

Global Optimal Solution to Quadratic Discrete Programming Problem with Inequality Constraints

Ning Ruan and David Yang Gao

Abstract This paper presents a canonical dual method for solving a quadratic discrete value selection problem subjected to inequality constraints. By using a linear transformation, the problem is first reformulated as a standard quadratic 0–1 integer programming problem. Then, by the canonical duality theory, this challenging problem is converted to a concave maximization over a convex feasible set in continuous space. It is proved that if this canonical dual problem has a solution in its feasible space, the corresponding global solution to the primal problem can be obtained directly by a general analytical form. Otherwise, the problem could be NP-hard. In this case, a quadratic perturbation method and an associated canonical primal-dual algorithm are proposed. Numerical examples are illustrated to demonstrate the efficiency of the proposed method and algorithm.

1 Introduction

Many decision-making problems, such as portfolio selection, capital budgeting, production planning, resource allocation, and computer networks, etc., can often be formulated as quadratic programming problems with discrete variables. See for examples, [4, 5, 9, 24]. In engineering applications, the decision variables can not have arbitrary values. Instead, either some or all of the variables must be selected from a list of integer or discrete values for practical reasons. For examples, structural members may have to be selected from selections available in standard sizes, member thicknesses may have to be selected from the commercially available ones, the number of bolts for a connection must be an integer, the number of reinforcing bars in a concrete member must be an integer, etc. [23]. However, these integer

N. Ruan (✉) · D.Y. Gao

Faculty of Science and Technology, Federation University Australia, Ballarat,
VIC 3353, Australia

e-mail: n.ruan@federation.edu.au

D.Y. Gao

e-mail: d.gao@federation.edu.au

© Springer International Publishing AG 2017

D.Y. Gao et al. (eds.), *Canonical Duality Theory*, Advances in Mechanics
and Mathematics 37, DOI 10.1007/978-3-319-58017-3_16

programming problems are computationally highly demanding. Nevertheless, some numerical methods are now available.

Several review articles on nonlinear optimization problems with discrete variables are available [1, 4, 28, 33, 37, 38], and some popular methods have been discussed, including branch and bound methods, a hybrid method that combines a branch and bound method with a dynamic programming technique [29], sequential linear programming, rounding-off techniques, cutting plane techniques [2], heuristic techniques, penalty function approaches, simulated annealing [25], and genetic algorithms, etc. The relaxation methods have also been proposed recently, leading to second order cone programming (SOC) [21] and improved linearization strategy [35].

Branch and bound is perhaps the most widely known and used deterministic method for discrete optimization problems. When applied to linear problems, this method can be implemented in a way to yield a global minimum point. However for nonlinear problems there is no such guarantee, unless the problem is convex. The branch and bound method has been used successfully to deal with problems with discrete design variables. However for the problem with a large number of discrete design variables, the number of subproblems (nodes) becomes large, making the method inefficient.

Simulated annealing (SA) and genetic algorithms (GA) belong to the category of stochastic search methods [22] which based on an element of random choice. Because of this, one has to sacrifice the possibility of an absolute guarantee of success within a finite amount of computation.

Canonical duality theory provides a new and potentially useful methodology for solving a large class of nonconvex/nonsmooth/discrete problems (see the review articles [13, 19]). It was shown in [8, 12] that the Boolean integer programming problems are actually equivalent to certain canonical dual problems in continuous space without duality gap, which can be solved deterministically under certain conditions. This theory has been generalized for solving multi-integer programming [39] and the well-known max cut problems [40]. It is also shown in [13, 16] that by the canonical duality theory, the NP-hard quadratic integer programming problem is identical to a continuous unconstrained Lipschitzian global optimization problem, which can be solved via deterministic methods (but not in polynomial times) (see [20]). The canonical duality theory has been used successfully for solving a large class of challenging problems not only in global optimization, but also in nonconvex analysis and continuum mechanics [17].

In this paper, our goal is to solve a general quadratic programming problem with its decision variables taking values from discrete sets. The elements from these discrete sets are not required to be binary or uniformly distributed. An effective numerical method is developed based on the canonical duality theory [10]. The rest of the paper is organized as follows. Section 2 presents a mathematical statement of the general discrete value quadratic programming problem and how it can be transformed into a standard 0–1 programming problem in higher dimensional space. Section 3 presents a brief review on the canonical duality theory. Detailed canonical dual transformation procedure is presented in Sect. 4 to show how the integer programming problem can be converted to a concave maximization in a convex space. A perturbed computational

method is developed in Sect. 5. Some numerical examples are illustrated in Sect. 6 to demonstrate the effectiveness and efficiency of the proposed method. The paper is ended with some concluding remarks.

2 Primal Problem and Equivalent Transformation

The discrete programming problem to be addressed is given below:

$$(\mathcal{P}_a) \quad \min P(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} - \mathbf{c}^T \mathbf{x} \tag{1}$$

$$\text{s.t. } \mathbf{g}(\mathbf{x}) = \mathbf{A} \mathbf{x} - \mathbf{b} \leq 0, \tag{2}$$

$$\mathbf{x} = [x_1, x_2, \dots, x_n]^T, x_i \in U_i, i = 1, \dots, n,$$

where $\mathbf{Q} = \{q_{ij}\} \in \mathbb{R}^{n \times n}$ is a symmetric matrix, $\mathbf{A} = \{a_{ij}\} \in \mathbb{R}^{m \times n}$ is a matrix with $\text{rank}(\mathbf{A}) = m < n$, $\mathbf{c} = [c_1, \dots, c_n]^T \in \mathbb{R}^n$ and $\mathbf{b} = [b_1, \dots, b_m]^T \in \mathbb{R}^m$ are given vectors. Here, for each $i = 1, \dots, n$,

$$U_i = \{u_{i,1}, \dots, u_{i,K_i}\},$$

where $u_{i,j}, j = 1, \dots, K_i$, are given real numbers. In this paper, we let $K = \sum_{i=1}^n K_i$.

Problem (\mathcal{P}_a) arises in many real-world applications, say, the pipe network optimization problems in water distribution systems, where the choices of pipelines are discrete values. Such problems have been studied extensively by traditional direct approaches (see [41]). Due to the constraint of discrete values, this problem is considered to be NP-hard and the traditional methods can only provide upper bound results. In this paper, we will show that the canonical duality theory will provide either a lower bound approach to this challenging problem, or the global optimal solution under certain conditions.

In order to convert the discrete value problem (\mathcal{P}_a) to the standard 0–1 programming problem, we introduce the following transformation,

$$x_i = \sum_{j=1}^{K_i} u_{i,j} y_{i,j}, i = 1, \dots, n, \tag{3}$$

where, for each $i = 1, \dots, n, u_{i,j} \in U_i, j = 1, \dots, K_i$. Then, the discrete programming problem (\mathcal{P}_a) can be written as the following 0–1 programming problem:

$$(\mathcal{P}_b) \quad \min P(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T \mathbf{B} \mathbf{y} - \mathbf{h}^T \mathbf{y} \tag{4}$$

$$\text{s.t. } \mathbf{g}(\mathbf{y}) = \mathbf{D} \mathbf{y} - \mathbf{b} \leq 0, \tag{5}$$

$$\sum_{j=1}^{K_i} y_{i,j} - 1 = 0, \quad i = 1, \dots, n, \tag{6}$$

$$y_{i,j} \in \{0, 1\}, \quad i = 1, \dots, n; \quad j = 1, \dots, K_i, \tag{7}$$

where

$$\mathbf{y} = [y_{1,1}, \dots, y_{1,K_1}, \dots, y_{n,1}, \dots, y_{n,K_n}]^T \in \mathbb{R}^K,$$

$$\mathbf{h} = [c_1 u_{1,1}, \dots, c_1 u_{1,K_1}, \dots, c_n u_{n,1}, \dots, c_n u_{n,K_n}]^T \in \mathbb{R}^K,$$

$$B = \begin{bmatrix} q_{1,1} u_{1,1}^2 & \cdots & q_{1,1} u_{1,1} u_{1,K_1} & \cdots & q_{1,n} u_{1,1} u_{n,K_n} & \cdots \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\ q_{1,1} u_{1,K_1} u_{1,1} & \cdots & q_{1,1} u_{1,K_1}^2 & \cdots & \cdots & \cdots \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\ q_{n,1} u_{n,K_n} u_{1,1} & \cdots & \cdots & \cdots & q_{n,n} u_{n,K_n}^2 & \cdots \end{bmatrix} \in \mathbb{R}^{K \times K},$$

$$D = \begin{bmatrix} a_{1,1} u_{1,1} & \cdots & a_{1,1} u_{1,K_1} & \cdots & a_{1,n} u_{n,K_n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m,1} u_{1,1} & \cdots & a_{m,1} u_{1,K_1} & \cdots & a_{m,n} u_{n,K_n} \end{bmatrix} \in \mathbb{R}^{m \times K}.$$

Theorem 1 *Problem (\mathcal{P}_b) is equivalent to Problem (\mathcal{P}_a) .*

Proof For any $i = 1, 2, \dots, n$, it is clear that constraints (6) and (7) are equivalent to the existence of only one $j \in \{1, \dots, K_i\}$, such that $y_{i,j} = 1$ while $y_{i,j} = 0$ for all other j . Thus, from the definition of \mathbf{y} , the conclusion follows readily. \square

Problem (\mathcal{P}_b) is a standard 0–1 quadratic programming problem with both equality and inequality constraints. Let

$$H = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n \times K}$$

and, for a given integer K , let

$$\mathbf{e}_K = [1, \dots, 1, \dots, 1, \dots, 1]^T \in \mathbb{R}^K.$$

Thus, on the feasible space

$$\mathcal{Y} = \{\mathbf{y} \in \mathbb{R}^K : D\mathbf{y} \leq \mathbf{b}, \quad H\mathbf{y} = \mathbf{e}_n, \quad \mathbf{y} \in \{0, 1\}^K\}, \quad (8)$$

the integer constrained problem (\mathcal{P}_b) can be reformulated as a standard constrained 0–1 programming problem:

$$(\mathcal{P}_c) : \min \left\{ P(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T B \mathbf{y} - \mathbf{h}^T \mathbf{y} \mid \mathbf{y} \in \mathcal{Y} \right\}. \quad (9)$$

3 Canonical Duality Theory: A Brief Review

The basic idea of the canonical duality theory can be demonstrated by solving the following general nonconvex problem (the primal problem (\mathcal{P}) in short)

$$(\mathcal{P}) : \min_{\mathbf{x} \in \mathcal{X}_a} \left\{ P(\mathbf{x}) = \frac{1}{2} \langle \mathbf{x}, \mathbf{A} \mathbf{x} \rangle - \langle \mathbf{x}, \mathbf{f} \rangle + W(\mathbf{x}) \right\}, \quad (10)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a given symmetric indefinite matrix, $\mathbf{f} \in \mathbb{R}^n$ is a given vector (input), $\langle \mathbf{x}, \mathbf{x}^* \rangle$ denotes the bilinear form between \mathbf{x} and its dual variable \mathbf{x}^* , $\mathcal{X}_a \subset \mathbb{R}^n$ is a given feasible space, and $W : \mathcal{X}_a \rightarrow \mathbb{R} \cup \{\infty\}$ is a general nonconvex objective function.

