mathscr{S}^n_{++} : LW + WL \in \mathscr{S}^n_{++}\}.$$

It has been proved in lemma 1 of [12] that

$$\mathscr{U} = \{(L, W) \in \mathscr{S}^n_{+} \times \mathscr{S}^n_{+} : LW + WL \in \mathscr{S}^n_{++}\}.$$

This alternative representation implies the $(A2).3$. Finally condition $(A3)$ is satisfied because of the assumption on $\Xi(x, \lambda)$ and the monotonicity of the operator $\Gamma_L(x, \sigma, L)$. □

Theorem 3. *the potential function $p : S \to \mathbb{R}$ defined as*

$$p(a, B, C, D) = \eta \log(\|a\|^2 + \|B\|_F^2 + \|C\|_F^2 + \|D\|_F^2) - \log(\det(B)) - \log(\det(C)) - \log(\det(D)), \tag{12}$$

where $\eta \geq 2n$, satisfies assumptions (A4)-(A6), with $o = (\mathbf{0}_{n+m}, \mathbf{0}_{n\times n}, I_n, \mathbf{0}_{n\times n})$ and $\bar{\beta} < 1/3$

Proof. It can be easily noticed that the value of p goes to ∞ as the sequence $\{a_k, B_k, C_k, D_k\}$ approaches the boundary of the feasible set. Considering that $\|Z\|_F = \sqrt{tr(Z^T Z)}$, then $\|Z\|_F^2$ is the sum of the squares of the n eigenvalues of Z and that $\det(Z)$ is the product of said eigenvalues, we have

$$p(a, B, C, D) = \eta \log\left(\sum_{i=1}^{n+m} \|a\|^2 + \sum_{i=1}^{n} b_i^2 + \sum_{i=1}^{n} c_i^2 + \sum_{i=1}^{n} d_i^2\right) - \sum_{i=1}^{n} \log b_i - \sum_{i=1}^{n} \log c_i - \sum_{i=1}^{n} \log d_i,$$

where $b_i = 1, \ldots, n$, $c_i = 1, \ldots, n$ and $d_i = 1, \ldots, n$ are the eigenvalues of B, C, and D respectively. Also considering that $n \log\left(\sum_{i=1}^{n} u_i\right) \geq \sum_{i=1}^{n} \log u_i + n \log n$ it is possible to write

$$p(a, B, C, D) > \left(\frac{2\eta}{3n} - 1\right)\left(\sum_{i=1}^{n} \log b_i + \sum_{i=1}^{n} \log c_i + \sum_{i=1}^{n} \log d_i\right),$$

therefore assumption (A4) is satisfied for $\eta > \frac{3}{2}n$.
If we define

$$\tau = \|a\|^2 + \|B\|_F^2 + \|C\|_F^2 + \|D\|_F^2,$$

it is possible to write the derivative of the potential function p as

$$\nabla p(a, B, C, D) = \begin{pmatrix} \frac{2\eta}{\tau} a \\ \frac{2\eta}{\tau} B - B^{-1} \\ \frac{2\eta}{\tau} C - C^{-1} \\ \frac{2\eta}{\tau} D - D^{-1} \end{pmatrix},$$

we have

$$(a, B, C, D) \bullet \nabla p(a, B, C, D) = 2\eta - 3n > 0,$$

and thus Assumption (A5) holds. For Assumption (A6), considering that $tr(Z)^2 \leq n\|Z\|_F^2$ and $n^2 \leq tr(Z^{-1})tr(Z)$ (for the arithmetic geometric mean inequality) we have

$$\frac{[\nabla p(a, B, C, D) \bullet (\mathbf{0}_{n+m}, \mathbf{0}_{n\times n}, I_n, \mathbf{0}_{n\times n})][(a, B, C, D) \bullet (\mathbf{0}_{n+m}, \mathbf{0}_{n\times n}, I_n, \mathbf{0}_{n\times n})]}{\|(\mathbf{0}_n, I_n, \mathbf{0}_{n\times n}, \mathbf{0}_{n\times n})\|_F^2} =$$
$$\frac{2\eta}{n}\frac{tr(C)^2}{\tau} - \frac{tr(C^{-1})tr(C)}{n} \leq$$
$$\frac{2\eta}{n}\frac{tr(C)^2}{\|C\|_F^2} - \frac{tr(C^{-1})tr(C)}{n} \leq$$
$$2\eta - n < \frac{1}{\beta}(2\eta - 3n) = \frac{1}{\beta}[(a, B, C, D) \bullet \nabla p(a, B, C, D)].$$

