Department of Mathematics and Statistics

# Global Optimization for Nonconvex Optimization Problems

Ning Ruan

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

August 2012

# Declaration

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgment has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

...........................................................

Ning Ruan

August 2012

# Abstract

Duality is one of the most successful ideas in modern science [46] [91]. It is essential in natural phenomena, particularly, in physics and mathematics [39] [94] [96]. In this thesis, we consider the canonical duality theory for several classes of optimization problems.

The first problem that we consider is a general sum of fourth-order polynomial minimization problem. This problem arises extensively in engineering and science, including database analysis, computational biology, sensor network communications, nonconvex mechanics, and ecology. We first show that this global optimization problem is actually equivalent to a discretized minimal potential variational problem in large deformation mechanics. Therefore, a general analytical solution is proposed by using the canonical duality theory.

The second problem that we consider is a nonconvex quadratic-exponential optimization problem. By using the canonical duality theory, the nonconvex primal problem in n-dimensional space can be converted into a one-dimensional canonical dual problem, which is either a concave maximization or a convex minimization problem with zero duality gap. Several examples are solved so as to illustrate the applicability of the theory developed.

The third problem that we consider is quadratic minimization problems subjected to either box or integer constraints. Results show that these nonconvex problems can be converted into concave maximization dual problems over convex feasible spaces without duality gap and the Boolean integer programming problem is actually equivalent to a critical point problem in continuous space. These dual problems can be solved under certain conditions. Both existence and uniqueness of the canonical dual solutions are presented. A canonical duality algorithm is presented and applications are illustrated.

The fourth problem that we consider is a quadratic discrete value selection problem subjected to inequality constraints. The problem is first transformed into a quadratic 0-1 integer programming problem. The dual problem is thus constructed by using the canonical duality theory. Under appropriate conditions, this dual problem is a maximization problem of a concave function over a convex continuous space. Theoretical results show that the canonical duality theory can either provide a global optimization solution, or an optimal lower bound approximation to this NP-hard problem. Numerical simulation studies, including some relatively large scale problems, are carried out so as to demon-

strate the effectiveness and efficiency of the canonical duality method. An open problem for understanding NP-hard problems is proposed.

The fifth problem that we consider is a mixed-integer quadratic minimization problem with fixed cost terms. We show that this well-known NP-hard problem in $\mathbb{R}^{2n}$ can be transformed into a continuous concave maximization dual problem over a convex feasible subset of $\mathbb{R}^n$ with zero duality gap. We also discuss connections between the proposed canonical duality theory approach and the classical Lagrangian duality approach. The resulting canonical dual problem can be solved under certain conditions, by traditional convex programming methods. Conditions for the existence and uniqueness of global optimal solutions are presented. An application to a decoupled mixed-integer problem is used to illustrate the derivation of analytic solutions for globally minimizing the objective function. Numerical examples for both decoupled and general mixed-integral problems are presented, and an open problem is proposed for future study.

The sixth problem that we consider is a general nonconvex quadratic minimization problem with nonconvex constraints. By using the canonical dual transformation, the nonconvex primal problem can be converted into a canonical dual problem (i.e., either a concave maximization problem with zero duality gap). Illustrative applications to quadratic minimization with multiple quadratic constraints, box/integer constraints, and general nonconvex polynomial constraints are discussed, along with insightful connections to classical Lagrangian duality. Conditions for ensuring the existence and uniqueness of global optimal solutions are presented. Several numerical examples are solved.

The seventh problem that we consider is a general nonlinear algebraic system. By using the least square method, the nonlinear system of m quadratic equations in $n$-dimensional space is first formulated as a nonconvex optimization problem. We then prove that, by using the canonical duality theory, this nonconvex problem is equivalent to a concave maximization problem in $\mathbb{R}^m$, which can be solved by well-developed convex optimization techniques. Both existence and uniqueness of global optimal solutions are discussed, and several illustrative examples are presented.

The eighth problem that we consider is a general sensor network localization problem. It is shown that by the canonical duality theory, this nonconvex minimization problem is equivalent to a concave maximization problem over a convex set in a symmetrical matrix space, and hence can be solved by combining a perturbation technique with existing optimization techniques. Applications are illustrated and results show that the proposed method is potentially a powerful one for large-scale sensor network localization problems.

# List of publications

The following papers (which have been published or accepted for publication) were completed during PhD candidature:

- N. Ruan and D. Y. Gao, "Canonical duality approach for nonlinear dynamical systems". *IMA Journal of Applied Mathematics*, 2013, accepted.
- N. Ruan and D. Y. Gao, "Canonical duality theory and algorithm for solving challenging problems in network optimisation", *Neural Information Processing*, 2012, T. W. Huang, Z. G. Zeng, C. D. Li and C. S. Leung (Eds.), ICONIP 2012, Part III, LNCS 7665, 702-709.
- N. Ruan, and D. Y. Gao, "Global optimal solutions to nonconvex Euclidean distance geometry problems". in *Proceedings of 20th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2012)*, University of Melbourne, Australia, 2012.
- D. Y. Gao, N. Ruan, and P. M. Pardalos, "Canonical dual solutions to sum of fourth-order polynomials minimization problems with applications to sensor network localization," Boginski, V. L., Commander, C.W., Pardalos, P.M. and Ye, Y., (Eds.) in *Sensors: Theory, Algorithms, and Applications*. Springer, Vol. 61, pp. 37-54, 2012.

The following papers were completed during the PhD candidature and are currently under review:

- N. Ruan and D. Y. Gao, "Global Optimal Solution to Quadratic Optimization Problem with Discrete Variables and Inequality Constraints". Submit to *Journal of Global Optimization*.
- N. Ruan, D. Y. Gao, and K. L. Teo, "Global optimal solutions to sensor network localization problem". Submit to *Performance Evaluation*.
- N. Ruan, K. L. Teo and D. Y. Gao,"The canonical dual method for the design of FIR filters with signed powers-of-two coefficients", to appear.

# Acknowledgements

The research reported in this thesis was carried out when I was a PhD student in the Department of Mathematics and Statistics, Curtin University. I am very grateful for the help I received from friends, family, teachers, and colleagues during this time. Without it, this thesis could not be possible.

I am especially indebted to my supervisors, Prof. Kok Lay Teo and Prof. David Gao. They have guided my research with remarkable patience and enthusiasm. Prof. Teo, in particular, deserves special thanks for helping me to apply for Curtin Strategic International Research Scholarship (CSIRS) (a PhD scholarship), which I was fortunate enough to receive.

I also thank Prof. Yong Hong Wu, Postgraduate Coordinator in the Department of Mathematics and Statistics. Prof. Wu was the Chairman of my Thesis Committee and he helped me on many occasions.

I thank all of the staff in the Department of Mathematics and Statistics for contributing to a friendly work environment. The past and present administrative staff Joyce Yang, Shuie Liu, Florence Wong, Carey Ryken-Rapp, Lisa Holling, and Shally Wang—deserve special thanks for providing kind and professional help on numerous occasions. I am also grateful to Prof. Lou Caccetta.

I have really enjoyed working with the other students in the Department of Mathematics and Statistics, especially Lin Qun, Xu Honglei, Tiffany Jones, Li Bin, Kyle Chow, Jonathan Blanchard, Zhou Jingyang, and Sarah Peursum. I thank each of them for being a wonderful friend and colleague.

Finally, on a more personal note, I sincerely thank everyone in my family for their support and encouragement during my PhD candidature.

# Contents

# CHAPTER 1

# Introduction

## 1.1 Motivation and background

In this thesis, canonical duality theory which was developed from nonconvex analysis [10] [27] and global optimization [38], is applied to study several interesting nonconvex optimization problems. Some concepts and basic results on canonical duality theory are briefly reviewed in this chapter.

## 1.2 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)

$$(\mathcal{P}): \min_{\mathbf{x}\in\mathcal{X}_a} \left\{ P(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, \mathbf{A}\mathbf{x}\rangle - \langle \mathbf{x}, \mathbf{f}\rangle + W(\mathbf{x}) \right\}, \tag{1.1}$$

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, $\langle \mathbf{x}, \mathbf{x}^*\rangle$ denotes the bilinear form between $\mathbf{x}$ and its dual variable $\mathbf{x}^*$, $W(\mathbf{x})$ is a general nonconvex function, and $\mathcal{X}_a \subset \mathbb{R}^n$ is a given feasible space.

The **key step** in the canonical dual transformation is to choose a nonlinear operator,

$$\boldsymbol{\xi} = \Lambda(\mathbf{x}) : \mathcal{X}_a \to \mathcal{E}_a \subset \mathbb{R}^p \tag{1.2}$$

and a *canonical function* $V : \mathcal{E}_a \to \mathbb{R}$ such that the nonconvex functional $W(\mathbf{x})$ 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:

$$(\mathcal{P}): \min_{\mathbf{x}\in\mathcal{X}_a} \{P(\mathbf{x}) = V(\Lambda(\mathbf{x})) - U(\mathbf{x})\}, \tag{1.3}$$

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 [38], a differentiable function $V(\boldsymbol{\xi})$ is said to be a *canonical function* on its domain $\mathcal{E}_a$ if the duality mapping $\boldsymbol{\varsigma} = \nabla V(\boldsymbol{\xi})$ from $\mathcal{E}_a$ to its range $\mathcal{S}_a \subset \mathbb{R}^p$ is invertible. Let $\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle$ denote the bilinear form on $\mathbb{R}^p$. Thus, for the given canonical function $V(\boldsymbol{\xi})$, its Legendre conjugate $V^*(\boldsymbol{\varsigma})$ can be defined uniquely by the Legendre transformation

$$V^*(\boldsymbol{\varsigma}) = \text{sta}\{\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle - V(\boldsymbol{\xi}) \mid \ \boldsymbol{\xi} \in \mathcal{E}_a\}, \tag{1.4}$$

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

$$\boldsymbol{\varsigma} = \nabla V(\boldsymbol{\xi}) \ \Leftrightarrow \ \boldsymbol{\xi} = \nabla V^*(\boldsymbol{\varsigma}) \ \Leftrightarrow V(\boldsymbol{\xi}) + V^*(\boldsymbol{\varsigma}) = \langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle. \tag{1.5}$$

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 $\langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - V^*(\boldsymbol{\varsigma})$ such that the nonconvex function $P(\mathbf{x})$ is reformulated as the so-called Gao-Strang total complementary function [38]:

$$\Xi(\mathbf{x}, \boldsymbol{\varsigma}) = \langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - V^*(\boldsymbol{\varsigma}) - U(\mathbf{x}). \tag{1.6}$$

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

$$\begin{aligned} P^d(\boldsymbol{\varsigma}) &= \text{sta}\{\Xi(\mathbf{x}, \boldsymbol{\varsigma}) \mid \mathbf{x} \in \mathcal{X}_a\} \\ &= U^\Lambda(\boldsymbol{\varsigma}) - V^*(\boldsymbol{\varsigma}), \end{aligned} \tag{1.7}$$

where $U^\Lambda(\mathbf{x})$ is defined by

$$U^\Lambda(\boldsymbol{\varsigma}) = \text{sta}\{\langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - U(\mathbf{x}) \mid \ \mathbf{x} \in \mathcal{X}_a\}. \tag{1.8}$$

In many applications, the geometrical nonlinear operator $\Lambda(\mathbf{x})$ is usually a quadratic function

$$\Lambda(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, D_k \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{b}_k \rangle, \tag{1.9}$$

where $D_k \in \mathbb{R}^{n \times n}$ and $\mathbf{b}_k \in \mathbb{R}^n$, $k = 1, \cdots, p$. Let $\boldsymbol{\varsigma} = [\varsigma_1, \cdots, \varsigma_p]^T$. In this case, the canonical dual function can be written in the following form:

$$P^d(\boldsymbol{\varsigma}) = -\frac{1}{2}\langle \mathbf{F}(\boldsymbol{\varsigma}), \mathbf{G}^+(\boldsymbol{\varsigma}) \mathbf{F}(\boldsymbol{\varsigma}) \rangle - V^*(\boldsymbol{\varsigma}), \tag{1.10}$$

where $\mathbf{G}(\boldsymbol{\varsigma}) = \mathbf{A} + \sum_{k=1}^p \varsigma_k D_k$, $\mathbf{F}(\boldsymbol{\varsigma}) = \mathbf{f} - \sum_{k=1}^p \varsigma_k \mathbf{b}_k$, and $\mathbf{G}^+$ denotes the Moore-Penrose generalized inverse of $\mathbf{G}$,

Let $\mathcal{S}_a^+ = \{\boldsymbol{\varsigma} \in \mathbb{R}^p | \, G(\boldsymbol{\varsigma}) \succeq 0\}$. Therefore, the canonical dual problem is proposed as:

$$(\mathcal{P}^d): \quad \max\{P^d(\boldsymbol{\varsigma}) | \ \boldsymbol{\varsigma} \in \mathcal{S}_a^+\}, \tag{1.11}$$

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

**Theorem 1.1** ( [38]). *Problem $(\mathcal{P}^d)$ is canonically dual to $(\mathcal{P})$ in the sense that if $\bar{\boldsymbol{\varsigma}}$ is a critical point of $P^d(\boldsymbol{\varsigma})$, then*

$$\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\varsigma}})\mathbf{F}(\bar{\boldsymbol{\varsigma}}) \tag{1.12}$$

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

$$P(\bar{\mathbf{x}}) = \Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\varsigma}}) = P^d(\bar{\boldsymbol{\varsigma}}). \tag{1.13}$$

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

$$\min_{\mathbf{x} \in \mathcal{X}_a} P(\mathbf{x}) = \Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\varsigma}}) = \max_{\boldsymbol{\varsigma} \in \mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}). \tag{1.14}$$

*Conversely, if $\bar{\mathbf{x}}$ is a solution to $(\mathcal{P})$, it must be in the form of (1.12) for critical solution $\bar{\boldsymbol{\varsigma}}$ of $P^d(\boldsymbol{\varsigma})$.*

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

$$\min \Pi(\mathbf{x}) = \frac{1}{2}\alpha(\frac{1}{2}\|\mathbf{x}\|^2 - \lambda)^2 - \mathbf{x}^T\mathbf{f}, \ \forall \mathbf{x} \in \mathbb{R}^n, \tag{1.15}$$

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

$$\alpha(\frac{1}{2}\|\mathbf{x}\|^2 - \lambda)\mathbf{x} = \mathbf{f}. \tag{1.16}$$

Clearly, to solve this system of nonlinear algebraic equations directly is very difficult. Let's make use of the canonical dual transformation. To do so, we let $\boldsymbol{\xi} = \Lambda(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 - \lambda \in \mathbb{R}$. 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 $V(\boldsymbol{\xi}) = \frac{1}{2}\alpha\boldsymbol{\xi}^2$. Its Legendre conjugate is given by $V^*(\varsigma) = \frac{1}{2}\alpha^{-1}\varsigma^2$, which is strictly convex. Thus, the total complementary function for this nonconvex optimization problem is

$$\Xi(\mathbf{x}, \varsigma) = (\frac{1}{2}\|\mathbf{x}\|^2 - \lambda)\varsigma - \frac{1}{2}\alpha^{-1}\varsigma^2 - \mathbf{x}^T\mathbf{f}. \tag{1.17}$$

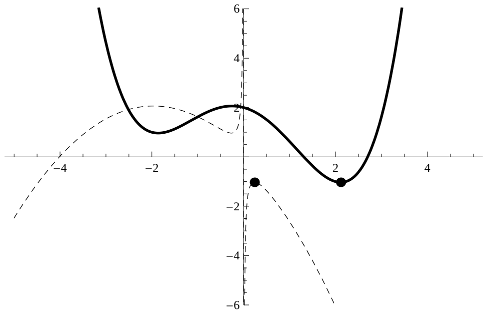



Figure 1.1: Graphs of the primal function $\Pi(\mathbf{x})$ (solid line) and its canonical dual function $\Pi^d(\varsigma)$ (dotted line).

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

$$\varsigma\mathbf{x} - \mathbf{f} = 0. \tag{1.18}$$

For each $\varsigma \neq 0$, equation (1.18) gives $\mathbf{x} = \mathbf{f}/\varsigma$ in vector form. Substituting this into the total complementary function $\Xi$, the canonical dual function can be easily obtained as

$$\begin{aligned} \Pi^d(\varsigma) &= \{\Xi(\mathbf{x},\varsigma)|\nabla_{\mathbf{x}}\Xi(\mathbf{x},\varsigma) = 0\} \\ &= -\frac{\mathbf{f}^T\mathbf{f}}{2\varsigma} - \frac{1}{2}\alpha^{-1}\varsigma^2 - \lambda\varsigma, \quad \forall\varsigma \neq 0. \end{aligned} \tag{1.19}$$

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

$$(\alpha^{-1}\varsigma + \lambda)\varsigma^2 = \frac{1}{2}\mathbf{f}^T\mathbf{f}. \tag{1.20}$$

For any given parameters $\alpha$, $\lambda$ and the vector $\mathbf{f} \in \mathbb{R}^n$, this cubic algebraic equation has at most three roots satisfying $\varsigma_1 \geq 0 \geq \varsigma_2 \geq \varsigma_3$, and each of these roots leads to a critical point of the nonconvex function $P(\mathbf{x})$, i.e., $\mathbf{x}_i = \mathbf{f}/\varsigma_i$, $i = 1, 2, 3$. By the fact that $\varsigma_1 \in \mathcal{S}_a^+ = \{\varsigma \in \mathbb{R} \mid \varsigma > 0\}$, it is clear from Theorem 1.1 that $\mathbf{x}_1$ is a global minimizer of $\Pi(\mathbf{x})$. Consider one dimensional problem with $\alpha = 1$, $\lambda = 2$, $f = \frac{1}{2}$. The primal function and canonical dual function are shown in Fig. 1.1, where $x_1 = 2.11491$ is the global minimizer of $P(\mathbf{x})$, $\varsigma_1 = 0.236417$ is global maximizer of $\Pi^d(\boldsymbol{\varsigma})$, and $\Pi(x_1) = -1.02951 = \Pi^d(\varsigma_1)$ (See the two black dots in Fig. 1.1).

## 1.3 Overview of this thesis

The primal goal of this thesis is to apply this newly developed canonical duality theory to eight nonconvex problems.

The rest of the thesis is organized as follows. Chapters 2 and 3 are concerned on canonical duality theory to unconstrained optimization problems. In Chapter 2, canonical duality theory is applied to study fourth-order polynomials minimization problems. Chapter 3 is on nonconvex quadratic exponential minimization problem. Chapters 4 to Chapter 7 are concerned on constrained problems. Chapter 4 is focused on the application of canonical duality theory to box and integer constrained problem. Chapter 5 is on general multi-integer constrained problem. Chapter 6 is on mixed integer programming. Chapter 7 is on general nonconvex constrained optimization. In Chapter 8, canonical duality theory is applied to algebraic system. Chapter 9 considers a real world application on sensor network localization. In Chapter 10, we summarize the contributions of the thesis and make comments on open problems. Some future research problems are also suggested.

## 1.4 Notation

The following is a list of notations used in this thesis.
$\mathbb{R}$: Real number.
$\mathbb{R}^n$: $n$-dimensional Euclidean space.
$\mathbb{R}^{n\times m}$: $n \times m$-dimensional real matrices space.
$\mathbb{Z}$: Integers.
$\mathbb{Z}^n$: $n$-dimensional integer space.
$|S|$: Cardinality of the set $\mathcal{S}$.
$\|\mathbf{x}\|$: Euclidean norm.
$\{x_i\}_{i=1}^n$: Column vector $(x_1, \cdots, x_n)^T$.
$\mathbf{e}$: Vector with all its components being 1.
$e_i$: Vector, where its $i$th component is 1, while the others are all 0.
$A^T$: Transpose of the matrix $A$.
$trace(A)$: Trace of the matrix $A$.
$rank(A)$: Rank of the matrix $A$.
$diag(A)$: Diagonal vector of the matrix $A$.
$diag(x)$: Diagonal matrix with diagonal elements $x_1, \cdots, x_n$.
$A \circ B$: Hadamard product of the matrices $A$ and $B$, i.e., $A \circ B = \{a_{ij}b_{ij}\}_{i,j=1}^n$.
$\langle A, B\rangle$: Inner product of the matrices $A$ and $B$, i.e., $\langle A, B\rangle = trace(AB)$.
$Q \succ 0$: $Q$ is a positive definite matrix.
$Q \succeq 0$: $Q$ is a positive semidefinite matrix.
$f^\star$: Conjugate function of $f$.
dom $f$: Domain of the function $f$.
$\delta f$: Derivative of the function $f$.
$\nabla f$: Gradient of the function $f$.

$\nabla^2 f$: Hessian of the function $f$.

$\partial f$: Set of subgradients of the function $f$.

$\exp(x)$: Exponential function $e^x$.

$sta\{\}$: finding stationary points of the statement in $\{\}$.

# CHAPTER 2

# Fourth-Order Polynomials Minimization Problems

## 2.1 Introduction

This chapter presents a canonical dual approach to solve a general sum of fourth-order polynomial minimization problem. This problem arises extensively in engineering and science, including chaotical dynamical systems [39], chemical database analysis [100], sensor network communications [17] [86], large deformation computational mechanics [35], and phase transitions of solids [45].

We first show that this global optimization problem is actually equivalent to a discretized minimal potential variational problem in large deformation mechanics. Therefore, a general analytical solution can be proposed by using the canonical duality theory.

## 2.2 Problem Statement

We are interested in solving the following general nonlinear programming problem

$$(\mathcal{P}): \quad \min \{P(\mathbf{x}) = \sum_{e=1}^{m} W_e(\mathbf{x}) + \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{x}^T\mathbf{f} \quad : \quad \mathbf{x} \in \mathbb{R}^n\}, \tag{2.1}$$

where

$$W_e(\mathbf{x}) = \frac{1}{2}\alpha_e \left(\frac{1}{2}\mathbf{x}^T \mathbf{A}_e \mathbf{x} + \mathbf{b}_e^T \mathbf{x} + c_e\right)^2, \tag{2.2}$$

and $\mathbf{A}_e = \mathbf{A}_e^T$, $Q = Q^T \in \mathbb{R}^{n\times n}$ are indefinite symmetrical matrixes, $\mathbf{f}$, $\mathbf{b}_e \in \mathbb{R}^n$ are given vectors, $c_e \in \mathbb{R}$ and $\alpha_e$ are given constants. Without loss of generality, we assume that $\alpha_e > 0, \quad \forall e = 1, \ldots, m$. The criticality condition $\delta P(\mathbf{x}) = 0$ leads to a nonlinear equilibrium equation:

$$\sum_{e=1}^{m} \alpha_e(\frac{1}{2}\mathbf{x}^T\mathbf{A}_e\mathbf{x} + \mathbf{b}_e^T\mathbf{x} + c_e)(\mathbf{A}_e\mathbf{x} + \mathbf{b}_e) + Q\mathbf{x} - \mathbf{f} = 0 \tag{2.3}$$

Direct methods for solving this coupled nonlinear algebraic system is very difficult. Also equation (2.3) is only a necessary condition for global minimizer of the problem ($\mathcal{P}$). A general sufficient condition for identifying the global minimizer is a fundamental task in global optimization.

## 2.3 Canonical Dual Transformation

Following the standard procedure of the canonical dual transformation, we introduce a differentiable *geometrical operator*

$$\boldsymbol{\xi} = \Lambda(\cdot) : \mathbb{R}^n \to \mathbb{R}^m, \tag{2.4}$$

which $\Lambda(\mathbf{x}) = \left\{\frac{1}{2}\mathbf{x}^T\mathbf{A}_k\mathbf{x} + \mathbf{b}_k^T\mathbf{x} + c_k\right\}$ is a map from $\mathbb{R}^n$ into $\mathcal{V}_a \subset \mathbb{R}^m$. Then, the nonconvex function $W(\mathbf{x})$ can be written in the canonical form

$$W(\mathbf{x}) = V(\Lambda(\mathbf{x})), \tag{2.5}$$

with

$$V(\boldsymbol{\xi}) = \sum_{k=1}^{m} \frac{1}{2}\alpha_k\xi_k^2 = \frac{1}{2}\boldsymbol{\alpha}^T(\boldsymbol{\xi} \circ \boldsymbol{\xi})$$

is a quadratic function, where $\boldsymbol{\alpha} = \{\alpha_k\} \in \mathbb{R}^m$, and $\boldsymbol{\xi} \circ \boldsymbol{\xi} = \{\xi_k\xi_k\} \in \mathbb{R}^m$ represents the Hadamard product. Thus, the duality relation

$$\boldsymbol{\varsigma} = \delta V(\boldsymbol{\xi}) = \boldsymbol{\alpha} \circ \boldsymbol{\xi} \tag{2.6}$$

is invertible for any given $\boldsymbol{\xi} \in \mathcal{V}_a$.

Let $\mathcal{V}_a^*$ be the range of the duality mapping $\boldsymbol{\varsigma} = \delta V(\boldsymbol{\xi}) : \mathcal{V}_a \to \mathcal{V}_a^* \subset \mathbb{R}^m$, i.e., $\boldsymbol{\varsigma} \in \mathbb{R}^m$. Then, for any given $\boldsymbol{\varsigma} \in \mathcal{V}_a^*$, the Legendre conjugate $V^*$ can be uniquely defined by

$$V^*(\boldsymbol{\varsigma}) = \text{sta}\{\boldsymbol{\xi}^T\boldsymbol{\varsigma} - V(\boldsymbol{\xi})\} = \sum_{k=1}^{m} \frac{1}{2}\alpha_k^{-1}\varsigma_k^2.$$

So $(\boldsymbol{\xi}, \boldsymbol{\varsigma})$ forms a *canonical duality pair* on $\mathcal{V}_a \times \mathcal{V}_a^*$ (cf. [38]) and the following canonical duality relations hold on $\mathcal{V}_a \times \mathcal{V}_a^*$:

$$\boldsymbol{\varsigma} = \delta V(\boldsymbol{\xi}) \quad \Leftrightarrow \quad \boldsymbol{\xi} = \delta V^*(\boldsymbol{\varsigma}) \quad \Leftrightarrow \quad \boldsymbol{\xi}^T \boldsymbol{\varsigma} = V(\boldsymbol{\xi}) + V^*(\boldsymbol{\varsigma}). \tag{2.7}$$

Replacing $W(\mathbf{x}) = V(\Lambda(\mathbf{x}))$ by $\Lambda(\mathbf{x})\boldsymbol{\varsigma} - V^*(\boldsymbol{\varsigma})$, the generalized complementary function [44] can be defined by

$$\begin{aligned} \Xi(\mathbf{x}, \boldsymbol{\varsigma}) &= \Lambda(\mathbf{x})\boldsymbol{\varsigma} - V^*(\boldsymbol{\varsigma}) + \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \mathbf{x}^T \mathbf{f} \\ &= \sum_{k=1}^{m} \left[ \left( \frac{1}{2}\mathbf{x}^T \mathbf{A}_k \mathbf{x} + \mathbf{b}_k^T \mathbf{x} + c_k \right) \varsigma_k - \frac{1}{2}\alpha_k^{-1}\varsigma_k^2 \right] + \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \mathbf{x}^T \mathbf{f}. \end{aligned} \tag{2.8}$$

For a fixed $\boldsymbol{\varsigma} \in \mathcal{V}_a^*$, the criticality condition $\delta_{\mathbf{x}}\Xi(\mathbf{x}, \boldsymbol{\varsigma}) = 0$ leads to the following *canonical equilibrium equation*:

$$\mathbf{G}(\boldsymbol{\varsigma})\mathbf{x} - F(\boldsymbol{\varsigma}) = 0, \tag{2.9}$$

where $F(\boldsymbol{\varsigma}) = \mathbf{f} - \sum_{k=1}^{m} \varsigma_k \mathbf{b}_k$, $\mathbf{G}(\boldsymbol{\varsigma}) = Q + \sum_{k=1}^{m} \varsigma_k \mathbf{A}_k$. Therefore, on the dual feasible space defined by

$$\mathcal{S}_a = \{\boldsymbol{\varsigma} \in \mathbb{R}^m\}, \tag{2.10}$$

the canonical dual function can be formulated as

$$\begin{aligned} P^d(\boldsymbol{\varsigma}) &= \mathrm{sta}\{\Xi(\mathbf{x}, \boldsymbol{\varsigma}) : \ \mathbf{x} \in \mathcal{X}_a\} \\ &= \sum_{k=1}^{m} \left( c_k \varsigma_k - \frac{1}{2}\alpha_k^{-1}\varsigma_k^2 \right) - \frac{1}{2} F^T(\boldsymbol{\varsigma}) \mathbf{G}^+(\boldsymbol{\varsigma}) F(\boldsymbol{\varsigma}), \end{aligned} \tag{2.11}$$

where $\mathbf{G}^+$ denotes the Moore-Penrose generalized inverse of $\mathbf{G}$, Thus, the canonical dual problem can be finally proposed as:

$$(\mathcal{P}^d): \quad \mathrm{sta}\left\{ P^d(\boldsymbol{\varsigma}) = \sum_{k=1}^{m} \left( c_k \varsigma_k - \frac{1}{2}\alpha_k^{-1}\varsigma_k^2 \right) - \frac{1}{2} F^T(\boldsymbol{\varsigma}) \mathbf{G}^+(\boldsymbol{\varsigma}) F(\boldsymbol{\varsigma}) \quad : \quad \boldsymbol{\varsigma} \in \mathcal{S}_a \right\}. \tag{2.12}$$

**Theorem 2.1** (Complementary-Dual Principle). *The problem $(\mathcal{P}^d)$ is canonically dual to the primal problem $(\mathcal{P})$ in the sense that if $\bar{\boldsymbol{\varsigma}}$ is a critical point of $(\mathcal{P}^d)$, then the vector*

$$\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\varsigma}}) F(\bar{\boldsymbol{\varsigma}}) \tag{2.13}$$

*is a critical point of $(\mathcal{P})$ and*

$$P(\bar{\mathbf{x}}) = P^d(\bar{\boldsymbol{\varsigma}}). \tag{2.14}$$

*Proof.* Suppose that $\bar{\boldsymbol{\varsigma}}$ is a critical point of $(\mathcal{P}^d)$. Then, we have

$$\frac{\partial P^d(\bar{\boldsymbol{\varsigma}})}{\partial \boldsymbol{\varsigma}_k} = c_k - \alpha_k^{-1}\boldsymbol{\varsigma}_k + \mathbf{b}_k^T\bar{\mathbf{x}} + \frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} = 0, \ k = 1, \cdots, m, \tag{2.15}$$

where $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\sigma}})F(\bar{\boldsymbol{\varsigma}})$. The criticality condition (2.15) is actually the canonical duality relation, i.e., $\boldsymbol{\varsigma}_k = \alpha_k(\frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} + \mathbf{b}_k^T\bar{\mathbf{x}} + c_k)$. Thus, we have

$$\begin{aligned}
\bar{\mathbf{x}} &= \mathbf{G}^+(\bar{\varsigma})F(\bar{\varsigma}) \\
&= \left[Q + \sum_{k=1}^m \alpha_k \left(\frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} + \mathbf{b}_k^T + c_k\right)\mathbf{A}_k\right]^+ \left[\mathbf{f} - \sum_{k=1}^m \alpha_k \left(\frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} + \mathbf{b}_k^T\bar{\mathbf{x}} + c_k\right)\mathbf{b}_k\right].
\end{aligned}$$

This shows that $\bar{\mathbf{x}}$ is a critical point of the primal problem $(\mathcal{P})$.

Moreover, in term of $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\varsigma})F(\bar{\varsigma})$, we have

$$\begin{aligned}
P^d(\bar{\varsigma}) &= \sum_{k=1}^m \left(c_k\bar{\varsigma}_k - \frac{1}{2}\alpha_k^+\bar{\varsigma}_k^2\right) - \frac{1}{2}F^T(\bar{\varsigma})\mathbf{G}^+(\bar{\varsigma})F(\bar{\varsigma}) \\
&= \sum_{k=1}^m \left(c_k\bar{\varsigma}_k - \frac{1}{2}\alpha_k^+\bar{\varsigma}_k^2\right) - \frac{1}{2}\left(\mathbf{f} - \sum_{k=1}^m \bar{\varsigma}_k b_k\right)^T \left(Q + \sum_{k=1}^m \bar{\varsigma}_k\mathbf{A}_k\right)^+ \left(\mathbf{f} - \sum_{k=1}^m \bar{\varsigma}_k b_k\right) \\
&= \sum_{k=1}^m \left(c_k\bar{\varsigma}_k - \frac{1}{2}\alpha_k^+\bar{\varsigma}_k^2\right) + \frac{1}{2}\bar{\mathbf{x}}^T \left(Q + \sum_{k=1}^m \bar{\varsigma}_k\mathbf{A}_k\right)\bar{\mathbf{x}} - \bar{\mathbf{x}}^T\left(\mathbf{f} - \sum_{k=1}^m \bar{\varsigma}_k b_k\right) \\
&= \sum_{k=1}^m \left[\left(\frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} + \mathbf{b}_k^T\bar{\mathbf{x}} + c_k\right)\bar{\varsigma}_k - \frac{1}{2}\alpha_k^+\bar{\varsigma}_k^2\right] + \frac{1}{2}\bar{\mathbf{x}}^T Q\bar{\mathbf{x}} - \bar{\mathbf{x}}^T\mathbf{f} \\
&= \sum_{k=1}^m \left[\left(\frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} + \mathbf{b}_k^T\bar{\mathbf{x}} + c_k\right)^2\alpha_k - \frac{1}{2}\alpha_k\left(\frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} + \mathbf{b}_k^T\bar{\mathbf{x}} + c_k\right)^2\right] \\
&\quad + \frac{1}{2}\bar{\mathbf{x}}^T Q\bar{\mathbf{x}} - \bar{\mathbf{x}}^T\mathbf{f} \\
&= \sum_{k=1}^m \frac{1}{2}\alpha_k\left(\frac{1}{2}\bar{\mathbf{x}}^T\mathbf{A}_k\bar{\mathbf{x}} + \mathbf{b}_k^T\bar{\mathbf{x}} + c_k\right)^2 + \frac{1}{2}\bar{\mathbf{x}}^T Q\bar{\mathbf{x}} - \bar{\mathbf{x}}^T\mathbf{f} \\
&= P(\bar{\mathbf{x}}).
\end{aligned}$$

This proves the theorem. □

Theorem 2.1 presents an analytic solution (2.13) for the critical point of the primal problem $(\mathcal{P})$. This solution is actually a special case of the general analytical solution form proposed in nonconvex variational problems [36].