It must be emphasized that, different from the objective function extensively used in mathematical optimization, a real-valued function $W(\mathbf{x})$ is called to be *objective* in continuum physics and the canonical duality theory only if (see [10] Chap.6, p. 288)

$$W(\mathbf{x}) = W(\mathbf{Q}\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}_a, \quad \forall \mathbf{Q} \in \mathcal{Q},$$

where $\mathcal{Q} = \{\mathbf{Q} \in \mathbb{R}^{n \times n} \mid \mathbf{Q}^{-1} = \mathbf{Q}^T, \det \mathbf{Q} = 1\}$ is a special rotation group.

Geometrically speaking, an objective function does not depend on the rotation, but only on certain measure of its variable. In Euclidean space \mathbb{R}^n , the simplest objective function is the ℓ_2 -norm $\|\mathbf{x}\|$ in \mathbb{R}^n since $\|\mathbf{Q}\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{Q}^T \mathbf{Q} \mathbf{x} = \|\mathbf{x}\|^2 \quad \forall \mathbf{Q} \in \mathcal{Q}$. By Cholesky factorization, any positive definite matrix has a unique decomposition $C = D^* D$. Thus, any convex quadratic function is objective. Physically, an objective function does not depend on observers [7], which is essential for any real-world mathematical modeling.

The **key step** in the canonical duality theory is to choose a nonlinear operator

$$\xi = \Lambda(\mathbf{x}) : \mathcal{X}_a \rightarrow \mathcal{E}_a \subset \mathbb{R}^p \quad (11)$$

and a *canonical function* $V : \mathcal{E}_a \rightarrow \mathbb{R}$ such that the nonconvex objective function $W(\mathbf{x})$ can be recast by adopting a canonical form $W(\mathbf{x}) = V(\Lambda(\mathbf{x}))$. Thus, the primal problem (\mathcal{P}) can be written in the following canonical form:

$$(\mathcal{P}) : \min_{\mathbf{x} \in \mathcal{X}_a} \{P(\mathbf{x}) = V(\Lambda(\mathbf{x})) - U(\mathbf{x})\}, \tag{12}$$

where $U(\mathbf{x}) = \langle \mathbf{x}, \mathbf{f} \rangle - \frac{1}{2} \langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle$. By the definition introduced in [10], a differentiable function $V(\xi)$ is said to be a *canonical function* on its domain \mathcal{E}_a if the duality mapping $\zeta = \nabla V(\xi)$ from \mathcal{E}_a to its range $\mathcal{S}_a \subset \mathbb{R}^p$ is invertible. Let $\langle \xi; \zeta \rangle$ denote the bilinear form on $\mathcal{E}_a \times \mathcal{S}_a$. Thus, for the given canonical function $V(\xi)$, its Legendre conjugate $V^*(\zeta)$ can be defined uniquely by the Legendre transformation (cf. Gao [10])

$$V^*(\zeta) = \text{sta}\{\langle \xi; \zeta \rangle - V(\xi) \mid \xi \in \mathcal{E}_a\}, \tag{13}$$

where the notation $\text{sta}\{g(\xi) \mid \xi \in \mathcal{E}_a\}$ stands for finding stationary point of $g(\xi)$ on \mathcal{E}_a . It is easy to prove that the following canonical duality relations hold on $\mathcal{E}_a \times \mathcal{S}_a$:

$$\zeta = \nabla V(\xi) \Leftrightarrow \xi = \nabla V^*(\zeta) \Leftrightarrow V(\xi) + V^*(\zeta) = \langle \xi; \zeta \rangle. \tag{14}$$

By this one-to-one canonical duality, the nonconvex term $W(\mathbf{x}) = V(\Lambda(\mathbf{x}))$ in the problem (\mathcal{P}) can be replaced by $\langle \Lambda(\mathbf{x}); \zeta \rangle - V^*(\zeta)$ such that the nonconvex function $P(\mathbf{x})$ is reformulated as the Gao-Strang total complementary function [10]:

$$\mathcal{E}(\mathbf{x}, \zeta) = \langle \Lambda(\mathbf{x}); \zeta \rangle - V^*(\zeta) - U(\mathbf{x}) : \mathcal{X}_a \times \mathcal{S}_a \rightarrow \mathbb{R}. \tag{15}$$

By using this total complementary function, the canonical dual function $P^d(\zeta)$ can be obtained as

$$\begin{aligned} P^d(\zeta) &= \text{sta}\{\mathcal{E}(\mathbf{x}, \zeta) \mid \mathbf{x} \in \mathcal{X}_a\} \\ &= U^\Lambda(\zeta) - V^*(\zeta), \end{aligned} \tag{16}$$

where $U^\Lambda(\zeta)$ is defined by

$$U^\Lambda(\zeta) = \text{sta}\{\langle \Lambda(\mathbf{x}); \zeta \rangle - U(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}_a\}. \tag{17}$$

In many applications, the geometrically nonlinear operator $\Lambda(\mathbf{x})$ is usually a quadratic function [3, 34]

$$\Lambda(\mathbf{x}) = \frac{1}{2} \langle \mathbf{x}, D_k \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{b}_k \rangle, \tag{18}$$

where $D_k \in \mathbb{R}^{n \times n}$ and $\mathbf{b}_k \in \mathbb{R}^n (k = 1, \dots, p)$. Let $\zeta = [\zeta_1, \dots, \zeta_p]^T$. In this case, the canonical dual function can be written in the following form:

$$P^d(\zeta) = -\frac{1}{2} \langle \mathbf{F}(\zeta), \mathbf{G}^{-1}(\zeta) \mathbf{F}(\zeta) \rangle - V^*(\zeta), \tag{19}$$

where

$$\mathbf{G}(\boldsymbol{\zeta}) = \mathbf{A} + \sum_{k=1}^p \zeta_k D_k, \quad \mathbf{F}(\boldsymbol{\zeta}) = \mathbf{f} - \sum_{k=1}^p \zeta_k \mathbf{b}_k.$$

Let

$$\mathcal{S}_a^+ = \{\boldsymbol{\zeta} \in \mathbb{R}^p \mid G(\boldsymbol{\zeta}) > 0\}.$$

It is easy to prove that \mathcal{S}_a^+ is convex. Moreover, \mathcal{S}_a^+ is nonempty as long as there exists one $D_k > 0$.

Therefore, the canonical dual problem can be proposed as

$$(\mathcal{P}^d) : \max\{P^d(\boldsymbol{\zeta}) \mid \boldsymbol{\zeta} \in \mathcal{S}_a^+\}. \tag{20}$$

which is a concave maximization problem over a convex set $\mathcal{S}_a^+ \subset \mathbb{R}^p$.

Theorem 2 ([10]). *Problem (\mathcal{P}^d) is canonically dual to (\mathcal{P}) in the sense that if $\bar{\boldsymbol{\zeta}}$ is a critical point of $P^d(\boldsymbol{\zeta})$, then*

$$\bar{\mathbf{x}} = \mathbf{G}^{-1}(\bar{\boldsymbol{\zeta}})\mathbf{F}(\bar{\boldsymbol{\zeta}}) \tag{21}$$

is a critical point of $\Pi(\mathbf{x})$ and

$$P(\bar{\mathbf{x}}) = \Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\zeta}}) = P^d(\bar{\boldsymbol{\zeta}}). \tag{22}$$

If $\bar{\boldsymbol{\zeta}}$ is a solution to (\mathcal{P}^d) , then $\bar{\mathbf{x}}$ is a global minimizer of (\mathcal{P}) and

$$\min_{\mathbf{x} \in \mathcal{X}_a} P(\mathbf{x}) = \Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\zeta}}) = \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} P^d(\boldsymbol{\zeta}). \tag{23}$$

Conversely, if $\bar{\mathbf{x}}$ is a solution to (\mathcal{P}) , it must be in the form of (21) for critical solution $\bar{\boldsymbol{\zeta}}$ of $P^d(\boldsymbol{\zeta})$.