□

We let

$$z = (x, \sigma, L, W), \quad \psi(z) = p(H(z)),$$

and report the following method that follows the same scheme of the interior point method presented in [13]:

Algorithm 1: CPRA: Complementarity Potential Reduction Algorithm

`(S.0)` : Choose $z^0 = (x^0, \sigma^0, L^0, W^0) \in \Omega$, $\gamma \in (0, 1)$, $\bar{\beta} < 1/3$, $\varepsilon > 0$, and set $k := 0$.

`(S.1)` : If $\|\Gamma(x, \sigma)\|^2 < \varepsilon$: STOP

`(S.2)` : Choose a scalar $\beta_k \in (0, \bar{\beta})$ and find a solution $d^k = (dx^k, d\sigma^k, dL^k, dW^k)$ of the following linear least squares problem:

$$\min_d \left\{ \frac{1}{2} \left\| Q(z^k, d) + H(z^k) - \beta_k \frac{o^T H(z^k)}{\|o\|^2} o \right\|^2 \right\}.$$

where

$$Q\ (z^k, d) = \left(\begin{array}{c} \nabla^2_{xx} \Xi(x^k, \sigma^k) dx + \nabla^2_{x\sigma} \Xi(x^k, \sigma^k) d\sigma \\ -\nabla^2_{x\sigma} \Xi(x^k, \sigma^k)^T dx - \nabla^2_{\sigma\sigma} \Xi(x^k, \sigma^k) d\sigma + \nabla_{\sigma L}(L^k \bullet G(\sigma^k)) dL \\ dW - G(d\sigma) \\ (dL) W^k + W^k (dL) + L^k (dW) + (dW) L^k \\ dL \end{array} \right)$$

`(S.3)` : find a step size α_k such that

$$z^k + \alpha_k d^k \in \Omega$$

and

$$\psi(z^k + \alpha_k d^k) \le \psi(z^k) + \gamma \nabla \psi(z^k) \bullet d^k$$

`(S.4)` : Set $z^{k+1} = z^k + \alpha_k d^k$, $k \leftarrow k + 1$, and go to (S.1).

Algorithm 1 is a modified, damped version of the Newton method. At Step `(S.0)` the initial values of the variables and parameters are set. In order to assure the feasibility of z^0, it generally suffices to put a large enough positive value of σ^0, such that $G(\sigma^0) \succ 0$. At Step `(S.1)` there is the stopping criterion that assures the final point is a good enough approximation of a stationary point of $\Xi(x, \sigma)$. At Step `(S.2)` the modified newton direction is calculated. As the linear system is not squared, the least squares solution to the system of equations is returned. One of the main features of the algorithm is the presence of the vector o that bends the direction toward the interior of the feasible set. It is important to underline that the calculated direction at every iteration is unique for Assumption $(A3)$ and always a descent direction of $\psi(\cdot)$ in z_k as shown in the following theorem:

Theorem 4. *Suppose that conditions* $(A5)$ *and* $(A6)$ *hold. Assume also that* $z \in \Omega_I$, $d^k = (dx^k, d\sigma^k, dL^k, dW^k) \in \mathbb{R}^{n+m} \times \mathscr{S}^n_+ \times \mathscr{S}^n_+$ *and* $\beta \in \mathbb{R}$ *are such that*

$$
\begin{aligned}
&H(z) \neq 0, \qquad 0 \le \beta < \bar{\beta}, \\
&d^k = arg\min_d \left\{ \tfrac{1}{2} \left\| Q(z,d) + H(z) - \beta_k \tfrac{o^T H(z)}{\|o\|^2} o \right\|^2 \right\},
\end{aligned} \tag{13}
$$

where $o \in S$ and $\bar{\beta} \in [0,1]$ are as in condition (A6). Then d^k is a descent direction for $\psi(\cdot)$ in z, that is $\nabla\psi(z) \bullet d^k < 0$