## 2.4 Global Optimality Criteria

In order to identify global extremality properties of the analytical solution (2.13), we need to introduce a useful feasible space

$$\mathcal{S}_a^+ = \{\boldsymbol{\varsigma} \in \mathcal{S}_a \mid \ \mathbf{G}(\boldsymbol{\varsigma}) \succeq 0\}. \tag{2.16}$$

By the canonical duality theory developed in [38], we have the following results.

**Theorem 2.2.** *Suppose that the vector $\bar{\boldsymbol{\varsigma}}$ is a critical point of the canonical dual function $P^d(\bar{\boldsymbol{\varsigma}})$. Let $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}})$.*

*If $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a^+$, then $\bar{\boldsymbol{\varsigma}}$ is a global maximizer of $P^d$ on $\mathcal{S}_a^+$ if and only if the vector $\bar{\mathbf{x}}$ is a global minimizer of $P$ on $\mathbb{R}^n$, i.e.,*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x}\in\mathbb{R}^n} P(\mathbf{x}) \ \Leftrightarrow \ \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}) = P^d(\bar{\boldsymbol{\varsigma}}). \tag{2.17}$$

*Proof.* By Theorem 2.1, we know that the vector $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a$ is a critical point of Problem $(\mathcal{P}^d)$ if and only if $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}})$ is a critical point of Problem $(\mathcal{P})$, and

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

By the fact that the canonical dual function $P^d(\boldsymbol{\varsigma})$ is concave on $\mathcal{S}_a^+$, the critical point $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a^+$ is a global maximizer of $P^d(\boldsymbol{\varsigma})$ over $\mathcal{S}_a^+$. Since $(\bar{\mathbf{x}}, \bar{\boldsymbol{\varsigma}})$ is a saddle point of the total complementary function $\Xi(\mathbf{x}, \boldsymbol{\varsigma})$ on $\mathbb{R}^n \times \mathcal{S}_a^+$, i.e., $\Xi$ is convex in $\mathbf{x} \in \mathbb{R}^n$ and concave in $\boldsymbol{\varsigma} \in \mathcal{S}_a^+$, we have

$$\begin{aligned}
P^d(\bar{\boldsymbol{\varsigma}}) &= \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}) = \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \min_{\mathbf{x}\in\mathbb{R}^n} \Xi(\mathbf{x}, \boldsymbol{\varsigma}) = \min_{\mathbf{x}\in\mathbb{R}^n} \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \Xi(\mathbf{x}, \boldsymbol{\varsigma}) \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \mathbf{f}^T\mathbf{x} + \sum_{k=1}^m \max_{\varsigma_k\in\mathcal{S}_a^+} \left\{ \left( \frac{1}{2}\mathbf{x}^T \mathbf{A}_k \mathbf{x} + \mathbf{b}_k^T \mathbf{x} + c_k \right) \varsigma_k - \frac{1}{2}\alpha_k^+ \varsigma_k^2 \right\} \right\} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \mathbf{f}^T\mathbf{x} + \sum_{k=1}^m \frac{1}{2}\alpha_k \left( \frac{1}{2}\mathbf{x}^T \mathbf{A}_k \mathbf{x} + \mathbf{b}_k^T \mathbf{x} + c_k \right)^2 \right\} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} P(\mathbf{x}) = P(\bar{\mathbf{x}})
\end{aligned}$$

This proves the statement (2.17). □

Theorem 2.2 shows that the extremality condition of the analytical solution (2.13) is controlled by the critical point of the canonical dual problem, i.e., if $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a^+$, the solution $\bar{\mathbf{x}}(\bar{\boldsymbol{\varsigma}})$ is a global minimizer of $(\mathcal{P})$.

## 2.5 Numerical Examples

We now present examples to illustrate the applications of the theory proposed in this chapter.

**Example 2.1 Unconstrained two-dimensional polynomial minimization.**

$$\min \left\{ P(x_1, x_2) = \sum_{k=1}^{2} \frac{1}{2}\alpha_k \left( \frac{1}{2}(a_{k1}x_1^2 + a_{k2}x_2^2) + c_k \right)^2 + \frac{1}{2}(q_1 x_1^2 + q_2 x_2^2) - \sum_{i=1}^{2} f_i x_i : x \in \mathbb{R}^2 \right\}.$$

On the dual feasible set

$$\mathcal{S}_a = \{\varsigma \in \mathbb{R}^2 \mid (q_1 + \varsigma_1 a_{11} + \varsigma_2 a_{21})(q_2 + \varsigma_1 a_{12} + \varsigma_2 a_{22}) \neq 0\},$$

the canonical dual function has the form of

$$\begin{aligned} P^d(\varsigma) &= \sum_{k=1}^{2} \left( c_k \varsigma_k - \frac{1}{2\alpha_k}\varsigma_k^2 \right) - \frac{1}{2}[f_1, f_2] \\ & \left[ \begin{array}{cc} (q_1 + \varsigma_1 a_{11} + \varsigma_2 a_{21})^{-1} & 0 \\ 0 & (q_2 + \varsigma_1 a_{12} + \varsigma_2 a_{22})^{-1} \end{array} \right] \left[ \begin{array}{c} f_1 \\ f_2 \end{array} \right]. \end{aligned}$$

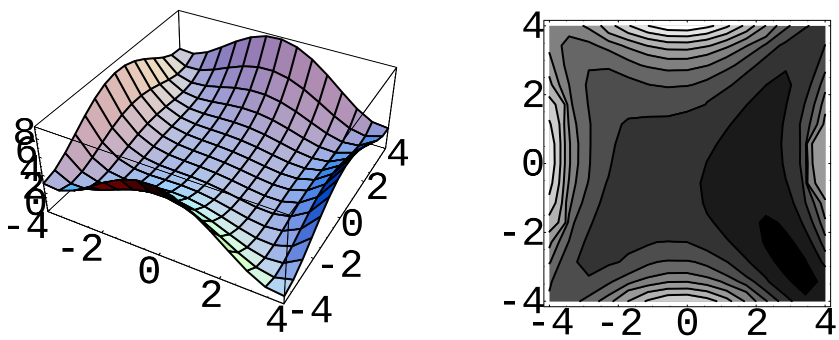


Figure 2.1: Graph of $P(\mathbf{x})$ (left) and contour of $P(\mathbf{x})$(right).

If we let $a_{11} = -0.4$, $a_{12} = 0.6$, $a_{21} = 0.5$, $a_{22} = -0.3$, $q_1 = -1$, $q_2 = 0.6$, $\mathbf{f} = [0.3, -0.2]^T$, $\mathbf{c} = [1, 2]^T$, $\boldsymbol{\alpha} = [0.2, 0.8]^T$, the graphs and contours of the primal and dual functions are illustrated in Figures 2.1 and 2.2. In this case, the dual problem has a unique critical point $\bar{\varsigma} = [0.3467, 2.4700]^T$ in the space

$$\mathcal{S}_a^+ = \{\varsigma \in \mathbb{R}^2 \mid (q_1 + \varsigma_1 a_{11} + \varsigma_2 a_{21})(q_2 + \varsigma_1 a_{12} + \varsigma_2 a_{22}) > 0\}.$$

Therefore, by Theorem 2.2, we know that

$$\bar{\mathbf{x}} = [f_1/(q_1 + \bar{\varsigma}_1 a_{11} + \bar{\varsigma}_2 a_{21}), f_2/(q_2 + \bar{\varsigma}_1 a_{12} + \bar{\varsigma}_2 a_{22})]^T = [3.1146, -2.9842]^T$$

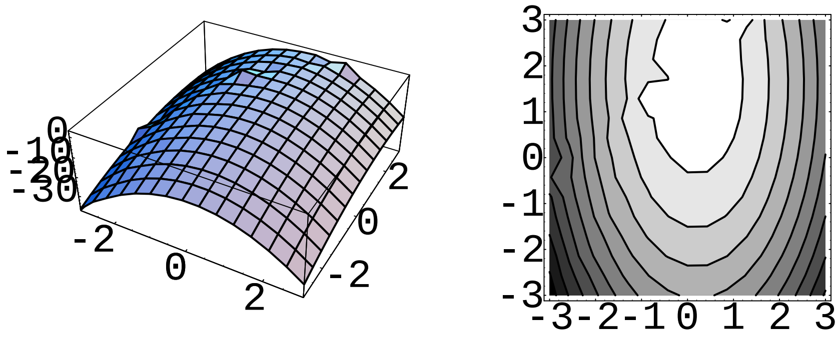



Figure 2.2: Graph of $P^d(\boldsymbol{\varsigma})$ (left) and contour of $P^d(\boldsymbol{\varsigma})$(right).

is a global minimization. It's easy to verify that

$$P(\bar{\mathbf{x}}) = 0.4075 = P^d(\bar{\boldsymbol{\varsigma}}).$$

**Example 2.2 Minimization problem of Colville Function.**

$$\begin{aligned} \min P(\mathbf{x}) = \quad & 100(x_2 - x_1^2)^2 + (1-x_1)^2 + 90(x_4 - x_3^2)^2 + (1-x_3)^2 + \\ & 10.1((x_2-1)^2 + (x_4-1)^2) + 19.8(x_2-1)(x_4-1) \\ s.t. \quad & -10 \le x_i \le 10, i = 1,2,3,4. \end{aligned}$$

This is a well-known test problem for global optimization. On the dual feasible set

$$\mathcal{S}_a = \{\boldsymbol{\varsigma} \in \mathbb{R}^2 \mid (1-\varsigma_1)(1-\varsigma_2) \neq 0\},$$

the canonical dual function has the form of

$$\begin{aligned} P^d(\boldsymbol{\varsigma}) = \quad & 42 - \frac{1}{400}\varsigma_1^2 - \frac{1}{360}\varsigma_2^2 \\ & -\frac{1}{2}\begin{bmatrix} 2 \\ 40-\varsigma_1 \\ 2 \\ 40-\varsigma_2 \end{bmatrix}^T \begin{bmatrix} 2-2\varsigma_1 & & & \\ & 20.2 & & 19.8 \\ & & 2-2\varsigma_2 & \\ & 19.8 & & 20.2 \end{bmatrix}^+ \begin{bmatrix} 2 \\ 40-\varsigma_1 \\ 2 \\ 40-\varsigma_2 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \Xi(\mathbf{x}, \boldsymbol{\varsigma}) = \quad & 42 + (x_2 - x_1^2)\boldsymbol{\varsigma}_1 + (x_4 - x_3^2)\boldsymbol{\varsigma}_2 - \frac{1}{400}\boldsymbol{\varsigma}_1^2 - \frac{1}{360}\boldsymbol{\varsigma}_2^2 \\ & +(x_1^2 + 10.1x_2^2 + x_3^2 + 10.1x_4^2 + 19.8x_2x_4) - (2x_1 + 40x_2 + 2x_3 + 40x_4) \end{aligned}$$

By solving the criticality condition $\nabla\Xi(\mathbf{x}, \boldsymbol{\varsigma}) = 0$, we get three critical points:

$\bar{\mathbf{x}}^1 = [1,1,1,1]^T$, $\bar{\boldsymbol{\varsigma}}^1 = (0,0)$,

$$\bar{\mathbf{x}}^2 = [-0.967974, 0.947139, -0.969516, 0.951248]^T, \ \ \bar{\boldsymbol{\varsigma}}^2 = [2.03309, 2.03144]^T,$$
$$\bar{\mathbf{x}}^3 = [-0.031251, 0.165971, -0.0312582, 0.184264]^T, \ \ \bar{\boldsymbol{\varsigma}}^3 = [32.999, 32.9916]^T,$$

and $\bar{\boldsymbol{\varsigma}}^1 \in \mathcal{S}_a^+$. By Theorem 2.2, we know that $\bar{\mathbf{x}}^1$ is global minimizer of $P(\mathbf{x})$. It is easy to check that $P(\bar{\mathbf{x}}^1) = \Xi(\bar{\mathbf{x}}^1, \bar{\varsigma}^1) = P^d(\bar{\boldsymbol{\varsigma}}^1) = 0$.

## 2.6 Conclusion

We have presented a detailed application of the canonical duality theory for solving general sum of fourth-order polynomial optimization problem. An analytical solution is obtained by the complementary-dual principle and its extremality property is classified by the duality theory. Results show that by using the canonical dual transformation, the nonconvex primal problem in $\mathbb{R}^n$ can be converted into a concave maximization dual problem $(\mathcal{P}^d_{\max})$ in $\mathbb{R}^m$, which can be solved by well-developed convex minimization techniques.

# CHAPTER 3

# Nonconvex Quadratic-Exponential Minimization Problem

## 3.1 Introduction

This chapter presents a set of complete solutions and optimality conditions for a nonconvex quadratic-exponential optimization problem. By using the canonical duality theory reported in Chapter 1, the nonconvex primal problem in n-dimensional space can be converted into an one-dimensional canonical dual problem, which is a concave maximization problem with zero duality gap. The global extrema of the nonconvex problem can be identified by the canonical duality theory. Several examples are solved so as to illustrate the applicability of the theory.

## 3.2 Problem Statement

The primal problem to be solved is given by

$$(\mathcal{P}_e) \quad \min \left\{ P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A \mathbf{x} - \mathbf{c}^T \mathbf{x} + W(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n \right\}, \tag{3.1}$$

where $A = A^T \in \mathbb{R}^{n \times n}$ is a given indefinite matrix, $\mathbf{c}$ is a given vector in $\mathbb{R}^n$, the nonconvex function $W(\mathbf{x})$ is an exponential function with quadratic function exponent:

$$W(\mathbf{x}) = \exp\left(\frac{1}{2}|B\mathbf{x}|^2 - \alpha\right), \tag{3.2}$$

where $B \in \mathbb{R}^{m \times n}$ is a matrix, $\alpha > 0$ is a positive constants, and $|\mathbf{v}|$ denotes the Euclidean norm of $\mathbf{v}$. The quadratic-exponential function can be used to model a large class of nonlinear phenomena, such as plant and insect growth [20], finite deformation elasticity [71], computational bio-chemistry [104], and bio-mechanics [58].

The criticality condition $\nabla P(\mathbf{x}) = 0$ leads to a nonlinear equilibrium equation:

$$A\mathbf{x} + \exp\left(\frac{1}{2}|B\mathbf{x}|^2 - \alpha\right) B^T B\mathbf{x} = \mathbf{c}. \tag{3.3}$$

Solving this coupled nonlinear algebraic system directly is very difficult. Also equation (3.3) is only a necessary condition for global minimizer of Problem $(\mathcal{P}_e)$. Due to the nonconvexity of the target function $P(\mathbf{x})$, Problem $(\mathcal{P}_e)$ may possess many local minimizers. A general sufficient condition for identifying the global minimizer is a fundamental task in global optimization.

We will show that by the use of the canonical dual transformation, the nonlinear coupled algebraic system in $\mathbb{R}^n$ can be converted into an algebraic equation in one-dimensional space. Therefore, a complete set of solutions is obtained.

## 3.3 Canonical Dual transformation

Following the standard procedure of the canonical dual transformation, we introduce a differentiable *geometrical operator*

$$\xi = \Lambda(\mathbf{x}) = \frac{1}{2}|B\mathbf{x}|^2 - \alpha, \tag{3.4}$$

which is a quadratic map from $\mathbb{R}^n$ into $\mathcal{V}_a = \{\xi \in \mathbb{R} \mid \ \xi \geq -\alpha\}$. Thus, the nonconvex function $W(\mathbf{x})$ can be written in the canonical form

$$W(\mathbf{x}) = V(\Lambda(\mathbf{x})), \tag{3.5}$$

where $V(\xi) = e^{\xi}$ is a canonical function on $\mathcal{V}_a$, i.e., the duality relation

$$\varsigma = \nabla V(\xi) = e^{\xi} \tag{3.6}$$

is invertible for any given $\xi \in \mathcal{V}_a$ (see the definition of the canonical function introduced in Chapter 1). It is clear that $\varsigma > 0$.

By letting $U(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{c}^T\mathbf{x}$, the primal problem $(\mathcal{P}_e)$ can be reformulated in the following canonical form:

$$(\mathcal{P}) \quad \min\{P(\mathbf{x}) = U(\mathbf{x}) + V(\Lambda(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^n\}. \tag{3.7}$$

Let $\mathcal{V}_a^* = \{\varsigma \in \mathbb{R} \mid \ \varsigma > 0\}$ be the range of the duality mapping $\varsigma = \nabla V(\xi) : \mathcal{V}_a \to \mathcal{V}_a^* \subset \mathbb{R}$. So $(\xi, \varsigma)$ forms a duality pair on $\mathcal{V}_a \times \mathcal{V}_a^*$ and the Legendre conjugate $V^*$ can be uniquely defined by

$$V^*(\varsigma) = \text{sta}\{\xi\varsigma - V(\xi) : \quad \xi \in \mathbb{R}\} = \varsigma \log \varsigma - \varsigma,$$

where sta{} denotes finding stationary points of the statement in {}. Thus, replacing $W(\mathbf{x}) = V(\Lambda(\mathbf{x}))$ by $\Lambda(\mathbf{x})\varsigma - V^*(\varsigma)$, the total complementary function (see Chapter 1) can be defined by

$$\begin{aligned} \Xi(\mathbf{x}, \varsigma) &= U(\mathbf{x}) + \Lambda(\mathbf{x})\varsigma - V^*(\varsigma) \\ &= \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{c}^T\mathbf{x} + (\frac{1}{2}|B\mathbf{x}|^2 - \alpha)\varsigma - (\varsigma \log \varsigma - \varsigma). \end{aligned} \tag{3.8}$$

For a fixed $\varsigma$, the criticality condition $\nabla_{\mathbf{x}}\Xi(\mathbf{x}, \varsigma) = 0$ leads to the following canonical equilibrium equation:

$$A\mathbf{x} - \mathbf{c} + \varsigma B^T B\mathbf{x} = 0. \tag{3.9}$$

Clearly, for any given $\varsigma > 0$, if the vector $\mathbf{c} \in \mathcal{C}_{ol}(A + \varsigma B^T B)$, i.e., $\mathbf{c}$ is in the column space of $(A + \varsigma B^T B)$, the general solution of equation (3.9) is

$$\mathbf{x} = (A + \varsigma B^T B)^+\mathbf{c}, \tag{3.10}$$

where $A^+$ denotes the Moore-Penrose generalized inverse of $A$. Substituting this result into the total complementary function $\Xi$, the canonical dual problem can be formulated as:

$$(\mathcal{P}^d): \quad \text{sta}\left\{P^d(\varsigma) = -\frac{1}{2}\mathbf{c}^T(A + \varsigma B^T B)^+\mathbf{c} - (\varsigma \log \varsigma - \varsigma) - \alpha\varsigma \quad : \quad \varsigma \in S_a\right\}, \tag{3.11}$$

where the dual feasible space is given by

$$S_a = \{\varsigma \in \mathbb{R} \mid \varsigma > 0\}.$$

Let

$$A_d(\varsigma) = A + \varsigma B^T B.$$

**Theorem 3.1.** *If $\bar{\varsigma}$ is a* KKT *point of $(\mathcal{P}^d)$, then the vector*

$$\bar{\mathbf{x}} = A_d^+(\bar{\varsigma})\mathbf{c}$$

*is a critical point of $(\mathcal{P}_e)$ and $P(\bar{\mathbf{x}}) = P^d(\bar{\boldsymbol{\varsigma}})$.*

*Proof.* Suppose that $\bar{\varsigma}$ is a KKT point of $(\mathcal{P}^d)$. Then, we have

$$\bar{\varsigma} > 0, \quad \nabla P^d(\bar{\varsigma}) = \frac{1}{2}|B\bar{x}|^2 - \log\bar{\varsigma} - \alpha \leq 0 \tag{3.12}$$

$$\bar{\varsigma}\nabla P^d(\bar{\varsigma}) = 0. \tag{3.13}$$

By the fact that $\bar{\varsigma} > 0$, the complementarity condition (3.13) leads to

$$\frac{1}{2}|B\bar{\mathbf{x}}|^2 - \log\bar{\varsigma} - \alpha = 0,$$

i.e., $\bar{\varsigma} = \exp\left(\frac{1}{2}|B\bar{\mathbf{x}}|^2 - \alpha\right)$. Thus, we have

$$\bar{\mathbf{x}} = A_d^+(\bar{\varsigma})\mathbf{c} = \left(A + \exp\left(\frac{1}{2}|B\bar{\mathbf{x}}|^2 - \alpha\right) B^T B\right)^+ \mathbf{c}.$$

Since $\nabla P(\bar{\mathbf{x}}) = 0$, therefore $\bar{\mathbf{x}}$ is a critical point of the primal problem $(\mathcal{P}_e)$.

Moreover, in term of $\bar{\mathbf{x}} = A_d^+(\bar{\varsigma})\mathbf{c}$, we have

$$\begin{aligned}
P^d(\bar{\varsigma}) &= -\frac{1}{2}\mathbf{c}^T A_d^+(\bar{\varsigma})\mathbf{c} - (\bar{\varsigma}\log\bar{\varsigma} - \bar{\varsigma}) - \alpha\bar{\varsigma} \\
&= \frac{1}{2}\bar{\mathbf{x}}^T(A + \bar{\varsigma}B^T B)\bar{\mathbf{x}} - \mathbf{c}^T\bar{\mathbf{x}} - (\bar{\varsigma}\log\bar{\varsigma} - \bar{\varsigma}) - \alpha\bar{\varsigma} \\
&= \frac{1}{2}\bar{\mathbf{x}}^T A\bar{\mathbf{x}} - \mathbf{c}^T\bar{\mathbf{x}} + (\frac{1}{2}|B\bar{\mathbf{x}}|^2 - \alpha)\bar{\varsigma} - (\bar{\varsigma}\log\bar{\varsigma} - \bar{\varsigma}) \\
&= \frac{1}{2}\bar{\mathbf{x}}^T A\bar{\mathbf{x}} - \mathbf{c}^T\bar{\mathbf{x}} + \bar{\varsigma} + (\frac{1}{2}|B\bar{\mathbf{x}}|^2 - \log\bar{\varsigma} - \alpha)\bar{\varsigma} \\
&= \frac{1}{2}\bar{\mathbf{x}}^T A\bar{\mathbf{x}} - \mathbf{c}^T\bar{\mathbf{x}} + \exp\left(\frac{1}{2}|B\bar{\mathbf{x}}|^2 - \alpha\right) \\
&= P(\bar{\mathbf{x}}).
\end{aligned}$$

This proves the theorem. □

The next section will show that the global extremum of the function $P : \mathbb{R}^n \to \mathbb{R}$ only rely on critical points of the canonical dual function $P^d(\varsigma)$.

## 3.4 Global Optimality Criteria

It is known that the criticality condition is only necessary for local minimization of the nonconvex problem $(\mathcal{P}_e)$. In order to identify global and local extrema among the critical points of Problem $(\mathcal{P}_e)$, we need to introduce a useful feasible space

$$\mathcal{S}_a^+ = \{\varsigma \in \mathcal{S}_a \mid \ A_d(\varsigma) \succ 0\}. \tag{3.14}$$

Clearly, $\mathcal{S}_a^+$ is an open convex subset of $\mathbb{R}$. By the canonical duality theory, we have the following result.

**Theorem 3.2.** *Suppose that the vector $\bar{\varsigma}$ is a critical point of the canonical dual function $P^d(\bar{\varsigma})$. Let $\bar{\mathbf{x}} = A_d^+(\bar{\varsigma})\mathbf{c}$. If $\bar{\varsigma} \in \mathcal{S}_a^+$, then $\bar{\varsigma}$ is a global maximizer of $P^d$ on $\mathcal{S}_a^+$, the vector $\bar{x}$ is a global minimizer of $P$ on $\mathbb{R}^n$, and*

$$P(\bar{x}) = \min_{x\in\mathbb{R}^n} P(x) = \max_{\varsigma\in\mathcal{S}_a^+} P^d(\varsigma) = P^d(\bar{\varsigma}). \tag{3.15}$$

*Proof.* By Theorem 3.1, we know that the vector $\bar{\varsigma} \in \mathcal{S}_a$ is a KKT point of Problem $(\mathcal{P}^d)$ if and only if $\bar{\mathbf{x}} = A_d^+(\bar{\varsigma})\mathbf{c}$ is a critical point of Problem $(\mathcal{P}_e)$, and

$$P(\bar{\mathbf{x}}) = \Xi(\bar{\mathbf{x}}, \bar{\varsigma}) = P^d(\bar{\varsigma}).$$

By the fact that the canonical dual function $P^d(\varsigma)$ is concave on $\mathcal{S}_a^+$ (which can be easily proved by $\nabla^2 P^d(\varsigma) < 0 \ \ \forall\varsigma \in \mathcal{S}_a^+$), the critical point $\bar{\varsigma} \in \mathcal{S}_a^+$ is a global maximizer of $P^d(\varsigma)$ over $\mathcal{S}_a^+$, and $(\bar{\mathbf{x}}, \bar{\varsigma})$ is a saddle point of the total complementary function $\Xi(\mathbf{x}, \varsigma)$ on $\mathbb{R}^n \times \mathcal{S}_a^+$, i.e., $\Xi$ is convex in $\mathbf{x} \in \mathbb{R}^n$ and concave in $\varsigma \in \mathcal{S}_a^+$. Thus, we have

$$\begin{aligned}
P^d(\bar{\varsigma}) &= \max_{\varsigma\in\mathcal{S}_a^+} P^d(\varsigma) = \max_{\varsigma\in\mathcal{S}_a^+}\min_{\mathbf{x}\in\mathbb{R}^n} \Xi(\mathbf{x},\varsigma) = \min_{\mathbf{x}\in\mathbb{R}^n}\max_{\varsigma\in\mathcal{S}_a^+} \Xi(\mathbf{x},\varsigma) \\
&= \min_{\mathbf{x}\in\mathbb{R}^n}\left\{\frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{c}^T\mathbf{x} + \max_{\varsigma\in\mathcal{S}_a^+}\left\{(\frac{1}{2}|B\mathbf{x}|^2 - \alpha)\varsigma - (\varsigma\log\varsigma - \varsigma)\right\}\right\} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n}\left\{\frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{c}^T\mathbf{x} + \max_{\varsigma\in\mathcal{S}_a^+}\left\{\varsigma + (\frac{1}{2}|B\mathbf{x}|^2 - \log\varsigma - \alpha)\varsigma\right\}\right\} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} P(\mathbf{x})
\end{aligned}$$

This proves the statement (3.15). □

This theorem shows that the extremality condition of the primal problem is controlled by the critical points of the canonical dual problem, i.e., if $\bar{\varsigma} \in \mathcal{S}_a^+$, the vector $\bar{\mathbf{x}}(\bar{\varsigma})$ is a global minimizer of $(\mathcal{P}_e)$.

In a special case when $A$ is a diagonal matrix and $B$ is an identity matrix, we have

$$A_d^+(\varsigma) = \left\{\frac{1}{a_i + \varsigma}\right\}. \tag{3.16}$$

In this case,

$$P^d(\varsigma) = -\frac{1}{2}\sum_{i=1}^n \frac{c_i^2}{a_i + \varsigma} - (\varsigma\log\varsigma - \varsigma) - \alpha\varsigma. \tag{3.17}$$

The criticality condition $\nabla P^d(\varsigma) = 0$ gives the canonical dual algebraic equation:

$$\frac{1}{2}\sum_{i=1}^{n}\left(\frac{c_i}{a_i+\varsigma}\right)^2 - \log\varsigma - \alpha = 0. \tag{3.18}$$

For the given $\alpha$, $\{c_i\}$, and $\{a_i\}$ such that $a_1 \le a_2 \le \ldots, \le a_n$, this dual algebraic equation (3.18) can be solved completely within each interval $-a_{i+1} < \varsigma < -a_i$ such that $a_i < a_{i+1}$ $(i = 1, 2, \cdots, n)$.

## 3.5 Numerical Examples

We now list a few examples to illustrate the applications of the theory presented above.

### 3.5.1 One-dimensional nonconvex minimization

First of all, let us consider one dimensional concave minimization problem:

$$\min\left\{P(x) = \frac{1}{2}ax^2 - cx + \exp\left(\frac{1}{2}x^2 - 2\right) : \ x \in \mathbb{R}\right\}. \tag{3.19}$$

In this case,

$$\mathcal{S}_a = \{\varsigma \in \mathbb{R} \mid \varsigma > 0, \ \ a + \varsigma \neq 0\}.$$

The dual function is

$$P^d(\varsigma) = -\frac{1}{2}c^2/(a+\varsigma) - \varsigma\log\varsigma - \varsigma. \tag{3.20}$$

If we choose $c = 0.5$, and $a = -2$, the dual solution $\varsigma_1 = 2.21$ is a unique global maximizer of $P^d$ on $\mathcal{S}_a^+ = \{\varsigma \in \mathbb{R}_+ \mid \ a + \varsigma > 0\}$. It gives the global minimizer $x_1 = 2.36$. It is easy to check that $P(x_1) = -4.56 = P^d(\varsigma_1)$. The graph of $P(x)$ and $P^d(\varsigma)$ are shown in Figures 3.1-3.2.

### 3.5.2 Two-dimensional nonconvex minimization

Consider

$$\min\left\{P(x_1,x_2) = \frac{1}{2}(a_1x_1^2 + a_2x_2^2) - c_1x_1 - c_2x_2 + \exp\left(\frac{1}{2}(x_1^2+x_2^2) - 2\right) \ : \ \ x \in \mathbb{R}^2\right\}.$$

The dual feasible set is given by

$$\mathcal{S}_a = \{\varsigma \in \mathbb{R}^2 \mid \ \varsigma > 0, \ \ (a_1+\varsigma)(a_2+\varsigma) \neq 0\}.$$

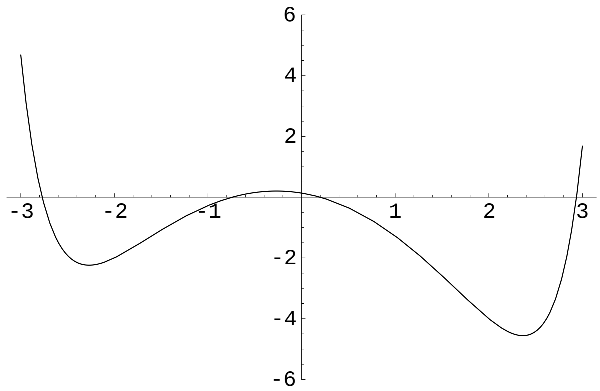



Figure 3.1: Graph of $P(x)$ for one dimensional problem which has global minimizer $x_1 = 2.36$.

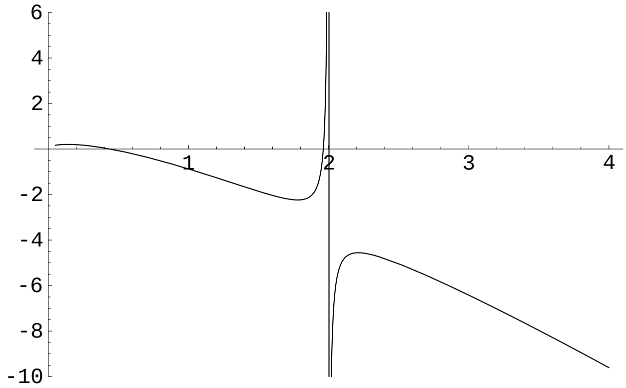



Figure 3.2: Graph of $P^d(\varsigma)$ for one dimensional problem which is concave on $\varsigma > 2$ and has global maximizer $\varsigma_1 = 2.21$.

The canonical dual function has the form of

$$P^d(\varsigma) = -\frac{1}{2}[c_1, c_2] \begin{bmatrix} \frac{1}{a_1+\varsigma} & \\ & \frac{1}{a_2+\varsigma} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \varsigma \log \varsigma - \varsigma. \tag{3.21}$$

**Case I.** $a_1 \leq 0$, $a_2 \leq 0$. We let $\mathbf{c} = [0.1, -0.3]^T$, $a_1 = -1$, $a_2 = -1.2$. The canonical dual problem has three critical points

$$\varsigma_1 = 1.34 \in \mathcal{S}_a^+ = \{\varsigma \in \mathbb{R}^2 | \ \varsigma > 1.2\},$$

and

$$\varsigma_2 = 0.94, \varsigma_3 = 0.14.$$

By Theorem 3.2, we know that $\mathbf{x}_1 = [c_1/(a_1+\varsigma_1), c_2/(a_2+\varsigma_1)]^T = [0.29, -2.12]^T$ is a global minimizer. It is easy to verify that

$$P(\mathbf{x}_1) = P^d(\varsigma_1) = -2.07$$

(see Figures 3.3-3.4).

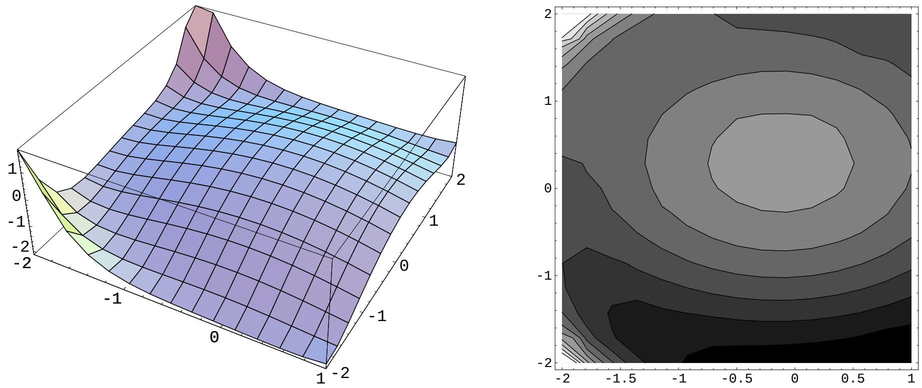


Figure 3.3: Graphs of $P(\mathbf{x})$ and its contour for two dimensional problem(Case I).

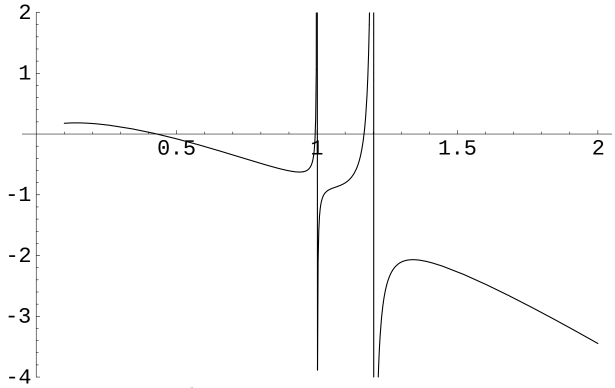


Figure 3.4: Graph of $P^d(\varsigma)$ for two dimensional problem(Case I).