To help explaining the theory, we consider a simple nonconvex optimization in \mathbb{R}^n :

$$\min P(\mathbf{x}) = \frac{1}{2}\alpha\left(\frac{1}{2}\|\mathbf{x}\|^2 - \lambda\right)^2 - \mathbf{x}^T \mathbf{f}, \quad \forall \mathbf{x} \in \mathbb{R}^n, \tag{24}$$

where $\alpha, \lambda > 0$ are given parameters. The criticality condition $\nabla P(\mathbf{x}) = 0$ leads to a nonlinear algebraic equation system in \mathbb{R}^n

$$\alpha\left(\frac{1}{2}\|\mathbf{x}\|^2 - \lambda\right)\mathbf{x} = \mathbf{f}. \tag{25}$$

Clearly, to solve this n-dimensional nonlinear algebraic equation directly is difficult. Also traditional convex optimization theory can not be used to identify global minimizer. However, by the canonical dual transformation, this problem can be

solved. To do so, we let $\xi = \Lambda(u) = \frac{1}{2}\|\mathbf{x}\|^2 - \lambda \in \mathbb{R}$. Then, the nonconvex function $W(\mathbf{x}) = \frac{1}{2}\alpha(\frac{1}{2}\|\mathbf{x}\|^2 - \lambda)^2$ can be written in canonical form $V(\xi) = \frac{1}{2}\alpha\xi^2$. Its Legendre conjugate is given by $V^*(\zeta) = \frac{1}{2}\alpha^{-1}\zeta^2$, which is strictly convex. Thus, the total complementary function for this nonconvex optimization problem is

$$\mathcal{E}(\mathbf{x}, \zeta) = (\frac{1}{2}\|\mathbf{x}\|^2 - \lambda)\zeta - \frac{1}{2}\alpha^{-1}\zeta^2 - \mathbf{x}^T \mathbf{f}. \quad (26)$$

For a fixed $\zeta \in \mathbb{R}$, the criticality condition $\nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x}, \zeta) = 0$ leads to

$$\zeta \mathbf{x} - \mathbf{f} = 0. \quad (27)$$

For each $\zeta \neq 0$, the Eq. (27) gives $\mathbf{x} = \mathbf{f}/\zeta$ in vector form. Substituting this into the total complementary function \mathcal{E} , the canonical dual function can be easily obtained as

$$\begin{aligned} P^d(\zeta) &= \{\mathcal{E}(\mathbf{x}, \zeta) | \nabla_{\mathbf{x}}\mathcal{E}(\mathbf{x}, \zeta) = 0\} \\ &= -\frac{\|\mathbf{f}\|^2}{2\zeta} - \frac{1}{2}\alpha^{-1}\zeta^2 - \lambda\zeta, \quad \forall \zeta \neq 0. \end{aligned} \quad (28)$$

The critical point of this canonical function is obtained by solving the following dual algebraic equation

$$(\alpha^{-1}\zeta + \lambda)\zeta^2 = \frac{1}{2}\|\mathbf{f}\|^2. \quad (29)$$

For any given parameters α , λ and the vector $\mathbf{f} \in \mathbb{R}^n$, this cubic algebraic equation has at most three roots satisfying $\zeta_1 \geq 0 \geq \zeta_2 \geq \zeta_3$, and each of these roots leads to a critical point of the nonconvex function $P(\mathbf{x})$, i.e., $\mathbf{x}_i = \mathbf{f}/\zeta_i$, $i = 1, 2, 3$. By the fact that $\zeta_1 \in \mathcal{S}_a^+ = \{\zeta \in \mathbb{R} | \zeta > 0\}$, then Theorem 1 tells us that \mathbf{x}_1 is a global minimizer of $P(\mathbf{x})$.

Consider one dimension problem with $\alpha = 1$, $\lambda = 2$, $f = \frac{1}{2}$, the primal function and canonical dual function are shown in Fig. 1, where, $x_1 = 2.11491$ is a global minimizer of $P(x)$, $\zeta_1 = 0.236417$ is a global maximizer of $P^d(\zeta)$, and $P(x_1) = -1.02951 = P^d(\zeta_1)$ (See the two black dots).

If we let $\mathbf{f} = 0$, the graph of $P(\mathbf{x})$ is symmetric (i.e., the so-called double-well potential or the Mexican hat for $n = 2$ [11]) with infinite number of global minimizers satisfying $\|\mathbf{x}\|^2 = 2\lambda$. In this case, the canonical dual $P^d(\zeta) = -\frac{1}{2}\alpha^{-1}\zeta^2 - \lambda\zeta$ is strictly concave with only one critical point (local maximizer) $\zeta_3 = -\alpha\lambda < 0$ (for $\alpha, \lambda > 0$). The corresponding solution $\mathbf{x}_3 = \mathbf{f}/\zeta_3 = 0$ is a local maximizer. By the canonical dual equation (29) we have $\zeta_1 = \zeta_2 = 0$ located on the boundary of \mathcal{S}_a^+ , which corresponding to the two global minimizers $x_{1,2} = \pm\sqrt{2\lambda}$ for $n = 1$, see Fig. 1b.

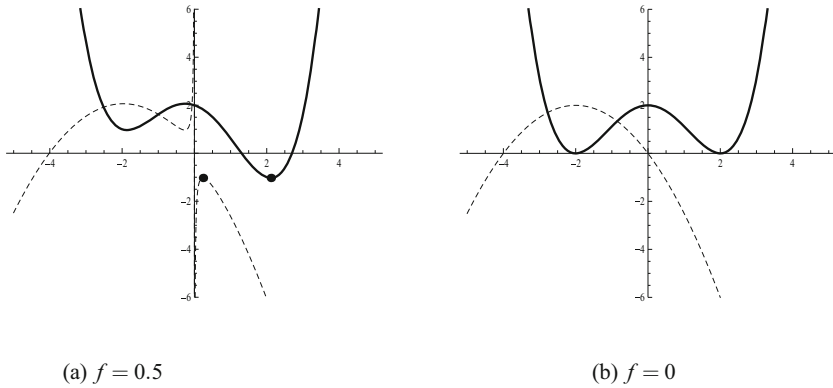


Fig. 1 Graphs of $P(\mathbf{x})$ (solid) and $P^d(\zeta)$ (dashed)

This simple example shows a fundamental issue in global optimization, i.e., the optimal solutions of a nonconvex problem depends sensitively on the linear term (input) \mathbf{f} . Geometrically speaking, the objective function $W(\mathbf{x})$ in $P(\mathbf{x})$ possesses certain symmetry. If there is no linear term, i.e., the *subjective function* in $P(\mathbf{x})$, the nonconvex problem usually has more than one global minimizer due to the symmetry. Traditional direct approaches and the popular SDP method are usually failed to deal with this situation. By the canonical duality theory, we understand that in this case the canonical dual function has no critical point in its open set \mathcal{S}_a^+ . Therefore, by adding a linear perturbation \mathbf{f} to break this symmetry, the canonical duality theory can be used to solve the nonconvex problems to obtain one of global optimal solutions. This idea was originally from Gao’s work (1996) on post-buckling analysis of large deformed beam. The potential energy of this beam model is a double-well function, similar to this example, without the force \mathbf{f} , the beam could have two buckling states (corresponding to two minimizers) and one un-buckled state (local maximizer). Later on (2008) in the Gao and Ogden work on analytical solutions in phase transformation [14], they further discovered that the nonconvex system has no phase transition unless the force distribution $f(x)$ vanished at certain points. They also discovered that if force field $f(x)$ changes dramatically, all the Newton type direct approaches failed even to find any local minimizer. This discovery is fundamentally important for understanding NP-hard problems in global optimization and chaos in nonconvex dynamical systems. The linear perturbation method has been used successfully for solving global optimization problems [16, 18, 32, 40]. Comprehensive reviews of the canonical duality theory and its applications in nonconvex analysis and global optimization can be found in [11, 13, 15].

4 Canonical Dual Problem

Now we are ready to apply the canonical duality theory for solving the integer programming problem (\mathcal{P}_c) presented in Sect.2. As indicated in [12, 13], the key step for solving this NP-hard problem is to use a so-called canonical measure $\rho = \{y_i(y_i - 1)\} \in \mathbb{R}^K$ such that the integer constraint $y_i \in \{0, 1\}$ can be equivalently written in the canonical form

$$\rho = \mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) = \{y_i(y_i - 1)\} = 0 \in \mathbb{R}^K$$

where the notation $\mathbf{s} \circ \mathbf{t} := [s_1t_1, s_2t_2, \dots, s_Kt_K]^T$ denotes the Hadamard product for any two vectors $\mathbf{s}, \mathbf{t} \in \mathbb{R}^K$. Thus, the so-called *geometrically admissible measure* Λ can be defined as

$$\begin{aligned} \xi &= \Lambda(\mathbf{y}) = \{D\mathbf{y} - \mathbf{b}, H\mathbf{y} - \mathbf{e}_n, \mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K)\} \\ &= \{\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\rho}\} \in \mathcal{E} = \mathbb{R}^{m+n+K}. \end{aligned}$$

Let

$$U(\mathbf{y}) = -P(\mathbf{y}) = \mathbf{h}^T \mathbf{y} - \frac{1}{2} \mathbf{y}^T B \mathbf{y},$$

and define

$$V(\xi) = \begin{cases} 0 & \text{if } \boldsymbol{\varepsilon} \leq 0, \boldsymbol{\delta} = 0, \boldsymbol{\rho} = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Clearly, the constraints in \mathcal{P} can be replaced by the canonical transformation $V(\Lambda(\mathbf{y}))$ and the primal problem (\mathcal{P}_c) can be equivalently written in the standard *canonical form* [13]

$$(\mathcal{P}) : \min \{\Pi(\mathbf{y}) = V(\Lambda(\mathbf{y})) - U(\mathbf{y}) : \mathbf{y} \in \mathbb{R}^K\}. \tag{30}$$