Proof. We introduce the following vector in $\mathbb{R}^{n+m+3n^2}$:

$$
\hat{H}(z) = \begin{pmatrix} \Gamma_L(x,\sigma,L) \\ vect\{W - G(\sigma)\} \\ vect\{(LW + WL)/2\} \\ vect\{L\} \end{pmatrix}. \tag{14}
$$

The Jacobian of $\hat{H}(z)$ is the following $(n+m+3n^2) \times (n+m+2n^2)$ matrix:

$$
J\hat{H}(z) = \begin{pmatrix} \nabla^2_{xx}\Xi(x,\sigma) & \nabla^2_{x\sigma}\Xi(x,\sigma) & \mathbf{0}_{n\times n^2} & \mathbf{0}_{n\times n^2} \\ -\nabla^2_{x\sigma}\Xi(x,\sigma) & \nabla^2_{\sigma\sigma}\Xi(x,\sigma) & C^T & \mathbf{0}_{m\times n^2} \\ \mathbf{0}_{n^2\times n} & C & \mathbf{0}_{n^2\times n^2} & I_{n^2} \\ \mathbf{0}_{n^2\times n} & \mathbf{0}_{n^2\times m} & W_{en} & L_{en} \\ \mathbf{0}_{n^2\times n} & \mathbf{0}_{n^2\times m} & I_{n^2} & \mathbf{0}_{n^2\times n^2} \end{pmatrix}. \tag{15}
$$

where

$$
W_{en} = \begin{pmatrix} W + I_n w_{11} & I_n w_{12} & \cdots & I_n w_{1n} \\ I_n w_{21} & W + I_n w_{22} & \cdots & I_n w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ I_n w_{n1} & I_n w_{n2} & \cdots & W + I_n w_{nn} \end{pmatrix}, \tag{16}
$$

$$
L_{en} = \begin{pmatrix} L + I_n l_{11} & I_n l_{12} & \cdots & I_n l_{1n} \\ I_n l_{21} & L + I_n l_{22} & \cdots & I_n l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ I_n l_{n1} & I_n l_{n2} & \cdots & L + I_n l_{nn} \end{pmatrix}, \tag{17}
$$

and $C \in \mathbb{R}^{n^2\times m}$ is $\nabla_{\sigma L}(L \bullet G(\sigma^k))^T$. Let $u \equiv \hat{H}(z)$, if we consider $\hat{d}^k \in \mathbb{R}^{n+m+2n^2}$, solution of the following least squares problem:

$$
\hat{d}^k = arg\min_d \left\{ \frac{1}{2} \left\| (Ju)d + u - \beta_k \frac{\hat{o}^T u}{\|\hat{o}\|^2} \hat{o} \right\|^2 \right\}, \tag{18}
$$

where $\hat{o}$ has been suitably changed from o to match the dimension of $\hat{H}(z)$, it is easy to notice that $\hat{d}^k$ is equivalent to d^k, solution of the least squares problem in (13), in the following sense:

$$\hat{d}^k = \begin{pmatrix} dx^k \\ d\sigma^k \\ vect\{dL^k\} \\ vect\{dW^k\} \end{pmatrix}.$$

Furthermore, if we define

$$\nabla\hat{\psi}(z) = \begin{pmatrix} \nabla_x \psi(z) \\ \nabla_\sigma \psi(z) \\ vect\{\nabla_L \psi(z)\} \\ vect\{\nabla_W \psi(z)\} \end{pmatrix}, \quad \nabla\hat{p}(u) = \begin{pmatrix} \nabla_x p(H(z)) \\ \nabla_\sigma p(H(z)) \\ vect\{\nabla_L p(H(z))\} \\ vect\{\nabla_W p(H(z))\} \end{pmatrix},$$

for the symmetry of the matrices involved in the calculations, we have

$$\nabla\psi(z^k) \bullet d^k = \nabla\hat{\psi}(z)^T \hat{d}^k, \quad \nabla\hat{\psi}(z) = Ju^T \nabla\hat{p}(u).$$

Another propriety of $\hat{d}^k$ is that it satisfies the normal equations of (18)

$$\hat{d}^k = \left(Ju^T Ju\right)^{-1} Ju^T \left(\beta_k \frac{\hat{o}^T u}{\|\hat{o}\|^2}\hat{o} - u\right). \tag{19}$$