**Case II.** $a_1 \leq 0$, $a_2 \geq 0$. We choose $\mathbf{c} = [0.1, -0.3]^T$, $a_1 = -1$, $a_2 = 0.6$. In this case, we have

$$\varsigma_1 = 1.05 \in \mathcal{S}_a^+ = \{\varsigma \in \mathbb{R}^2 | \ \varsigma > 1\},$$

and

$$\varsigma_2 = 0.95, \varsigma_3 = 0.15.$$

Thus, $\mathbf{x}_1$ =$[2.02, -0.18]^T \in \mathbb{R}^2$ is a global minimizer. It is easy to verify that

$$P(\mathbf{x}_1) = P^d(\varsigma_1) = -1.23$$

(see Figures 3.5-3.6).

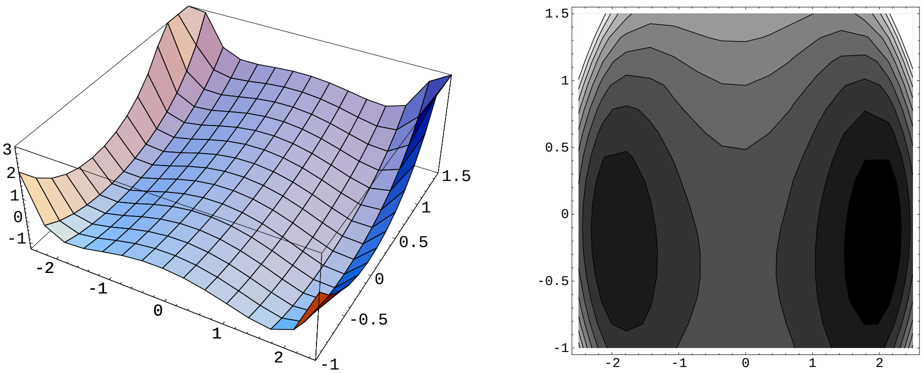



Figure 3.5: Graphs of $P(x)$ and its contour for two dimensional problem(Case II).

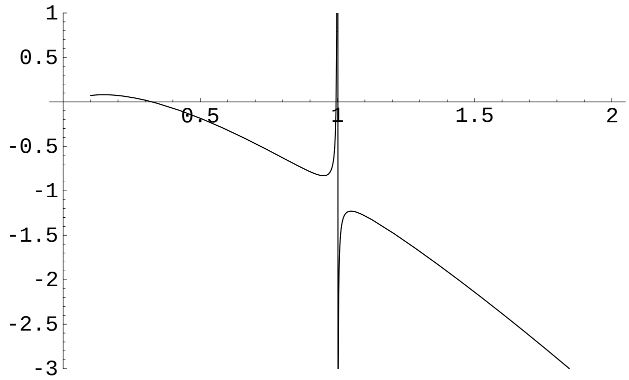



Figure 3.6: Graph of $P^d(\varsigma)$ for two dimensional problem(Case II).

### 3.5.3 Two-dimensional general nonconvex minimization

Let $A$ be a diagonal matrix and let $B$ be a $3 \times 2$ matrix. The primal problem is

$$\min \left\{ P(x_1, x_2) = \frac{1}{2}(a_1 x_1^2 + a_2 x_2^2) - c_1 x_1 - c_2 x_2 + \exp\left(\frac{1}{2}|B\mathbf{x}|^2 - 4\right) \ : \ \mathbf{x} \in \mathbb{R}^2 \right\}.$$

Suppose that $B^T B = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$. Then, on the dual feasible set, we have

$$\mathcal{S}_a = \left\{ \boldsymbol{\varsigma} \in \mathbb{R}^2 | \ \varsigma > 0, \mathbf{c} \in \mathcal{C}_{ol}\left( \begin{bmatrix} a_1 + \varsigma c_{11} & \varsigma c_{12} \\ \varsigma c_{21} & a_2 + \varsigma c_{22} \end{bmatrix} \right) \right\},$$

The canonical dual function has the form of

$$P^d(\boldsymbol{\varsigma}) = -\frac{1}{2}[c_1, c_2] \begin{bmatrix} a_1 + \varsigma c_{11} & \varsigma c_{12} \\ \varsigma c_{21} & a_2 + \varsigma c_{22} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ c_2 \end{bmatrix} - \varsigma \log \varsigma - 3\varsigma. \tag{3.22}$$

Let $\mathbf{c} = [0.5, -0.5]^T$, $a_1 = -2$, $a_2 = 1.2$, $B = \begin{bmatrix} -1 & -1 \\ -1 & -2 \\ 2 & 1 \end{bmatrix}$. The critical points of the canonical dual problem inside $\mathcal{S}_a^+$ is $\varsigma_1 = 0.94$, where

$$\mathcal{S}_a^+ = \{\varsigma \in \mathbb{R} | \ \varsigma > 0.28\}.$$

By Theorem 3.2, we know that $\mathbf{x}_1 = [2.03, -1.47]^T \in \mathbb{R}^2$ is a global minimizer. It is easy to verify that

$$P(\mathbf{x}_1) = P^d(\varsigma_1) = -3.65$$

(see Figures 3.7-3.8).

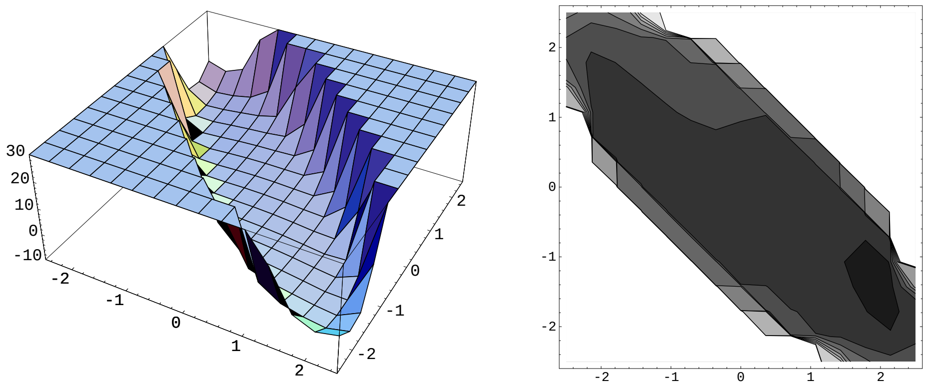


Figure 3.7: Graphs of $P(\mathbf{x})$ and its contour for two dimensional general problem.

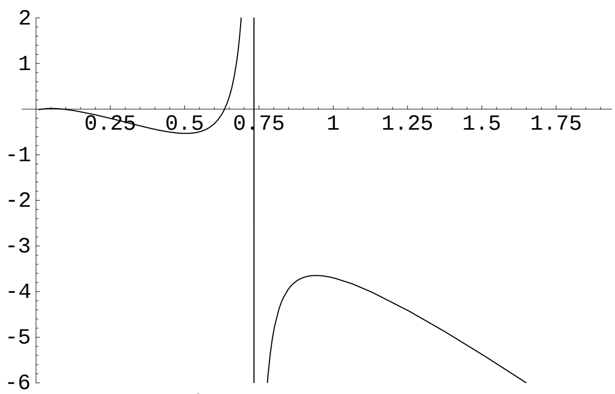


Figure 3.8: Graph of $P^d(\varsigma)$ for two dimensional general problem.

Comparing Fig. 3.7 with Fig. 3.8, we can see clearly that the graph of the primal function is very flat, indicating a very slow convergent rate of any numerical method used for solving this problem directly. On the contrary, the dual problem with only one variable can be solved very easily to obtain all extreme points and the largest dual solution $\varsigma_1$ leads to the global minimizer of the primal problem.

## 3.6 Conclusions

In this chapter, we have presented an application of the canonical duality theory to the nonconvex optimization problem $(\mathcal{P}_e)$. Generally speaking, the nonconvex quadratic form with an exponential objective function can be used to model many nonconvex systems. By using the canonical dual transformation, the nonconvex primal problem in n-dimensional space can be converted into a one-dimensional canonical dual problem, which can be solved completely. The global extrema can be identified by Theorem 3.2. As indicated in [38], for any given nonconvex problem, as long as the geometrical operator can be chosen properly and the canonical duality pairs can be identified correctly, the canonical dual transformation can be used to formulate perfect duality pair.

# CHAPTER 4

# Box and Integer Constrained Problem

## 4.1 Introduction

This chapter applies the canonical duality theory for solving quadratic minimization problems subjected to either box or integer constraints. Results show that these nonconvex problems can be converted into concave maximization of dual problems over convex feasible spaces without duality gap. Furthermore, the Boolean integer programming problem [25] is actually equivalent to a critical point problem in continuous space. These dual problems can be solved under certain conditions and an analytic solution for integer programming problem is obtained. Both existence and uniqueness of the canonical dual solutions are presented.

## 4.2 Problem Statement

Let us consider the following constrained nonconvex quadratic minimization problem:

$$(\mathcal{P}): \quad \min \left\{ P(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, Q\mathbf{x} \rangle - \langle \mathbf{x}, \mathbf{f} \rangle \mid \mathbf{x} \in \mathcal{X}_a \right\}, \tag{4.1}$$

where $Q = Q^T \in \mathbb{R}^{n \times n}$ is a given indefinite matrix, $\mathbf{f}$ is a given vector in $\mathbb{R}^n$, $\mathcal{X}_a \subset \mathbb{R}^n$ is a feasible space, and $\langle *, * \rangle$ represents a bilinear form on $\mathbb{R}^n \times \mathbb{R}^n$. For box constrained problem, $\mathcal{X}_a$ is defined by

$$\mathcal{X}_a = \{\mathbf{x} \in \mathbb{R}^n \mid -1 \le x_i \le 1, \ \forall i = 1, \ldots, n\}. \tag{4.2}$$

Problem $(\mathcal{P})$ is probably the most simple global optimization problem, which appears in many applications [32]. Replacing the inequality constraints in $\mathcal{X}_a$ by equality constraints $x_i = \pm 1$ $(i = 1, 2, \ldots, n)$, Problem $(\mathcal{P})$ is reduced to the well-known integer programming:

$$(\mathcal{P}_{ip}): \quad \min \left\{ P(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, Q\mathbf{x} \rangle - \langle \mathbf{x}, \mathbf{f} \rangle \ \mid \ \mathbf{x} \in \partial\mathcal{X}_a \right\}, \tag{4.3}$$

where the feasible set $\partial \mathcal{X}_a$ denotes the boundary of $\mathcal{X}_a$, i.e.,

$$\partial \mathcal{X}_a = \{\mathbf{x} \in \mathbb{R}^n | \ \mathbf{x} \in \{-1, 1\}^n\}. \tag{4.4}$$

Due to the nonconvexity of the quadratic function $P(\mathbf{x})$, quadratic minimization problems with either box or integer constraints are known to be NP-hard [72] [75] [76].

## 4.3 Canonical Dual Transformation

Following the standard procedure of the canonical dual transformation, we rewrite the inequality constraints $-1 \leq x_i \leq 1,\ i = 1, \ldots, n$, in $\mathcal{X}_a$ in the canonical form: $\mathbf{x} \circ \mathbf{x} \leq \mathbf{e}$, where the notation $\mathbf{s} \circ \mathbf{t} := [s_1 t_1, s_2 t_2, \ldots, s_n t_n]^T$ denotes the Hadamard product for any two vectors $\mathbf{s}, \mathbf{t} \in \mathbb{R}^n$. We introduce the geometrical operator

$$\boldsymbol{\xi} = \{\xi, \boldsymbol{\epsilon}\} = \Lambda(\mathbf{x}) = \frac{1}{2}[\mathbf{x}^T Q \mathbf{x}, \mathbf{x} \circ \mathbf{x}]^T : \mathbb{R}^n \to \mathcal{E} = \mathbb{R}^{1+n}, \tag{4.5}$$

and let

$$V(\boldsymbol{\xi}) = \begin{cases} \xi & \text{if } \boldsymbol{\xi} \leq \frac{1}{2}\mathbf{e}, \\ +\infty & \text{otherwise.} \end{cases} \tag{4.6}$$

Then the box constrained problem $(\mathcal{P})$ can be reformulated as the following unconstrained canonical form:

$$\min\{\Pi(\mathbf{x}) = V(\Lambda(\mathbf{x})) - \langle \mathbf{x}, \mathbf{f} \rangle \mid \ \mathbf{x} \in \mathbb{R}^n\}. \tag{4.7}$$

Let $\partial f(\mathbf{x})$ denote the set of subgradient of the function $f$ at the point $\mathbf{x}$, i.e.,

$$\partial f(x) := \{\mathbf{u} | f(\mathbf{x}) + \mathbf{u}^T(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y})\}.$$

By the fact that $V(\boldsymbol{\xi})$ is convex and lower semi-continuous on $\mathcal{E}$, the canonical dual variable $\boldsymbol{\xi}^*$ can be defined as:

$$\boldsymbol{\xi}^* \in \partial V(\boldsymbol{\xi}) = \{1, \boldsymbol{\sigma}\} \in \mathcal{E}^* = \mathbb{R}^{1+n}. \tag{4.8}$$

Let $\langle \boldsymbol{\xi}; \boldsymbol{\xi}^* \rangle$ denote the bilinear form on $\mathcal{E} \times \mathcal{E}^*$, the so-called complementary function $V^\sharp(\boldsymbol{\xi}^*)$ can be defined by the Fenchel transformation:

$$V^\sharp(\boldsymbol{\xi}^*) = \sup_{\boldsymbol{\xi} \in \mathcal{E}} \{\langle \boldsymbol{\xi}; \boldsymbol{\xi}^* \rangle - V(\boldsymbol{\xi})\} = \begin{cases} \frac{1}{2}\langle \mathbf{e}, \boldsymbol{\sigma} \rangle & \text{if } \boldsymbol{\lambda} \geq \mathbf{0} \in \mathbb{R}^n, \\ +\infty & \text{otherwise.} \end{cases}$$

Since both $V(\boldsymbol{\xi})$ and $V^\sharp(\boldsymbol{\xi}^*)$ are proper convex functions over their effective domains $\mathcal{E}_a = \{\boldsymbol{\xi} = \{\xi, \boldsymbol{\xi}\} \in \mathcal{E} | \ \ \boldsymbol{\xi} \leq \frac{1}{2}\mathbf{e}\}$ and $\mathcal{E}_a^* = \{\boldsymbol{\xi}^* = \{1, \boldsymbol{\sigma}\} \in \mathcal{E}^* | \ \boldsymbol{\sigma} \geq 0\}$, respectively, the following canonical duality relations hold on $\mathcal{E}_a \times \mathcal{E}_a^*$ :

$$\boldsymbol{\xi}^* \in \partial V(\boldsymbol{\xi}) \quad \Leftrightarrow \quad \boldsymbol{\xi}^* \in \partial V^\sharp(\boldsymbol{\xi}) \quad \Leftrightarrow \quad V(\boldsymbol{\xi}) + V^\sharp(\boldsymbol{\xi}^*) = \langle \boldsymbol{\xi}; \boldsymbol{\xi}^* \rangle. \tag{4.9}$$

Replacing $V(\Lambda(\mathbf{x}))$ in the canonical primal problem (4.7) by $V(\Lambda(\mathbf{x})) = \langle \Lambda(\mathbf{x}); \boldsymbol{\xi}^* \rangle - V^\sharp(\boldsymbol{\xi}^*)$. Then the *total complementary function* $\Xi(\mathbf{x}, \boldsymbol{\sigma}) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ associated with the problem $(\mathcal{P})$ can be defined as:

$$\begin{aligned} \Xi(\mathbf{x}, \boldsymbol{\sigma}) &= \langle \Lambda(\mathbf{x}); \boldsymbol{\xi}^* \rangle - V^\sharp(\boldsymbol{\xi}^*) - \langle \mathbf{x}, \mathbf{f} \rangle & (4.10) \\ &= \frac{1}{2}\langle \mathbf{x}, \mathbf{G}(\boldsymbol{\sigma})\mathbf{x} \rangle - \frac{1}{2}\langle \mathrm{e}, \boldsymbol{\sigma} \rangle - \langle \mathbf{x}, \mathbf{f} \rangle \quad \text{s.t. } \boldsymbol{\sigma} \in \mathbb{R}^n_+, & (4.11) \end{aligned}$$

where $\mathbb{R}^n_+ := \{\boldsymbol{\sigma} \in \mathbb{R}^n | \ \boldsymbol{\sigma} \geq \mathbf{0}\}$. For a fixed $\boldsymbol{\sigma} \in \mathbb{R}^n_+$, the criticality condition $\nabla_{\mathbf{x}} \Xi(\mathbf{x}, \boldsymbol{\sigma}) = 0$ leads to

$$\mathbf{G}(\boldsymbol{\sigma})\bar{\mathbf{x}} = \mathbf{f}. \tag{4.12}$$

Clearly, if the matrix $\mathbf{G}(\boldsymbol{\sigma})$ is invertible on $\mathcal{S}_a$, the primal variable $\bar{\mathbf{x}}$ can be uniquely defined by $\bar{\mathbf{x}} = \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{f}$.

On the other hand, for a given matrix $Q$ and $\boldsymbol{\sigma} \in \mathbb{R}^n_+$, if the vector $\mathbf{f}$ is in the column space $\mathcal{C}_{ol}(\mathbf{G}(\boldsymbol{\sigma}))$ of the matrix $\mathbf{G}(\boldsymbol{\sigma})$, i.e., a linear space spanned by the columns of $\mathbf{G}(\boldsymbol{\sigma})$, the generalized solution $\bar{\mathbf{x}}$ of the canonical equilibrium equation (4.12) is given by

$$\bar{\mathbf{x}} = \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{f},$$

where $\mathbf{G}^+(\boldsymbol{\sigma})$ denotes the Moore-Penrose generalized inverse of $\mathbf{G}(\boldsymbol{\sigma})$. Substituting this generalized solution into the total complementary function $\Xi$ and let $\mathcal{S}_g$ be a generalized canonical dual feasible space defined by

$$\mathcal{S}_g = \{\boldsymbol{\sigma} \in \mathbb{R}^n | \ \boldsymbol{\sigma} \geq 0\}, \tag{4.13}$$

the generalized canonical dual function $P_g : \mathcal{S}_g \to \mathbb{R}$ can be formulated as

$$\begin{aligned} P^g(\boldsymbol{\sigma}) &= \text{sta}\{\Xi(\mathbf{x}, \boldsymbol{\sigma}) | \ \mathbf{x} \in \mathbb{R}^n\} \\ &= -\frac{1}{2}\langle \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{f}, \mathbf{f} \rangle - \frac{1}{2}\langle \mathrm{e}, \boldsymbol{\sigma} \rangle. & (4.14) \end{aligned}$$

Therefore, the generalized canonical dual problem $(\mathcal{P}^g)$ can be formulated as

$$(\mathcal{P}^g): \quad \max\left\{P^g(\boldsymbol{\sigma}) = -\frac{1}{2}\langle \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{f}, \mathbf{f} \rangle - \frac{1}{2}\langle \mathrm{e}, \boldsymbol{\sigma} \rangle \ | \ \boldsymbol{\sigma} \in \mathcal{S}_g\right\}. \tag{4.15}$$

Similarly, the canonical dual problem for the integer programming problem $(\mathcal{P}_{ip})$ can be formulated as

$$(\mathcal{P}^g_{ip}): \quad \max\left\{ P^g(\boldsymbol{\sigma}) = -\frac{1}{2}\langle \mathbf{G}^{\dagger}(\boldsymbol{\sigma})\mathbf{f}, \mathbf{f}\rangle - \frac{1}{2}\langle \mathbf{e}, \boldsymbol{\sigma}\rangle \mid \boldsymbol{\sigma} \neq 0 \right\}. \tag{4.16}$$

Then we have the following result.

**Theorem 4.1** (Complementary-Dual Principle)**.** *Problem* $(\mathcal{P}^g)$ *is canonically dual to* $(\mathcal{P})$ *in the sense that if* $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_g$ *is a feasible solution of* $(\mathcal{P}^g)$*, then the vector*

$$\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) = \mathbf{G}^{+}(\bar{\boldsymbol{\sigma}})\mathbf{f} \tag{4.17}$$

*is a feasible solution of Problem* $(\mathcal{P})$ *and*

$$P(\bar{\mathbf{x}}) = P^g(\bar{\boldsymbol{\sigma}}). \tag{4.18}$$

*Moreover, if* $\bar{\boldsymbol{\sigma}} \neq 0$ *is a critical point of* $(\mathcal{P}^g)$*, then* $\bar{\mathbf{x}} \in \partial\mathcal{X}_a$ *is a KKT point of* $(\mathcal{P}_{ip})$*.*

*Proof.* By introducing a Lagrange multiplier $\boldsymbol{\epsilon} \in \mathbb{R}^n_- := \{\boldsymbol{\epsilon} \in \mathbb{R}^n | \ \boldsymbol{\epsilon} \leq 0\}$ to relax the inequality condition $\boldsymbol{\sigma} \geq 0$ in $\mathcal{S}_g$, the Lagrangian $L : \mathcal{S}_g \times \mathbb{R}^n_- \to \mathbb{R}$ associated with Problem $(\mathcal{P}^g)$ is

$$L(\boldsymbol{\sigma}, \boldsymbol{\epsilon}) = P^g(\boldsymbol{\sigma}) - \langle \boldsymbol{\epsilon}, \boldsymbol{\sigma}\rangle. \tag{4.19}$$

It is easy to prove that the criticality condition $\nabla_{\boldsymbol{\sigma}} L(\bar{\boldsymbol{\sigma}}, \boldsymbol{\epsilon}) = 0$ leads to

$$\boldsymbol{\epsilon} = \nabla P^g(\bar{\boldsymbol{\sigma}}) = \frac{1}{2}(\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) \circ \bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) - \mathbf{e}) = \boldsymbol{\xi}(\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}})) - \frac{1}{2}\mathbf{e} \tag{4.20}$$

and the KKT conditions

$$0 \leq \bar{\boldsymbol{\sigma}} \ \perp \ \boldsymbol{\epsilon}(\bar{\mathbf{x}}) \leq 0, \tag{4.21}$$

where $\bar{\mathbf{x}} = \mathbf{G}^{+}(\bar{\boldsymbol{\sigma}})\mathbf{f}$, and $\bar{\boldsymbol{\sigma}} \perp \boldsymbol{\epsilon}$ denotes the complementarity condition, i.e.,

$$\boldsymbol{\epsilon}(\mathbf{x}) \perp \bar{\boldsymbol{\sigma}} \quad \Leftrightarrow \quad \frac{1}{2}(x_i^2 - 1)\bar{\sigma}_i = 0, \ \ \forall i = 1, \ldots, n.$$

This shows that if $\bar{\boldsymbol{\sigma}}$ is a KKT point of the problem $(\mathcal{P}^g)$, then $\bar{\mathbf{x}} = \mathbf{G}^{+}(\bar{\boldsymbol{\sigma}})\mathbf{f}$ is a KKT point of the primal problem $(\mathcal{P})$.

By the complementarity condition in (4.21) and $\bar{\mathbf{x}} = \mathbf{G}^{+}(\bar{\boldsymbol{\sigma}})\mathbf{f}$, we have

$$\begin{aligned} P^g(\bar{\boldsymbol{\sigma}}) &= \frac{1}{2}\langle \mathbf{G}^{+}(\bar{\boldsymbol{\sigma}})\mathbf{f}, \mathbf{f}\rangle - \langle \mathbf{G}^{+}(\bar{\boldsymbol{\sigma}})\mathbf{f}, \mathbf{f}\rangle - \frac{1}{2}\langle \mathbf{e}, \bar{\boldsymbol{\sigma}}\rangle \\ &= \frac{1}{2}\langle \bar{\mathbf{x}}, Q\bar{\mathbf{x}}\rangle - \langle \bar{\mathbf{x}}, \mathbf{f}\rangle + \frac{1}{2}\langle \bar{\mathbf{x}} \circ \bar{\mathbf{x}} - \mathbf{e}, \bar{\boldsymbol{\sigma}}\rangle = P(\bar{\mathbf{x}}). \end{aligned}$$

Moreover, if $\bar{\boldsymbol{\sigma}} \neq 0$, the complementarity condition $\boldsymbol{\epsilon}(\bar{\mathbf{x}}) \perp \bar{\boldsymbol{\sigma}}$ in (4.21) leads to $\boldsymbol{\epsilon}(\bar{\mathbf{x}}) = \frac{1}{2}(\bar{\mathbf{x}} \circ \bar{\mathbf{x}} - \mathbf{e}) = 0$, i.e., $\nabla P^g(\bar{\boldsymbol{\sigma}}) = 0$. This shows that if $\bar{\boldsymbol{\sigma}} \neq 0$ is a critical point of $P^g(\boldsymbol{\sigma})$, the associated vector $\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) = \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{f} \in \{-1,1\}^n$ is a KKT point of the integer programming problem $(\mathcal{P}_{ip})$. □

**Corollary 4.1.** *If $\bar{\boldsymbol{\sigma}} \neq 0$ is a critical point of $P^g(\boldsymbol{\sigma})$, then the vector $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{f} \in \{-1,1\}^n$ is a feasible solution to the integer programming problem $(\mathcal{P}_{ip})$.*

*Proof.* By the criticality condition $\delta P^g(\bar{\boldsymbol{\sigma}}, \boldsymbol{\sigma}) = \langle \nabla P^g(\bar{\boldsymbol{\sigma}}), \boldsymbol{\sigma}\rangle = 0 \; \forall \boldsymbol{\sigma} \neq 0 \in \mathbb{R}^n$, where $\delta P^g(\bar{\boldsymbol{\sigma}}, \boldsymbol{\sigma})$ denotes the derivative of $P^g$ at $\bar{\boldsymbol{\sigma}}$ in the direction $\boldsymbol{\sigma}$, we have the canonical complementarity equation

$$\langle \bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) \circ \bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) - \mathbf{e}, \boldsymbol{\sigma}\rangle = 0 \;\; \forall \boldsymbol{\sigma} \in \mathbb{R}^n, \tag{4.22}$$

where $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{f}$. Therefore, under the condition $\boldsymbol{\sigma} \neq 0$, the canonical solution $\mathbf{x} = \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{f}$ is a feasible solution of $(\mathcal{P}_{ip})$. □

For the given indefinite matrix $\mathbf{Q}$, the inequality constraint $\boldsymbol{\sigma} \neq 0$ is essential for the canonical dual integer programming problem $(\mathcal{P}^g_{ip})$. But this condition, as well as the condition $\boldsymbol{\sigma} \in \mathcal{C}_{ol}(\mathbf{G}(\boldsymbol{\sigma}))$ in $\mathcal{S}^+_g$ can also be relaxed by perturbation methods.

## 4.4 Global Optimality Criteria

In this section, we shall present global optimality conditions for the nonconvex problems $(\mathcal{P})$ and $(\mathcal{P}_{ip})$. We let

$$\mathcal{S}^+_g = \{\boldsymbol{\sigma} \in \mathcal{S}_g |\; \mathbf{G}(\boldsymbol{\sigma}) \succeq 0\}, \tag{4.23}$$

and consider the following canonical dual problem:

$$(\mathcal{P}^+_g): \;\; \max\left\{ P^g(\boldsymbol{\sigma}) = -\frac{1}{2}\langle \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{f}, \mathbf{f}\rangle - \frac{1}{2}\langle \mathbf{e}, \boldsymbol{\sigma}\rangle \;|\;\; \boldsymbol{\sigma} \in \mathcal{S}^+_g \right\}. \tag{4.24}$$

**Theorem 4.2.** *For any given matrix $Q \in \mathbb{R}^{n\times n}$ and a vector $\mathbf{f} \in \mathbb{R}^n$, the canonical dual problem $(\mathcal{P}^+_g)$ has at least one KKT point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}^+_g$ and the following weak duality relation holds*

$$\min_{\mathbf{x}\in\mathcal{X}_a} P(\mathbf{x}) \geq \max_{\boldsymbol{\sigma}\in\mathcal{S}^+_g} P^g(\boldsymbol{\sigma}) = P^g(\bar{\boldsymbol{\sigma}}). \tag{4.25}$$

*If the KKT point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}^+_g$ is a critical point of $P^g(\boldsymbol{\sigma})$, then the vector $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{f}$ is a global minimizer to the primal problem $(\mathcal{P})$ and the following strong duality relation holds*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x}\in\mathcal{X}_a} P(\mathbf{x}) = \max_{\boldsymbol{\sigma}\in\mathcal{S}^+_g} P^g(\boldsymbol{\sigma}) = P^g(\bar{\boldsymbol{\sigma}}). \tag{4.26}$$

*Proof.* Since $\mathcal{S}_g^+$ is a closed convex set, for any given $\boldsymbol{\sigma} \in \mathcal{S}_g^+$ such that $\mathbf{x} = \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{f}$, the Hessian matrix of $P^d(\boldsymbol{\sigma})$

$$\nabla^2 P^g(\boldsymbol{\sigma}) = -\text{Diag }(\mathbf{x}(\boldsymbol{\sigma}))\ \mathbf{G}^+(\boldsymbol{\sigma})\ \text{Diag }(\mathbf{x}(\boldsymbol{\sigma})) \tag{4.27}$$

is negative semi-definite on $\mathcal{S}_g^+$. Thus, the canonical dual function $P^g(\boldsymbol{\sigma})$ is concave on $\mathcal{S}_g^+$. By the fact that, for any given $\boldsymbol{\sigma} \geq 0 \in \mathbb{R}^n$,

$$\lim_{\alpha\to\infty} P^g(\alpha\boldsymbol{\sigma}) = -\infty, \tag{4.28}$$

we know that the canonical dual function $P^g(\boldsymbol{\sigma})$ is coercive on the closed convex set $\mathcal{S}_g^+$. Therefore, the canonical dual problem $(\mathcal{P}^g_{\max})$ has at least one maximizer $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_g^+$ by the theory of convex analysis [28] [80]. Since the total complementary function $\Xi(\mathbf{x}, \boldsymbol{\sigma})$ is a saddle function on $\mathbb{R}^n \times \mathcal{S}_g^+$, we have

$$\min_{\mathbf{x}\in\mathcal{X}_a} P(\mathbf{x}) = \min_{\mathbf{x}\in\mathbb{R}^n} \max_{\boldsymbol{\sigma}\in\mathcal{S}_g^+} \Xi(\mathbf{x}, \boldsymbol{\sigma}) \geq \max_{\boldsymbol{\sigma}\in\mathcal{S}_g^+} \min_{\mathbf{x}\in\mathbb{R}^n} \Xi(\mathbf{x}, \boldsymbol{\sigma}) = \max_{\boldsymbol{\sigma}\in\mathcal{S}_g^+} P^g(\boldsymbol{\sigma}).$$

This leads to the weak duality relation (4.25).

By Theorem 4.1 we know that if the vector $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_g$ is a critical point of the canonical dual function $(\mathcal{P}^g)$, then $\bar{\mathbf{x}} = \mathbf{G}^\dagger(\bar{\boldsymbol{\sigma}})\mathbf{f}$ is a KKT point of Problem $(\mathcal{P})$ and

$$P(\bar{\mathbf{x}}) = \Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}) = P^g(\bar{\boldsymbol{\sigma}}). \tag{4.29}$$

Since the geometrical operator $\Lambda(\mathbf{x})$ defined in (4.5) is a (pure) quadratic operator, the quadratic function

$$\mathbf{G}_a(\mathbf{x}, \boldsymbol{\sigma}) = \frac{1}{2}\langle \mathbf{x}, \mathbf{G}(\boldsymbol{\sigma})\mathbf{x}\rangle \tag{4.30}$$

is a convex function of $\mathbf{x} \in \mathbb{R}^n$ for any given $\boldsymbol{\sigma} \in \mathcal{S}_g^+$. Therefore, the total complementary function $\Xi : \mathbb{R}^n \times \mathcal{S}_g^+ \to \mathbb{R}$ is a saddle function which is convex in $\mathbf{x} \in \mathbb{R}^n$ and concave in $\boldsymbol{\sigma} \in \mathcal{S}_g^+$. Thus, we have (4.26). □

**Corollary 4.2.** *Suppose that* $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_g^+$ *is a critical point of the canonical dual problem* $(\mathcal{P}_g^+)$ *and* $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{f}$.

*If* $\mathbf{G}(\bar{\boldsymbol{\sigma}}) \succ 0$, *then* $\bar{\mathbf{x}}$ *is a unique global minimizer of Problem* $(\mathcal{P})$.

*If* $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_g^+$ *and* $\bar{\boldsymbol{\sigma}} \neq 0$, *then* $\bar{\mathbf{x}}$ *is a global minimizer of the integer programming problem* $(\mathcal{P}_{ip})$.

Theorem 4.2 shows that a vector $\bar{\mathbf{x}} = \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{f}$ is a global minimizer of Problem $(\mathcal{P})$ if $\bar{\boldsymbol{\sigma}}$ is a KKT point of $(\mathcal{P}_g^+)$.

We will illustrate the advantage of using canonical duality theory through following example.