By the fact that $V(\xi)$ is convex, lower, semi-continuous on \mathcal{E} , its sub-differential leads to the canonical dual variable $\boldsymbol{\varsigma} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \partial V(\xi) \in \mathcal{E}^* = \mathbb{R}^{m+n+K}$, and its Fenchel super-conjugate (cf. Rockafellar [30])

$$\begin{aligned} V^\sharp(\boldsymbol{\varsigma}) &= \sup\{\langle \xi; \boldsymbol{\varsigma} \rangle - V(\xi) : \xi \in \mathcal{E}\} \\ &= \begin{cases} 0 & \text{if } \boldsymbol{\sigma} \geq 0, \boldsymbol{\tau} \neq 0, \boldsymbol{\mu} \neq 0 \\ +\infty & \text{otherwise} \end{cases} \end{aligned} \tag{31}$$

is also convex, l.s.c. on \mathcal{E}^* . By convex analysis, the following generalized canonical duality relations

$$\boldsymbol{\varsigma} \in \partial V(\xi) = \mathcal{E}_a^* \Leftrightarrow \xi \in \partial V^\sharp(\boldsymbol{\varsigma}) = \mathcal{E}_a \Leftrightarrow V(\xi) + V^\sharp(\boldsymbol{\varsigma}) = \langle \xi; \boldsymbol{\varsigma} \rangle \tag{32}$$

hold on $\mathcal{E} \times \mathcal{E}^*$, where

$$\mathcal{E}_a = \{\boldsymbol{\xi} = \{\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\rho}\} \in \mathcal{E} \mid \boldsymbol{\varepsilon} \leq 0, \boldsymbol{\delta} = 0, \boldsymbol{\rho} = 0\},$$

$$\mathcal{E}_a^* = \{\boldsymbol{\zeta} = \{\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}\} \in \mathcal{E}^* \mid \boldsymbol{\sigma} \geq 0, \boldsymbol{\tau} \neq 0, \boldsymbol{\mu} \neq 0\}$$

are effective domains of V and V^\sharp , respectively. The last equality in (32) is equivalent to the following KKT complementarity conditions:

$$\boldsymbol{\varepsilon}^T \boldsymbol{\sigma} = 0, \quad \boldsymbol{\delta}^T \boldsymbol{\tau} = 0, \quad \boldsymbol{\rho}^T \boldsymbol{\mu} = 0. \quad (33)$$

Clearly, the condition $\boldsymbol{\mu} \neq 0$ leads to the integer condition $\boldsymbol{\rho} = \{y_i(y_i - 1)\} = 0 \in \mathbb{R}^K$. Let

$$\mathbf{F}(\boldsymbol{\zeta}) = \mathbf{h} - D^T \boldsymbol{\sigma} - H^T \boldsymbol{\tau} + \boldsymbol{\mu}, \quad (34)$$

$$\mathbf{G}(\boldsymbol{\mu}) = B + 2\text{Diag}(\boldsymbol{\mu}). \quad (35)$$

Thus, on $\mathbb{R}^K \times \mathcal{E}_a^*$, the total complementary function \mathcal{E} associated with $\Pi(\mathbf{y})$ can be written as

$$\begin{aligned} \mathcal{E}(\mathbf{y}, \boldsymbol{\zeta}) &= \langle \Lambda(\mathbf{y}); \boldsymbol{\zeta} \rangle - V^\sharp(\boldsymbol{\zeta}) - U(\mathbf{y}) \\ &= \frac{1}{2} \mathbf{y}^T \mathbf{G}(\boldsymbol{\mu}) \mathbf{y} - \mathbf{F}^T(\boldsymbol{\zeta}) \mathbf{y} - \boldsymbol{\sigma}^T \mathbf{b} - \boldsymbol{\tau}^T \mathbf{e}_n. \end{aligned}$$

The criticality condition $\nabla_{\mathbf{y}} \mathcal{E}(\mathbf{y}, \boldsymbol{\zeta}) = 0$ leads to the canonical equilibrium equation

$$\mathbf{G}(\boldsymbol{\mu}) \mathbf{y} - \mathbf{F}(\boldsymbol{\zeta}) = 0. \quad (36)$$

Let $\mathcal{S}_a \subset \mathcal{E}_a^*$ be a canonical dual space:

$$\mathcal{S}_a = \{\boldsymbol{\zeta} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{E}_a^* : \det \mathbf{G}(\boldsymbol{\zeta}) \neq 0\}. \quad (37)$$

Then on \mathcal{S}_a , the canonical dual function can be finally formulated as

$$\begin{aligned} \Pi^d(\boldsymbol{\zeta}) &= \text{sta}\{\mathcal{E}(\mathbf{y}, \boldsymbol{\zeta}) : \mathbf{y} \in \mathbb{R}^K\} \\ &= -\frac{1}{2} \mathbf{F}^T(\boldsymbol{\zeta}) \mathbf{G}^{-1}(\boldsymbol{\mu}) \mathbf{F}(\boldsymbol{\zeta}) - \boldsymbol{\sigma}^T \mathbf{b} - \boldsymbol{\tau}^T \mathbf{e}_n. \end{aligned} \quad (38)$$

Theorem 3 (Complementary-Dual Principle). *If $\bar{\boldsymbol{\zeta}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$ is a KKT point of $\Pi^d(\boldsymbol{\zeta})$ on \mathcal{S}_a , then the vector*

$$\bar{\mathbf{y}}(\bar{\boldsymbol{\zeta}}) = \mathbf{G}^{-1}(\bar{\boldsymbol{\mu}}) \mathbf{F}(\bar{\boldsymbol{\zeta}}) \quad (39)$$

is a KKT point of Problem (\mathcal{P}) and

$$\Pi(\bar{\mathbf{y}}) = \Pi^d(\bar{\boldsymbol{\zeta}}). \tag{40}$$

Proof By introducing the Lagrange multiplier vectors $\boldsymbol{\xi} = \{\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\rho}\} \in \mathcal{E}_a$ to relax the inequality constraints¹ in \mathcal{E}_a^* , the Lagrangian function associated with the dual function $\Pi^d(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu})$ becomes

$$L(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}, \boldsymbol{\rho}) = \Pi^d(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) - \boldsymbol{\varepsilon}^T \boldsymbol{\sigma} - \boldsymbol{\delta}^T \boldsymbol{\tau} - \boldsymbol{\rho}^T \boldsymbol{\mu}.$$

Then, in terms of $\mathbf{y} = G^{-1}(\boldsymbol{\mu})\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu})$, the criticality condition $\nabla_{\boldsymbol{\zeta}} L(\boldsymbol{\zeta}, \boldsymbol{\xi}) = 0$ leads to

$$\begin{aligned} \frac{\partial L(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\rho})}{\partial \boldsymbol{\sigma}} &= D\mathbf{y} - \mathbf{b} - \boldsymbol{\varepsilon} = 0, \\ \frac{\partial L(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\rho})}{\partial \boldsymbol{\tau}} &= H\mathbf{y} - \mathbf{e}_n - \boldsymbol{\delta} = 0, \\ \frac{\partial L(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{\rho})}{\partial \boldsymbol{\mu}} &= \mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) - \boldsymbol{\rho} = 0, \end{aligned}$$

as well as the KKT conditions

$$\boldsymbol{\sigma} \geq 0, \quad \boldsymbol{\varepsilon} \leq 0, \quad \boldsymbol{\sigma}^T \boldsymbol{\varepsilon} = 0, \tag{41}$$

$$\boldsymbol{\tau} \neq 0, \quad \boldsymbol{\delta} = 0, \quad \boldsymbol{\delta}^T \boldsymbol{\tau} = 0. \tag{42}$$

$$\boldsymbol{\mu} \neq 0, \quad \boldsymbol{\rho} = 0, \quad \boldsymbol{\rho}^T \boldsymbol{\mu} = 0. \tag{43}$$

They can be written as:

$$D\mathbf{y} - \mathbf{b} \leq 0, \tag{44}$$

$$H\mathbf{y} - \mathbf{e}_n = 0, \tag{45}$$

$$\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) = 0, \tag{46}$$

This proves that if $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$ is a KKT point of $\Pi^d(\boldsymbol{\zeta})$, then the vector

$$\bar{\mathbf{y}}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}}) = \mathbf{G}^{-1}(\bar{\boldsymbol{\mu}})\mathbf{F}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$$

is a KKT point of Problem (\mathcal{P}).

¹The inequality $\det \mathbf{G}(\boldsymbol{\zeta}) \neq 0$ is not a constraint since the Lagrange multiplier for this inequality is identical zero.

Again, by the complementary conditions (41)–(43) and (39), we have

$$\begin{aligned}
 \Pi^d(\bar{\sigma}, \bar{\tau}, \bar{\mu}) &= -\frac{1}{2}\mathbf{F}(\bar{\sigma}, \bar{\tau}, \bar{\mu})^T \mathbf{G}(\bar{\mu})^{-1} \mathbf{F}(\bar{\sigma}, \bar{\tau}, \bar{\mu}) - \bar{\sigma}^T \mathbf{b} - \bar{\tau}^T \mathbf{e}_n \\
 &= \frac{1}{2}\bar{\mathbf{y}}^T B \bar{\mathbf{y}} - \mathbf{h}^T \bar{\mathbf{y}} + \bar{\sigma}^T (D \bar{\mathbf{y}} - \mathbf{b}) + \bar{\tau}^T (H \bar{\mathbf{y}} - \mathbf{e}_n) + \bar{\mu}^T (\bar{\mathbf{y}} \circ (\bar{\mathbf{y}} - \mathbf{e}_K)) \\
 &= \frac{1}{2}\bar{\mathbf{y}}^T B \bar{\mathbf{y}} - \mathbf{h}^T \bar{\mathbf{y}} = \Pi(\bar{\mathbf{y}}).
 \end{aligned}$$

Therefore, the theorem is proved. \square

Theorem 3 shows that the strong duality (40) holds for all KKT points of the primal and dual problems. In continuum mechanics, this theorem solved a 50-year-old problem and is known as the Gao principle [27]. In nonconvex analysis, this theorem can be used for solving a large class of fully nonlinear partial differential equations.

Remark 1. As we have demonstrated that by the generalized canonical duality (32), all KKT conditions can be recovered for both equality and inequality constraints. Generally speaking, the nonzero Lagrange multiplier condition for the linear equality constraint is usually ignored in optimization textbooks. But it can not be ignored for nonlinear constraints. It is proved recently [26] that the popular augmented Lagrange multiplier method can be used mainly for linear constrained problems. Since the inequality constraint $\boldsymbol{\mu} \neq 0$ produces a nonconvex feasible set \mathcal{E}_a^* , this constraint can be replaced by either $\boldsymbol{\mu} < 0$ or $\boldsymbol{\mu} > 0$. But the condition $\boldsymbol{\mu} < 0$ is corresponding to $\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) \geq 0$, this leads to a nonconvex open feasible set for the primal problem. By the fact that the integer constraints $y_i(y_i - 1) = 0$ are actually a special case (boundary) of the boxed constraints $0 \leq y_i \leq 1$, which is corresponding to $\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) \leq 0$, we should have $\boldsymbol{\mu} > 0$ (see [8] and [12, 16]). In this case, the KKT condition (43) should be replaced by

$$\boldsymbol{\mu} > 0, \quad \mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) \leq 0, \quad \boldsymbol{\mu}^T [\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K)] = 0. \quad (47)$$

Therefore, as long as $\boldsymbol{\mu} \neq 0$ is satisfied, the complementarity condition in (47) leads to the integer condition $\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) = 0$. Similarly, the inequality $\boldsymbol{\tau} \neq 0$ can be replaced by $\boldsymbol{\tau} > 0$.