Therefore, from the assumptions of the theorem and by exploiting the (19) it is possible to obtain

$$\begin{aligned}
\nabla\hat{\psi}(z)^T \hat{d}^k &= \nabla\hat{p}(u)^T (Ju)\hat{d}^k \\
&\overset{(19)}{=} \nabla\hat{p}(u)^T Ju \left(Ju^T Ju\right)^{-1} Ju^T \left(\beta_k \tfrac{\hat{o}^T u}{\|\hat{o}\|^2}\hat{o} - u\right) \\
&= \nabla\hat{p}(u)^T Ju Ju^{-1} (Ju^T)^{-1} Ju^T \left(\beta_k \tfrac{\hat{o}^T u}{\|\hat{o}\|^2}\hat{o} - u\right) \\
&= \nabla\hat{p}(u)^T \left(\beta_k \tfrac{\hat{o}^T u}{\|\hat{o}\|^2}\hat{o} - u\right) \le -\nabla\hat{p}(u)^T u(1 - \tfrac{\beta_k}{\bar{\beta}}) \\
&= -\nabla p(H(z)) \bullet H(z)(1 - \tfrac{\beta_k}{\bar{\beta}}) \overset{(A5)}{<} 0,
\end{aligned}$$

where with Ju^{-1} and $(Ju^T)^{-1}$ are the Moore Penrose pseudo inverses of Ju and Ju^T, respectively. The third equality derives from the propriety

$$(AB)^{-1} = B^{-1}A^{-1},$$

valid for the Moore Penrose pseudo inverse in the case we are considering (interested readers can refer to [8]). The last equality follows from the definition of $\hat{H}(z)$ and $\hat{p}(u)$. □

At step `(S.3)` the potential function (12) is used to measure the progress of the algorithm. Finally at Step `(S.4)` the value of k is updated and the loop is completed.

It is possible to observe that the sequence generated by Algorithm 1 necessarily belongs to Ω. We now present the convergence result:

Theorem 5. *Assume that $CE(\Omega, H)$ has a solution. Let $\{z^k\}$ be the sequence generated by Algorithm 1, then*

(a) the sequence $\{H(z^k)\}$ is bounded;
(b) any accumulation point of $\{z^k\}$, if it exists, solves CE(Ω, H);
(c) $\lim_{k\to\infty} H(z^k) = 0$;
(d) the sequence $\{z^k\} = \{(x^k, \sigma^k, L^k, W^k)\}$ is bounded.

Proof. The proof of statements (a) and (b) follows from Theorem 3 of [13].

In order to prove the (c) we first have to prove the (d), that is the boundedness of $\{z^k\}$. To prove the boundless of $\{z^k\}$ we have to prove the boundedness of the sequences $\{x^k\}$, $\{\sigma^k\}$, $\{L^k\}$, and $\{W^k\}$. The boundedness of $\{L^k\}$ is a direct consequence of the boundedness of $\{H(z^k)\}$.

To prove the boundedness of the sequences $\{x^k\}$ and $\{\sigma^k\}$ we use the operator Γ. In detail, from the (4) we obtain

$$\nabla_x \Xi(x, \sigma) = \quad G(\sigma)x - F(\sigma), \tag{20}$$

$$-\nabla_\sigma \Xi(x, \sigma) = \nabla V^*(\sigma) - \nabla V(\Lambda(x)). \tag{21}$$

It is easy to see that if one of the two sequences goes to infinity while the other converges, $\|\Gamma(x^k, \sigma^k)\| \to \infty$ contradicting the (a).
We consider the case in which $\{x^k\}$ and $\{\sigma^k\}$ go to infinity simultaneously. It is possible to notice from the (4) that $F(\sigma)$ is linear in σ, and therefore if both the variables go to infinity we have $\|\nabla_x \Xi(x^k, \sigma^k)\| \to \infty$. Finally if we suppose that $\{W^k\} \to \infty$, from the boundedness of $\{\sigma^k\}$ and constraint $W - G(\sigma)$ we obtain the desired contradiction with the (a).