**Example 1.** *For a given vector* $\mathbf{f} \in \mathbb{R}^n$*, we consider the following constrained convex maximization problem:*

$$\max\{\|\mathbf{x}+\mathbf{f}\|_2 \;:\; \|\mathbf{x}\|_\infty \leq 1\}, \tag{4.31}$$

*which is equivalent to the following concave quadratic minimization problem*

$$\min\left\{P(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T\mathbf{x} - \mathbf{x}^T\mathbf{f} \;:\; |x_i| \leq 1,\ \forall i = 1,\dots,n,\ \mathbf{x}^T\mathbf{x} \leq r\right\}, \tag{4.32}$$

*where* $r > n$ *to ensure that the additional quadratic constraint* $\mathbf{x}^T\mathbf{x} \leq r$ *in the feasible space* $\mathcal{X}_c = \{\mathbf{x} \in \mathbb{R}^n |\ -1 \leq x_i \leq 1,\ \ \forall i = 1,\dots,n,\ \ \mathbf{x}^T\mathbf{x} \leq r\}$ *is never active. It is known that for high dimensional nonconvex constrained optimization problems, to check which constraints are active is fundamentally difficult.*

*If we let* $n = 2$*,* $r = 100$*, and* $\mathbf{f} = (1,1)^T$*, the optimal solution is* $\bar{\mathbf{x}} = (1,1)^T$ *with objective value* $P(\bar{\mathbf{x}}) = -3$*. To illustrate the difficulty of applying the classical Lagrangian duality theory directly to (4.32), we first introduce Lagrange multipliers* $(\sigma_1,\sigma_2,\sigma_3,\sigma_4)^T \in \mathbb{R}^4_+$ *to relax the linear box constraints* $-1 \leq x_i \leq 1,\ i = 1,2,$ *and* $\sigma_5 \geq 0$ *to relax the quadratic constraint* $\frac{1}{2}\mathbf{x}^T\mathbf{x} \leq 50$*. The Lagrangian associated with (4.32) is*

$$L(\mathbf{x},\boldsymbol{\sigma}) = -\frac{1}{2}(x_1^2+x_2^2)-(x_1+x_2)+\sum_{i=1}^{2}[\sigma_i(x_i-1)-\sigma_{i+2}(x_i+1)]+\frac{1}{2}\sigma_5\left(x_1^2+x_2^2-100\right),$$

*with the lagrangian dual function given by*

$$P^*(\boldsymbol{\sigma}) = \min_{\mathbf{x}\in\mathbb{R}^2} L(\mathbf{x},\boldsymbol{\sigma}).$$

*when* $\sigma_5 < 1$*, we get* $P^*(\boldsymbol{\sigma}) = -\infty$*. When* $\sigma_5 = 1$*, we obtain* $\max_{\boldsymbol{\sigma}\geq 0}\{P^*(\boldsymbol{\sigma}) :\ \sigma_5 = 1\} = -52$ *at the solution* $\sigma_0 = (1,1,0,0,1)$*. Finally, for any given* $\boldsymbol{\sigma} \in \mathcal{S}_r = \{\boldsymbol{\sigma} \in \mathbb{R}^5_+ |\ \sigma_5 > 1\}$*, the Lagrangian dual function can be obtained as*

$$P^*(\boldsymbol{\sigma}) = -\frac{1}{2(\sigma_5-1)}\left[(1-\sigma_1+\sigma_3)^2+(1-\sigma_2+\sigma_4)^2\right]-\sum_{i=1}^{4}\sigma_i - 50\sigma_5.$$

*It is easy to check that the solution to the Lagrangian dual problem*

$$\sup\{P^*(\boldsymbol{\sigma}) :\ \boldsymbol{\sigma} \in \mathcal{S}_r\} = -52,$$

*realized as* $\boldsymbol{\sigma} \to \boldsymbol{\sigma}_o = (1,1,0,0,1)^T$*. Hence, the optimal dual value is given by* $P^*(\boldsymbol{\sigma}) = -52$*, and there exists a duality gap between the primal and the Lagrangian dual problem,* *i.e.,*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x}\in\mathcal{X}_c} P(\mathbf{x}) = -3 > -52 = \max_{\boldsymbol{\sigma}\in\mathbb{R}^5_+} P^*(\boldsymbol{\sigma}) = P^*(\boldsymbol{\sigma}_o).$$

*To close this duality gap, we rewrite the constraints in the canonical form* $\mathbf{g}(\mathbf{x}) = \Lambda(\mathbf{x}) \leq \mathbf{d}$ *with*

$$B^\alpha_{ij} = \begin{cases} 1 & \text{if } i = j = \alpha, \\ 0 & \text{otherwise}, \end{cases} \quad i, j, \alpha = 1, 2, \quad B^3_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad i, j = 1, 2$$

*and* $\mathbf{d} = 0.5(1, 1, 100)^T$. *Then, on the canonical dual feasible space* $\mathcal{S}^+_q = \{\boldsymbol{\sigma} \in \mathbb{R}^3_+ | \ \sigma_i + \sigma_3 - 1 > 0, \ \ i = 1, 2\}$ *the canonical dual problem* $(\mathcal{P}^d_q)$ *is*

$$\max\left\{ P^d_q(\boldsymbol{\sigma}) = -\frac{1}{2}\left(\frac{1}{\sigma_1 + \sigma_3 - 1} + \frac{1}{\sigma_2 + \sigma_3 - 1}\right) - \frac{1}{2}(\sigma_1 + \sigma_2) - 50\sigma_3 \ : \ \boldsymbol{\sigma} \in \mathcal{S}^+_q \right\}. \tag{4.33}$$

*The optimal solution for this concave maximization problem is* $\bar{\boldsymbol{\sigma}} = (2, 2, 0)^T$ *with the optimal value* $P^d_q(\bar{\boldsymbol{\sigma}}) = -3$. *Observed that* $\bar{\sigma}_3 = 0$ *reflects the fact that the quadratic constraint* $\mathbf{x}^T\mathbf{x} \leq r$ *is inactive. Since* $\bar{\boldsymbol{\sigma}} \in \mathcal{S}^+_q$ *is a critical point of* $P^d_q(\boldsymbol{\sigma})$*, therefore, the vector* $\bar{\mathbf{x}} = G^{-1}_q(\bar{\boldsymbol{\sigma}})\mathbf{f} = (1, 1)^T$ *is a global minimizer of the primal problem with zero duality gap.*

**Remark 1.** This example shows the difficulty of directly applying the classical Lagrangian duality for solving nonconvex minimization problem with linear (including both box and integer) constraints. The classical Lagrangian duality theory was originally developed for linearly constrained convex problems in analytical mechanics, where the Lagrange multipliers and the linear constraints possess certain perfect duality. The primal problem in above example has both linear and nonlinear (quadratic) constraints, the Lagrange multipliers $\sigma_i, \ i = 1, 2, 3, 4$ are dual to the linear constraints, while $\sigma_5$ is dual to the quadratic constraints. Since the linear and nonlinear constraints are different geometrical measures, their corresponding dual variables, i.e., the Lagrange multipliers $\sigma_i$, $i = 1, 2, 3, 4$, and $\sigma_5$ are in different metric spaces with different (physical) units. Therefore, the classical Lagrangian dual problem in this case does not make physical sense. The weak Lagrangian duality theory leads to various duality gaps.

## 4.5 Existence and Uniqueness Conditions

The weak duality theorem (4.25) shows that the canonical dual problem $(\mathcal{P}^d_{\max})$ provides a lower bound for the box/integer constrained problems. In order to study existence and uniqueness of the canonical dual problems, we introduce a singular hyper-surface defined by

$$\mathcal{G}_a = \{\boldsymbol{\sigma} \in \mathbb{R}^n | \ \det \mathbf{G}(\boldsymbol{\sigma}) = 0\}. \tag{4.34}$$

Then, we have the following theorem.

**Theorem 4.3** (Existence and Uniqueness Criterion)**.** *Suppose that for a given symmetric matrix $Q$ and a vector $\mathbf{f}$ such that $\mathcal{S}_g^+ \neq \emptyset$ and $\mathcal{G}_a \subset \mathcal{S}_g^+$. If for any given $\boldsymbol{\sigma}_o \in \mathcal{G}_a$ and $\boldsymbol{\sigma} \in \mathcal{S}_a^+$,*

$$\lim_{\alpha \to 0^+} P^g(\boldsymbol{\sigma}_o + \alpha \boldsymbol{\sigma}) = -\infty, \tag{4.35}$$

*then the canonical dual problem $(\mathcal{P}_g^+)$ has a unique critical point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_a^+$ and $\bar{\mathbf{x}} = \mathbf{G}^{-1}(\bar{\boldsymbol{\sigma}})\mathbf{f}$ is a global minimizer to the primal problem $(\mathcal{P})$. If $\bar{\boldsymbol{\sigma}} \neq 0$, then $\bar{\mathbf{x}}$ is a global minimizer to the integer programming problem $(\mathcal{P}_{ip})$.*

*Proof.* If $\mathcal{G}_a \subset \mathcal{S}_g^+$, then $\mathcal{S}_g^+$ is a closed convex subset of $\mathbb{R}^n_+$. Since $P^g : \mathcal{S}_g^+ \to \mathbb{R}$ is concave, if (4.35) holds, the canonical dual function $P^g(\boldsymbol{\sigma})$ is coercive on the open convex set $\mathcal{S}_a^+$. Therefore, the canonical dual problem $(\mathcal{P}^g)$ has a unique maximizer $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_a^p$. □

Clearly, if $Q \succ 0$, the quadratic objective function $P(\mathbf{x})$ is convex and the solution to the box constrained primal problem $(\mathcal{P})$ could be a stationary point in the box $\mathcal{X}_a$. If $Q \prec 0$, the primal function $P(\mathbf{x})$ is concave and its global minimizer $\bar{\mathbf{x}}$ must be located on the boundary of the feasible space $\mathcal{X}_a$. In this case, the box constrained problem $(\mathcal{P})$ is identical to the integer constrained problem $(\mathcal{P}_{ip})$, and both of them are considered to be NP-hard. However, by the fact that $\mathcal{G}_a \subset \mathcal{S}_g^+$ and for any given $\mathbf{f} \in \mathbb{R}^n$, the dual feasible space $\mathcal{S}_g^+ \neq \emptyset$, the canonical dual problem $(\mathcal{P}^d_{\max})$ could be much easier to solve.

In the case that $Q = \text{Diag}\,(\mathbf{q})$ is a diagonal matrix with $\mathbf{q} = \{q_i\} \in \mathbb{R}^n$ being its diagonal elements, the canonical dual function $P^d(\boldsymbol{\lambda})$ has a simple form

$$P^d(\boldsymbol{\sigma}) = -\sum_{i=1}^{n} \left( \frac{c_i^2}{2(q_i + \sigma_i)} + \frac{1}{2}\sigma_i \right). \tag{4.36}$$

The criticality condition $\delta P^d(\boldsymbol{\sigma}) = 0$ leads to the dual solutions

$$\sigma_i = -q_i \pm |c_i|, \ \ \forall i = 1, 2, \ldots, n. \tag{4.37}$$

Clearly, for any given $\mathbf{q} \in \mathbb{R}^n$, if $c_i \neq 0 \ \ \forall i = 1, \ldots, n$, the condition (4.35) holds. Therefore, by Theorems 4.2 and 4.3, we have the following result.

**Corollary 4.3.** *For any given diagonal matrix $Q = Diag\,(\mathbf{q})$ and a vector $\mathbf{f} \in \mathbb{R}^n$ such* *that* $c_i \neq 0 \ \forall i = 1, \dots, n$, *it holds that*

$$\mathbf{x} = \left\{ \frac{c_i}{|c_i|} \right\} \text{ is a global minimizer of if } \boldsymbol{\lambda} = \{-q_i + |c_i|\} > 0;$$

$$\mathbf{x} = \left\{ -\frac{c_i}{|c_i|} \right\} \begin{cases} \text{is a local minimizer if } \boldsymbol{\lambda} = \{-q_i - |c_i|\} > 0, \\ \text{is a local maximizer of if } \boldsymbol{\lambda} = \{-q_i - |c_i|\} < 0. \end{cases}$$

## 4.6 Perturbations and Analytical Solutions

For any given indefinite symmetrical matrix $Q \in \mathbb{R}^{n \times n}$, there exists a parametrical vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $Q + \text{Diag}\ (\boldsymbol{\alpha})$ is either positive definite or negative definite. By the fact that $\mathbf{x} \circ \mathbf{x} = \mathbf{e}$, the integer programming problem $(\mathcal{P}_{ip})$ is identical to the following perturbed problem

$$(\mathcal{P}_\alpha): \quad \min \left\{ P_\alpha(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, (Q + \text{Diag}\ (\boldsymbol{\alpha}))\mathbf{x}\rangle - \langle \mathbf{x}, \mathbf{f}\rangle - d_\alpha \mid \mathbf{x} \in \partial \mathcal{X}_a \right\}, \tag{4.38}$$

where $d_\alpha = \frac{1}{2}\langle \mathbf{e}, \boldsymbol{\alpha}\rangle$.

Clearly, if we choose $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $Q_\alpha = Q + \text{Diag}\ (\boldsymbol{\alpha}) \prec 0$, the primal function $P_\alpha(\mathbf{x})$ is strictly concave and its global minimizers must be located on the boundary $\partial \mathcal{X}_a$. In this case, the condition $\mathbf{G}_\alpha(\boldsymbol{\sigma}) = Q + \text{Diag}\ (\boldsymbol{\alpha} + \boldsymbol{\sigma}) \succeq 0$ implies $\boldsymbol{\sigma} > 0$. Therefore, on the perturbed dual feasible space

$$\mathcal{S}_\alpha^+ = \{\boldsymbol{\sigma} \in \mathbb{R}^n |\ \mathbf{G}_\alpha(\boldsymbol{\sigma}) \succeq 0\}, \tag{4.39}$$

the *perturbed canonical dual problem* is

$$(\mathcal{P}_\alpha^g): \quad \max \left\{ P_\alpha^g(\boldsymbol{\sigma}) = -\frac{1}{2}\langle \mathbf{G}_\alpha^+(\boldsymbol{\sigma})\mathbf{f}, \mathbf{f}\rangle - \frac{1}{2}\langle \mathbf{e}, \boldsymbol{\sigma}\rangle - d_\alpha \mid \boldsymbol{\sigma} \in \mathcal{S}_\alpha^+ \right\}. \tag{4.40}$$

Since the inequality constraint $\boldsymbol{\sigma} \neq 0$ is relaxed by the $\alpha$-concave perturbation $Q + \text{Diag}\ (\boldsymbol{\alpha}) \prec 0$, this perturbed canonical dual problem is easier than $(\mathcal{P}_{ip}^g)$.

**Theorem 4.4** (Analytic Solution to Integer Programming Problem $(\mathcal{P}_{ip})$)**.** *For a given* $\boldsymbol{\alpha} \in \mathbb{R}^n$ *such that* $\det Q_\alpha \neq 0$. *Then the problem* $(\mathcal{P}_\alpha^d)$ *is canonically dual to the integer programming* $(\mathcal{P}_{ip})$ *in the sense that if* $\bar{\boldsymbol{\sigma}} = \{\bar{\sigma}_i\}^n$ *is a solution to* $(\mathcal{P}_\alpha^d)$, *then the vector* $\bar{\mathbf{x}} = \{\bar{x}_i\}^n$ *defined by*

$$\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) = \left\{ \frac{c_i - \bar{\sigma}_i}{|c_i - \bar{\sigma}_i|} \right\}^n \tag{4.41}$$

*is a feasible solution to* $(\mathcal{P}_{ip})$, *and* $P(\bar{\mathbf{x}}) = P_\alpha^d(\bar{\boldsymbol{\sigma}})$.

*If* $Q_\alpha \succ 0$, *the dual problem* $(\mathcal{P}_\alpha^d)$ *has at most one solution* $\bar{\boldsymbol{\sigma}}$, *which is a global* *maximizer of* $P^d_\alpha(\boldsymbol{\sigma})$*, the vector* $\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}})$ *is a unique global minimizer of* $(\mathcal{P}_{ip})$*, and*

$$P_\alpha(\bar{\mathbf{x}}) = \min_{\mathbf{x} \in \{-1,1\}^n} P_\alpha(\mathbf{x}) = \max_{\boldsymbol{\sigma} \in \mathbb{R}^n} P^d_\alpha(\boldsymbol{\sigma}) = P^d_\alpha(\bar{\boldsymbol{\sigma}}). \tag{4.42}$$

*Proof.* The first part of the theorem can be proved easily by the complementary-dual principle. If $Q_\alpha \succ 0$, then $P^d_\alpha(\boldsymbol{\sigma})$ is strictly concave and the canonical dual problem $(\mathcal{P}^d_\alpha)$ is equivalent to

$$\max \left\{ P^d_\alpha(\boldsymbol{\sigma}) \mid \boldsymbol{\sigma} \in \mathbb{R}^n \right\}, \tag{4.43}$$

which has at most one solution $\bar{\boldsymbol{\sigma}}$ over $\mathbb{R}^n$. By the canonical duality theory, the feasible solution $\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}})$ is a global minimizer of $(\mathcal{P}_{ip})$.

□

Theorem 4.4 shows that for convex perturbation $Q_\alpha \succ 0$, the canonical dual problem $(\mathcal{P}^d_\alpha)$ is a unconstrained concave maximization problem (4.43). Therefore, if the primal problem has a unique global minimizer, it can be obtained by solving the convex perturbation canonical dual problem (4.43). However, for certain given $Q$ and $\mathbf{f}$, this problem may have no critical solution.

Combining Theorems 4.2 and 4.4, the condition for the existence of unique solution is given in the following theorem.

**Theorem 4.5** (Unique Analytic Solution)**.** *For a given matrix* $Q = \{q_{ij}\} \in \mathbb{R}^{n \times n}$ *and a vector* $\mathbf{f} = \{c_i\} \in \mathbb{R}^n$*, let* $\boldsymbol{\alpha} = \{\alpha_i\} \in \mathbb{R}^n$ *be a parametrical vector such that either* $Q + Diag\ (\boldsymbol{\alpha}) \succ 0$ *or* $Q + Diag\ (\boldsymbol{\alpha}) \prec 0$*. If*

$$|c_i| > \sum_{j=1}^n |\alpha_i \delta_{ij} + q_{ij}| \ \ \forall i = 1, \dots, n, \tag{4.44}$$

*where* $\delta_{ij} = 1$ *if* $i = j$*, 0 if* $i \neq j$ *is Kronecker delta, then the integer programming problem* $(\mathcal{P}_{ip})$ *has a unique global minimizer* $\bar{\mathbf{x}} = \{\bar{x}_i\}^n$ *given by*

$$\bar{x}_i = \begin{cases} 1 & \text{if } c_i > \sum_{j=1}^n |\alpha_i \delta_{ij} + q_{ij}|, \\ -1 & \text{if } c_i < -\sum_{j=1}^n |\alpha_i \delta_{ij} + q_{ij}|. \end{cases} \tag{4.45}$$

Proof. By the criticality condition $\nabla P^\alpha(\bar{\boldsymbol{\sigma}}) = 0$, we have

$$\left( G^+_\alpha(\bar{\boldsymbol{\sigma}}) \mathbf{f} \right) \circ \left( G^+_\alpha(\bar{\boldsymbol{\sigma}}) \mathbf{f} \right) = \mathbf{e}, \tag{4.46}$$

or in the component form $(G^+_\alpha(\bar{\boldsymbol{\sigma}})\mathbf{f})^2_i = 1$. Thus, we have $G^+_\alpha(\bar{\boldsymbol{\sigma}})\mathbf{f} = \mathbf{t}$, where $\mathbf{t} = \{\pm 1\}^n$.

This leads to the linear equation $\bar{\boldsymbol{\sigma}} \circ \mathbf{t} = \mathbf{f} - \boldsymbol{\alpha} \circ \mathbf{t} - Q\mathbf{t}$, or equivalently,

$$\bar{\boldsymbol{\sigma}} = (\mathbf{f} - \boldsymbol{\alpha} \circ \mathbf{t} - Q\mathbf{t}) \circ \mathbf{t}.$$

If condition (4.44) holds and let $\mathbf{t} = \bar{\mathbf{x}} = \{\bar{x}_i\}^n$, where $\bar{x}_i$ is defined by (4.45), then we have $\bar{\boldsymbol{\sigma}} > 0$. This leads to $G_\alpha(\bar{\boldsymbol{\sigma}}) \succ 0$ since $Q + \text{Diag}\ (\boldsymbol{\alpha}) \succ 0$. By Corollary 4.2 we know that $\bar{\mathbf{x}} = G_\alpha^+(\bar{\boldsymbol{\sigma}})\mathbf{f} = \{\bar{x}_i\}^n$ given by (4.45) is a global minimizer to the integer minimization problem $(\mathcal{P}_{ip})$.

On the other hand, if $Q + \text{Diag}\ (\boldsymbol{\alpha}) \prec 0$ and the condition (4.44) holds, the dual problem $(\mathcal{P}_\alpha^d)$ has a unique critical point $\bar{\boldsymbol{\sigma}} = \mathbf{G}_\alpha \bar{\mathbf{x}}$. Therefore, the vector defined by (4.45) must be a unique solution of $(\mathcal{P}_{ip})$. □

Theorem 4.5 shows that the existence of a unique analytical solution depends mainly on the given input $\mathbf{f}$. If $\mathbf{f}$ is very small or even zero (for example, max-cut problems), the primal problem $(\mathcal{P}_{ip})$ is usually NP-hard and has more than one global minimizer.

## 4.7 Numerical Examples

**Example 4.1. One-dimensional Concave Minimization.** First of all, let us consider one dimensional concave minimization problem:

$$\min \left\{ P(x) = \frac{1}{2} q x^2 - cx \ \mid \ -1 \leq x \leq 1 \right\}. \tag{4.47}$$

Clearly, if $q < 0$, the global minimizer of $P(x)$ has to be one of boundary points $\bar{x} = \pm 1$. Since $q \neq 0$, the canonical dual function $P^d(\sigma) = P^g(\sigma)$ is

$$P^d(\sigma) = -\frac{1}{2} c^2/(q + \sigma) - \frac{1}{2}\sigma. \tag{4.48}$$

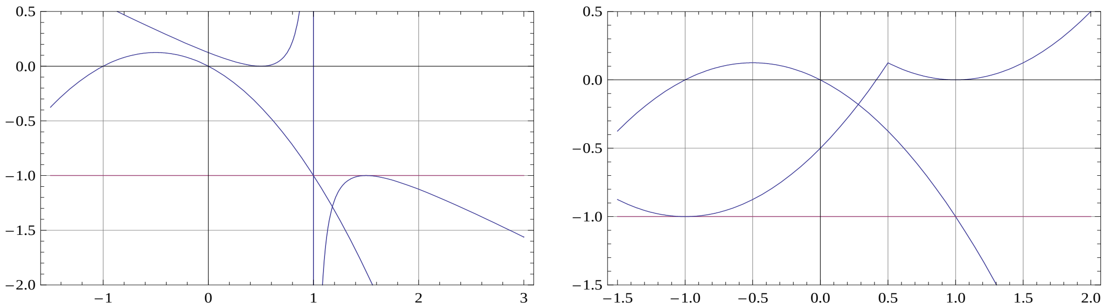
The criticality condition $\delta P^d(\sigma) = \frac{1}{2} c^2/(q+\sigma)^2 - \frac{1}{2} = 0$ has two roots: $\bar{\sigma}_{1,2} = -q \pm |c|$, and $\bar{x}_{1,2} = \pm c/|c|$ are two KKT points of $(\mathcal{P}_{ip})$. By Theorem 4.5 we know that $\bar{\sigma}_1 = -q + |c| > -q > 0$ is a unique global maximizer of $P^d$ on $\mathcal{S}_a^+ = \{\sigma \in \mathbb{R}\ |\sigma \geq 0,\ \ q + \sigma > 0\}$.

The canonical dual function $P_\alpha^d(\sigma)$ for this example is a nonconvex/nonsmooth function

$$P_\alpha^d(\sigma) = -\frac{1}{2} q^{-1} \sigma^2 - |c - \sigma|, \tag{4.49}$$

which has at most two critical points: $\bar{\sigma}_1 = q$ if $c > q$ and $\bar{\sigma}_2 = -q$ if $c < -q$.

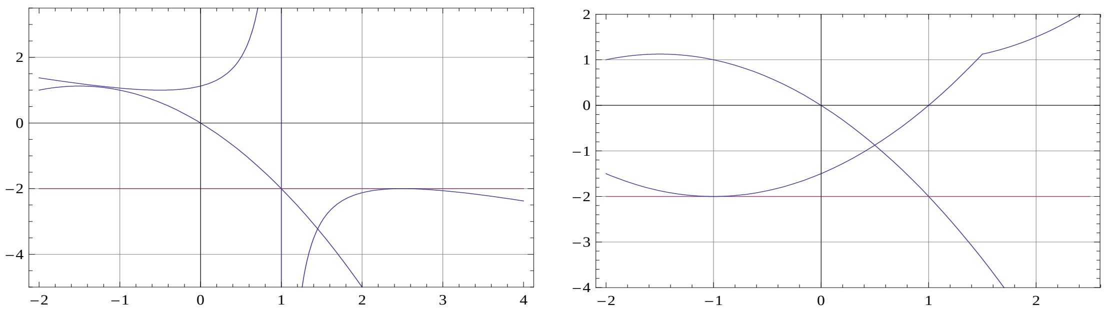
If we choose $c = 0.5,\ \ q = -1$, the dual solution $\bar{\sigma}_1 = 1.5$ of Problem $(\mathcal{P}^d)$ gives the global minimizer $\bar{x}_1 = c/(q + \bar{\sigma}_1) = 1$. It is easy to check that $P(\bar{x}_1) = -1 = P^d(\bar{\sigma}_1)$. The graphs of $P(x)$ and $P^d(\sigma)$ are shown in Fig. 4.1 (a).




(a) $P(x)$ and $P^d(\sigma)$ (b)$P(x)$ and $P^d_\alpha(\sigma)$

Figure 4.1: Graphs of $P(x)$ and its dual functions for Example 4.1 with $c = 0.5$

The graphs of $P(x)$ and $P^d_\alpha(\sigma)$ are shown in Fig. 4.1 (b). As we can see, the graph of $P^d_\alpha(\sigma)$ is nonconvex/nonsmooth and has two critical points: $\bar{\boldsymbol{\sigma}}_1 = -1$ and $\bar{\boldsymbol{\sigma}}_2 = 1$. By the analytical solution of (4.41), we have $\bar{x}_1 = 1$ and $\bar{x}_2 = -1$. It is easy to verify that $P(\bar{x}_1) = P^d_\alpha(\bar{\sigma}_1) = -1$ and $P(\bar{x}_2) = P^d_\alpha(\bar{\sigma}_2) = 0$.




(a) $P(x)$ and $P^d(\sigma)$ (b)$P(x)$ and $P^d_\alpha(\sigma)$

Figure 4.2: Graphs of $P(x)$ and its dual functions for Example 4.1 with $c = 1.5$

If we choose $c = 1.5, \quad q = -1$, the canonical dual problem $(\mathcal{P}^d)$ has two critical points: $\bar{\sigma}_1 = 2.5$ and $\bar{\sigma}_2 = -0.5$. By the fact that $\bar{\sigma}_1 \in \mathcal{S}^+_a$, $\bar{x}_1 = c/(q + \bar{\sigma}_1) = 1$ is a global minimizer and $P(\bar{x}_1) = -2 = P^d(\bar{\sigma}_1)$. In this case, the canonical dual problem $(\mathcal{P}^d_\alpha)$ has only one critical point $\bar{\sigma} = -1$ which is a global minimizer of $P^d_\alpha(\sigma)$. By Theorem 4.4, we know that $\bar{x} = 1$ is a global minimizer of $(\mathcal{P}_{ip})$ and $P(\bar{x}) = -2 = P^d_\alpha(\bar{\sigma})$.

**Example 4.2. Two-dimensional Nonconvex Programming Problem**. We now consider the following quadratic programming within a convex set:

$$\min \quad P(x_1, x_2) = \frac{1}{2}(q_1 x_1^2 + q_2 x_2^2 + 2q_3 x_1 x_2) - c_1 x_1 - c_2 x_2 \tag{4.50}$$

$$s.t. \quad -1 \le x_i \le 1, \quad i = 1, 2. \tag{4.51}$$

The canonical dual function has the form of

$$P^g(\sigma_1, \sigma_2) = -\frac{1}{2}[c_1, c_2] \begin{bmatrix} q_1 + \sigma_1 & q_3 \\ q_3 & q_2 + \sigma_2 \end{bmatrix}^+ \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \frac{1}{2}(\sigma_1 + \sigma_2). \tag{4.52}$$

Three cases to be considered.

**Case I.** $q_1 \le 0$, $q_2 \le 0$, and $q_3 = 0$. In this case, $P(\mathbf{x})$ is concave. If we let $\mathbf{f} = [0.1, -0.3]^T$, $q_1 = -0.5$, $q_2 = -0.6$, the dual function $P^g(\boldsymbol{\sigma}) = P^d(\boldsymbol{\sigma})$ has four critical points:

$$\boldsymbol{\sigma}_1 = [0.6, 0.9]^T, \ \boldsymbol{\sigma}_2 = [0.4, 0.3]^T, \ \boldsymbol{\sigma}_3 = [0.4, 0.9]^T, \ \boldsymbol{\sigma}_4 = [0.6, 0.3]^T.$$

Since

$$\boldsymbol{\sigma}_1 = [0.6, 0.9]^T \in \mathcal{S}_a^+ = \{\boldsymbol{\sigma} \in \mathbb{R}^2 | \ \sigma_1 > 0.5, \ \sigma_2 > 0.6\},$$

by Theorem 4.4, $\mathbf{x}_1 = [c_1/(q_1 + \sigma_1), c_2/(q_2 + \sigma_2)]^T = [1.0, -1.0]^T$ is a global minimizer, and

$$P(\mathbf{x}_1) = P^d(\boldsymbol{\sigma}_1) = -0.95.$$

See Fig. 4.3.

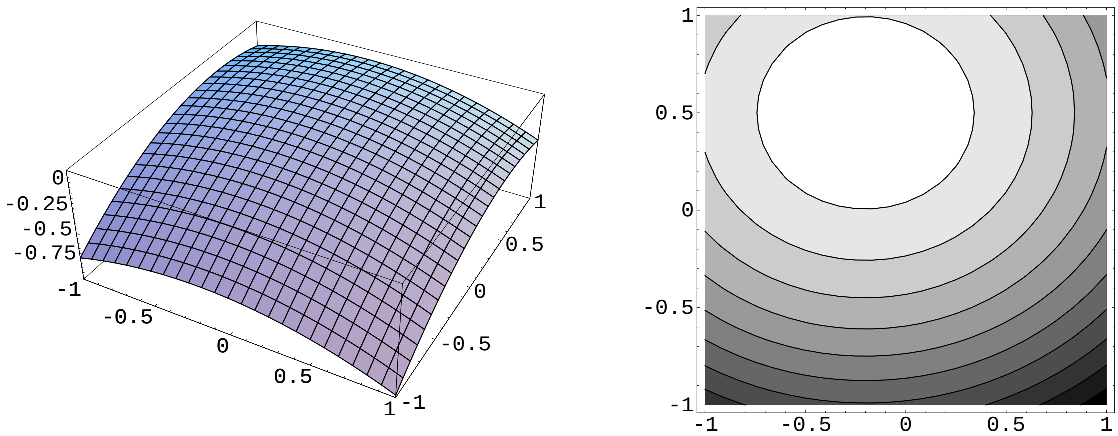



Figure 4.3: Graph of the concave function $P(x_1, x_2)$ and its contour for Example 4.2 (I)

**Case II**. $q_1 \le 0$, $q_2 \ge 0$, and $q_3 = 0$. In this case, $P(\mathbf{x})$ is a saddle function. If we let $\mathbf{f} = [0.1, -0.3]^T$, $q_1 = -0.5$, $q_2 = 0.3$, the dual function $P^d$ has four critical points

$$\boldsymbol{\sigma}_1 = [0.6, 0.0]^T, \ \boldsymbol{\sigma}_2 = [0.4, 0.0]^T, \ \boldsymbol{\sigma}_3 = [0.4, -0.6]^T, \ \boldsymbol{\sigma}_4 = [0.6, -0.6]^T.$$

Since $\boldsymbol{\sigma}_1 \in \mathcal{S}_a^+ = \{\boldsymbol{\sigma} \in \mathbb{R}^2 | \ \sigma_1 > 0.5, \ \ \sigma_2 \geq 0\}$ is a KKT point, by Theorem 4.4, we know that $\mathbf{x}_1 = [1.0, -1.0]^T \in \mathcal{X}_a$ is a global minimizer. It is easy to verify that

$$P(\mathbf{x}_1) = P^d(\boldsymbol{\sigma}_1) = -0.5.$$

See Fig. 4.4.