By this remark, we can introduce a convex subset of the dual feasible space \mathcal{S}_a :

$$\mathcal{S}_a^+ = \{\boldsymbol{\zeta} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{E}^* : \boldsymbol{\sigma} \geq 0, \boldsymbol{\tau} > 0, \boldsymbol{\mu} > 0, \mathbf{G}(\boldsymbol{\mu}) > 0\}. \quad (48)$$

Then the canonical dual problem can be eventually proposed as the following

$$(\mathcal{P}^d) \quad \max \left\{ \Pi^d(\boldsymbol{\zeta}) = -\frac{1}{2}\mathbf{F}^T(\boldsymbol{\zeta})\mathbf{G}^{-1}(\boldsymbol{\mu})\mathbf{F}(\boldsymbol{\zeta}) - \boldsymbol{\sigma}^T \mathbf{b} - \boldsymbol{\tau}^T \mathbf{e}_n \mid \boldsymbol{\zeta} \in \mathcal{S}_a^+ \right\}. \quad (49)$$

It is easy to check that $\Pi^d(\boldsymbol{\zeta})$ is concave on the convex open set \mathcal{S}_a^+ . Therefore, if \mathcal{S}_a^+ is not empty, this canonical dual problem can be solved easily by convex minimization techniques.

Theorem 4 Assume that $\bar{\boldsymbol{\zeta}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$ is a KKT point of $\Pi^d(\boldsymbol{\zeta})$ and $\bar{\mathbf{y}} = \mathbf{G}^{-1}(\bar{\boldsymbol{\mu}})$ $\mathbf{F}(\bar{\boldsymbol{\zeta}})$. If $\bar{\boldsymbol{\zeta}} \in \mathcal{S}_a^+$, then $\bar{\mathbf{y}}$ is a global minimizer of $\Pi(\mathbf{y})$ and $\bar{\boldsymbol{\zeta}}$ is a global maximizer of $\Pi^d(\boldsymbol{\zeta})$ with

$$\Pi(\bar{\mathbf{y}}) = \min_{\mathbf{y} \in \mathbb{R}^K} \Pi(\mathbf{y}) = \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \Pi^d(\boldsymbol{\zeta}) = \Pi^d(\bar{\boldsymbol{\zeta}}) \tag{50}$$

Proof It is easy to check that the total complementary function $\Xi(\mathbf{y}, \boldsymbol{\zeta})$ is a saddle function on the open set $\mathbb{R}^K \times \mathcal{S}_a^+$, i.e., convex (quadratic) in $\mathbf{y} \in \mathbb{R}^K$ and concave (linear) in $\boldsymbol{\zeta} \in \mathcal{S}_a^+$. Therefore, if $(\bar{\mathbf{y}}, \bar{\boldsymbol{\zeta}})$ is a critical point of $\Xi(\mathbf{y}, \boldsymbol{\zeta})$, we must have

$$\begin{aligned} \Pi^d(\bar{\boldsymbol{\zeta}}) &= \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} P^d(\boldsymbol{\zeta}) = \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \min_{\mathbf{y} \in \mathbb{R}^K} \Xi(\mathbf{y}, \boldsymbol{\zeta}) = \min_{\mathbf{y} \in \mathbb{R}^K} \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \Xi(\mathbf{y}, \boldsymbol{\zeta}) \\ &= \min_{\mathbf{y} \in \mathbb{R}^K} \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \left\{ \frac{1}{2} \mathbf{y}^T \mathbf{G}(\boldsymbol{\mu}) \mathbf{y} - (\mathbf{h} - D^T \boldsymbol{\sigma} - H^T \boldsymbol{\tau} + \boldsymbol{\mu})^T \mathbf{y} - \boldsymbol{\sigma}^T \mathbf{b} - \boldsymbol{\tau}^T \mathbf{e}_n \right\} \\ &= \min_{\mathbf{y} \in \mathbb{R}^K} \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \left\{ \frac{1}{2} \mathbf{y}^T B \mathbf{y} - \mathbf{h}^T \mathbf{y} + \boldsymbol{\sigma}^T (D \mathbf{y} - \mathbf{b}) + \boldsymbol{\tau}^T (H \mathbf{y} - \mathbf{e}_n) + \boldsymbol{\mu}^T [\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K)] \right\} \\ &= \min_{\mathbf{y} \in \mathbb{R}^K} \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \left\{ \frac{1}{2} \mathbf{y}^T B \mathbf{y} - \mathbf{h}^T \mathbf{y} + \langle \Lambda(\mathbf{y}); \boldsymbol{\zeta} \rangle \right\} \end{aligned} \tag{51}$$

Note that

$$\min_{\boldsymbol{\zeta} \in \mathcal{E}^*} \{V^\sharp(\boldsymbol{\zeta})\} = V^\sharp(\bar{\boldsymbol{\zeta}}) = 0, \quad \min_{\boldsymbol{\xi} \in \mathcal{E}} \{V(\boldsymbol{\xi})\} = V(\bar{\boldsymbol{\xi}}) = 0.$$

Thus, it follows from (51) that

$$\begin{aligned} \Pi^d(\bar{\boldsymbol{\zeta}}) &= \min_{\mathbf{y} \in \mathbb{R}^K} \max_{\boldsymbol{\zeta} \in \mathcal{E}^*} \left\{ \frac{1}{2} \mathbf{y}^T B \mathbf{y} - \mathbf{h}^T \mathbf{y} + \langle \Lambda(\mathbf{y}); \boldsymbol{\zeta} \rangle - V^\sharp(\boldsymbol{\zeta}) \right\} \\ &= \min_{\mathbf{y} \in \mathbb{R}^K} \left\{ \frac{1}{2} \mathbf{y}^T B \mathbf{y} - \mathbf{h}^T \mathbf{y} \right\} + \max_{\boldsymbol{\zeta} \in \mathcal{E}^*} \{ \langle \Lambda(\mathbf{y}); \boldsymbol{\zeta} \rangle - V^\sharp(\boldsymbol{\zeta}) \} \\ &= \min_{\mathbf{y} \in \mathbb{R}^K} \left\{ \frac{1}{2} \mathbf{y}^T B \mathbf{y} - \mathbf{h}^T \mathbf{y} + V(\Lambda(\mathbf{y})) \right\} \\ &= \min_{\mathbf{y} \in \mathbb{R}^K} \Pi(\mathbf{y}) = \min_{\mathbf{y} \in \mathcal{Y}} P(\mathbf{y}). \end{aligned}$$

This completes the proof. □

Remark 2. By the fact that \mathcal{S}_a^+ is an open convex set, if the problem (\mathcal{P}) has multiple global minimizers, then its canonical dual solutions could be located on the boundary of \mathcal{S}_a^+ as illustrated in Sect. 3 and in [12, 31]. In order to solve this problem, we let

$$\mathcal{S}_c^+ = \{\boldsymbol{\zeta} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{S}_a^+ : \boldsymbol{\mu} \geq 0, \boldsymbol{\tau} \geq 0, \mathbf{G}(\boldsymbol{\mu}) \succeq 0\}.$$

Then on this closed convex domain, the relaxed concave maximization problem

$$(\mathcal{P}^\sharp) \max\{\Pi^d(\boldsymbol{\zeta}) : \boldsymbol{\zeta} \in \mathcal{S}_c^+\} \tag{52}$$

has at least one solution $\bar{\boldsymbol{\zeta}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$. If the corresponding $\bar{\mathbf{y}} = \mathbf{G}^{-1}(\bar{\boldsymbol{\mu}})\mathbf{F}(\bar{\boldsymbol{\zeta}})$ is feasible, then $\bar{\mathbf{y}}$ is a global minimizer of the primal problem (\mathcal{P}) . If $\mathbf{G}(\bar{\boldsymbol{\mu}})$ is singular, then $\mathbf{G}^{-1}(\bar{\boldsymbol{\mu}})$ can be replaced by the Moore–Penrose generalized inverse \mathbf{G}^\dagger (see [31]). Otherwise, the relaxed canonical dual (\mathcal{P}^\sharp) provides a lower bound approach to the primal problem (\mathcal{P}) , i.e.,

$$\min_{\mathbf{y} \in \mathcal{Y}} P(\mathbf{y}) \geq \max_{\boldsymbol{\zeta} \in \mathcal{S}_c^+} \Pi^d(\boldsymbol{\zeta}).$$

This is one of the main advantages of the canonical duality theory.