The (c) is a direct consequence of conditions (b) and (d). □

5 Conclusions

We presented an interior points method framework for canonical duality theory that converges under mild assumptions. The framework in this paper not only has really favorable convergence proprieties, but it is also general and potentially able to handle large-sized problems efficiently with a good level of reliability.

In our view, these results constitute an important step for several topics in optimization. The new findings of this paper indicate that it is possible to adapt interior points methods to the problems reformulated with canonical duality. Therefore, other popular interior points methods such as primal–dual methods could be used to solve problem (1) and find the global solution of many nonconvex optimization problems efficiently.

There are also several applications that can be investigated with the presented framework. In detail, the maximum cut problem and the radial basis function neural networks problems can also be solved with canonical duality [10, 16], and the proposed algorithm could be useful to find their global solutions for large-sized instances.

Acknowledgements The Author would like to thank Simone Sagratella for his help. Without his suggestions in the initial conception of this method, it would have been quite difficult to understand the right path to take in order to create the presented framework. The author would also like to thank Professor Stefano Lucidi for his suggestions for improving the paper.

References

1. Dreves, A., Facchinei, F., Kanzow, C., Sagratella, S.: On the solution of the KKT conditions of generalized Nash equilibrium problems. SIAM J. Optim. **21**, 1082–1108 (2011)
2. Facchinei, F., Kanzow, C., Sagratella, S.: Solving quasi-variational inequalities via their KKT conditions. Math. Prog. Ser. A **144**, 369–412 (2014). doi:10.1007/s10107-013-0637-0
3. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. Volume. I and II. Springer, New York (2003)
4. Gao, D.Y.: Duality principles in nonconvex systems: Theory, methods and applications. Kluwer Academic Publishers, Dordrecht (2000)
5. Gao, D.Y.: Canonical dual transformation method and generalized triality theory in nonsmooth global optimization. J. Glob. Optim. **17**(1/4), 127–160 (2000)
6. Gao, D.Y.: Canonical duality theory: theory, method, and applications in global optimization. Comput. Chem. **33**, 1964–1972 (2009)
7. Gao, D.Y., Ruan, N., Pardalos, P.M.: Canonical dual solutions to sum of fourth-order polynomials minimization problems with applications to sensor network localization. In: Pardalos, P.M., Ye, Y.Y., Boginski, V., Commander, C. (eds.) Sensors: Theory, Algorithms and Applications, vol. 61, pp. 37–54. Springer (2012)
8. Greville, T.N.E.: Note on the generalized inverse of a matrix product. SIAM Review **8**(4), 518–521 (1966)
9. Latorre, V., Gao, D.Y.: Canonical duality for solving general nonconvex constrained problems. Optim. Lett. **10**(8), 1763–1779 (2016). doi:10.1007/s11590-015-0860-0
10. Latorre, V., Gao, D.Y.: Canonical dual solution to nonconvex radial basis neural network optimization problem. Neurocomputing **134**, 189–197 (2014)
11. Latorre, V., Sagratella, S.: A Canonical Duality Approach for the Solution of Affine Quasi-Variational Inequalities. J. Glob. Opt. (2014). doi:10.1007/s10898-014-0236-5
12. Monteiro, R.D.C., Pang, J.S.: On two interior-point mappings for nonlinear semidefinite complementarity problems. Math. Oper. Res. **23**, 39–60 (1998)
13. Monteiro, R.D.C., Pang, J.S.: A potential reduction Newton method for constrained equations. SIAM J. Optim. **9**(3), 729–754 (1999)
14. Rockafellar, R.T.: Monotone operators associated with saddle-functions and minimax problems. Proc. Pure Math **18**, 241–250 (1970)
15. Ruan, N., Gao, D.Y.: Global optimal solutions to a general sensor network localization problem. Perform. Eval. **75–76**, 1–16 (2014)
16. Wang, Z.B., Fang, S.C., Gao, D.Y., Xing, W.X.: Canonical dual approach to solving the maximum cut problem. J. Glob. Optim. **54**, 341–352 (2012)
17. Wu, C., Gao, D. Y.: Canonical primal-dual method for solving non-convex minimization problems. In: Gao, D.Y., Latorre, V., Ruan, N. (eds.) Canonical Duality Theory. Springer (2017)