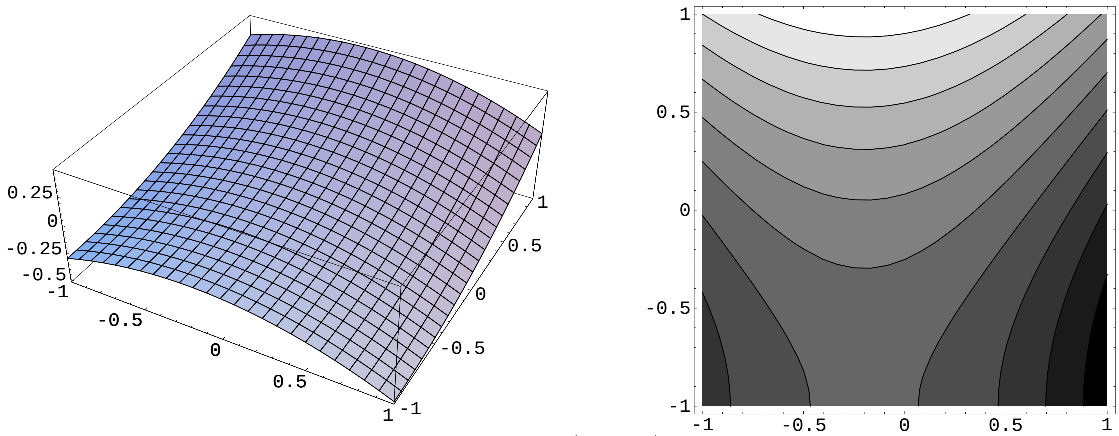



Figure 4.4: Graph of the saddle function $P(x_1, x_2)$ and its contour for Example 4.2 (II)

**Case III**. General matrix $Q \in \mathbb{R}^{2\times 2}$ with integer solutions.

Let $\mathbf{f} = [1, -2]^T$, $\ q_1 = -2.$, $\ q_2 = -1$, $\ q_3 = -3$. In this case, the eigenvalues of $Q$ are $\{-4.54138, 1.54138\}$. This implies that the primal problem is nonconvex. The dual problem has four critical points

$$\boldsymbol{\sigma}_1 = [4, 6]^T, \ \ \boldsymbol{\sigma}_2 = [6, 2]^T, \ \ \boldsymbol{\sigma}_3 = [0, 0]^T, \ \ \boldsymbol{\sigma}_4 = [-2, -4]^T,$$

from which, we have

$$\mathbf{x}_1 = [-1, -1]^T, \ \ \mathbf{x}_2 = [1, 1]^T, \ \ \mathbf{x}_3 = [1, -1]^T, \ \ \mathbf{x}_4 = [-1, 1]^T.$$

on the four corners of the box $\mathcal{X}_a = \{\mathbf{x} \in \mathbb{R}^2 | \ \ -1 \leq x_1 \leq 1, \ \ -1 \leq x_2 \leq 1\}$. Since $\boldsymbol{\sigma}_1 \in \mathcal{S}_a^+$, we know that $\mathbf{x}_1 \in \mathcal{X}_a$ is a global minimizer (see Fig. 4.5), and

$$P(\mathbf{x}_1) = -5.5 < P(\mathbf{x}_2) = -3.5 < P(\mathbf{x}_3) = -1.5 < P(\mathbf{x}_4) = 4.5.$$

**Case IV**. General matrix $Q \in \mathbb{R}^{2\times 2}$ with mixed solutions.

We choose $q_1 = -4, \ \ q_2 = 10, \ \ q_3 = -2$, the eigenvalues of $Q$ are $\{10.3, -4.3\}$, i.e. the primal problem is nonconvex. If we let $\mathbf{f} = [-8, 10]^T$, the dual solution is $\boldsymbol{\sigma} = [10.4, 0]^T \in \mathcal{S}_a^+$. Since $\sigma_2 = 0$, the constraint $-1 \leq x_2 \leq 1$ is inactive. The corresponding primal solution $\mathbf{x} = [-1.0, 0.8]^T$ is not on the corner of the feasible set $\mathcal{X}_a$

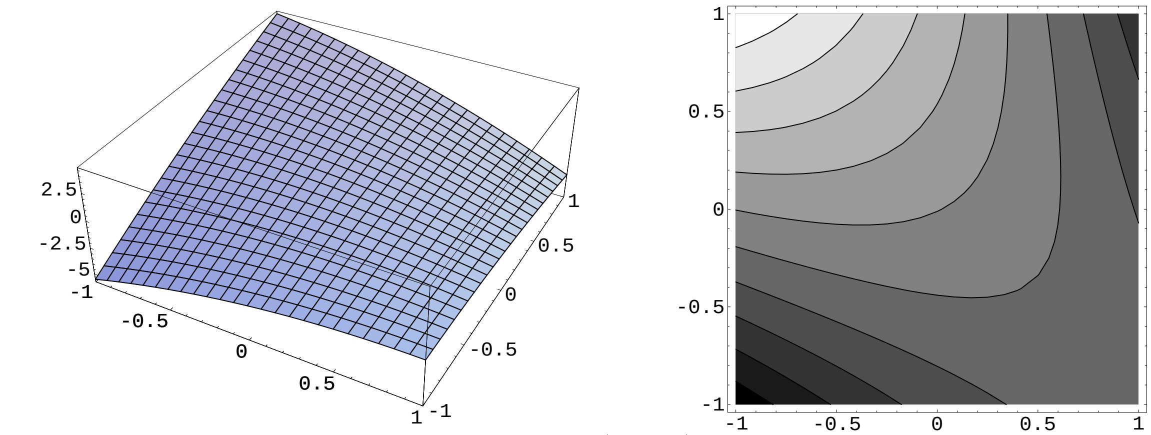


Figure 4.5: Graph of the saddle function $P(x_1, x_2)$ and its contour for Example 4.2 (III)

(see Fig. 4.6), but we still have $P(\mathbf{x}) = -1.3 = P^d(\boldsymbol{\sigma})$.

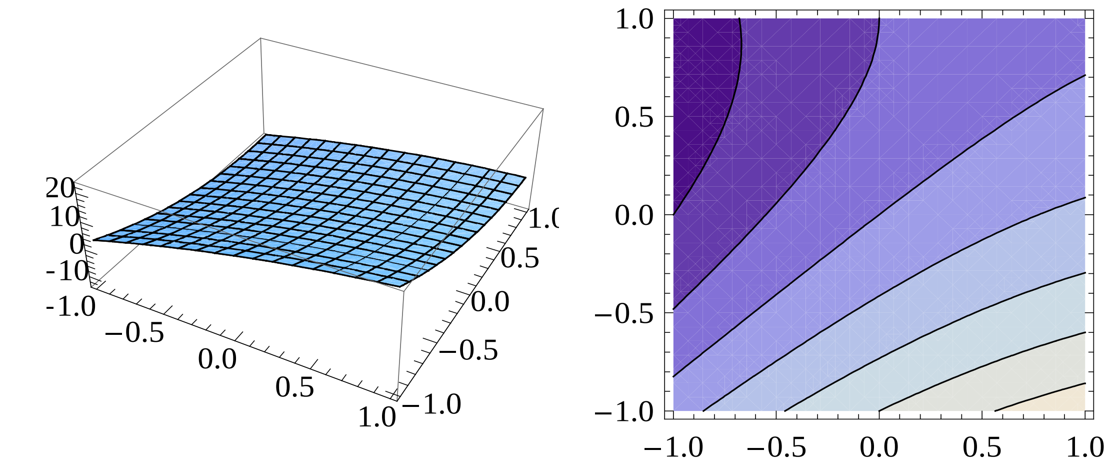


Figure 4.6: Graph of the function $P(x_1, x_2)$ and its contour for Example 4.2 (IV)

**Example 4.3. High Dimensional Integer Programming Problem**
We now let $n = 10$ and randomly choose $Q$ and $\mathbf{f}$ as given below:

$$Q = \begin{bmatrix} -6 & 2 & -1 & -3 & 1 & 1 & -3 & -3 & 0 & -1 \\ 2 & -10 & -1 & 2 & 1 & 0 & 2 & 1 & -3 & -4 \\ -1 & -1 & -5 & 0 & 3 & -1 & 1 & 0 & -1 & -4 \\ -3 & 2 & 0 & -6 & 1 & 1 & 1 & -2 & 0 & 0 \\ 1 & 1 & 3 & 1 & -7 & 0 & -4 & -1 & -1 & 2 \\ 1 & 0 & -1 & 1 & 0 & -6 & -2 & 1 & 3 & -1 \\ -3 & 2 & 1 & 1 & -4 & -2 & -8 & -1 & 0 & 0 \\ -3 & 1 & 0 & -2 & -1 & 1 & -1 & -3 & 0 & 0 \\ 0 & -3 & -1 & 0 & -1 & 3 & 0 & 0 & -7 & -4 \\ -1 & -4 & -4 & 0 & 2 & -1 & 0 & 0 & -4 & -6 \end{bmatrix},$$

$$\mathbf{f} = [-9.49, 6.14, 9.13, 0.0525, -2.54, 6.69, 0.847, -8.36, 6.31, -2.69]^T.$$

To use the direct enumeration method to solve this problem, it is required $2^{10}$ times of enumerations. However, by using the canonical dual problem, it takes few iterations to obtain the global maximizer:

$$\boldsymbol{\sigma} = 2[12.2, 16.0, 12.0, 6.0, 8.8, 6.3, 7.6, 10.2, 8.7, 8.7]^T.$$

The global minimizer of the primal problem $(\mathcal{P})$ is then

$$\mathbf{x} = [-1, 1, 1, -1, -1, 1, -1, -1, 1, 1]^T$$

and $P^d(\boldsymbol{\sigma}) = -119.1 = P(\mathbf{x})$.

## 4.8 Conclusions

We have presented a detailed application of the canonical duality theory to solving box and integer constrained quadratic optimization problems. By using the canonical dual transformation, several canonical dual problems and their perturbations are proposed. Since the canonical dual problem $(\mathcal{P}^g_{\max})$ is a smooth concave maximization problem [19] over convex feasible spaces, it is not difficult to solve for certain given $\mathbf{Q}$ and $\mathbf{f}$. Existence and uniqueness criteria are established. If $\mathbf{Q}$ and $\mathbf{f}$ satisfy certain appropriate conditions, the unique analytical solution can be obtained.

Theorem 4.4 is particularly useful, which shows that for any given $\mathbf{Q}$ and $\mathbf{f}$, the discrete integer constrained problem $(\mathcal{P}_{ip})$ is equivalent to the continuous unconstrained canonical dual problem $(\mathcal{P}^d_\alpha)$. For convex-perturbation $\mathbf{Q} + \text{Diag}\ (\boldsymbol{\alpha}) \succ 0$, if the concave maximization problem

$$(\mathcal{P}^\sharp_\alpha): \quad \max\{P^d_\alpha(\boldsymbol{\sigma})|\ \boldsymbol{\sigma} \in \mathbb{R}^n\} \tag{4.53}$$

has a critical solution, the discrete problem $(\mathcal{P}_{ip})$ can be solved uniquely. Otherwise, the nonsmooth problem $(\mathcal{P}^\sharp_\alpha)$ provides a lower bound for the box constrained problem $(\mathcal{P})$.

# CHAPTER 5

# Discrete Value Selection Problem

## 5.1 Introduction

Many decision making problems, such as portfolio selection, capital budgeting, production planning, resource allocation, and computer networks can often be formulated as quadratic programming problems with discrete variables. See for examples, [8] [21] [31] [62]. In some engineering applications, the variables of these optimization problems are not allowed to have arbitrary values. Instead, some or all of the variables must be selected from a set of integers or discrete values [99]. For examples, structural members may have to be selected from selections available in standard sizes, their thicknesses are required 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 structure must be an integer. However, these integer programming problems are computationally highly demanding.

Several survey articles on nonlinear optimization problems with discrete variables have been published [66] [83] [95]. Furthermore, some popular methods have been proposed, which include branch and bound methods [15] [18] [52], branch and cut method [63] [93], a hybrid method that combines a branch-and-bound method with a dynamic programming technique [68], sequential linear programming, rounding-off techniques, cutting plane techniques [11] [77], heuristic techniques, penalty function approach and sequential linear programming. The relaxation method has also been proposed, leading to second order cone programming (SOC) [47]. More recently, simulated annealing [54] [59] [69] and genetic algorithms [57] have been proposed.

In this chapter, our goal is to solve a general quadratic programming problem with its decision variables taking values from discrete sets [30]. 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.

## 5.2 Problem Statement

The discrete programming problem to be addressed is given below:

$$(\mathcal{P}_a) \quad \text{Minimize } P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{c}^T\mathbf{x} \tag{5.1}$$

$$\text{subject to } \mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b} \leq 0, \tag{5.2}$$

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

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

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

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

Problem $(\mathcal{P}_a)$ arises in many real-world applications, such as the pipe network optimisation problems in water distribution systems, where the choices of pipelines are discrete values. Such problems have been studied extensively [101]. Due to the constraint of discrete values, this problem is considered to be NP-hard. In this chapter, we will show that the canonical duality theory will provide a lower bound approach to this challenging problem. Furthermore, the global optimal solution could be obtained under certain conditions.

## 5.3 Equivalent Transformation

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

$$x_i = \sum_{j=1}^{K_i} u_{i,j} y_{i,j},\ i = 1, \cdots, n, \tag{5.3}$$

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

$$
\begin{aligned}
(\mathcal{P}_b) \quad & \text{Minimize } P(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} && (5.4)\\
& \text{subject to } \mathbf{g}(\mathbf{y}) = D\mathbf{y} - \mathbf{b} \leq 0, && (5.5)\\
& \qquad \sum_{j=1}^{K_i} y_{i,j} - 1 = 0,\ i = 1, \cdots, n, && (5.6)\\
& \qquad y_{i,j} \in \{0, 1\},\ i = 1, \ldots, n;\ j = 1, \cdots, K_i, && (5.7)
\end{aligned}
$$

where

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

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

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

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

**Theorem 5.1.** *Problem $(\mathcal{P}_b)$ is equivalent to Problem $(\mathcal{P}_a)$.*

*Proof.* For any $i = 1, 2, \cdots, n$, it is clear that constraints (5.6) and (5.7) are equivalent to the existence of only one $j \in \{1, \cdots, K_i\}$, 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. □

Problem $(\mathcal{P}_b)$ is a standard 0-1 quadratic programming problem with both equality and inequality constraints. In order to use the canonical duality theory for solving this

NP-hard problem, we need to reform the integer constraint in the canonical form. 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 any integer $N$, let

$$\mathbf{e}_N = [1, \cdots, 1, \cdots, 1, \cdots, 1]^T \in \mathbb{R}^N.$$

By the fact that the solution to the quadratic equation $y_i(y_i - 1) = 0$ must be either 0 or 1, the integer constrained problem $(\mathcal{P}_b)$ can be reformulated to the following quadratic programming problem:

$$(\mathcal{P}) \quad \text{Minimize } P(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} \tag{5.8}$$

$$\text{subject to } \mathbf{g}(\mathbf{y}) = D\mathbf{y} - \mathbf{b} \leq 0, \tag{5.9}$$

$$H\mathbf{y} - \mathbf{e}_n = 0, \tag{5.10}$$

$$\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) = 0, \tag{5.11}$$

where the notation $\mathbf{s} \circ \mathbf{t} := [s_1 t_1, s_2 t_2, \ldots, s_K t_K]^T$ denotes the Hadamard product for any two vectors $\mathbf{s}, \mathbf{t} \in \mathbb{R}^K$.

## 5.4 Canonical Dual Transformation

Now we apply the canonical duality theory to integer programming problem presented in Section 5.2. Let

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

and define

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

where $\Lambda$ is the so-called geometric operator. Let

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

Let $\boldsymbol{\varsigma} = [(\boldsymbol{\sigma})^T, (\boldsymbol{\tau})^T, (\boldsymbol{\mu})^T]^T \in \mathcal{S} = \mathbb{R}^{m+n+K}$ be the canonical dual variable corresponding to $\boldsymbol{\xi} \in Z = \{(\boldsymbol{\epsilon}, \boldsymbol{\delta}, \boldsymbol{\rho}) : \boldsymbol{\epsilon} \leq 0, \boldsymbol{\delta} = 0, \boldsymbol{\rho} = 0\}$. Then, the Fenchel super-conjugate of the function $W(\boldsymbol{\xi})$ is defined by

$$\begin{aligned} W^\sharp(\boldsymbol{\varsigma}) &= \sup\{\boldsymbol{\xi}^T\boldsymbol{\varsigma} - W(\boldsymbol{\xi}) : \ \boldsymbol{\xi} \in Z\} \\ &= \begin{cases} 0 & \text{if } \boldsymbol{\sigma} \geq 0 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned} \tag{5.12}$$

Let

$$\mathbf{G}(\boldsymbol{\mu}) = B + 2\text{Diag }(\boldsymbol{\mu}), \tag{5.13}$$

and

$$\mathbf{F}(\boldsymbol{\varsigma}) = \mathbf{h} - D^T\boldsymbol{\sigma} - H^T\boldsymbol{\tau} + \boldsymbol{\mu}. \tag{5.14}$$

Then, the total complementary function can be obtained as:

$$\begin{aligned} \Xi(\mathbf{y}, \boldsymbol{\varsigma}) &= \langle \Lambda(\mathbf{y}), \boldsymbol{\varsigma} \rangle - W^\sharp(\boldsymbol{\varsigma}) - U(\mathbf{y}) \\ &= \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} + \boldsymbol{\sigma}^T(D\mathbf{y} - \mathbf{b}) \\ &\quad + \boldsymbol{\tau}^T(H\mathbf{y} - \mathbf{e}_n) + \boldsymbol{\mu}^T(\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K)) \\ &= \frac{1}{2}\mathbf{y}^T\mathbf{G}(\boldsymbol{\mu})\mathbf{y} - \mathbf{F}^T(\boldsymbol{\varsigma})\mathbf{y} - \boldsymbol{\sigma}^T\mathbf{b} - \boldsymbol{\tau}^T\mathbf{e}_n. \end{aligned}$$

The critical condition $\nabla_{\mathbf{y}}\Xi(\mathbf{y}, \boldsymbol{\varsigma}) = 0$ leads to

$$\mathbf{G}(\boldsymbol{\mu})\mathbf{y} = \mathbf{F}(\boldsymbol{\varsigma}). \tag{5.15}$$

Let

$$\mathcal{S}_a = \{\boldsymbol{\varsigma} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{S} = \mathbb{R}^{m+n+K} : \ \ \boldsymbol{\sigma} \geq 0, \ \boldsymbol{\mu} \neq 0\}. \tag{5.16}$$

Therefore, the canonical dual problem can be formulated as follows:

$$\begin{aligned} (\mathcal{P}^d) \quad & \text{Maximize } \left\{ P^d(\boldsymbol{\varsigma}) = -\frac{1}{2}\mathbf{F}^T(\boldsymbol{\varsigma})\mathbf{G}^+(\boldsymbol{\mu})\mathbf{F}(\boldsymbol{\varsigma}) - \boldsymbol{\sigma}^T\mathbf{b} - \boldsymbol{\tau}^T\mathbf{e}_n \right\}, \\ & \text{subject to } \ \boldsymbol{\varsigma} \in \mathcal{S}_a. \end{aligned}$$

**Theorem 5.2** (Complementary-Dual Principle)**.** *Problem $(\mathcal{P}^d)$ is a canonically dual to Problem $(\mathcal{P})$ in the sense that if $\bar{\boldsymbol{\varsigma}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$ is a KKT solution of Problem $(\mathcal{P}^d)$, then the vector*

$$\bar{\mathbf{y}}(\bar{\boldsymbol{\varsigma}}) = \mathbf{G}^+(\bar{\boldsymbol{\mu}})\mathbf{F}(\bar{\boldsymbol{\varsigma}}) \tag{5.17}$$

*is a KKT solution of Problem* $(\mathcal{P})$ *and*

$$P(\bar{\mathbf{y}}) = P^d(\bar{\boldsymbol{\varsigma}}).$$

*Proof.* By introducing the Lagrange multiplier vectors $\boldsymbol{\epsilon} \leq 0 \in \mathbb{R}^m$, and $\boldsymbol{\rho} \in \mathbb{R}^K$ to relax the inequality constraints $\boldsymbol{\sigma} \geq 0$ and $\boldsymbol{\mu} \neq 0$, respectively, the Lagrangian function associated with the dual function $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu})$ becomes

$$L(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\epsilon}, \boldsymbol{\rho}) = P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) - \boldsymbol{\epsilon}^T \boldsymbol{\sigma} - \boldsymbol{\rho}^T \boldsymbol{\mu}.$$

Then, in terms of $\mathbf{y} = G^+(\boldsymbol{\mu})\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu})$ the KKT conditions of the dual problem become

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

They can be written as:

$$D\mathbf{y} \leq \mathbf{b}, \tag{5.18}$$

$$H\mathbf{y} - \mathbf{e}_n = 0, \tag{5.19}$$

$$\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) = 0, \tag{5.20}$$

$$\boldsymbol{\sigma} \geq 0, \ \ \boldsymbol{\sigma}^T (D\mathbf{y} - \mathbf{b}) = 0, \tag{5.21}$$

$$\boldsymbol{\mu} \neq 0, \ \ \mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) = 0. \tag{5.22}$$

This proves that if $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$ is a KKT solution of $(\mathcal{P}^d)$, then (5.18)-(5.20) are the so-called primal feasibility conditions, while (5.21)-(5.22) are the so-called dual feasibility conditions. Therefore, the vector

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

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

Again, by the complementary conditions and (5.17), we have

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

Therefore, the theorem is proved. □

**Remark 5.1.** Since the inequality constraint $\boldsymbol{\mu} \neq 0$ in the canonical dual problem $(\mathcal{P}^d)$ produces a nonconvex feasible set, this constraint can be replaced by either $\boldsymbol{\mu} < 0$ or $\boldsymbol{\mu} > 0$. Since 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, it is reasonable to let $\boldsymbol{\mu} > 0$. In this case, the KKT condition (5.22) should be replaced by

$$\boldsymbol{\mu} > 0, \;\; \mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) \leq 0, \;\; \boldsymbol{\mu}^T [\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K)] = 0. \tag{5.23}$$

Therefore, as long as $\boldsymbol{\mu} \neq 0$ is satisfied, the complementarity condition in (5.23) leads to the integer condition $\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K) = 0$.

## 5.5 Global Optimality Criteria

To continue, let the feasible space $\mathcal{Y}$ of Problem $(\mathcal{P})$ and the dual feasible space $\mathcal{Z}$ be defined by

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

and

$$\mathcal{S}_a = \{\boldsymbol{\varsigma} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{S} : \boldsymbol{\sigma} \geq 0, \boldsymbol{\mu} > 0\},$$

respectively. Furthermore, we introduce a subset of the dual feasible space:

$$\mathcal{S}_a^+ := \{\boldsymbol{\varsigma} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{S}_a : \mathbf{G}(\boldsymbol{\mu}) \succ 0\}. \tag{5.24}$$

We have the following theorem.

**Theorem 5.3.** *Assume that* $\bar{\boldsymbol{\varsigma}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$ *is a KKT point of* $P^d(\boldsymbol{\varsigma})$ *and* $\bar{\mathbf{y}} = \mathbf{G}^+(\bar{\boldsymbol{\mu}})\mathbf{F}(\bar{\boldsymbol{\varsigma}})$.

*If $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a^+$, then $\bar{\mathbf{y}}$ is a global minimizer of $P(\mathbf{y})$ and $\bar{\boldsymbol{\varsigma}}$ is a global maximizer of $P^d(\boldsymbol{\varsigma})$ with*

$$P(\bar{\mathbf{y}}) = \min_{\mathbf{y}\in\mathcal{Y}} P(\mathbf{y}) = \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}) = P^d(\bar{\boldsymbol{\varsigma}}). \tag{5.25}$$

*Proof* The canonical dual function $P^d(\boldsymbol{\varsigma})$ is concave on $\mathcal{S}_a^+$. Therefore, a KKT point $\bar{\boldsymbol{\varsigma}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}}) \in \mathcal{S}_a^+$ must be a global maximizer of $P^d(\boldsymbol{\varsigma})$ on $\mathcal{S}_a^+$. For any given $\boldsymbol{\varsigma} \in \mathcal{S}_a^+$, the complementary function $\Xi(\mathbf{y}, \boldsymbol{\varsigma})$ is convex in $\mathbf{y}$ and concave in $(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu})$, the critical point $(\bar{\mathbf{y}}, \bar{\boldsymbol{\varsigma}})$ is a saddle point of the complementary function. More specifically, we have

$$\begin{aligned}
P^d(\bar{\boldsymbol{\varsigma}}) &= \max_{(\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\mu})\in\mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}) \\
&= \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \min_{\mathbf{y}\in\mathcal{Y}} \Xi(\mathbf{y}, \boldsymbol{\varsigma}) = \min_{\mathbf{y}\in\mathcal{Y}} \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \Xi(\mathbf{y}, \boldsymbol{\varsigma}) \\
&= \min_{\mathbf{y}\in\mathcal{Y}} \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \left\{ \frac{1}{2}\mathbf{y}^T \mathbf{G}(\boldsymbol{\mu})\mathbf{y} - (\mathbf{h} - D^T\boldsymbol{\sigma} - H^T\boldsymbol{\tau} + \boldsymbol{\mu})^T\mathbf{y} - \boldsymbol{\sigma}^T\mathbf{b} - \boldsymbol{\tau}^T\mathbf{e}_n \right\} \\
&= \min_{\mathbf{y}\in\mathcal{Y}} \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \left\{ \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} + \boldsymbol{\sigma}^T(D\mathbf{y} - \mathbf{b}) + \boldsymbol{\tau}^T(H\mathbf{y} - \mathbf{e}_n) + \boldsymbol{\mu}^T[\mathbf{y} \circ (\mathbf{y} - \mathbf{e}_K)] \right\} \\
&= \min_{\mathbf{y}\in\mathcal{Y}} \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \{ \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} + \boldsymbol{\varsigma}^T\boldsymbol{\xi} \}.
\end{aligned}$$

Note that

$$\max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \{W^\sharp(\boldsymbol{\varsigma})\} = 0$$

and

$$\max_{\boldsymbol{\xi}\in Z} \{W(\boldsymbol{\xi})\} = 0.$$

Thus, it follows from (5.26) that

$$\begin{aligned}
P^d(\bar{\boldsymbol{\varsigma}}) &= \min_{\mathbf{y}\in\mathcal{Y}} \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \{ \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} + \boldsymbol{\varsigma}^T\boldsymbol{\xi} - W^\sharp(\boldsymbol{\varsigma}) \} \\
&= \min_{\mathbf{y}\in\mathcal{Y}} \{ \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} \} + \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \{ \boldsymbol{\varsigma}^T\boldsymbol{\xi} - W^\sharp(\boldsymbol{\varsigma}) \} \\
&= \min_{\mathbf{y}\in\mathcal{Y}} \{ \frac{1}{2}\mathbf{y}^T B\mathbf{y} - \mathbf{h}^T\mathbf{y} \} \\
&= \min_{\mathbf{y}\in\mathcal{Y}} P(\mathbf{y}).
\end{aligned}$$

This completes the proof. □

**Remark 5.2.** By the fact that the inequality $\boldsymbol{\mu} \neq 0$ in Problem $(\mathcal{P}^d)$ is replaced by the unilateral inequality $\boldsymbol{\mu} > 0$ in the convex feasible set $\mathcal{S}_a^+$, the canonical dual function

$P^d(\boldsymbol{\varsigma})$ may have no KKT point in the (semi) open convex domain $\mathcal{S}_a^+$. If we let

$$\mathcal{S}_c^+ = \{\boldsymbol{\varsigma} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathcal{S}_a^+ : \quad \boldsymbol{\mu} \geq 0\},$$

then on this closed convex domain, the concave maximization problem

$$(\mathcal{P}^\sharp) \quad \max\{P^d(\boldsymbol{\varsigma}) | \ \boldsymbol{\varsigma} \in \mathcal{S}_c^+\} \tag{5.26}$$

has at least one solution $\bar{\boldsymbol{\varsigma}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}})$. If the corresponding $\bar{\mathbf{y}} = \mathbf{G}^+(\bar{\boldsymbol{\mu}})\mathbf{F}(\bar{\boldsymbol{\varsigma}})$ is feasible, then $\bar{\mathbf{y}}$ is a global minimizer of the primal problem $(\mathcal{P})$. Otherwise, the value $P(\bar{\mathbf{y}})$ provides a lower bound to the primal problem $(\mathcal{P})$. This is one of the main advantages of the canonical duality theory.

## 5.6 Numerical Examples

All data and computational results presented in this section are produced within Matlab environment. For proper display, all the elements of a matrix are rounded to two decimals.

**Example 5.1. 5-dimensional problem.**

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

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

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

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

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

Under the transformation (5.3), this problem is transformed into the 0-1 programming

Problem $(\mathcal{P})$, where

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

$$B = \begin{bmatrix}
13.71 & 20.56 & 34.27 & 2.40 & 3.61 & 6.01 & 1.58 & 2.37 & 3.95 & 0.39 & 0.58 & 0.97 & 2.38 & 3.57 & 5.95 \\
20.56 & 30.84 & 51.41 & 3.61 & 5.41 & 9.01 & 2.37 & 3.55 & 5.92 & 0.58 & 0.88 & 1.46 & 3.57 & 5.36 & 8.93 \\
34.27 & 51.41 & 85.68 & 6.01 & 9.01 & 15.02 & 3.95 & 5.92 & 9.87 & 0.97 & 1.46 & 2.43 & 5,95 & 8.93 & 14.88 \\
2.40 & 3.61 & 6.01 & 11.05 & 16.57 & 27.61 & 1.27 & 1.91 & 3.18 & 2.61 & 3.91 & 6.52 & 1.95 & 2.93 & 4.88 \\
3.61 & 5.41 & 9.01 & 16.57 & 24.85 & 41.42 & 1.91 & 2.86 & 4.77 & 3.91 & 5.87 & 9.78 & 2.93 & 4.39 & 7.32 \\
6.01 & 9.01 & 15.02 & 27.61 & 41.42 & 69.03 & 3.18 & 4.77 & 7.96 & 6.52 & 9.78 & 16.31 & 4.88 & 7.32 & 12.20 \\
1.58 & 2.37 & 3.95 & 1.27 & 1.91 & 3.18 & 8.27 & 12.40 & 20.67 & 2.37 & 3.55 & 5.92 & 1.57 & 2.36 & 3.93 \\
2.37 & 3.55 & 5.92 & 1.91 & 2.86 & 4.77 & 12.40 & 18.60 & 31.00 & 3.55 & 5.33 & 8.89 & 2.36 & 3.53 & 5.90 \\
3.95 & 5.92 & 9.87 & 3.18 & 4.77 & 7.96 & 20.67 & 31.00 & 51.67 & 5.92 & 8.86 & 14.81 & 3.93 & 5.90 & 9.83 \\
0.39 & 5.58 & 0.97 & 2.61 & 3.91 & 6.52 & 2.37 & 3.55 & 5.92 & 10.50 & 15.74 & 26.24 & 1.20 & 1.80 & 3.00 \\
0.58 & 0.88 & 1.46 & 3.91 & 5.87 & 9.78 & 3.55 & 5.33 & 8.89 & 15.74 & 23.62 & 39.36 & 1.80 & 2.70 & 4.50 \\
0.97 & 1.46 & 2.43 & 6.52 & 9.78 & 16.31 & 5.92 & 8.89 & 14.81 & 26.24 & 39.36 & 65.60 & 3.00 & 4.50 & 7.51 \\
2.38 & 3.57 & 5.95 & 1.95 & 2.93 & 4.88 & 1.57 & 2.36 & 3.93 & 1.20 & 1.80 & 3.00 & 13.35 & 20.02 & 33.37 \\
3.57 & 5.36 & 8.93 & 2.93 & 4.39 & 7.32 & 2.36 & 3.54 & 5.90 & 1.80 & 2.70 & 4.50 & 20.02 & 30.04 & 50.06 \\
5.95 & 8.93 & 14.88 & 4.88 & 7.32 & 12.20 & 3.93 & 5.90 & 9.83 & 3.00 & 4.50 & 7.51 & 33.37 & 50.06 & 83.43
\end{bmatrix},$$

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

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

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

The canonical dual problem can be stated as follows:

$$\begin{aligned}
(\mathcal{P}^d) \quad & \text{Maximize } \; P^d(\boldsymbol{\varsigma}) = -\frac{1}{2}\mathbf{F}(\boldsymbol{\varsigma})^T\mathbf{G}^+(\boldsymbol{\mu})\mathbf{F}(\boldsymbol{\varsigma}) - \boldsymbol{\sigma}^T\mathbf{b} - \boldsymbol{\tau}^T\mathbf{e}_5 \\
& \text{subject to } \; \boldsymbol{\varsigma} = (\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\mu}) \in \mathbb{R}^{4+5+15}, \;\; \boldsymbol{\sigma} \geq 0, \boldsymbol{\mu} > 0,
\end{aligned}$$

where $F(\boldsymbol{\varsigma})$ and $G(\boldsymbol{\mu})$ are as defined by (5.13) and (5.14), respectively.

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

$$\begin{aligned} \bar{\boldsymbol{\sigma}} &= [0, 0, 0, 0]^T, \\ \bar{\boldsymbol{\tau}} &= [73.90, -106.70, 111.95, -59.27, -0.01]^T, \end{aligned}$$

and

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

It is clear that $\bar{\boldsymbol{\varsigma}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}}) \in \mathcal{S}_c^+$. Thus, from Theorem 5.3,

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

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

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

to give

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

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

**Example 5.2. 10-dimensional problem**.

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

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

$$\mathbf{f} = [0.89, 0.03, 0.49, 0.17, 0.98, 0.71, 0.50, 0.47, 0.06, 0.68]^T,$$

$$\mathbf{A} = \begin{bmatrix} 0.04 & 0.82 & 0.97 & 0.83 & 0.83 & 0.42 & 0.02 & 0.20 & 0.05 & 0.94 \\ 0.07 & 0.72 & 0.65 & 0.08 & 0.80 & 0.66 & 0.98 & 0.49 & 0.74 & 0.42 \\ 0.52 & 0.15 & 0.80 & 0.13 & 0.06 & 0.63 & 0.17 & 0.34 & 0.27 & 0.98 \\ 0.10 & 0.66 & 0.45 & 0.17 & 0.40 & 0.29 & 0.11 & 0.95 & 0.42 & 0.30 \\ 0.82 & 0.52 & 0.43 & 0.39 & 0.53 & 0.43 & 0.37 & 0.92 & 0.55 & 0.70 \end{bmatrix},$$

$$\mathbf{b} = [33.76, 37.07, 26.75, 25.46, 37.36]^T.$$

By solving the canonical dual problem of Problem $(\mathcal{P}_a)$, we obtain

$$\begin{aligned} \bar{\boldsymbol{\sigma}} &= [0, 0, 0, 0, 0]^T, \\ \bar{\boldsymbol{\tau}} &= [-19.99, -20.12, -18.13, -18.37, -14.32, \\ &\quad -17.13, -18.46, -19.73, -17.65, -16.55]^T, \end{aligned}$$

and

$$\begin{aligned} \bar{\boldsymbol{\mu}} = \; & [9.51, 0.97, 21.93, 53.36, 74.34, 9.95, 0.21, 20.53, 51.01, 71.35 \\ & 8.68, 0.77, 19.68, 48.03, 66.94, 8.30, 1.77, 21.91, 52.13, 72.27 \\ & 6.40, 1.54, 17.39, 41.19, 57.04, 7.57, 1.98, 21.10, 49.77, 68.90 \\ & 9.15, 0.16, 18.79, 46.72, 65.34, 9.82, 0.09, 19.90, 49.63, 69.45 \\ & 8.76, 0.13, 17.92, 44.60, 62.39, 6.26, 4.03, 24.60, 55.48, 76.04]^T, \end{aligned}$$

It is clear that $\bar{\boldsymbol{\varsigma}} = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \bar{\boldsymbol{\mu}}) \in \mathcal{S}_c^+$. Therefore,

$$\begin{aligned} \bar{\mathbf{y}} = [&1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0, \\ &1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0,1,0,0,0,0]^T \end{aligned}$$

is the global minimizer of the problem $(\mathcal{P})$ with $P^d(\bar{\boldsymbol{\varsigma}}) = 45.54 = P(\bar{\mathbf{y}})$. The solution to the original primal problem is

$$\bar{\mathbf{x}} = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]^T$$

with $P(\bar{\mathbf{x}}) = 45.54$.