5 Canonical Perturbation Method

In fact, Problem (\mathcal{P}^d) can be rewritten as a convex minimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{F}^T(\boldsymbol{\zeta}) \mathbf{G}^{-1}(\boldsymbol{\mu}) \mathbf{F}(\boldsymbol{\zeta}) + \boldsymbol{\sigma}^T \mathbf{b} + \boldsymbol{\tau}^T \mathbf{e}_n, \\ \text{s.t.} \quad & \boldsymbol{\zeta} \in \mathcal{S}_a^+. \end{aligned}$$

If the primal problem has a unique global minimal solution, this canonical dual problem may have a unique critical point in \mathcal{S}_a^+ which can be obtained easily by well-developed nonlinear minimization techniques. Otherwise, the canonical dual function $\Pi^d(\boldsymbol{\zeta})$ may have critical point $\bar{\boldsymbol{\zeta}}$ located on the boundary of \mathcal{S}_a^+ , where the matrix $\mathbf{G}(\boldsymbol{\mu})$ is singular. In order to handle this issue, (\mathcal{P}^d) can be relaxed to a semi-definite programming problem:

$$\begin{aligned} \min \quad & g + \boldsymbol{\sigma}^T \mathbf{b} + \boldsymbol{\tau}^T \mathbf{e}_n, \\ \text{s.t.} \quad & g \geq \frac{1}{2} \mathbf{F}^T(\boldsymbol{\zeta}) \mathbf{G}^\dagger(\boldsymbol{\mu}) \mathbf{F}(\boldsymbol{\zeta}), \end{aligned} \tag{53}$$

$$\mathbf{G}(\boldsymbol{\mu}) \succeq 0, \tag{54}$$

$$\boldsymbol{\zeta} \in \mathcal{E}^*, \boldsymbol{\sigma} \geq 0, \boldsymbol{\mu} > 0, \tag{55}$$

where the parameter g is actually the Gao–Strang pure complementary gap function [19], and \mathbf{G}^\dagger represents the Moore–Penrose generalized inverse of \mathbf{G} . Since $\boldsymbol{\tau}$ is a Lagrange multiplier for the linear equality $H\mathbf{y} = \mathbf{e}_n$, the condition $\boldsymbol{\tau} \neq 0$ can be ignored in this section as long as the final solution \mathbf{y} is feasible.

Lemma 1 (Schur Complementary Lemma). *Let*

$$A = \begin{bmatrix} B & C^T \\ C & D \end{bmatrix},$$

If $B \succ 0$, then A is positive (semi) definite if and only if the matrix $D - CB^{-1}C^T$ is positive (semi) definite. If $B \succeq 0$, then, A is positive semi-definite if and only if the matrix $D - CB^{-1}C^T$ is positive semi-definite and $(I - BB^{-1})C = 0$.

According to Lemma 1, (53) is equivalent to

$$\begin{bmatrix} \mathbf{G}(\boldsymbol{\mu}) & \mathbf{F}(\boldsymbol{\zeta}) \\ \mathbf{F}^T(\boldsymbol{\zeta}) & 2g \end{bmatrix} \succeq 0.$$

Thus, the canonical dual problem (\mathcal{P}^d) can be further relaxed to the following standard semi-definite problem (SDP):

$$\begin{aligned} \min \quad & g + \boldsymbol{\sigma}^T \mathbf{b} + \boldsymbol{\tau}^T \mathbf{e}_n, \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{G}(\boldsymbol{\mu}) & \mathbf{F}(\boldsymbol{\zeta}) \\ \mathbf{F}^T(\boldsymbol{\zeta}) & 2g \end{bmatrix} \succeq 0, \quad \mathbf{G}(\boldsymbol{\mu}) \succeq 0, \\ & \boldsymbol{\zeta} \in \mathcal{E}^*, \quad \boldsymbol{\sigma} \geq 0, \quad \boldsymbol{\mu} > 0. \end{aligned}$$

Although the SDP relaxation can be used theoretically to solve the canonical dual problem for the case that Π^d has critical points on the boundary $\partial \mathcal{S}_a^+$, in practice, the matrix $\mathbf{G}(\boldsymbol{\mu})$ will be ill-conditioning when the dual solution approaches to $\partial \mathcal{S}_a^+$. In order to solve this type of challenging problems, a canonical perturbation method has been suggested [16, 32]. Let

$$\begin{aligned} \mathcal{E}_{\delta_k}(\mathbf{y}, \boldsymbol{\zeta}) &= \mathcal{E}(\mathbf{y}, \boldsymbol{\zeta}) + \frac{\delta_k}{2} \|\mathbf{y} - \mathbf{y}_k\|^2 \\ &= \frac{1}{2} \mathbf{y}^T \mathbf{G}_{\delta_k}(\boldsymbol{\mu}) \mathbf{y} - \mathbf{F}_{\delta_k}^T(\boldsymbol{\zeta}) \mathbf{y} - \boldsymbol{\sigma}^T \mathbf{b} - \boldsymbol{\tau}^T \mathbf{e}_n + \frac{\delta_k}{2} \mathbf{y}_k^T \mathbf{y}_k, \end{aligned}$$

where, $\{\delta_k\}$ is a bounded sequence of positive real numbers, $\{\mathbf{y}_k\} \in \mathbb{R}^K$ is a set of given vectors, $\mathbf{G}_{\delta_k}(\boldsymbol{\mu}) = \mathbf{G}(\boldsymbol{\mu}) + \delta_k I$, $\mathbf{F}_{\delta_k}^T(\boldsymbol{\zeta}) = \mathbf{F}^T(\boldsymbol{\zeta}) + \delta_k \mathbf{y}_k$. Let

$$\mathcal{S}_{\delta_k}^+ = \{\boldsymbol{\zeta} \in \mathcal{S}_a : \mathbf{G}_{\delta_k}(\boldsymbol{\mu}) \succeq 0\}.$$

Clearly, we have $\mathcal{S}_a^+ \subset \mathcal{S}_{\delta_k}^+$. Therefore, the perturbed canonical dual problem can be expressed as

$$\begin{aligned} (\mathcal{P}_{\delta_k}^d) \quad \max \quad & \Pi_{\delta_k}^d(\boldsymbol{\zeta}) = -\frac{1}{2} \mathbf{F}_{\delta_k}^T(\boldsymbol{\zeta}) \mathbf{G}_{\delta_k}^\dagger(\boldsymbol{\mu}) \mathbf{F}_{\delta_k}(\boldsymbol{\zeta}) - \boldsymbol{\sigma}^T \mathbf{b} - \boldsymbol{\tau}^T \mathbf{e}_n, \\ \text{s.t.} \quad & \boldsymbol{\zeta} \in \mathcal{S}_{\delta_k}^+. \end{aligned}$$

Based on this perturbed problem, the following canonical primal-dual algorithm can be proposed for solving the nonconvex problem (\mathcal{P}).

Algorithm 1 (Canonical Primal-Dual Algorithm)

Given initial data $\delta_0 > 0$, $\mathbf{y}_0 \in \mathbb{R}^K$, and error allowance $\varepsilon > 0$, let $k = 0$.

1. Solve the perturbed canonical dual problem ($\mathcal{P}_{\delta_k}^d$) to obtain $\boldsymbol{\varsigma}_k \in \mathcal{S}_{\delta_k}^+$.
2. Compute $\tilde{\mathbf{y}}_{k+1} = [\mathbf{G}_{\delta_k}(\boldsymbol{\varsigma}_k)]^\dagger \mathbf{F}_{\delta_k}(\boldsymbol{\varsigma}_k)$ and let $\mathbf{y}_{k+1} = \mathbf{y}_k + \beta_k(\tilde{\mathbf{y}}_{k+1} - \mathbf{y}_k)$, $\beta_k \in [0, 1]$.
3. If $|P(\mathbf{y}_{k+1}) - P(\mathbf{y}_k)| \leq \varepsilon$, then stop, \mathbf{y}_{k+1} is the optimal solution. Otherwise, let $k = k + 1$, go back to step 1.

In this algorithm, $\{\beta_k\} \in [0, 1]$ are given parameters, which change the search directions. Clearly, if $\beta_k = 1$, we have $\mathbf{y}_{k+1} = \tilde{\mathbf{y}}_{k+1}$.

The key step in this algorithm is to solve the perturbed canonical dual problem ($\mathcal{P}_{\delta_k}^d$), which is equivalent to

$$\begin{aligned} \min \quad & g + \boldsymbol{\sigma}^T \mathbf{b} + \boldsymbol{\tau}^T \mathbf{e}_n, \\ \text{s.t.} \quad & \begin{bmatrix} \mathbf{G}(\boldsymbol{\mu}) + \delta_k I & \mathbf{F}(\boldsymbol{\varsigma}) + \delta_k \mathbf{y}_k \\ \mathbf{F}^T(\boldsymbol{\varsigma}) + \delta_k \mathbf{y}_k & 2g \end{bmatrix} \succeq 0, \\ & \mathbf{G}(\boldsymbol{\mu}) \succeq 0, \\ & \boldsymbol{\varsigma} \in \mathcal{S}, \quad \boldsymbol{\sigma} \geq 0, \quad \boldsymbol{\mu} > 0. \end{aligned}$$

This problem can be solved by a well-known software package named SeDuMi [36].

6 Numerical Experience

All data and computational results presented in this section are produced by Matlab. In order to save space and fit the matrix in the paper, we round our these results up to two decimals.

Example 1. 5-dimensional problem.

Consider Problem (\mathcal{P}_a) with $\mathbf{x}=[x_1, \dots, x_5]^T$, while $x_i \in \{2, 3, 5\}, i=1, \dots, 5$,

$$Q = \begin{bmatrix} 3.43 & 0.60 & 0.39 & 0.10 & 0.60 \\ 0.60 & 2.76 & 0.32 & 0.65 & 0.49 \\ 0.39 & 0.32 & 2.07 & 0.59 & 0.39 \\ 0.10 & 0.65 & 0.59 & 2.62 & 0.30 \\ 0.60 & 0.49 & 0.39 & 0.30 & 3.34 \end{bmatrix},$$

$$\mathbf{c} = [38.97, -24.17, 40.39, -9.65, 13.20]^T,$$

$$\mathbf{A} = \begin{bmatrix} 0.94 & 0.23 & 0.04 & 0.65 & 0.74 \\ 0.96 & 0.35 & 0.17 & 0.45 & 0.19 \\ 0.58 & 0.82 & 0.65 & 0.55 & 0.69 \\ 0.06 & 0.02 & 0.73 & 0.30 & 0.18 \end{bmatrix},$$

$$\mathbf{b} = [11.49, 9.32, 14.43, 5.66]^T.$$

Under the transformation (3), this problem is transformed into the 0–1 programming Problem (\mathcal{P}), where

$$\mathbf{y} = [y_{1,1}, y_{1,2}, y_{1,3}, \dots, y_{5,1}, y_{5,1}, y_{5,3}]^T \in \mathbb{R}^{15},$$

$$B = \begin{bmatrix} 13.71 & 20.56 & 34.27 & 2.40 & 3.61 & 6.01 & 1.58 & 2.37 & 3.95 & 0.39 & 0.58 & 0.97 & 2.38 & 3.57 & 5.95 \\ 20.56 & 30.84 & 51.41 & 3.61 & 5.41 & 9.01 & 2.37 & 3.55 & 5.92 & 0.58 & 0.88 & 1.46 & 3.57 & 5.36 & 8.93 \\ 34.27 & 51.41 & 85.68 & 6.01 & 9.01 & 15.02 & 3.95 & 5.92 & 9.87 & 0.97 & 1.46 & 2.43 & 5.95 & 8.93 & 14.88 \\ 2.40 & 3.61 & 6.01 & 11.05 & 16.57 & 27.61 & 1.27 & 1.91 & 3.18 & 2.61 & 3.91 & 6.52 & 1.95 & 2.93 & 4.88 \\ 3.61 & 5.41 & 9.01 & 16.57 & 24.85 & 41.42 & 1.91 & 2.86 & 4.77 & 3.91 & 5.87 & 9.78 & 2.93 & 4.39 & 7.32 \\ 6.01 & 9.01 & 15.02 & 27.61 & 41.42 & 69.03 & 3.18 & 4.77 & 7.96 & 6.52 & 9.78 & 16.31 & 4.88 & 7.32 & 12.20 \\ 1.58 & 2.37 & 3.95 & 1.27 & 1.91 & 3.18 & 8.27 & 12.40 & 20.67 & 2.37 & 3.55 & 5.92 & 1.57 & 2.36 & 3.93 \\ 2.37 & 3.55 & 5.92 & 1.91 & 2.86 & 4.77 & 12.40 & 18.60 & 31.00 & 3.55 & 5.33 & 8.89 & 2.36 & 3.53 & 5.90 \\ 3.95 & 5.92 & 9.87 & 3.18 & 4.77 & 7.96 & 20.67 & 31.00 & 51.67 & 5.92 & 8.86 & 14.81 & 3.93 & 5.90 & 9.83 \\ 0.39 & 5.58 & 0.97 & 2.61 & 3.91 & 6.52 & 2.37 & 3.55 & 5.92 & 10.50 & 15.74 & 26.24 & 1.20 & 1.80 & 3.00 \\ 0.58 & 0.88 & 1.46 & 3.91 & 5.87 & 9.78 & 3.55 & 5.33 & 8.89 & 15.74 & 23.62 & 39.36 & 1.80 & 2.70 & 4.50 \\ 0.97 & 1.46 & 2.43 & 6.52 & 9.78 & 16.31 & 5.92 & 8.89 & 14.81 & 26.24 & 39.36 & 65.60 & 3.00 & 4.50 & 7.51 \\ 2.38 & 3.57 & 5.95 & 1.95 & 2.93 & 4.88 & 1.57 & 2.36 & 3.93 & 1.20 & 1.80 & 3.00 & 13.35 & 20.02 & 33.37 \\ 3.57 & 5.36 & 8.93 & 2.93 & 4.39 & 7.32 & 2.36 & 3.54 & 5.90 & 1.80 & 2.70 & 4.50 & 20.02 & 30.04 & 50.06 \\ 5.95 & 8.93 & 14.88 & 4.88 & 7.32 & 12.20 & 3.93 & 5.90 & 9.83 & 3.00 & 4.50 & 7.51 & 33.37 & 50.06 & 83.43 \end{bmatrix},$$

$$\mathbf{h} = [77.95, 116.92, 194.87, -48.34, -72.51, -120.85, 80.78, 121.17, 201.96, -19.29, -28.94, -48.23, 26.39, 39.59, 65.99]^T,$$

$$D = \begin{bmatrix} 1.88 & 2.83 & 4.71 & 0.47 & 0.70 & 1.17 & 0.09 & 0.12 & 0.22 & 1.30 & 1.94 & 3.24 & 1.49 & 2.23 & 3.72 \\ 1.91 & 2.87 & 4.78 & 0.71 & 1.06 & 1.77 & 0.34 & 0.51 & 0.85 & 0.90 & 1.35 & 2.25 & 0.38 & 0.57 & 0.94 \\ 1.15 & 1.72 & 2.88 & 1.64 & 2.46 & 4.11 & 1.30 & 1.95 & 3.25 & 1.09 & 1.64 & 2.74 & 1.37 & 2.06 & 3.43 \\ 0.12 & 0.18 & 0.30 & 0.03 & 0.05 & 0.08 & 1.46 & 2.20 & 3.66 & 0.59 & 0.89 & 1.48 & 0.37 & 0.55 & 0.92 \end{bmatrix},$$

$$H = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{5 \times 15}.$$

The canonical dual problem can be stated as follows:

$$(\mathcal{P}^d) \text{Maximize } \Pi^d(\boldsymbol{\zeta}) = -\frac{1}{2} \mathbf{F}(\boldsymbol{\zeta})^T \mathbf{G}^\dagger(\boldsymbol{\mu}) \mathbf{F}(\boldsymbol{\zeta}) - \boldsymbol{\sigma}^T \mathbf{b} - \boldsymbol{\tau}^T \mathbf{e}_5$$

$$\text{subject to } \boldsymbol{\zeta} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathbb{R}^{4+5+15}, \quad \boldsymbol{\sigma} \geq 0, \boldsymbol{\mu} > 0.$$

By solving this dual problem with the sequential quadratic programming method in the optimization Toolbox within the Matlab environment, we obtain

$$\bar{\sigma} = [0, 0, 0, 0]^T,$$

$$\bar{\tau} = [73.90, -106.70, 111.95, -59.27, -0.01]^T,$$

and

$$\bar{\mu} = [39.34, 22.07, 12.49, 33.56, 3.01, 76.14, 61.00, 35.52, 18.78, 1.47, 41.96, 0.001, 0.001, 0.006]^T.$$

It is clear that $\bar{\zeta} = (\bar{\sigma}, \bar{\tau}, \bar{\mu}) \in \mathcal{S}_a^+$. Thus, from Theorem 4,

$$\begin{aligned} \bar{y} &= (B + 2\text{Diag}(\bar{\mu}))^\dagger (\mathbf{h} - D^T \bar{\sigma} - H^T \bar{\tau} + \bar{\mu}) \\ &= [0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0]^T \end{aligned}$$

is the global minimizer of Problem (\mathcal{P}) with $\Pi^d(\bar{\zeta}) = -227.87 = \Pi(\bar{y})$. The solution to the original primal problem can be calculated by using the transformation

$$\bar{x}_i = \sum_{j=1}^{K_i} u_{i,j} \bar{y}_{i,j}, \quad i = 1, 2, 3, 4, 5,$$

to give

$$\bar{\mathbf{x}} = [5, 2, 5, 2, 2]^T$$

with $P(\bar{\mathbf{x}}) = -227.87$.

Example 2. 10-dimensional problem. Consider Problem (\mathcal{P}_a), with $\mathbf{x} = [x_1, \dots, x_{10}]^T$, while $x_i \in \{1, 2, 4, 7, 9\}$, $i = 1, \dots, 10$,

$$Q = \begin{bmatrix} 6.17 & 0.62 & 0.46 & 0.37 & 0.56 & 0.66 & 0.67 & 0.85 & 0.57 & 0.44 \\ 0.62 & 5.63 & 0.29 & 0.56 & 0.79 & 0.29 & 0.43 & 0.69 & 0.49 & 0.39 \\ 0.46 & 0.29 & 5.81 & 0.55 & 0.22 & 0.55 & 0.36 & 0.27 & 0.51 & 0.91 \\ 0.37 & 0.56 & 0.55 & 6.10 & 0.28 & 0.42 & 0.44 & 0.34 & 0.75 & 0.44 \\ 0.56 & 0.79 & 0.22 & 0.28 & 4.75 & 0.40 & 0.55 & 0.42 & 0.49 & 0.44 \\ 0.66 & 0.29 & 0.55 & 0.42 & 0.40 & 5.71 & 0.32 & 0.57 & 0.65 & 0.70 \\ 0.67 & 0.43 & 0.36 & 0.44 & 0.55 & 0.32 & 5.27 & 0.56 & 0.37 & 0.85 \\ 0.85 & 0.69 & 0.27 & 0.34 & 0.42 & 0.57 & 0.56 & 5.91 & 0.15 & 0.62 \\ 0.57 & 0.49 & 0.51 & 0.75 & 0.49 & 0.65 & 0.37 & 0.15 & 4.51 & 0.46 \\ 0.44 & 0.39 & 0.91 & 0.44 & 0.44 & 0.70 & 0.85 & 0.62 & 0.46 & 5.73 \end{bmatrix},$$

Table 1 Numerical results for large scale integer programming problems

n	m	CPU Time (s)
20	5	1.77
50	5	6.23
100	5	26.05
200	5	136.29
300	5	408.59

$a_{ij} \in (0, 1)$, for $i = 1, \dots, n; j = 1, \dots, n$, and $c_i \in (0, 1), x_i \in \{1, 2, 3, 4, 5\}$, for $i = 1, \dots, n$. Without loss of generality, we ensure that the constructed Q is a symmetric matrix. Otherwise, we let $Q = \frac{Q+Q^T}{2}$. Furthermore, let Q be diagonally dominated. For each x_i , its lower bound is $l_i = 1$, and its upper bound is $u_i = 5$. Let $l = [l_1, \dots, l_n]^T$ and $u = [u_1, \dots, u_n]^T$. The right-hand sides of the linear constraints are chosen such that the feasibility of the test problem is satisfied. More specifically, we set $\mathbf{b} = \sum_j a_{ij}l_j + 0.5 \cdot (\sum_j a_{ij}u_j - \sum_j a_{ij}l_j)$.

We then construct the canonical problem of each of the five problems. It is solved by using the sequential quadratic programming method with active set strategy from the Optimization Toolbox within the Matlab environment. The specifications of the personal notebook computer used are: Window 7 Enterprise, Intel(R), Core(TM)(2.50 GHZ). Table 1 presents the numerical results, where m is number of linear constraints in Problem $I(\mathcal{P}_a)$.