**Example 5.3. Relatively large size problems**.

Consider Problem $(\mathcal{P}_a)$ with $n = 20$, 50, 100, 200 and 300. Let these five problems be referred to as Problem (1), $\cdots$, Problem (5), respectively. Their coefficients are generated randomly with uniform distribution. For each problem, $q_{ij} \in (0,1)$, $a_{ij} \in (0,1)$, for $i = 1, \cdots, n$; $j = 1, \cdots, n$, and $c_i \in (0,1)$, $x_i \in \{1,2,3,4,5\}$, for $i = 1, \cdots n$. 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 such that it is diagonally dominated. For each $x_i$, its lower bound is $l_i = 1$, and its upper bound is $u_i = 5$. Let $l = [l_1, \cdots, l_n]^T$ and $u = [u_1, \cdots, u_n]^T$. 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 5.1 presents the numerical results, where $m$ is the number of linear constraints in Problem $(\mathcal{P}_a)$.

Table 5.1: Numerical results for large scale integer programming problems

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

From Table 5.1, 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.

## 5.7 Conclusion

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

# CHAPTER 6

# Fix Charge Problem

This chapter presents a canonical dual approach for solving a mixed-integer quadratic minimization problem with fixed cost terms [65]. We show that this well-known NP-hard problem in $\mathbb{R}^{2n}$ can be transformed into a continuous concave maximization dual problem over a convex feasible subset of $\mathbb{R}^n$ with zero duality gap. We also discuss connections between the proposed canonical duality theory approach and the classical Lagrangian duality approach. The resulting canonical dual problem can be solved, under certain conditions, by traditional convex programming methods. It turns out that an analytical solution for the mixed integer programming problem is obtained. Conditions for the existence and uniqueness of global optimal solutions are presented. An application to a decoupled mixed-integer problem is used to illustrate the derivation of analytic solutions for both globally minimizing and maximizing the objective function. Numerical examples for both decoupled and general mixed-integral problems are presented, and an open problem is proposed for future study.

## 6.1 Problem Statement

In this chapter, we address the following quadratic, mixed-integer fixed-charge problem [7] [53]:

$$(\mathcal{P}_\flat): \quad \min\left\{ P(\mathbf{x},\mathbf{v}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{c}^T\mathbf{x} + \mathbf{f}^T\mathbf{v} \;\mid\; (\mathbf{x},\mathbf{v}) \in \mathcal{X}_v \right\} \tag{6.1}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a given (generally indefinite) symmetric matrix, $\mathbf{c},\ \mathbf{f} \in \mathbb{R}^n$ are given vectors, the binary variable vector $\mathbf{v} \in \{0,1\}^n$ represents fixed-cost variables, and the feasible space $\mathcal{X}_v$ is defined by

$$\mathcal{X}_v = \{(\mathbf{x},\mathbf{v}) \in \mathbb{R}^n \times \{0,1\}^n \mid -\mathbf{v} \le \mathbf{x} \le \mathbf{v}\}. \tag{6.2}$$

Problem $(\mathcal{P}_\flat)$ arises in mathematical economics, facility location, and lot-sizing application contexts [2] [12] [48], where the constraints of the form $\mathbf{x} \in [-\mathbf{v},\mathbf{v}]$ with $\mathbf{v} \in \{0,1\}^n$ being referred to as fixed-charge constraints [73]. These types of constraints have received a great deal of attention in the integer programming [26] literature , and many different types of valid inequalities have been developed to deal with this structure [9] [51]. Since we do not assume that the matrix $\mathbf{A}$ is positive semidefinite, the problem remains NP-hard, even with all the fixed-cost variables $v_i$ $(i = 1, \ldots, n)$ fixed to one [55] [72] [75] [87]. In order to numerically solve the latter continuous, box constrained quadratic program, many effective methods have been developed [3] [4] [32] [33] [34] [49] [88] [89] [90]. Naturally, the problem becomes even more challenging with the addition of the fixed-charge feature.

## 6.2 Canonical Dual Transformation

In order to formulate a canonical dual problem that exhibits a zero duality gap, the key step is to rewrite the variable box constraints $-\mathbf{v} \leq \mathbf{x} \leq \mathbf{v},\ \mathbf{v} \in \{0, 1\}^n$ in the (relaxed) quadratic form:

$$\mathbf{x} \circ \mathbf{x} \leq \mathbf{v},\ \mathbf{v} \circ (\mathbf{v} - \mathbf{e}) \leq 0, \tag{6.3}$$

where $\mathbf{e} = \{1\}^n$ is an $n$-vector of all ones and the notation $\mathbf{x} \circ \mathbf{v} := [x_1 v_1,\ x_2 v_2, \ldots,\ x_n v_n]^T$ denotes the Hadamard product for any two vectors $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$. Accordingly, consider the following (continuous relaxation) reformulation of the primal problem $(\mathcal{P}_\flat)$:

$$(\mathcal{P}_r) \quad \min\{P(\mathbf{x}, \mathbf{v}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{c}^T \mathbf{x} + \mathbf{f}^T \mathbf{v} \ :\ \mathbf{x} \circ \mathbf{x} \leq \mathbf{v},\ \mathbf{v} \circ (\mathbf{v} - \mathbf{e}) \leq 0\}. \tag{6.4}$$

Introducing a nonlinear transformation (i.e., the so-called *geometrical mapping*):

$$\mathbf{y} = \Lambda(\mathbf{x}, \mathbf{v}) = \begin{pmatrix} \boldsymbol{\epsilon}(\mathbf{x}) \\ \boldsymbol{\xi}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mathbf{x} \circ \mathbf{x} - \mathbf{v} \\ \mathbf{v} \circ \mathbf{v} - \mathbf{v} \end{pmatrix} \in \mathbb{R}^{2n},$$

the constraints (6.3) can be replaced identically by $\Lambda(\mathbf{x}, \mathbf{v}) \leq 0$. Let

$$V(\mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{y} \leq 0 \in \mathbb{R}^{2n} \\ +\infty & \text{otherwise} \end{cases}$$

and let $\mathbf{y}^* = \begin{pmatrix} \boldsymbol{\sigma} \\ \boldsymbol{\tau} \end{pmatrix} \in \mathbb{R}^{2n}$ be the vector of dual variables associated with the corresponding restrictions $\mathbf{y} \leq 0$. The sup-Fenchel conjugate of $V(\mathbf{y})$ can be defined by

$$\begin{aligned} V^\sharp(\mathbf{y}^*) &= \sup_{\mathbf{y}\in\mathbb{R}^{2n}} \{\langle \mathbf{y}, \mathbf{y}^* \rangle - V(\mathbf{y})\} \\ &= \sup_{\boldsymbol{\epsilon}\in\mathbb{R}^n} \sup_{\boldsymbol{\xi}\in\mathbb{R}^n} \{\boldsymbol{\epsilon}^T\boldsymbol{\sigma} + \boldsymbol{\xi}^T\boldsymbol{\tau} - V(\boldsymbol{\epsilon}, \boldsymbol{\xi})\} \\ &= \begin{cases} 0 & \text{if } \boldsymbol{\sigma} \geq 0 \in \mathbb{R}^n, \boldsymbol{\tau} \geq 0 \in \mathbb{R}^n, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

By the theory of convex analysis, the following extended canonical duality relations hold:

$$\mathbf{y}^* \in \partial V(\mathbf{y}) \quad \Leftrightarrow \quad \mathbf{y} \in \partial V^\sharp(\mathbf{y}^*) \quad \Leftrightarrow \quad V(\mathbf{y}) + V^\sharp(\mathbf{y}^*) = \mathbf{y}^T\mathbf{y}^*, \tag{6.5}$$

or equivalently:

$$\boldsymbol{\epsilon} \leq 0 \quad \Leftrightarrow \quad \boldsymbol{\sigma} \geq 0 \quad \Leftrightarrow \quad \boldsymbol{\epsilon}^T\boldsymbol{\sigma} = 0, \tag{6.6}$$

$$\boldsymbol{\xi} \leq 0 \quad \Leftrightarrow \quad \boldsymbol{\tau} \geq 0 \quad \Leftrightarrow \quad \boldsymbol{\xi}^T\boldsymbol{\tau} = 0. \tag{6.7}$$

Observe that the complementarity condition $\boldsymbol{\xi}^T\boldsymbol{\tau} = \boldsymbol{\tau}^T(\mathbf{v}\circ\mathbf{v} - \mathbf{v}) = 0$, $\forall\boldsymbol{\tau} > 0$ in (6.7) leads to the integrality condition $\mathbf{v}\circ\mathbf{v} - \mathbf{v} = 0$.

Letting $U(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{c}^T\mathbf{x} - \mathbf{f}^T\mathbf{v}$, the relaxed primal problem $(\mathcal{P}_r)$ can be written in the following unconstrained canonical form:

$$(\mathcal{P}_c): \quad \min\{\Pi(\mathbf{x}, \mathbf{v}) = V(\Lambda(\mathbf{x}, \mathbf{v})) - U(\mathbf{x}, \mathbf{v}) \mid \ \mathbf{x} \in \mathbb{R}^n, \ \ \mathbf{v} \in \mathbb{R}^n\}. \tag{6.8}$$

Firstly, we replace $V(\Lambda(\mathbf{x}, \mathbf{v}))$ in (6.8) by the Fenchel-Young equality

$$V(\Lambda(\mathbf{x}, \mathbf{v})) = \Lambda(\mathbf{x}, \mathbf{v})^T\mathbf{y}^*(\boldsymbol{\sigma}, \boldsymbol{\tau}) - V^\sharp(\mathbf{y}^*(\boldsymbol{\sigma}, \boldsymbol{\tau})).$$

Then the *total complementary function*

$$\Xi(\mathbf{x}, \mathbf{v}, \boldsymbol{\sigma}, \boldsymbol{\tau}) : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$$

associated with Problem $(\mathcal{P}_c)$ can be defined as given below:

$$\begin{aligned} \Xi(\mathbf{x}, \mathbf{v}, \boldsymbol{\sigma}, \boldsymbol{\tau}) &= \Lambda(\mathbf{x}, \mathbf{v})^T\mathbf{y}^*(\boldsymbol{\sigma}, \boldsymbol{\tau}) - V^\sharp(\mathbf{y}^*(\boldsymbol{\sigma}, \boldsymbol{\tau})) - U(\mathbf{x}, \mathbf{v}) \\ &= \frac{1}{2}\mathbf{x}^T\mathbf{G}(\boldsymbol{\sigma})\mathbf{x} + \mathbf{c}^T\mathbf{x} + \mathbf{v}^T\text{Diag}\,(\boldsymbol{\tau})\mathbf{v} - (\boldsymbol{\sigma} + \boldsymbol{\tau} - \mathbf{f})^T\mathbf{v} - V^\sharp(\mathbf{y}^*(\boldsymbol{\sigma}, \boldsymbol{\tau})), \end{aligned}$$

where

$$\mathbf{G}(\boldsymbol{\sigma}) = \mathbf{A} + 2\text{Diag }(\boldsymbol{\sigma}), \tag{6.9}$$

and the notation Diag $(\boldsymbol{\sigma})$ stands for a diagonal matrix with $\sigma_i$, $i = 1, ..., n$, being its diagonal elements. From this complementary function, we obtain the canonical dual function $\Pi^d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ as:

$$P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \text{sta}\{\Xi(\mathbf{x}, \mathbf{v}, \boldsymbol{\sigma}, \boldsymbol{\tau}) \mid \mathbf{x} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n\} = U^\Lambda(\boldsymbol{\sigma}, \boldsymbol{\tau}) - V^\sharp(\boldsymbol{\sigma}, \boldsymbol{\tau}), \tag{6.10}$$

where $U^\Lambda(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is the $\Lambda$-conjugate transformation defined by

$$U^\Lambda(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \text{sta}\{\Lambda(\mathbf{x}, \mathbf{v})^T \mathbf{y}^*(\boldsymbol{\sigma}, \boldsymbol{\tau}) - U(\mathbf{x}, \mathbf{v}) |\ \mathbf{x} \in \mathbb{R}^n, \mathbf{v} \in \mathbb{R}^n\}. \tag{6.11}$$

Accordingly, introducing a dual feasible space

$$\mathcal{S}_\sharp = \{(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathbb{R}^n \times \mathbb{R}^n |\ \ \boldsymbol{\sigma} \geq 0, \ \ \boldsymbol{\tau} > 0, \ \ \mathbf{c} \in \mathcal{C}_{ol}(\mathbf{G}(\boldsymbol{\sigma}))\}, \tag{6.12}$$

where $\mathcal{C}_{ol}(\mathbf{G})$ denotes the column space of $\mathbf{G}$ (i.e., a vector space spanned by the columns of the matrix $\mathbf{G}$), the canonical dual function can be formulated as

$$P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = U^\Lambda(\boldsymbol{\sigma}, \boldsymbol{\tau}) = -\frac{1}{2}\mathbf{c}^T \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{c} - \frac{1}{4}\sum_{i=1}^n \frac{1}{\tau_i}(\sigma_i + \tau_i - f_i)^2, \ \ \forall (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\sharp, \tag{6.13}$$

where $\mathbf{G}^+$ denotes the Moore-Penrose generalized inverse of $\mathbf{G}$. Denoting

$$P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = -\frac{1}{2}\mathbf{c}^T \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{c} - \frac{1}{4}\sum_{i=1}^n \frac{1}{\tau_i}(\sigma_i + \tau_i - f_i)^2 \ \ : \ \ \mathcal{S}_\sharp \to \mathbb{R}, \tag{6.14}$$

the proposed dual to $(\mathcal{P}_\flat)$ is then stated as follows:

$$(\mathcal{P}^\sharp): \quad \max \left\{ P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = -\frac{1}{2}\mathbf{c}^T \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{c} - \frac{1}{4}\sum_{i=1}^n \frac{1}{\tau_i}(\sigma_i + \tau_i - f_i)^2 \ |\ \ (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\sharp \right\}. \tag{6.15}$$

For any given $n$-vectors $\mathbf{t} = \{t_i\}^n$ and $\mathbf{s} = \{s_i\}^n$, we denote $\mathbf{t} \oslash \mathbf{s} = \{t_i/s_i\}^n$.

**Theorem 6.1** (Complementary-Dual Principle)**.** *Problem $(\mathcal{P}^\sharp)$ is canonically (i.e., perfectly) dual to the primal problem $(\mathcal{P}_\flat)$ in the sense that if $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) \in \mathcal{S}_\sharp$ is a KKT point of $(\mathcal{P}^\sharp)$, then the vector $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$ defined by*

$$\bar{\mathbf{x}} = -\mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c}, \tag{6.16}$$

$$\bar{\mathbf{v}} = \frac{1}{2}(\bar{\boldsymbol{\sigma}} + \bar{\boldsymbol{\tau}} - \mathbf{f}) \oslash \bar{\boldsymbol{\tau}} \tag{6.17}$$

*is a local optimal solution to the primal problem* $(\mathcal{P}_\flat)$, *and*

$$P(\bar{\mathbf{x}}, \bar{\mathbf{v}}) = \Xi(\bar{\mathbf{x}}, \bar{\mathbf{v}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}). \tag{6.18}$$

*Proof.* By introducing Lagrange multipliers $(\boldsymbol{\epsilon}, \boldsymbol{\xi}) \in \mathbb{R}^n_- \times \mathbb{R}^n_-$ associated with the respective inequalities in (6.12) (where $\mathbb{R}^n_-$ is the nonpositive orthant of $\mathbb{R}^n$), the Lagrangian $\Theta : \mathcal{S}_\sharp \times \mathbb{R}^n_- \times \mathbb{R}^n_- \to \mathbb{R}$ for Problem $(\mathcal{P}^\sharp)$ is given by

$$\Theta(\boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\epsilon}, \boldsymbol{\xi}) = P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) - \boldsymbol{\epsilon}^T \boldsymbol{\sigma} - \boldsymbol{\xi}^T \boldsymbol{\tau}. \tag{6.19}$$

It is easy to prove that the criticality conditions

$$\nabla_{\boldsymbol{\sigma}} \Theta(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \boldsymbol{\epsilon}, \boldsymbol{\xi}) = 0, \;\; \nabla_{\boldsymbol{\tau}} \Theta(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}, \boldsymbol{\epsilon}, \boldsymbol{\xi}) = 0$$

lead to

$$\boldsymbol{\epsilon} = \nabla_{\boldsymbol{\sigma}} P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) = \bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) \circ \bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) - \bar{\mathbf{v}}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}), \tag{6.20}$$

$$\boldsymbol{\xi} = \nabla_{\boldsymbol{\tau}} P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) = \bar{\mathbf{v}}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) \circ \bar{\mathbf{v}}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) - \bar{\mathbf{v}}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}), \tag{6.21}$$

and the accompanying KKT conditions, which include

$$0 \le \bar{\boldsymbol{\sigma}} \;\perp\; \boldsymbol{\epsilon} \le 0, \tag{6.22}$$

$$0 < \bar{\boldsymbol{\tau}} \;\perp\; \boldsymbol{\xi} \le 0, \tag{6.23}$$

where $\bar{\mathbf{x}}(\bar{\boldsymbol{\sigma}}) = -\mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c}$, and $\bar{\mathbf{v}}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) = \frac{1}{2}(\bar{\boldsymbol{\sigma}} + \bar{\boldsymbol{\tau}} - \mathbf{f}) \oslash \bar{\boldsymbol{\tau}}$. By the strict inequality condition $\bar{\boldsymbol{\tau}} > 0$, the complementarity condition $\bar{\boldsymbol{\tau}}^T(\bar{\mathbf{v}} \circ \bar{\mathbf{v}} - \bar{\mathbf{v}}) = 0$ in (6.23 ) leads to the integrality condition $(\bar{\mathbf{v}} \circ \bar{\mathbf{v}} - \bar{\mathbf{v}}) = 0$. This shows that if $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}})$ is a KKT point of the problem $(\mathcal{P}^\sharp)$, then $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$ is a local optimal solution to the discrete primal problem $(\mathcal{P}_\flat)$.

By using (6.16) and (6.17), we have:

$$\begin{aligned}
P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) &= \frac{1}{2}\mathbf{c}^T \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c} - \mathbf{c}^T \mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c} - 2\bar{\mathbf{v}}^T \text{Diag }(\bar{\boldsymbol{\tau}})\bar{\mathbf{v}} + \bar{\mathbf{v}}^T \text{Diag }(\bar{\boldsymbol{\tau}})\bar{\mathbf{v}} \\
&= \frac{1}{2}\bar{\mathbf{x}}^T \mathbf{A}\bar{\mathbf{x}} + \bar{\mathbf{x}}^T \text{Diag }(\bar{\boldsymbol{\sigma}})\bar{\mathbf{x}} + \mathbf{c}^T \bar{\mathbf{x}} - \bar{\mathbf{v}}^T(\bar{\boldsymbol{\sigma}} + \bar{\boldsymbol{\tau}} - \mathbf{f}) + \bar{\boldsymbol{\tau}}^T(\bar{\mathbf{v}} \circ \bar{\mathbf{v}}) \\
&= \Xi(\bar{\mathbf{x}}, \bar{\mathbf{v}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) = P(\bar{\mathbf{x}}, \bar{\mathbf{v}}) + \bar{\boldsymbol{\sigma}}^T(\bar{\mathbf{x}} \circ \bar{\mathbf{x}} - \bar{\mathbf{v}}) + \bar{\boldsymbol{\tau}}^T(\bar{\mathbf{v}} \circ \bar{\mathbf{v}} - \bar{\mathbf{v}}) \\
&= P(\bar{\mathbf{x}}, \bar{\mathbf{v}})
\end{aligned}$$

due to the complementarity conditions (6.22) and (6.23). This proves the theorem. □

**Remark 6.1.** *Theorem 6.1 shows that by the canonical duality theory, the NP-hard discrete primal problem* $(\mathcal{P}_\flat)$ *is actually equivalent to a continuous dual problem* $(\mathcal{P}^\sharp)$ *with zero duality gap. If* $\mathbf{G}(\bar{\boldsymbol{\sigma}})$ *is invertible, then the KKT point* $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}})$ *of the canonical dual* *problem* $(\mathcal{P}^\sharp)$ *is a critical point of the canonical dual function* $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$.

□

## 6.3 Global Optimality Criteria

Theorem 6.1 shows that any KKT point of the canonical dual problem $(\mathcal{P}^d)$ leads to a KKT point of the continuously reformulated primal problem $(\mathcal{P}_\flat)$. In this section, we present global optimality conditions for the nonconvex problem $(\mathcal{P}_\flat)$. We first introduce a useful feasible space:

$$\mathcal{S}_\sharp^+ = \{(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathbb{R}^n \times \mathbb{R}^n \mid \ \boldsymbol{\sigma} \geq 0, \ \boldsymbol{\tau} > 0, \ \ \mathbf{G}(\boldsymbol{\sigma}) \succ 0\} \tag{6.24}$$

By the *triality theory* developed in [38], we have the following results, where $\mathbf{y}^* = (\boldsymbol{\sigma}, \boldsymbol{\tau})$.

**Theorem 6.2.** *Suppose that the vector* $\bar{\mathbf{y}}^* = (\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) \in \mathcal{S}_\sharp^+$ *is a critical point of the dual function* $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$. *Let* $(\bar{\mathbf{x}}, \bar{\mathbf{v}}) = \left(-\mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c}, \ \frac{1}{2}(\bar{\boldsymbol{\sigma}} + \bar{\boldsymbol{\tau}} - \mathbf{f}) \oslash \bar{\boldsymbol{\tau}}\right)$.

*If* $\bar{\mathbf{y}}^* \in \mathcal{S}_\sharp^+$, *then* $\bar{\mathbf{y}}^*$ *is a global maximizer of* $P^d$ *on* $\mathcal{S}_\sharp^+$. *The vector* $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$ *is a global minimizer of* $P$ *on* $\mathcal{X}_v$, *and*

$$P(\bar{\mathbf{x}}, \bar{\mathbf{v}}) = \min_{(\mathbf{x},\mathbf{v})\in\mathcal{X}_v} P(\mathbf{x}, \mathbf{v}) = \max_{(\boldsymbol{\sigma},\boldsymbol{\tau})\in\mathcal{S}_\sharp^+} P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}). \tag{6.25}$$

*Proof.* By Theorem 6.1, we know that if the vector $\bar{\mathbf{y}}^*$ is a critical point of the problem $(\mathcal{P}^\sharp)$, then the vector $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$ defined by (6.16) and (6.17) is a local optimal solution to the problem $(\mathcal{P}_\flat)$, and

$$P(\bar{\mathbf{x}}, \bar{\mathbf{v}}) = \Xi(\bar{\mathbf{x}}, \bar{\mathbf{v}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}).$$

By the fact that the canonical dual function $P^d(\mathbf{y}^*)$ is concave on $\mathcal{S}_\sharp^+$, the critical point $\bar{\mathbf{y}}^* \in \mathcal{S}_\sharp^+$ is a global maximizer of $P^d(\mathbf{y}^*)$ over $\mathcal{S}_\sharp^+$, and $(\bar{\mathbf{x}}, \bar{\mathbf{v}}, \bar{\mathbf{y}}^*)$ is a saddle point of the total complementary function $\Xi(\mathbf{x}, \mathbf{v}, \mathbf{y}^*)$ on $\mathbb{R}^{2n} \times \mathcal{S}_\sharp^+$, i.e., $\Xi$ is convex in $(\mathbf{x}, \mathbf{v}) \in \mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ and concave in $\mathbf{y}^* \in \mathcal{S}_\sharp^+$. Thus, by the (right) saddle min-max duality theory (see [38]), we have

$$\begin{aligned}
P^d(\bar{\mathbf{y}}^*) &= \max_{\mathbf{y}^*\in\mathcal{S}_\sharp^+} P^d(\mathbf{y}^*) = \max_{\mathbf{y}^*\in\mathcal{S}_\sharp^+} \min_{(\mathbf{x},v)\in\mathbb{R}^{2n}} \Xi(\mathbf{x}, \mathbf{v}, \mathbf{y}^*) = \min_{(\mathbf{x},v)\in\mathbb{R}^{2n}} \max_{\mathbf{y}^*\in\mathcal{S}_\sharp^+} \Xi(\mathbf{x}, \mathbf{v}, \mathbf{y}^*) \\
&= \min_{(\mathbf{x},v)\in\mathbb{R}^{2n}} \left\{ P(\mathbf{x}, v) + \max_{(\boldsymbol{\sigma},\boldsymbol{\tau})\in\mathcal{S}_\sharp^+} \left\{ \Lambda(\mathbf{x}, v)^T \mathbf{y}^* - V^\sharp(\mathbf{y}^*)) \right\} \right\} \\
&= \min_{(\mathbf{x},v)\in\mathcal{X}_v} P(\mathbf{x}, v) = P(\bar{\mathbf{x}}, \bar{\mathbf{v}})
\end{aligned}$$

due to the fact that

$$V(\Lambda(\mathbf{x}, v)) = \sup_{\mathbf{y}^* \in \mathcal{S}_\sharp^+} \{\Lambda(\mathbf{x}, v)^T \mathbf{y}^* - V^\sharp(\mathbf{y}^*)\} = \begin{cases} 0 & \text{if } (\mathbf{x}, v) \in \mathcal{X}_v, \\ +\infty & \text{otherwise.} \end{cases}$$

This proves the statement (6.25).

□

Theorem 6.2 shows that, under the stated conditions, the nonconvex quadratic mixed-integer minimization problem $(\mathcal{P}_\flat)$ is canonically dual to the following concave maximization problem:

$$(\mathcal{P}_+^\sharp): \quad \max \left\{ P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) \; : \; (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\sharp^+ \right\}. \tag{6.26}$$

Since $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is a continuous concave function over a convex feasible space $\mathcal{S}_\sharp^+$, if $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) \in \mathcal{S}_\sharp^+$ is a critical point of $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$, it must be a global maximizer of the problem $(\mathcal{P}_+^\sharp)$, and the vector $(\bar{\mathbf{x}}, \bar{\mathbf{v}}) = \left(-\mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c}, \; \frac{1}{2}(\bar{\boldsymbol{\sigma}} + \bar{\boldsymbol{\tau}} - \mathbf{f}) \oslash \bar{\boldsymbol{\tau}}\right)$ is a global minimizer of the problem $(\mathcal{P}_\flat)$. By the fact that for a fixed $\boldsymbol{\sigma}$, the criticality condition $\nabla_{\boldsymbol{\tau}} P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = 0$ leads to $\boldsymbol{\tau} = |\boldsymbol{\sigma} - \mathbf{f}| > 0 \in \mathbb{R}^n$. Substituting this into $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$, we have

$$P^g(\boldsymbol{\sigma}) = -\frac{1}{2}\mathbf{c}^T \mathbf{G}^+(\boldsymbol{\sigma})\mathbf{c} - \sum_{i=1}^n (\sigma_i - f_i)^+, \tag{6.27}$$

where $(t_i)^+ = \max\{t_i, 0\}$. Furthermore, let $\boldsymbol{\delta}(\mathbf{t})^+ = \{\delta_i(t_i)^+\}^n \in \mathbb{R}^n$, where

$$\delta_i(t_i)^+ = \begin{cases} 1 & \text{if } t_i > 0 \\ 0 & \text{if } t_i < 0, \end{cases} \quad i = 1, \cdots, n, \tag{6.28}$$

and

$$\mathcal{S}_\sigma^+ = \{\boldsymbol{\sigma} \in \mathbb{R}^n | \; \boldsymbol{\sigma} \geq 0, \;\; \boldsymbol{\sigma} \neq \mathbf{f}, \;\; \mathbf{G}(\boldsymbol{\sigma}) \succ 0\}. \tag{6.29}$$

Then, the canonical dual problem $(\mathcal{P}_+^\sharp)$ can be written in the following simple form:

$$(\mathcal{P}_+^g): \quad \max \left\{ P^g(\boldsymbol{\sigma}) \; : \; \boldsymbol{\sigma} \in \mathcal{S}_\sigma^+ \right\}. \tag{6.30}$$

**Theorem 6.3** (Analytic solution to $(\mathcal{P}_\flat)$)**.** *For given $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{c}, \; \mathbf{f} \in \mathbb{R}^n$, if $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_\sigma^+$ is a critical point of $P^g(\boldsymbol{\sigma})$, then the vector*

$$(\bar{\mathbf{x}}, \bar{\mathbf{v}}) = (-\mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c}, \; \boldsymbol{\delta}(\bar{\boldsymbol{\sigma}} - \mathbf{f})^+) \tag{6.31}$$

*is a global minimizer of $(\mathcal{P}_\flat)$.*

This theorem can be proved easily by using Theorem 6.2. In the next section, we will study certain existence and uniqueness conditions for the canonical dual problem to have a critical point in $\mathcal{S}_\sigma^+$.

## 6.4 Existence and Uniqueness Criteria

**Definition 6.1.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be coercive if for every sequence $\{x^v\} \subset \mathbb{R}^n$ for which $\|x^v\| \to \infty$ it must be the case that $f(x^v) \to \infty$ as well.*

Let

$$\partial\mathcal{S}_\sigma^+ = \{\boldsymbol{\sigma} \in \mathcal{S}_\sigma^+ : \ \det \mathbf{G}(\boldsymbol{\sigma}) = 0\}. \tag{6.32}$$

Then, we have the following theorem:

**Theorem 6.4** (Existence and Uniqueness Criteria)**.** *For a given matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ and vectors $\mathbf{c}$, $\mathbf{f} \in \mathbb{R}^n$, if for any given $\boldsymbol{\sigma} \in \mathcal{S}_\sigma^+$,*

$$\lim_{\alpha\to 0^+} \mathbf{c}^T[G(\boldsymbol{\sigma}_o + \alpha\boldsymbol{\sigma})]^+\mathbf{c} = \infty \ \text{ and } \ \lim_{\alpha\to\infty} \mathbf{c}^T[G(\boldsymbol{\sigma}_o + \alpha\boldsymbol{\sigma})]^+\mathbf{c} \geq 0, \ \ \forall \boldsymbol{\sigma}_o \in \partial\mathcal{S}_\sigma^+, \tag{6.33}$$

*then the canonical dual problem $(\mathcal{P}_+^g)$ has at least one critical point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_\sigma^+$ and the vector*

$$(\bar{\mathbf{x}}, \bar{\mathbf{v}}) = \left(-\mathbf{G}^+(\bar{\boldsymbol{\sigma}})\mathbf{c}, \ \ \boldsymbol{\delta}(\bar{\boldsymbol{\sigma}} - \mathbf{f})^+\right)$$

*is a global optimizer of the primal problem $(\mathcal{P}_\flat)$. Moreover, if*

$$c_i \neq 0, \ \ \bar{\sigma}_i - f_i \neq 0, \ \ \forall i = 1, \dots, n, \tag{6.34}$$

*then the vector $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$ is a unique global minimizer of $(\mathcal{P}_\flat)$.*

*Proof.* By the fact that, on $\mathcal{S}_\sigma^+$, we have

$$\frac{\partial \mathbf{G}^+(\boldsymbol{\sigma})}{\partial \sigma_k} = -\mathbf{G}^+(\boldsymbol{\sigma})\frac{\partial \mathbf{G}(\boldsymbol{\sigma})}{\partial \sigma_k}\mathbf{G}^+(\boldsymbol{\sigma}),$$

the Hessian of the quadratic form $-\frac{1}{2}\mathbf{c}^T G^+(\boldsymbol{\sigma})\mathbf{c}$ is:

$$H_{1\sigma^2}(\boldsymbol{\sigma}) = \{-4x_i(\boldsymbol{\sigma})G_{ij}^+(\boldsymbol{\sigma})x_j(\boldsymbol{\sigma})\}, \tag{6.35}$$

where $\mathbf{x}(\boldsymbol{\sigma}) = -G^+(\boldsymbol{\sigma})\mathbf{c}$. Therefore, the Hessian matrix of the dual objective function $P^d$ is:

$$H(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \nabla^2 P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \begin{pmatrix} H_{1\sigma^2} + H_{2\sigma^2} & H_{\sigma\tau} \\ H_{\tau\sigma} & H_{\tau^2} \end{pmatrix},$$

where

$$H_{2\sigma^2} = \text{Diag}\left\{-\frac{1}{2\tau_i}\right\},$$
$$H_{\sigma\tau} = H_{\tau\sigma} = \text{Diag}\left\{\frac{(\sigma_i - f_i)}{2\tau_i^2}\right\},$$
$$H_{\tau^2} = \text{Diag}\left\{-\frac{(\sigma_i - f_i)^2}{2\tau_i^3}\right\}.$$

It is clear that

$$H_{1\sigma^2}(\boldsymbol{\sigma}) \preceq 0, \; H_{2\sigma^2}(\boldsymbol{\tau}) \prec 0, \quad H_{\tau^2}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \preceq 0, \quad \forall (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\sharp^+, \tag{6.36}$$
$$H_{1\sigma^2}(\boldsymbol{\sigma}) \succeq 0, \; H_{2\sigma^2}(\boldsymbol{\tau}) \succ 0, \quad H_{\tau^2}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \succeq 0, \quad \forall (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\flat^-. \tag{6.37}$$

For any given non-zero vector $\mathbf{w} = (\mathbf{s}, \mathbf{t}) \in \mathbb{R}^{2n}$, we have

$$\mathbf{w}^T H(\boldsymbol{\sigma}, \boldsymbol{\tau})\mathbf{w} = \mathbf{s}^T H_{1\sigma^2}(\boldsymbol{\sigma})\mathbf{s} + \sum_{i=1}^{n} -\frac{1}{2\tau_i}\left(s_i - t_i \frac{\sigma_i - f_i}{\tau_i}\right)^2. \tag{6.38}$$

Thus

$$\nabla^2 P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) \preceq 0 \quad \text{if } (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\sharp^+,$$
$$\nabla^2 P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) \succeq 0 \quad \text{if } (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\flat^-.$$

Therefore, $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is concave on $\mathcal{S}_\sharp^+$, convex on $\mathcal{S}_\flat^-$, and $P^g(\boldsymbol{\sigma})$ is concave on $\mathcal{S}_\sigma^+$. From the conditions in (6.33), we have, for any $\boldsymbol{\sigma}_0 \in \partial\mathcal{S}_\sigma^+$,

$$\lim_{\alpha \to 0^+} P^g(\boldsymbol{\sigma}_o + \alpha\boldsymbol{\sigma}) = -\infty, \quad \forall \boldsymbol{\sigma} \in \mathcal{S}_\sigma^+, \tag{6.39}$$

and

$$\lim_{\alpha \to \infty} P^g(\boldsymbol{\sigma}_o + \alpha\boldsymbol{\sigma}) = -\infty, \quad \forall \boldsymbol{\sigma} \in \mathcal{S}_\sigma^+. \tag{6.40}$$