From Table 1, we see that the algorithm based on the canonical dual method can solve large scale problems with reasonable computational time. Furthermore, for each of the five problems, the solution obtained is a global optimal solution. For the case of $n = 300$, the equivalent problem in the form of Problem (\mathcal{P}_b) has 1500 variables. For such a problem, there are 2^{1500} possible combinations.

7 Conclusion

We have presented a canonical duality approach for solving a general quadratic discrete value selection problem with linear constraints. Our results show that this NP-hard problem can be converted to a continuous concave dual maximization problem over a convex space without duality gap. For certain given data, if this canonical dual has a KKT point in the dual feasible space \mathcal{S}_a^+ , the problem can be solved easily by well-developed convex optimization methods. Otherwise, a canonical perturbation method is proposed, which can be used to deal with challenging cases when the primal problem has multiple global minimizers. Several examples, including some relatively large scale ones, were solved effectively by using the method proposed.

Remanning open problems include how to solve the canonical dual problem (\mathcal{P}^d) more efficiently instead of using the SDP approximation. Also, for the given data $Q, \mathbf{c}, \mathbf{A}, \mathbf{b}$, the existence condition for the canonical dual problem having KKT point

in \mathcal{S}_a^+ is fundamentally important for understanding NP-hard problems. If the canonical dual (\mathcal{P}^d) has no KKT point in the closed set $\mathcal{S}_c^+ = \mathcal{S}_a^+ \cup \partial\mathcal{S}_a^+$, the primal problem is equivalent to the following canonical dual problem (see Eq. (67) in [16])

$$\min \text{sta}\{\Pi^d(\boldsymbol{\zeta}) \mid \boldsymbol{\zeta} \in \mathcal{S}_a\}, \quad (56)$$

i.e., to find the minimal stationary value of Π^d on \mathcal{S}_a . Since the feasible set \mathcal{S}_a is nonconvex, to solve this canonical dual problem is very difficult. Therefore, it is a conjecture that the primal problem (\mathcal{P}) could be NP-hard if its canonical dual (\mathcal{P}^d) has no KKT point in the closed set \mathcal{S}_a^+ [12]. In this case, one alternative approach for solving (\mathcal{P}) is the canonical dual relaxation (\mathcal{P}^\sharp). Although the relaxed problem (\mathcal{P}^\sharp) is convex, by Remark 2 we know that there exists a duality gap between the primal problem (\mathcal{P}). It turns out that the associated SDP method provides only a lower bound approach for solving the primal problem. Further researches are needed to know how big is this duality gap, how much does this relaxation lose, and how to solve the nonconvex canonical dual problem (56).

Acknowledgements The research is supported by US Air Force Office of Scientific Research under grants FA2386-16-1-4082 and FA9550-17-1-0151.

References

1. Arora, J.S., Huang, M.W., Hsieh, C.C.: Methods for optimization of nonlinear problems with discrete variables: a review. *Struct. Optim.* **8**, 69–85 (1994)
2. Balas, E., Ceria, S., Cornuéjols, G.: A lift-and-project cutting plane algorithm for mixed 0–1 programs. *Math. Program.* **58**, 295–324 (1993)
3. Belmiloudi, A.: *Stabilization, Optimal and Robust Control: Theory and Applications in Biological and Physical Science*. Springer, London (2008)
4. Bertsimas, D., Shioda, R.: Classification and regression via integer optimization. *Oper. Res.* **55**(2), 252–271 (2007)
5. Chen, D.S., Batson, R.G., Dang, Y.: *Applied Integer Programming: Modeling and Solution*. Wiley, New Jersey (2010)
6. Ciarlet, P.G.: *Linear and Nonlinear Functional Analysis with Applications*. SIAM, Philadelphia (2013)
7. Marsden, J.E., Hughes, T.J.R.: *Mathematical Foundations of Elasticity*. Prentice-Hall, Upper Saddle River (1983)
8. Fang, S.C., Gao, D.Y., Sheu, R.L., Wu, S.Y.: Canonical dual approach to solving 0–1 quadratic programming problems. *J. Ind. Manag. Optim.* **4**(4), 125–142 (2008)
9. Floudas, C.A.: *Nonlinear and Mixed-Integer Optimization: Theory Methods and Applications*. Kluwer Academic Publishers, Dordrecht (2000)
10. Gao, D.Y.: *Duality Principles in Nonconvex Systems: Theory, Methods and Applications*, p. 454. Kluwer Academic Publishers, Dordrecht (2000)
11. Gao, D.Y.: Nonconvex semi-linear problems and canonical duality solutions, In: *Advances in Mechanics and Mathematics*, vol. II, pp. 261–312. Kluwer Academic Publishers, Dordrecht (2003)
12. Gao, D.Y.: Solutions and optimality criteria to box constrained nonconvex minimization problem. *J. Ind. Manag. Optim.* **3**(2), 293–304 (2007)

13. Gao, D.Y.: Canonical duality theory: unified understanding and generalized solution for global optimization problems. *Comput. Chem. Eng.* **33**, 1964–1972 (2009)
14. Gao, D.Y., Ogden, R.W.: Multiple solutions to non-convex variational problems with implications for phase transitions and numerical computation. *Q. J. Mech. Appl. Math* **61**(4), 497–522 (2008)
15. Gao, D.Y., Ruan, N.: Solutions and optimality criteria for nonconvex quadratic-exponential minimization problem. *Math. Method. Oper. Res.* **67**, 479–496 (2008)
16. Gao, D.Y., Ruan, N.: Solutions to quadratic minimization problems with box and integer constraints. *J. Glob. Optim.* **47**, 463–484 (2010)
17. 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: Boginski, V.L., Commander, C.W., Pardalos, P.M., Ye, Y.Y. (eds.) *Sensors: Theory, Algorithms and Applications*, vol. 61, pp. 37–54. Springer, Berlin (2012)
18. Gao, D.Y., Ruan, N., Sherali, H.D.: Solutions and optimality criteria for nonconvex constrained global optimization problems. *J. Glob. Optim.* **45**, 473–497 (2009)
19. Gao, D.Y., Sherali, H.D.: Canonical duality: connection between nonconvex mechanics and global optimization. In: *Advances in Application Mathematics and Global Optimization*, pp. 249–316. Springer, Berlin (2009)
20. Gao, D.Y., Watson, L.T., Easterling, D.R., Thacker, W.I., Billups, S.C.: Solving the canonical dual of box- and integer-constrained nonconvex quadratic programs via a deterministic direct search algorithm. *Optim. Methods Softw.* **28**(2), 313–326 (2013)
21. Ghaddar, B., Vera, J.C., Anjos, M.: Second-order cone relaxations for binary quadratic polynomial programs. *SIAM J. Optim.* **21**, 391–414 (2011)
22. Holland, J.H.: *Adaptation in Natural and Artificial System*. The University of Michigan Press, Ann, Arbor, MI (1975)
23. Huang, M.W., Arora, J.S.: Optimal design with discrete variables: Some numerical experiments. *INT. J. Numer. Meth. Eng.* **40**, 165–188 (1997)
24. Karlof, J.K.: *Integer Programming: Theory and Practice*. CRC Press, Boca Raton (2006)
25. Kincaid, R.K., Padula, S.L.: Minimizing distortion and internal forces in truss structures by simulated annealing. In: *Proceedings of the 31st AIAA SDM Conference*, Long Beach, CA, pp. 327–333 (1990)
26. Latorrie, V., Gao, D.Y.: Canonical duality for solving general nonconvex constrained problems. *Optim. Lett.* **10**(8), 1–17 (2015)
27. Li, S.F., Gupta, A.: On dual configuration forces. *J. Elast.* **84**, 13–31 (2006)
28. Loh, H.T., Papalambros, P.Y.: Computational implementation and tests of a sequential linearization approach for solving mixed-discrete nonlinear design optimization. *J. Mech. Des. ASME* **113**, 335–345 (1991)
29. Marsten, R.E., Morin, T.L.: A hybrid approach to discrete mathematical programming. *Math. Program.* **14**, 21–40 (1978)
30. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
31. Ruan, N., Gao, D.Y.: Global optimal solutions to a general sensor network localization problem. *Perform. Eval.* **75–76**, 1–16 (2014)
32. Ruan, N., Gao, D.Y., Jiao, Y.: Canonical dual least square method for solving general nonlinear systems of equations. *Comput. Optim. Appl.* **47**, 335–347 (2010)
33. Sandgren, E.: Nonlinear integer and discrete programming in mechanical design optimization. *J. Mech. Des. ASME* **112**, 223–229 (1990)
34. Santos, H.A.F.A., Moitinho De Almeida, J.P.: Dual extremum principles for geometrically exact finite strain beams. *Int/ J. Nonlinear Mech.* Elsevier **46**(1), 151 (2010)
35. Sherali, H.D., Smith, J.C.: An improved linearization strategy for zero-one quadratic programming problems. *Optim. Lett.* **1**(1), 33–47 (2007)
36. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cone. *Optim. Meth. Softw.* **11**, 625–653 (1999)
37. Thanedar, P.B., Vanderplaats, G.N.: A survey of discrete variable optimization for structural design. *J. Struct. Eng. ASCE* **121**, 301–306 (1994)

38. Wang, S., Teo, K.L., Lee, H.W.J.: A new approach to nonlinear mixed discrete programming problems. *Eng. Optim.* **30**(3), 249–262 (1998)
39. Wang, Z.B., Fang, S.C., Gao, D.Y., Xing, W.X.: Global extremal conditions for multi-integer quadratic programming. *J. Ind. Manag. Optim.* **4**(2), 213–225 (2008)
40. 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**(2), 341–351 (2012)
41. Zecchin, A.C., Simpson, A.R., Maier, H.R., Nixon, J.B.: Parametric study for an ant algorithm applied to water distribution system optimization. *IEEE Trans. Evol. Comput.* **9**(2), 175–191 (2005)