(6.39) and (6.40) show that the canonical dual function $P^g(\boldsymbol{\sigma})$ is concave and coercive on the convex set $\mathcal{S}_\sigma^+$. Therefore, by the theory of convex analysis [80], we know that the canonical dual problem $(\mathcal{P}_+^g)$ has at least one critical point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_\sigma^+$. Because for convex problem, critical points are always global optimizer [1], so $\bar{\boldsymbol{\sigma}}$ is a global maximizer of $P^g(\boldsymbol{\sigma})$ over $\mathcal{S}_\sigma^+$. By Theorem 6.2, the corresponding vector $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$ is a global optimizer of the primal problem $(\mathcal{P}_\flat)$. Moreover, if the conditions in (6.34) hold, then $H_{1\sigma^2}(\boldsymbol{\sigma}) \prec 0$; $H_{\tau^2}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \prec 0$, $\forall (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \mathcal{S}_\sharp^+$, and the Hessian $\nabla^2 P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) \prec 0$, i.e., $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is strictly concave on $\mathcal{S}_\sharp^+$. Therefore, $(\mathcal{P}^d)$ has a unique critical point in $\mathcal{S}_\sharp^+$, which implies that $(\mathcal{P}_+^g)$ has a unique critical point in $\mathcal{S}_\sigma^+$ and that the primal problem has a unique global minimizer. □

## 6.5 Application to Decoupled Problem

We now apply the theory presented in this chapter to a decoupled system. For simplicity, let $\mathbf{A} = \text{Diag}\ (\mathbf{a})$ be a diagonal matrix with $\mathbf{a} = \{a_i\} \in \mathbb{R}^n$ being its diagonal elements. Now, consider the following extremal problem:

$$\max \quad \left\{ P(\mathbf{x}, \mathbf{v}) = \sum_{i=1}^{n} \left( \frac{1}{2} a_i x_i^2 + c_i x_i + f_i v_i \right) \right\} \tag{6.41}$$

$$\text{s.t.} \quad -v_i \leq x_i \leq v_i,\ v_i \in \{0, 1\},\ i = 1, \ldots, n. \tag{6.42}$$

For this decoupled problem, the canonical dual function has a simple form given by

$$P^d(\boldsymbol{\sigma}, \boldsymbol{\tau}) = -\frac{1}{2} \sum_{i=1}^{n} \left( \frac{c_i^2}{a_i + 2\sigma_i} + \frac{(\sigma_i + \tau_i - f_i)^2}{2\tau_i} \right). \tag{6.43}$$

By Theorem 6.2, if the critical point $(\boldsymbol{\varsigma}, \boldsymbol{\tau}) \in \mathcal{S}_\sharp^+$, then the corresponding primal solution is:

$$(x_i, v_i) = \left( -\tfrac{c_i}{a_i + 2\sigma_i}, \tfrac{f_i + \sigma_i + \tau_i}{2\tau_i} \right),\ \forall\, i = 1, \cdots, n. \tag{6.44}$$

The global extrema of the primal problem can be determined by the following theorem:

**Theorem 6.5.** *For any given* $\mathbf{a}, \mathbf{c}, \mathbf{f} \in \mathbb{R}^n$*, if* $c_i \neq 0$*,*

$$-\frac{1}{2}[a_i - |c_i|] \geq 0,\ \textit{and}\ f_i - \frac{1}{2}[a_i - |c_i|] > 0, \forall\, i = 1, \cdots, n, \tag{6.45}$$

*then the canonical dual function* $P^d$ *has a unique critical point*

$$(-\frac{1}{2}[a_i - |c_i|], \forall\, i = 1, \ldots, n; f_i - \frac{1}{2}[a_i - |c_i|], \forall\, i = 1, \ldots, n) \tag{6.46}$$

*which is a global maximizer of* $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$ *on* $\mathcal{S}_\sharp^+$*, and*

$$(\mathbf{x}_\sharp, \mathbf{v}_\sharp) = \left( \{-\frac{c_i}{|c_i|}, i = 1, \cdots, n\}, \mathbf{e} \right) \tag{6.47}$$

*is a global minimizer of* $P(\mathbf{x}, \mathbf{v})$ *on* $\mathcal{X}_v$*.*

*On the other hand, if* $c_i \neq 0$*,*

$$\min_{i=1,\cdots,n} \left\{ -\frac{1}{2}(a_i \pm c_i) \right\} < 0, \quad \min_{i=1,\cdots,n} \left\{ f_i - \frac{1}{2}(a_i \pm c_i) \right\} < 0, \tag{6.48}$$

*then the canonical dual function $P^d$ has a unique critical point*

$$(\boldsymbol{\sigma}_\flat, \boldsymbol{\tau}_\flat) = \left( \min_{i=1,\cdots,n} \left\{ -\frac{1}{2}(a_i \pm c_i) \right\}, \min_{i=1,\cdots,n} \left\{ f_i - \frac{1}{2}(a_i \pm c_i) \right\} \right) \in \mathcal{S}_\flat^-, \tag{6.49}$$

*which is a global minimizer of $P^d(\boldsymbol{\sigma}, \boldsymbol{\tau})$ on $\mathcal{S}_\flat^-$ and*

$$(\mathbf{x}_\flat, \mathbf{v}_\flat) = \left( \left\{ \frac{c_i}{|c_i|}, i = 1, \cdots, n \right\}, \ \mathbf{e} \right) \tag{6.50}$$

*is a global maximizer of $P(\mathbf{x}, \mathbf{v})$ on $\mathcal{X}_v$.*

## 6.6 Numerical Examples

### 6.6.1 Two-dimensional decoupled problem

Let $a_1 = -3$, $a_2 = 2$, $c_1 = 5$, $c_2 = -8$, $f_1 = -2$, and $f_2 = 2$. The canonical dual function $P^d$ has a total of nine critical points $(\boldsymbol{\sigma}, \boldsymbol{\tau})_k$, $k = 1, \ldots, 9$, and the corresponding results are listed below:

$$\begin{array}{lll}
(\boldsymbol{\sigma}, \boldsymbol{\tau})_1 = (4, 3, 2, 5), & (\mathbf{x}, \mathbf{v})_1 = (-1, 1, 1, 1), & P_1^d = -13.5; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_2 = (2, 3, 0, 5), & (\mathbf{x}, \mathbf{v})_2 = (0, 1, 0, 1), & P_2^d = -9.0; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_3 = (4, -2, 2, 0), & (\mathbf{x}, \mathbf{v})_3 = (-1, 0, 1, 0), & P_3^d = -4.5; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_4 = (-1, 3, -3, 5), & (\mathbf{x}, \mathbf{v})_4 = (1, 1, 1, 1), & P_4^d = -3.5; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_5 = (2, -2, 0, 0), & (\mathbf{x}, \mathbf{v})_5 = (0, 0, 0, 0), & P_5^d = 0; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_6 = (4, -5, 2, -3), & (\mathbf{x}, \mathbf{v})_6 = (-1, -1, 1, 1), & P_6^d = 2.5; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_7 = (-1, -2, -3, 0), & (\mathbf{x}, \mathbf{v})_7 = (1, 0, 1, 0), & P_7^d = 5.5; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_8 = (2, -5, 0, -3), & (\mathbf{x}, \mathbf{v})_8 = (0, -1, 0, 1), & P_8^d = 7; \\
(\boldsymbol{\sigma}, \boldsymbol{\tau})_9 = (-1, -5, -3, -3), & (\mathbf{x}, \mathbf{v})_9 = (1, -1, 1, 1), & P_9^d = 12.5.
\end{array}$$

By the fact that $(\boldsymbol{\sigma}, \boldsymbol{\tau})_1 \in \mathcal{S}_\sharp^+$, we can tell that $(\mathbf{x}, \mathbf{v})_1$ is a global minimizer of $P(\mathbf{x}, \mathbf{v})$.

### 6.6.2 General nonconvex problem

Let $n = 10$ and let $\mathbf{c}$, $\mathbf{f}$ and $\mathbf{A}$ be chosen randomly as follows:

$$\mathbf{c} = \{16, -13, -12, -18, -11, 7, 11, 16, -4, 18\}^T,$$

$$\mathbf{f} = \{11, 5, 13, 18, 6, 4, -16, 16, -20, -3\}^T,$$

$$
\mathbf{A} = \begin{bmatrix}
10 & 9 & 9 & 9 & 1 & 9 & 4 & 1 & 5 & 9 \\
2 & 5 & 7 & 3 & 2 & 10 & 7 & 2 & 8 & 2 \\
7 & 2 & 6 & 6 & 2 & 2 & 6 & 1 & 7 & 5 \\
5 & 5 & 2 & 9 & 6 & 3 & 9 & 5 & 7 & 8 \\
2 & 9 & 1 & 9 & 8 & 10 & 9 & 4 & 4 & 5 \\
8 & 2 & 1 & 9 & 7 & 3 & 7 & 3 & 1 & 4 \\
4 & 2 & 8 & 2 & 2 & 6 & 6 & 2 & 4 & 2 \\
4 & 7 & 7 & 10 & 2 & 5 & 7 & 5 & 6 & 3 \\
3 & 6 & 9 & 10 & 1 & 8 & 6 & 5 & 9 & 5 \\
7 & 7 & 2 & 7 & 7 & 3 & 7 & 7 & 8 & 6
\end{bmatrix}.
$$

By solving the canonical dual problem $(\mathcal{P}^g_+)$, we obtain the global maximizer

$$\bar{\boldsymbol{\sigma}} = [6.9, 6.9, 6.3, 9.8, 3.9, 2.9, 15.995, 11.5, 116, 8.0]^T,$$

and

$$\bar{\boldsymbol{\tau}} = [17.9, 11.9, 19.3, 27.8, 9.9, 6.9, 0.005, 27.5, 8.4, 5.0]^T.$$

The global minimizer of the primal problem $(\mathcal{P})$ is then given by

$$\bar{\mathbf{x}} = [-1.0, 1.0, 1.0, 1.0, 1.0, -1.0, 0, -1.0, 0, -1.0]^T,$$

and

$$\bar{\mathbf{v}} = [1, 1, 1, 1, 1, 1, 0, 1, 0, 1]^T,$$

with $P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\tau}}) = -181 = P(\bar{\mathbf{x}}, \bar{\mathbf{v}})$.

## 6.7 Conclusions

We have studied in this chapter an application of canonical duality theory to solve the mixed-integer quadratic optimization problem $(\mathcal{P}_\flat)$ and its co-problem $(\mathcal{P}_\sharp)$. Using an appropriate quadratic mapping $\mathbf{y} = \Lambda(\mathbf{x}, \mathbf{v}) = (\mathbf{x} \circ \mathbf{x} - \mathbf{v},\ \mathbf{v} \circ \mathbf{v} - \mathbf{v})$, the given nonconvex mixed-integer primal problem was converted into a canonical dual problem in continuous space, and its relationship with the classical Lagrangian duality under a similar transformation was revealed. Theorem 6.2 shows that the canonical dual problem $(\mathcal{P}^\sharp)$ is a concave maximization over the convex dual feasible space $\mathcal{S}^+_\sharp$ and the co-dual $(\mathcal{P}^\flat)$ is a convex minimization problem on $\mathcal{S}^-_\flat$. Therefore, both problems can be solved via convex programming optimization methods. Theorem 6.3 shows that the mixed-integer programming problem in $\mathbb{R}^{2n}$ is canonically dual to a simplified concave maximization problem $(\mathcal{P}^g_+)$ over a convex feasible set $\mathcal{S}^+_\sigma \subset \mathbb{R}^n$, which can be solved by well-developed convex minimization techniques. Certain existence and uniqueness conditions related to critical points belonging to a derived dual feasible space for yielding a zero duality gap were established in Theorem 6.4.

# CHAPTER 7

# Nonconvex Constrained Optimization

This chapter presents a canonical duality theory for solving a general nonconvex quadratic minimization problem with nonconvex constraints. By using the *canonical dual transformation*, the nonconvex primal problem can be converted into a canonical dual problem (i.e., either a concave maximization or a convex minimization problem with zero duality gap). The global extremum of the nonconvex problem can be identified by the triality theory associated with the canonical duality theory. Illustrative applications to quadratic minimization with multiple quadratic constraints, box/integer constraints, and general nonconvex polynomial constraints are discussed, along with insightful connections to classical Lagrangian duality. Criteria for the existence and uniqueness of optimal solutions are presented. Several numerical examples are provided.

## 7.1 Introduction

We are interested in solving the following general constrained nonlinear programming problem:

$$(\mathcal{P}): \quad \min \{P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{f} \quad : \quad \mathbf{x} \in \mathcal{X}_k\}, \tag{7.1}$$

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

$$\mathcal{X}_k = \{\mathbf{x} \in \mathcal{X}_a | \ \mathbf{g}(\mathbf{x}) \le \mathbf{d} \in \mathbb{R}^m\}, \tag{7.2}$$

where $\mathbf{g}(\mathbf{x}) = \{g_\alpha(\mathbf{x})\} : \mathcal{X}_a \to \mathbb{R}^m$ is a given vector-valued differentiable (not necessary convex) function, $\mathcal{X}_a$ is a convex open set in $\mathbb{R}^n$, and $\mathbf{d} \in \mathbb{R}^m$ is a given vector.

The problem $(\mathcal{P})$ involves minimizing a nonconvex quadratic function over a nonconvex feasible space. By introducing a Lagrangian multiplier vector $\boldsymbol{\sigma} \in \mathbb{R}^m_+ = \{\boldsymbol{\sigma} \in \mathbb{R}^m | \ \boldsymbol{\sigma} \ge 0\}$ to relax the inequality constraints in $\mathcal{X}_k$, the classical Lagrangian $L : \mathcal{X}_a \times \mathbb{R}^m_+ \to \mathbb{R}$ is given by

$$L(\mathbf{x}, \boldsymbol{\sigma}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{f} + \boldsymbol{\sigma}^T(\mathbf{g}(\mathbf{x}) - \mathbf{d}). \tag{7.3}$$

If all the components of $\mathbf{g}(\mathbf{x})$ are convex functions, and $\mathbf{A} \succeq 0$, i.e., positive semidefinite (PSD), then Problem $(\mathcal{P})$ has a convex quadratic objective function and convex constraints, and the Lagrangian is a saddle function, i.e., $L(\mathbf{x}, \boldsymbol{\sigma})$ is convex in the primal variables $\mathbf{x}$, concave (linear) in the dual variables (Lagrange multipliers) $\boldsymbol{\sigma}$, and the Lagrangian dual problem can be easily defined by the Fenchel-Moreau-Rockafellar transformation

$$P^*(\boldsymbol{\sigma}) = \inf_{\mathbf{x} \in \mathcal{X}_a} L(\mathbf{x}, \boldsymbol{\sigma}), \tag{7.4}$$

where, under certain constraint qualifications that insure the existence of a Karush-Kuhn-Tucker (KKT) solution [14], we have the following strong min-max duality relation:

$$\inf_{\mathbf{x} \in \mathcal{X}_k} P(\mathbf{x}) = \sup_{\boldsymbol{\sigma} \in \mathbb{R}^m_+} P^*(\boldsymbol{\sigma}). \tag{7.5}$$

In this case, the problem can be solved easily by any well-developed convex programming technique.

However, due to the assumed nonconvexity of Problem $(\mathcal{P})$, the Lagrangian $L(\mathbf{x}, \boldsymbol{\sigma})$ is no longer a saddle function and the Fenchel-Young inequality leads to only the following weak duality relation in general:

$$\inf_{\mathbf{x} \in \mathcal{X}_k} P(\mathbf{x}) \geq \sup_{\boldsymbol{\sigma} \in \mathbb{R}^m_+} P^*(\boldsymbol{\sigma}). \tag{7.6}$$

The slack $\theta = \inf P(\mathbf{x}) - \sup P^*(\boldsymbol{\sigma})$ in the inequality (7.6) is called the *duality gap* in global optimization. Very often, we have $\theta = \pm\infty$. This duality gap shows that the well-developed Fenchel-Moreau-Rockafellar duality theory can be used only for solving convex minimization problems. Also, due to the nonconvexity of the objective function and/or constraints, the problem may have multiple local solutions. The identification of a global minimizer has been a fundamentally challenging task in global optimization.

In the next section, we will show how to use the canonical dual transformation to convert the nonconvex constrained problem into a canonical dual problem, in order to derive related global optimality conditions.

## 7.2 Canonical Dual Transformation

For convenience, we introduce an indicator function of the feasible set $\mathcal{X}_k$:

$$W(\boldsymbol{\epsilon}) = \begin{cases} 0 & \text{if } \boldsymbol{\epsilon} \leq \mathbf{d} \\ +\infty & \text{otherwise} \end{cases} \tag{7.7}$$

and let

$$U(\mathbf{x}) = \mathbf{x}^T\mathbf{f} - \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}.$$

Then the primal problem $(\mathcal{P})$ can be written in the following unconstrained form:

$$\min\{P(\mathbf{x}) = W(\mathbf{g}(\mathbf{x})) - U(\mathbf{x}) : \quad \mathbf{x} \in \mathcal{X}_a\}. \tag{7.8}$$

By the Fenchel transformation, the conjugate function $W^\sharp(\boldsymbol{\sigma})$ of $W(\boldsymbol{\epsilon})$ can be defined by

$$W^\sharp(\boldsymbol{\sigma}) = \sup_{\boldsymbol{\epsilon}\in\mathbb{R}^m}\{\boldsymbol{\epsilon}^T\boldsymbol{\sigma} - W(\boldsymbol{\epsilon})\} = \begin{cases} \mathbf{d}^T\boldsymbol{\sigma} & \text{if } \boldsymbol{\sigma} \geq 0 \\ +\infty & \text{otherwise,} \end{cases} \tag{7.9}$$

which is convex and l.s.c. (lower semi-continuous) on $\mathbb{R}^m$. From convex analysis [80], the following relations hold for $(\boldsymbol{\epsilon}, \boldsymbol{\sigma}) \in \mathbb{R}^m \times \mathbb{R}^m$:

$$\boldsymbol{\sigma} \in \partial W(\boldsymbol{\epsilon}) \quad \Leftrightarrow \quad \boldsymbol{\epsilon} \in \partial W^\sharp(\boldsymbol{\sigma}) \quad \Leftrightarrow \quad W(\boldsymbol{\epsilon}) + W^\sharp(\boldsymbol{\sigma}) = \boldsymbol{\epsilon}^T\boldsymbol{\sigma}.$$

Replacing $W(\mathbf{g}(\mathbf{x}))$ in $\Pi(\mathbf{x})$ by the Fenchel-Young equality $W(\mathbf{g}(\mathbf{x})) = \mathbf{g}^T(\mathbf{x})\boldsymbol{\sigma} - W^\sharp(\boldsymbol{\sigma})$, the *extended Lagrangian* $\Xi_o : \mathcal{X}_a \times \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ associated with Problem (7.8) can be given as:

$$\Xi_o(\mathbf{x}, \boldsymbol{\sigma}) = \mathbf{g}^T(\mathbf{x})\boldsymbol{\sigma} - W^\sharp(\boldsymbol{\sigma}) - U(\mathbf{x}). \tag{7.10}$$

Clearly, we have $\Xi_o(\mathbf{x}, \boldsymbol{\sigma}) = L(\mathbf{x}, \boldsymbol{\sigma}), \;\; \forall(\mathbf{x}, \boldsymbol{\sigma}) \in \mathcal{X}_a \times \mathbb{R}^m_+$.

Since $\mathbf{g}(\mathbf{x})$ is a nonconvex function, following the standard procedure of the canonical dual transformation, we assume that there exists a *geometrical operator*

$$\boldsymbol{\xi} = \{\xi^\alpha_\beta\} = \Lambda(\mathbf{x}) : \mathcal{X}_a \subset \mathbb{R}^n \to \mathcal{E}_a \subset \mathbb{R}^{m\times p_\alpha}, \tag{7.11}$$

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

$$\mathbf{g}(\mathbf{x}) = V(\Lambda(\mathbf{x})), \tag{7.12}$$

and the duality mapping

$$\boldsymbol{\varsigma} = \{\varsigma^\beta_\alpha\} = \nabla V(\boldsymbol{\xi}) = \left\{\frac{\partial V_\alpha(\boldsymbol{\xi})}{\partial \xi^\alpha_\beta}\right\} : \mathcal{E}_a \to \mathcal{E}^*_a \subset \mathbb{R}^{p_\alpha \times m} \tag{7.13}$$

is invertible. We note that the geometric variable $\boldsymbol{\xi} = \{\xi^\alpha_\beta\}$ is an $m \times p_\alpha$ matrix, while its canonical dual variable $\boldsymbol{\varsigma} = \{\varsigma^\beta_\alpha\}$ is a $p_\alpha \times m$ matrix. For the constrained problem $(\mathcal{P})$ considered in this chapter, the dimension $p_\alpha$ of the geometrical variable $\boldsymbol{\xi} = \{\xi^\alpha_\beta\}$ depends on each given constraint $g_\alpha(\mathbf{x}) \leq d_\alpha, \;\alpha = 1, \ldots, m$. Let $I_\alpha = \{\beta |\; \beta \in \{1, \ldots, p_\alpha\}\}$ be an index set and let

$$\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle = \left\{ \sum_{\beta \in I_\alpha} \xi_\beta^\alpha \varsigma_\alpha^\beta \right\} : \mathcal{E}_a \times \mathcal{E}_a^* \to \mathbb{R}^m$$

denote the *partial bilinear form* on the product space $\mathcal{E}_a \times \mathcal{E}_a^*$. Thus, the Legendre conjugate $V^* : \mathcal{E}_a^* \to \mathbb{R}^m$ of $V$ can be defined by

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

where the notation sta$\{*\}$ denotes computing the stationary points of $\{*\}$. By the assumption that the duality relation (7.13) is invertible (i.e., canonical), the Legendre conjugate $V^*(\boldsymbol{\varsigma})$ is uniquely defined on $\mathcal{E}_a^*$ and the inverse duality relation can be written as:

$$\boldsymbol{\xi} = \nabla V^*(\boldsymbol{\varsigma}) = \left\{ \frac{\partial V_\alpha^*(\boldsymbol{\varsigma})}{\partial \varsigma_\alpha^\beta} \right\}. \tag{7.14}$$

It is easy to verify that the following equivalent relations hold on $\mathcal{E}_a \times \mathcal{E}_a^*$:

$$\boldsymbol{\varsigma} = \nabla V(\boldsymbol{\xi}) \ \Leftrightarrow \ \boldsymbol{\xi} = \nabla V^*(\boldsymbol{\varsigma}) \ \Leftrightarrow \ \langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle = V(\boldsymbol{\xi}) + V^*(\boldsymbol{\varsigma}). \tag{7.15}$$

Noting that (7.12) and (7.15) are used to replace $\mathbf{g}(\mathbf{x})$ in (7.10), we obtain

$$V(\Lambda(\mathbf{x})) = \langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - V^*(\boldsymbol{\varsigma}).$$

Define the generalized total complementary function $\Xi : \mathcal{X}_a \times \mathbb{R}_+^m \times \mathcal{E}_a^* \to \mathbb{R}$ as:

$$\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \boldsymbol{\sigma}^T[\langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - V^*(\boldsymbol{\varsigma}) - \mathbf{d}] + \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{f}, \tag{7.16}$$

where $\boldsymbol{\sigma} \in \mathbb{R}_+^m$ is the dual variable vector associated with $\mathbf{g}(\mathbf{x}) \leq \mathbf{d} \in \mathbb{R}^m$. Through this total complementary function, the *canonical dual function* is defined by

$$P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \text{sta}\{\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) \ : \ \ \mathbf{x} \in \mathcal{X}_a\} = U^\Lambda(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) - \boldsymbol{\sigma}^T(V^*(\boldsymbol{\varsigma}) + \mathbf{d}), \tag{7.17}$$

where $U^\Lambda(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$ is the *parametric* $\Lambda$*-conjugate function* of the quadratic function $U(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{x}^T\mathbf{f}$ defined by the following conjugate transformation [37]:

$$U^\Lambda(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \text{sta}\{\boldsymbol{\sigma}^T\langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - U(\mathbf{x}) \ : \ \mathbf{x} \in \mathcal{X}_a\}. \tag{7.18}$$

Let $\mathcal{S}_k \subset \mathbb{R}_+^m \times \mathcal{E}_a^*$ be a canonical dual feasible space on which the canonical dual function $P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$ is well defined. Then, the canonical dual problem can be posed as follows:

$$(\mathcal{P}^d) : \ \ \max\{P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \ \ : \ \ (\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k\}. \tag{7.19}$$

**Theorem 7.1** (Complementary-Dual Principle). *Problem* $(\mathcal{P}^d)$ *is canonically dual to the primal problem* $(\mathcal{P})$ *in the sense that if* $(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\varsigma})$ *is a critical point of* $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$ *over* $(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{X}_a \times \mathbb{R}^m_+ \times \mathcal{E}^*_a$*, then* $\bar{\mathbf{x}}$ *is a KKT point of* $(\mathcal{P})$*,* $(\bar{\boldsymbol{\sigma}}, \bar{\varsigma})$ *is a KKT point of* $(\mathcal{P}^d)$*, and*

$$P(\bar{\mathbf{x}}) = \Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\varsigma}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\varsigma}). \tag{7.20}$$

*Proof.* If $(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ is a critical point of $\Xi$, then we have the following criticality conditions

$$\nabla_{\mathbf{x}}\Xi(\bar{\mathbf{x}}, \bar{\varsigma}, \bar{\boldsymbol{\sigma}}) = \bar{\boldsymbol{\sigma}}^T \langle \Lambda_t(\bar{\mathbf{x}}); \bar{\varsigma} \rangle + \mathbf{A}\bar{\mathbf{x}} - \mathbf{f} = 0, \tag{7.21}$$

$$\nabla_{\boldsymbol{\varsigma}}\Xi(\bar{\mathbf{x}}, \bar{\varsigma}, \bar{\boldsymbol{\sigma}}) = \Lambda(\bar{\mathbf{x}}) - \nabla V^*(\bar{\varsigma}) = 0, \tag{7.22}$$

where $\Lambda_t(\mathbf{x}) = \nabla\Lambda(\mathbf{x})$ denotes the derivative of $\Lambda$, along with the conditions

$$0 \leq \bar{\boldsymbol{\sigma}} \quad \perp \quad (\langle \Lambda(\bar{\mathbf{x}}); \bar{\varsigma} \rangle - V^*(\bar{\varsigma}) - \mathbf{d}) \leq 0, \tag{7.23}$$

where the notation $\perp$ represents the complementarity or orthogonality condition. Since $(\boldsymbol{\xi}, \boldsymbol{\varsigma})$ is a canonical duality pair on $\mathcal{E}_a \times \mathcal{E}^*_a$, the criticality condition (7.22) is equivalent to $\bar{\varsigma} = \nabla_{\boldsymbol{\xi}} V(\Lambda(\bar{\mathbf{x}})) = \partial V(\boldsymbol{\xi}(\bar{\mathbf{x}}))/\partial \boldsymbol{\xi}$. Substituting this into (7.21) and using the chain rule to deduce $\nabla \mathbf{g}(\bar{\mathbf{x}}) = \langle \Lambda_t(\bar{\mathbf{x}}); \nabla_{\boldsymbol{\xi}} V(\Lambda(\bar{\mathbf{x}})) \rangle$, we have

$$\mathbf{A}\bar{\mathbf{x}} - \mathbf{f} + \bar{\boldsymbol{\sigma}}^T \nabla \mathbf{g}(\bar{\mathbf{x}}) = \nabla_{\mathbf{x}} L(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}) = 0.$$

This is the criticality condition of the primal problem $(\mathcal{P})$. By the Legendre equality $\langle \Lambda(\bar{\mathbf{x}}); \bar{\varsigma} \rangle - V^*(\bar{\varsigma}) = V(\Lambda(\bar{\mathbf{x}}))$, the condition (7.23) can be written as:

$$0 \leq \bar{\boldsymbol{\sigma}} \perp (\mathbf{g}(\bar{\mathbf{x}}) - \mathbf{d}) \leq 0.$$

This shows that $\bar{\mathbf{x}}$ is a KKT point of the primal problem $(\mathcal{P})$. From the complementarity condition (7.23), we have

$$\Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\varsigma}) = P(\bar{\mathbf{x}}).$$

On the other hand, by the definition of the canonical dual function, if $(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ is a KKT point, the criticality condition (7.21) leads to

$$U^\Lambda(\bar{\boldsymbol{\sigma}}, \bar{\varsigma}) = \bar{\boldsymbol{\sigma}}^T \langle \Lambda(\bar{\mathbf{x}}); \bar{\varsigma} \rangle - U(\bar{\mathbf{x}}).$$

Therefore, $\Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\varsigma}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\varsigma})$ and $(\bar{\boldsymbol{\sigma}}, \bar{\varsigma})$ is a KKT point of the dual problem $(\mathcal{P}^d)$. □

This theorem shows that there is no duality gap between the primal problem and its canonical dual. In order to identify the global minimizer, we need to study the convexity of the generalized complementary function $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$. Without losing much generality, we introduce the following assumptions:

(A1) the geometrical operator $\Lambda(\mathbf{x}) : \mathbb{R}^n \to \mathcal{E}_a$ is twice differentiable; and
(A2) the canonical function $V : \mathcal{E}_a \to \mathbb{R}^m$ is convex.

By Assumptions (A1) and (A2), we know that the conjugate function $V^*(\boldsymbol{\varsigma}) : \mathcal{E}_a^* \to \mathbb{R}^m$ is also convex, and for any given $(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k$, the generalized complementary function $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$ is twice differentiable on $\mathbf{x}$. Let $G_a(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \nabla^2_{\mathbf{x}} \Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$ denote the Hessian matrix of $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$ and let

$$\mathcal{S}_k^+ = \{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k \mid \; G_a(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) \succeq 0, \;\; \forall \mathbf{x} \in \mathcal{X}_a\} \tag{7.24}$$

be a subset of $\mathcal{S}_k$. We have the following theorem.

**Theorem 7.2** (Global Optimality Condition)**.** *Suppose that Assumptions (A1) and (A2) hold and that* $(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ *is a critical point of* $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$*. If* $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}) \in \mathcal{S}_k^+$*, then* $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ *is a global maximizer of* $P^d$ *on* $\mathcal{S}_k^+$ *and* $\bar{\mathbf{x}}$ *is a global minimizer of* $P$ *on* $\mathcal{X}_k$*, i.e.,*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x} \in \mathcal{X}_k} P(\mathbf{x}) = \max_{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k^+} P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}). \tag{7.25}$$

*Proof.* By the convexity of $V(\boldsymbol{\xi})$, its Legendre conjugate $V^* : \mathcal{E}_a^* \to \mathbb{R}^m$ is also convex. Thus, for any given $\boldsymbol{\sigma} \in \mathbb{R}^m_+$, the linear combination $\boldsymbol{\sigma}^T V^*(\boldsymbol{\varsigma}) : \mathcal{E}_a^* \to \mathbb{R}$ is convex and the generalized complementary function $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$ is concave in $\boldsymbol{\varsigma}$. By considering $\boldsymbol{\sigma} \in \mathbb{R}^m_+$ as a Lagrange multiplier for the inequality constraint in $\mathcal{X}_c$, the complementary function $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$ can be viewed as a concave (linear) function of $\boldsymbol{\sigma} \in \mathbb{R}^m_+$ for any given $(\mathbf{x}, \boldsymbol{\varsigma}) \in \mathcal{X}_a \times \mathcal{E}_a^*$. Therefore, for any given $\mathbf{x} \in \mathbb{R}^n$, we have

$$\max_{\boldsymbol{\sigma} \in \mathbb{R}^m_+} \max_{\boldsymbol{\varsigma} \in \mathcal{E}_a^*} \Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \max_{\boldsymbol{\sigma} \in \mathbb{R}^m_+} L(\mathbf{x}, \boldsymbol{\sigma}) = \begin{cases} P(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{X}_c, \\ \infty & \text{otherwise}. \end{cases}$$

Moreover, if $(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k^+$, then $\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma})$ is convex in $\mathbf{x} \in \mathcal{X}_a$ and concave in $\boldsymbol{\varsigma}$ for any given $\boldsymbol{\sigma} \in \mathbb{R}^m_+$. Therefore, if $(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ is a critical point of $\Xi$, we have

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}_k} P(\mathbf{x}) &= \min_{\mathbf{x} \in \mathcal{X}_a} \max_{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k^+} \Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \Xi(\bar{\mathbf{x}}, \bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}) \\ &= \max_{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k^+} \min_{\mathbf{x} \in \mathcal{X}_a} \Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \max_{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k^+} P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma}). \end{aligned}$$

By Theorem 7.1, we have (7.25). □

This theorem provides a sufficient condition for a global minimizer of the nonconvex primal problem. In many applications, the geometrical mapping $\Lambda(\mathbf{x}) : \mathcal{X}_a \to \mathcal{E}_a$ is usually a quadratic operator

$$\Lambda(\mathbf{x}) = \left\{ \frac{1}{2}\mathbf{x}^T\mathbf{B}^\alpha_\beta\mathbf{x} + \mathbf{x}^T\mathbf{C}^\alpha_\beta \right\} : \mathbb{R}^n \to \mathcal{E}_a \subset \mathbb{R}^{m\times p_\alpha}, \tag{7.26}$$

where $\mathbf{B}^\alpha_\beta = \{B^\alpha_{ij\beta}\} = \{B^\alpha_{ji\beta}\} \in \mathbb{R}^{n\times n}$, $\mathbf{C}^\alpha_\beta = \{C^\alpha_{i\beta}\} \in \mathbb{R}^n$, and the range $\mathcal{E}_a$ depends on both $\mathbf{B}^\alpha_\beta$ and $\mathbf{C}^\alpha_\beta$. In this case, the generalized complementary function has the form:

$$\Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \frac{1}{2}\mathbf{x}^T G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{x} - \boldsymbol{\sigma}^T(V^*(\boldsymbol{\varsigma}) + \mathbf{d}) - \mathbf{x}^T\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma}), \tag{7.27}$$

where

$$G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \mathbf{A} + \sum_{\alpha=1}^m \sum_{\beta\in I_\alpha} \sigma_\alpha \mathbf{B}^\alpha_\beta \varsigma^\beta_\alpha \tag{7.28}$$

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

$$\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \mathbf{f} - \sum_{\alpha=1}^m \sum_{\beta\in I_\alpha} \sigma_\alpha \mathbf{C}^\alpha_\beta \varsigma^\beta_\alpha. \tag{7.29}$$

The criticality condition (7.21) in this case is a linear equation of $\mathbf{x}$, i.e.,

$$G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{x} = \mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma}). \tag{7.30}$$

Clearly, for a given $(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$, if $\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$ is in the column space of $G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$, denoted by $\mathcal{C}_{ol}(G_a)$, the solution of the equation (7.30) can be written in the form:

$$\mathbf{x} = G_a^+(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma}), \tag{7.31}$$

where $G_a^+$ is the Moore-Penrose generalized inverse of $G_a$. Thus, the canonical dual feasible space $\mathcal{S}_k$ can be defined as:

$$\mathcal{S}_k = \{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathbb{R}^m_+ \times \mathcal{E}^*_a |\ \mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{C}_{ol}(G_a)\}, \tag{7.32}$$

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

$$P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) = -\frac{1}{2}\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma})^T G_a^+(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) - \boldsymbol{\sigma}^T(V^*(\boldsymbol{\varsigma}) + \mathbf{d}). \tag{7.33}$$

Since $\Lambda(\mathbf{x})$ is a quadratic operator, its derivative is an affine operator

$$\Lambda_t(\mathbf{x}) = \nabla\Lambda(\mathbf{x}) = \mathbf{x}^T\mathbf{B}^\alpha_\beta + \mathbf{C}^\alpha_\beta.$$

The complementary operator $\Lambda_c(\mathbf{x})$ of $\Lambda_t$ is defined by

$$\Lambda_c(\mathbf{x}) = \Lambda(\mathbf{x}) - \Lambda_t(\mathbf{x})\mathbf{x} = -\frac{1}{2}\mathbf{x}^T\mathbf{B}^{\alpha}_{\beta}\mathbf{x}. \tag{7.34}$$

Thus, the complementary gap function can be defined as:

$$G(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = \boldsymbol{\sigma}^T\langle -\Lambda_c(\mathbf{x}); \boldsymbol{\varsigma}\rangle + \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} = \frac{1}{2}\mathbf{x}^T G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{x}. \tag{7.35}$$

This gap function plays an important role in nonconvex analysis and global optimization. Clearly, $G(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) \geq 0 \; \forall \mathbf{x} \in \mathcal{X}_a$ if $G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \succeq 0$. Let

$$\mathcal{S}^+_k = \{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k \mid G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \succeq 0\}, \tag{7.36}$$

**Theorem 7.3.** *Suppose that $\Lambda(\mathbf{x})$ is a quadratic operator defined by (7.26) and Assumption (A2) holds.*

*If $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}) \in \mathcal{S}_k$ is a critical point of $(\mathcal{P}^d)$, then $\bar{\mathbf{x}} = G^+_a(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})\mathbf{F}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ is a KKT point of $(\mathcal{P})$ and $P(\bar{\mathbf{x}}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$.*

*If the critical point $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}) \in \mathcal{S}^+_c$, then $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ is a global maximizer of $P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$ on $\mathcal{S}^+_k$, the vector $\bar{\mathbf{x}}$ is a global minimizer of $P(\mathbf{x})$ on $\mathcal{X}_k$, and*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x} \in \mathcal{X}_k} P(\mathbf{x}) = \max_{(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}^+_k} P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) = P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}). \tag{7.37}$$

*Proof.* If $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}) \in \mathcal{S}_k$ is a critical point of $(\mathcal{P}^d)$, we have

$$\delta_{\boldsymbol{\varsigma}} P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}) = \bar{\boldsymbol{\sigma}}^T \left(\Lambda(\bar{\mathbf{x}}) - \nabla V^*(\bar{\boldsymbol{\varsigma}})\right) = 0, \tag{7.38}$$

$$0 \leq \bar{\boldsymbol{\sigma}} \quad \perp \quad \delta_{\boldsymbol{\sigma}} P^d(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}}) = \langle \Lambda(\bar{\mathbf{x}}); \bar{\boldsymbol{\varsigma}}\rangle - V^*(\bar{\boldsymbol{\varsigma}}) - \mathbf{d} \leq 0, \tag{7.39}$$

where $\bar{\mathbf{x}} = G^+_a(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})\mathbf{F}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$. Equation (7.38) asserts that if $\bar{\sigma}_\alpha \neq 0$, then the corresponding $\xi_\alpha(\bar{\mathbf{x}}) = \Lambda_\alpha(\bar{\mathbf{x}}) = \nabla_{\varsigma_\alpha} V^*(\bar{\boldsymbol{\varsigma}})$. By the fact that $(\Lambda(\bar{\mathbf{x}}), \bar{\boldsymbol{\varsigma}})$ is a canonical duality pair on $\mathcal{E}_a \times \mathcal{E}^*_a$, from the equivalent relations in (7.15), we have $\langle \Lambda(\bar{\mathbf{x}}); \bar{\boldsymbol{\varsigma}}\rangle - V^*(\bar{\boldsymbol{\varsigma}}) = V(\Lambda(\bar{\mathbf{x}})) = \mathbf{g}(\bar{\mathbf{x}})$. Therefore, the complementarity condition in (7.39) leads to $\bar{\sigma}_\alpha(g_\alpha(\bar{\mathbf{x}}) - d_\alpha) = 0$. If $\bar{\sigma}_\alpha \neq 0$, we have the criticality condition $g_\alpha(\bar{\mathbf{x}}) - d_\alpha = 0$. This shows that if $(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ is a critical point of $P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$, the vector $\bar{\mathbf{x}} = G^+_a(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})\mathbf{F}(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varsigma}})$ is a KKT point of $(\mathcal{P})$.

□

## 7.3 Quadratic Constrained Problems

We begin by considering the following nonconvex quadratic minimization problem with quadratic inequality constraints, denoted by $(\mathcal{P}_q)$.

$$(\mathcal{P}_q): \quad \begin{cases} \min \left\{ P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{f} \right\} \\ s.t. \ \frac{1}{2}\mathbf{x}^T\mathbf{B}^\alpha\mathbf{x} + \mathbf{x}^T\mathbf{C}^\alpha \le d_\alpha, \quad \alpha = 1, \dots, m, \end{cases} \tag{7.40}$$

where $\mathbf{B}^\alpha = \{B_{ij}^\alpha\} = \{B_{ji}^\alpha\} \in \mathbb{R}^{n\times n}, \ \mathbf{C}^\alpha = \{C_i^\alpha\} \in \mathbb{R}^n, \forall\ \alpha = 1, \cdots, m$, and $\mathbf{d} = \{d_\alpha\} \in \mathbb{R}^m$ is a vector. Due to the nonconvex cost function and nonconvex inequality constraints, this problem is known to be NP-hard.

Since the constraint $\mathbf{g}(\mathbf{x})$ is a vector-valued quadratic function defined on $\mathcal{X}_a = \mathbb{R}^n$, we simply let

$$\mathbf{g}(\mathbf{x}) = \Lambda(\mathbf{x}) = \left\{ \frac{1}{2}\mathbf{x}^T\mathbf{B}^\alpha\mathbf{x} + \mathbf{x}^T\mathbf{C}^\alpha \right\} : \mathbb{R}^n \to \mathbb{R}^m. \tag{7.41}$$

Compared with (7.26), we have $p_\alpha = 1$ and the canonical function $V(\boldsymbol{\xi}) = \boldsymbol{\xi}$ is a self-mapping. Therefore, the canonical dual variable $\boldsymbol{\varsigma} = \nabla V(\boldsymbol{\xi}) = I$ is an identity matrix in $\mathbb{R}^{m\times m}$ and $V^*(\boldsymbol{\varsigma}) = \text{sta}\{\langle \boldsymbol{\xi}; \boldsymbol{\varsigma}\rangle - \boldsymbol{\xi} |\ \boldsymbol{\xi} \in \mathbb{R}^m\} = 0$. In this case, the generalized complementary function (7.27) has a very simple form:

$$\Xi_q(\mathbf{x}, \boldsymbol{\sigma}) = \frac{1}{2}\mathbf{x}^T G_q(\boldsymbol{\sigma})\mathbf{x} - \mathbf{x}^T\mathbf{F}_q(\boldsymbol{\sigma}) - \boldsymbol{\sigma}^T\mathbf{d}, \tag{7.42}$$

where

$$G_q(\boldsymbol{\sigma}) = \mathbf{A} + \sum_{\alpha=1}^m \sigma_\alpha \mathbf{B}^\alpha, \quad \text{and} \quad \mathbf{F}_q(\boldsymbol{\sigma}) = \mathbf{f} - \sum_{\alpha=1}^m \sigma_\alpha \mathbf{C}^\alpha. \tag{7.43}$$

Therefore, on the dual feasible space

$$\mathcal{S}_q = \{\boldsymbol{\sigma} \in \mathbb{R}^m_+ |\ \mathbf{F}_q(\boldsymbol{\sigma}) \in \mathcal{C}_{ol}(G_q)\}, \tag{7.44}$$

the canonical dual function $P_q^d$ can be formulated as

$$P_q^d(\boldsymbol{\sigma}) = -\frac{1}{2}\mathbf{F}_q(\boldsymbol{\sigma})^T G_q^+(\boldsymbol{\sigma})\mathbf{F}_q(\boldsymbol{\sigma}) - \boldsymbol{\sigma}^T\mathbf{d}. \tag{7.45}$$

In this case, the complementary gap function has a simple form:

$$G(\mathbf{x}, \boldsymbol{\sigma}) = \frac{1}{2}\mathbf{x}^T G_q(\boldsymbol{\sigma})\mathbf{x}, \tag{7.46}$$

which is nonnegative on $\mathbb{R}^n$ if $G_q(\boldsymbol{\sigma}) \succeq 0$. Let

$$\mathcal{S}_q^+ = \{\boldsymbol{\sigma} \in \mathcal{S}_q |\ G_q(\boldsymbol{\sigma}) \succeq 0\}, \ \ \mathcal{S}_q^- = \{\boldsymbol{\sigma} \in \mathcal{S}_q |\ G_q(\boldsymbol{\sigma}) \prec 0\}. \tag{7.47}$$

Then the canonical dual problem for this quadratic constrained problem is given by

$$(\mathcal{P}_q^d): \quad \max\{P_q^d(\boldsymbol{\sigma}): \ \boldsymbol{\sigma} \in \mathcal{S}_q^+\}. \tag{7.48}$$

And we have the following result.

**Theorem 7.4.** *The problem $(\mathcal{P}_q^d)$ is canonically dual to the primal problem $(\mathcal{P}_q)$ in the sense that for each critical point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_q$ of $(\mathcal{P}_q^d)$, the vector $\bar{\mathbf{x}} = G_q^+(\bar{\boldsymbol{\sigma}})\mathbf{F}_q(\bar{\boldsymbol{\sigma}})$ is a KKT point of $(\mathcal{P}_q)$ and $P(\bar{\mathbf{x}}) = P^d(\bar{\boldsymbol{\sigma}})$.*

*Particularly, if the critical point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_q^+$, then $\bar{\boldsymbol{\sigma}}$ is a global maximizer of $(\mathcal{P}_q^d)$. The vector $\bar{\mathbf{x}}$ is a global minimizer of $(\mathcal{P}_q)$, and*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x} \in \mathcal{X}_k} P(\mathbf{x}) = \max_{\boldsymbol{\sigma} \in \mathcal{S}_q^+} P_q^d(\boldsymbol{\sigma}) = P_q^d(\bar{\boldsymbol{\sigma}}). \tag{7.49}$$

*If $G_q(\bar{\boldsymbol{\sigma}}) \succ 0$, then $\bar{\boldsymbol{\sigma}}$ is the unique global maximizer of $(\mathcal{P}_q^d)$ and the vector $\bar{\mathbf{x}} = G_q^{-1}(\bar{\boldsymbol{\sigma}})\mathbf{F}_q(\bar{\boldsymbol{\sigma}})$ is the unique global minimizer of $(\mathcal{P}_q)$.*

Theorem 7.4 shows that the Hessian matrix of the complementary gap function $G(\mathbf{x}, \boldsymbol{\sigma})$ provides sufficient and uniqueness conditions for globally minimizing the quadratic constrained problem $(\mathcal{P}_q)$. In order to study the existence theory, we need to introduce the following sets:

$$\partial\mathcal{S}_q = \{\boldsymbol{\sigma} \in \mathbb{R}^m | \ \det G_q(\boldsymbol{\sigma}) = 0\}, \tag{7.50}$$

$$\partial\mathcal{S}_q^+ = \{\boldsymbol{\sigma} \in \mathcal{S}_q | \ \det G_q(\boldsymbol{\sigma}) = 0\}. \tag{7.51}$$

**Theorem 7.5.** *Suppose that for given matrices $\mathbf{A}$, $\{\mathbf{B}^\alpha\}$, $\{\mathbf{C}^\alpha\}$ and vectors $\mathbf{f}$, $\mathbf{d}$, there exists at least one $\boldsymbol{\sigma}_0 \in \mathcal{S}_q^+$ such that $G_q(\boldsymbol{\sigma}_0) \succeq 0$ and*

$$\lim_{\substack{\|\boldsymbol{\sigma}\| \to \infty \\ \boldsymbol{\sigma} \in \mathcal{S}_q^+}} P_q^d(\boldsymbol{\sigma}) = -\infty. \tag{7.52}$$

*Then the canonical dual problem $(\mathcal{P}_q^d)$ has at least one KKT point $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_q^+$. If $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_q^+$ is also a critical point of $P_q^d(\boldsymbol{\sigma})$, then $\bar{\mathbf{x}} = G_q^+(\bar{\boldsymbol{\sigma}})\mathbf{F}_q(\bar{\boldsymbol{\sigma}})$ is a global minimizer for the primal problem $(\mathcal{P}_q)$.*

*Moreover, if $\partial\mathcal{S}_q \subset \mathbb{R}^m_+$, there exists at least one $\boldsymbol{\sigma}_0 \in \mathcal{S}_q^+$ such that $G_q(\boldsymbol{\sigma}_0) \succ 0$, and*

$$\lim_{\substack{\boldsymbol{\sigma} \to \partial\mathcal{S}_q^+ \\ \boldsymbol{\sigma} \in \mathcal{S}_q^+}} P_q^d(\boldsymbol{\sigma}) = -\infty, \tag{7.53}$$

*then the canonical dual problem $(\mathcal{P}_q^d)$ has a unique global maximizer $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_q^+$ and $\bar{\mathbf{x}} = G_q^{-1}(\bar{\boldsymbol{\sigma}})\mathbf{F}_q(\bar{\boldsymbol{\sigma}})$ is a unique global minimizer for the primal problem $(\mathcal{P}_q)$.*

*Proof.* By the fact that the feasible space $\mathcal{S}_q^+$ is a semi-closed convex set whose boundary $\partial\mathcal{S}_q^+$ is a hyper surface in $\mathbb{R}^m$, if there exists a $\boldsymbol{\sigma}_0$ such that $G_q(\boldsymbol{\sigma}_0) \succeq 0$, then $\mathcal{S}_q^+$ is not empty. Since the canonical dual function $P_q^d(\boldsymbol{\sigma})$ is continuous and concave on $\mathcal{S}_q^+$, which is finite on $\partial\mathcal{S}_q^+$, if the condition (7.52) holds, then $P_q^d(\boldsymbol{\sigma})$ has at least one maximizer on $\mathcal{S}_q^+$.

Moreover, if $\partial\mathcal{S}_q \subset \mathbb{R}_+^m$, then $\mathcal{S}_q^+ \subset \mathbb{R}_+^m$. If there exists a $\boldsymbol{\sigma}_0$ such that $G_q(\boldsymbol{\sigma}_0) \succ 0$, then $\mathcal{S}_q^+$ is non-empty and has at least one interior point. Under the conditions (7.52) and (7.53), the canonical dual function $P_q^d(\boldsymbol{\sigma})$ is strictly concave and coercive on the open convex set $\mathcal{S}_q^+ \setminus \partial\mathcal{S}_q^+$. Therefore, the canonical dual problem $(\mathcal{P}_q)$ has a unique maximizer $\bar{\boldsymbol{\sigma}} \in \mathcal{S}_q^+$ that is a critical point of $P_q^d(\boldsymbol{\sigma})$. □

Theorem 7.5 shows that under conditions (7.52) and (7.53), the canonical dual function $P_q^d(\boldsymbol{\sigma})$ has a unique maximizer $\bar{\boldsymbol{\sigma}}$ on the open feasible space

$$\mathcal{S}_q^\ddagger = \{\boldsymbol{\sigma} \in \mathcal{S}_q |\ G_q(\boldsymbol{\sigma}) \succ 0\}. \tag{7.54}$$

In this case, the matrix $G_q(\boldsymbol{\sigma})$ is invertible on $\mathcal{S}_q^\ddagger$ and the canonical dual function $P_q^d$ can be written as:

$$P_q^d(\boldsymbol{\sigma}) = -\frac{1}{2}\mathbf{F}_q(\boldsymbol{\sigma})^T G_q^{-1}(\boldsymbol{\sigma})\mathbf{F}_q(\boldsymbol{\sigma}) - \boldsymbol{\sigma}^T\mathbf{d}. \tag{7.55}$$

Particularly, if $m = 1$ and $\mathbf{C} = \mathbf{0}$, Problem $(\mathcal{P}_q)$ has only one quadratic constraint $g(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T B\mathbf{x} \le d$. Therefore, the canonical dual function has only one variable:

$$P_q^d(\sigma) = -\frac{1}{2}\mathbf{f}^T(\mathbf{A} + \sigma B)^{-1}\mathbf{f} - \sigma d, \tag{7.56}$$

and the criticality condition $\nabla P_q^d(\sigma) = 0$ leads to a nonlinear algebraic equation

$$\frac{1}{2}\mathbf{f}^T(\mathbf{A} + \sigma B)^{-1}B(\mathbf{A} + \sigma B)^{-1}\mathbf{f} = d, \tag{7.57}$$

which can be solved easily to obtain all dual solutions. Moreover, if $B = I$ is an identity matrix in $\mathbb{R}^n$, then the constraint $\frac{1}{2}\mathbf{x}^T B\mathbf{x} = \frac{1}{2}\|\mathbf{x}\|^2 \le d$ is an $n$-dimensional sphere.

## 7.4 Nonconvex Polynomial Constrained Problems

We now assume that $\mathbf{g}(\mathbf{x})$ is a general fourth order polynomial constraint given by

$$\mathbf{g}(\mathbf{x}) = \left\{ \sum_{\beta \in I_\alpha} \frac{1}{2} D_\alpha^\beta \left( \frac{1}{2}\mathbf{x}^T \mathbf{B}_\beta^\alpha \mathbf{x} + \mathbf{x}^T \mathbf{C}_\beta^\alpha - E_\beta^\alpha \right)^2 \right\} \le \mathbf{d}, \tag{7.58}$$

where $\mathbf{B}^\alpha_\beta = \{B^\alpha_{ij\beta}\} \in \mathbb{R}^{n\times n}$ and $\mathbf{C}^\alpha_\beta = \{C^\alpha_{i\beta}\} \in \mathbb{R}^n$ are given as before, $I_\alpha$ is a (finite) index set that depends on each index $\alpha = 1,\ldots,m$, and $\{D^\beta_\alpha\}$ and $\{E^\alpha_\beta\}$ are two given second order tensors. We assume that $D^\beta_\alpha > 0,\ \ \forall \alpha \in \{1,\ldots,m\},\ \beta \in I_\alpha$.

By introducing a geometrical measure

$$\boldsymbol{\xi} = \Lambda(\mathbf{x}) = \{\xi^\alpha_\beta\} = \left\{\frac{1}{2}\mathbf{x}^T\mathbf{B}^\alpha_\beta\mathbf{x} + \mathbf{x}^T\mathbf{C}^\alpha_\beta\right\} : \mathbb{R}^n \to \mathcal{E}_a,$$

where the range of $\mathcal{E}_a$ depends on the tensors $\{\mathbf{B}^\alpha_\beta\}$ and $\{\mathbf{C}^\alpha_\beta\}$, the canonical function

$$V(\boldsymbol{\xi}) = \left\{\sum_{\beta\in I_\alpha}\frac{1}{2}D^\beta_\alpha\left(\xi^\alpha_\beta - E^\alpha_\beta\right)^2\right\} : \mathcal{E}_a \to \mathbb{R}^m$$

is a quadratic function. Thus, the canonical duality relation

$$\boldsymbol{\varsigma} = \nabla V(\boldsymbol{\xi}) = \{D^\beta_\alpha(\xi^\alpha_\beta - E^\alpha_\beta)\} : \mathcal{E}_a \to \mathcal{E}^*_a$$

is a linear mapping, where the range of $\mathcal{E}^*_a$ depends on the tensors $\{D^\beta_\alpha\}$ and $\{E^\alpha_\beta\}$. The Legendre conjugate $V^*$ can be defined uniquely as:

$$V^*(\boldsymbol{\varsigma}) = \left\{\sum_{\beta\in I_\alpha}\left(\frac{1}{2D^\beta_\alpha}(\varsigma^\beta_\alpha)^2 + E^\alpha_\beta\varsigma^\beta_\alpha\right)\right\}. \tag{7.59}$$

Substituting this into (7.33), the canonical dual function has the following form:

$$P^d(\boldsymbol{\sigma},\boldsymbol{\varsigma}) = -\frac{1}{2}\mathbf{F}(\boldsymbol{\sigma},\boldsymbol{\varsigma})^T G^+_a(\boldsymbol{\sigma},\boldsymbol{\varsigma})\mathbf{F}(\boldsymbol{\sigma},\boldsymbol{\varsigma}) - \sum_{\alpha=1}^m\sum_{\beta\in I_\alpha}\left(\sigma_\alpha\left(\frac{1}{2D^\beta_\alpha}(\varsigma^\beta_\alpha)^2 + E^\alpha_\beta\varsigma^\beta_\alpha + d_\alpha\right)\right), \tag{7.60}$$

which is concave on the dual feasible space

$$\mathcal{S}^+_c = \{(\boldsymbol{\sigma},\boldsymbol{\varsigma}) \in \mathbb{R}^m_+ \times \mathcal{E}^*_a |\ \mathbf{F}(\boldsymbol{\sigma},\boldsymbol{\varsigma}) \in \mathcal{C}_{ol}(G_a),\ \ G_a(\boldsymbol{\sigma},\boldsymbol{\varsigma}) \succeq 0\}. \tag{7.61}$$

**Remark 7.1.** It is again insightful to view the connection between the canonical dual (7.60) and the classical Lagrangian dual. In this case, we have

$$L(\mathbf{x},\boldsymbol{\sigma}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{f} + \boldsymbol{\sigma}^T\left(\mathbf{g}(\mathbf{x}) - \mathbf{d}\right). \tag{7.62}$$

Clearly, without introducing the canonical dual pair $(\boldsymbol{\xi},\boldsymbol{\varsigma})$, the Fenchel-Moreau-Rockafellar dual $P^*(\boldsymbol{\sigma}) = \inf_{\mathbf{x}\in\mathcal{X}_a} L(\mathbf{x},\boldsymbol{\sigma})$ cannot be defined explicitly due to the high order nonlinearity of the constraint $\mathbf{g}(\mathbf{x}) \le \mathbf{d}$. However, by using the canonical dual transformation

$\boldsymbol{\xi} = \Lambda(\mathbf{x})$ and the chain rule, the necessary condition $\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\sigma}) = 0$, leads to

$$\mathbf{A}\mathbf{x} - \mathbf{f} + \sum_{\alpha=1}^{m} \sum_{\beta \in I_\alpha} \left( \sigma_\alpha \varsigma_\alpha^\beta (\mathbf{B}_\beta^\alpha \mathbf{x} + \mathbf{C}_\beta^\alpha) \right) = 0. \tag{7.63}$$

This is the canonical equilibrium equation $G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{x} = \mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$, where

$$\boldsymbol{\varsigma} \equiv \frac{\partial V(\Lambda(\mathbf{x}))}{\partial \boldsymbol{\xi}} \tag{7.64}$$

is as defined above. If $G_a(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$ is invertible for $(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_k$ as defined in (7.32), it follows from (7.63) that $\mathbf{x}$ uniquely satisfies $\mathbf{x} = G_a^{-1}(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$. Furthermore, by (7.63), we get

$$\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} = \frac{1}{2}\mathbf{x}^T \mathbf{f} - \frac{1}{2}\sum_{\alpha=1}^{m} \sum_{\beta \in I_\alpha} \boldsymbol{\sigma}_\alpha \boldsymbol{\varsigma}_\alpha^\beta [\mathbf{x}^T B_\beta^\alpha \mathbf{x} + \mathbf{x}^T C_\beta^\alpha]. \tag{7.65}$$

Substituting for $\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x}$ in (7.62) using (7.65), and then applying the optimality condition $\mathbf{x} = G_a^{-1}(\boldsymbol{\sigma}, \boldsymbol{\varsigma})\mathbf{F}(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$, the Lagrangian dual function reduces precisely to $P^d(\boldsymbol{\sigma}, \boldsymbol{\varsigma})$ defined in (7.60) under the global optimality condition $(\boldsymbol{\sigma}, \boldsymbol{\varsigma}) \in \mathcal{S}_c^+$. Note that $\nabla_{\boldsymbol{\varsigma}} \Xi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\varsigma}) = 0$ produces the inverse of the identity (7.64) under the relevant case when $\boldsymbol{\sigma}^\alpha \neq 0, \ \forall \alpha = 1, \dots, m$, thus validating the foregoing derivation.

We now present some special cases.

### 7.4.1 Quadratic minimization with one nonconvex polynomial constraint

We first assume that the primal problem has only one nonconvex constraint

$$g(\mathbf{x}) = \frac{1}{2}\left( \frac{1}{2}\mathbf{x}^T B \mathbf{x} + \mathbf{x}^T \mathbf{c} - \eta \right)^2 \leq d, \tag{7.66}$$

where $B$ is an $n \times n$ matrix, $\mathbf{c} \in \mathbb{R}^n$ is a vector, and $\eta > 0$ is a constant. In this case, $m = |I_\alpha| = 1$, and

$$G_a(\sigma, \varsigma) = \mathbf{A} + \sigma\varsigma B, \quad \mathbf{F}(\sigma, \varsigma) = \mathbf{f} - \sigma\varsigma \mathbf{c}.$$

The canonical dual function is

$$P^d(\sigma, \varsigma) = -\frac{1}{2}\mathbf{F}^T(\sigma, \varsigma) G_a^+(\sigma, \varsigma) \mathbf{F}(\sigma, \varsigma) - \sigma \left( \frac{1}{2}\varsigma^2 + \eta\varsigma + d \right). \tag{7.67}$$

### 7.4.2 Combined quadratic and nonconvex polynomial constraints

We now consider the problem with the following two constraints:

$$g_1(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{B}^1\mathbf{x} + \mathbf{x}^T\mathbf{C}^1 \le d_1,$$
$$g_2(\mathbf{x}) = \frac{1}{2}\left(\frac{1}{2}\mathbf{x}^T\mathbf{B}^2\mathbf{x} + \mathbf{x}^T\mathbf{C}^2 - \eta\right)^2 \le d_2.$$

In this case, $m = 2$, $I_\alpha = \{1\}$, for $\alpha = 1, 2$, and the geometrical operator

$$\boldsymbol{\xi} = \Lambda(\mathbf{x}) = \left\{\frac{1}{2}\mathbf{x}\mathbf{B}^\alpha\mathbf{x} + \mathbf{x}^T\mathbf{C}^\alpha\right\} : \mathbb{R}^n \to \mathbb{R}^2$$

is a 2-vector. The canonical function $V(\boldsymbol{\xi})$ is a vector-valued function

$$V(\boldsymbol{\xi}) = \{\xi_1, \frac{1}{2}(\xi_2 - \eta)^2\}.$$

The canonical dual variable is $\boldsymbol{\varsigma} = \nabla V(\boldsymbol{\xi}) = [1, \xi_2 - \eta]^T$. Since $\varsigma_1 = 1$, we let $\varsigma_2 = \varsigma$. Thus, the canonical dual function has only three variables $(\sigma_1, \sigma_2, \varsigma) \in \mathbb{R}^3$, i.e.,

$$P^d(\sigma_1, \sigma_2, \varsigma) = -\frac{1}{2}\mathbf{F}(\sigma_1, \sigma_2, \varsigma)^T G_a^+(\sigma_1, \sigma_2, \varsigma)\mathbf{F}(\sigma_1, \sigma_2, \varsigma) - \sigma_1 d_1 - \sigma_2\left(\frac{1}{2}\varsigma^2 + \eta\varsigma + d_2\right), \tag{7.68}$$

where

$$G_a(\sigma_1, \sigma_2, \varsigma) = \mathbf{A} + \sigma_1\mathbf{B}^1 + \sigma_2\varsigma\mathbf{B}^2, \quad \mathbf{F}(\sigma_1, \sigma_2, \varsigma) = \mathbf{f} - \sigma_1\mathbf{C}^1 - \sigma_2\varsigma\mathbf{C}^2.$$

## 7.5 Numerical Examples

**Example 7.1** In 2-D space, let

$$\mathbf{A} = \begin{pmatrix} 3 & 0.5 \\ 0.5 & -2.0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix}, \quad \text{and} \quad \mathbf{f} = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix}.$$

Clearly, the matrix $\mathbf{A}$ is indefinite, and $B$ is positive definite. Setting $d = 2$, the graph of the primal function $P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{f}$ is a saddle surface (see Fig. 7.1), and the boundary of the feasible set $\mathcal{X}_k = \{\mathbf{x} \in \mathbb{R}^2 | \ \frac{1}{2}\mathbf{x}^T B\mathbf{x} \le d\}$ is an ellipse (see Fig. 7.1). In this case, the canonical dual function (7.56) can be formulated as:

$$P_q^d(\sigma) = -\frac{1}{2}(1 \ \ 1.5)\begin{pmatrix} 3+\sigma & 0.5 \\ 0.5 & -2+0.5\sigma \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ 1.5 \end{pmatrix} - 2\sigma,$$

which has four critical points (see Fig. 7.2):

$$\bar{\sigma}_1 = 5.08 > \bar{\sigma}_2 = 3.06 > \bar{\sigma}_3 = -2.46 > \bar{\sigma}_4 = -3.68.$$

Since $G_q(\bar{\sigma}_1) \succ 0$ , it yields that $\mathbf{x}_1 = (-0.05, 2.83)^T$ is a global minimizer located on the boundary of $\mathcal{X}_k$. We have

$$P(\mathbf{x}_1) = -12.25 = P^d(\bar{\sigma}_1).$$

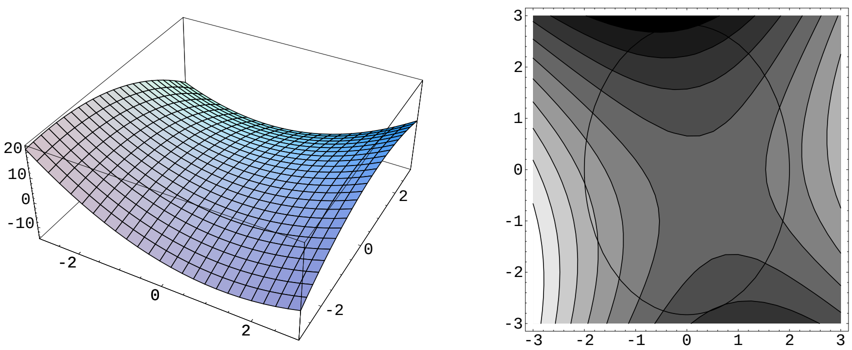



Figure 7.1: Graph of $P(\mathbf{x})$ (left); contours of $P(\mathbf{x})$ and boundary of $\mathcal{X}_k$ (right) for Example 7.1

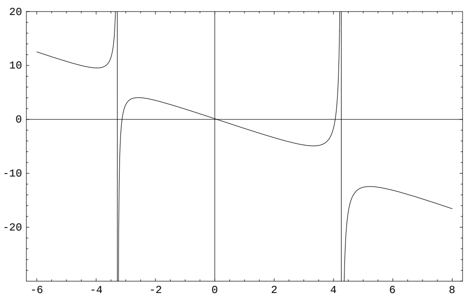



Figure 7.2: Graph of $P^d(\sigma)$.

**Example 7.2** In 2-D space, let $B$ be an identity matrix, $\mathbf{c} = \mathbf{0}$, and $\mathbf{A}$ a diagonal matrix with $a_{11} = 0.6, a_{12} = a_{21} = 0$, and $a_{22} = -0.5$. Setting $\mathbf{f} = [0.2, -0.1]^T$, $d = 1$, and $\eta = 1.5$, the constraint $g(\mathbf{x}) \leq d$ is an annulus (see Fig. 7.3 (right)). Solving the dual problem, we get

$$\bar{\sigma} = 0.3829201, \quad \text{and} \quad \bar{\varsigma} = 1.4142136.$$

The primal solution

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.1752033 \\ -2.4078477 \end{bmatrix}$$

is located on the boundary of the feasible set $\mathcal{X}_k$ (see Fig. 7.3) and we have

$$P(\bar{\mathbf{x}}) = -1.7160493 = P^d(\bar{\sigma}, \bar{\varsigma}).$$

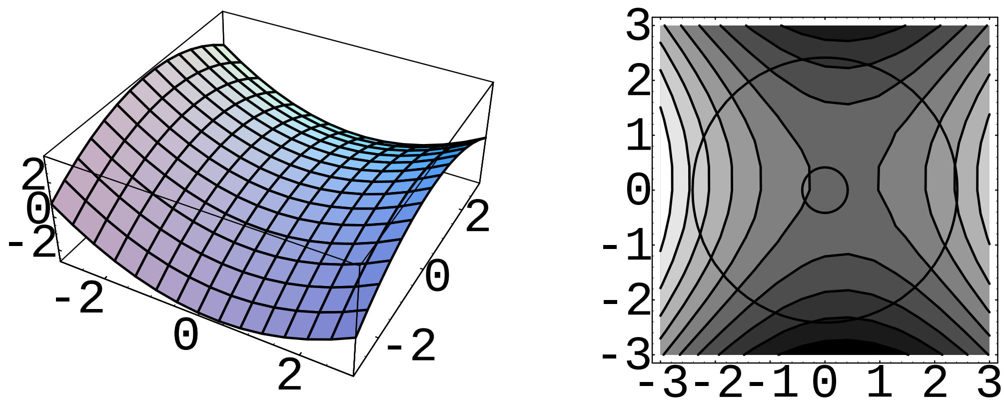



Figure 7.3: Graph of $P(\mathbf{x})$ (left); contours of $P(\mathbf{x})$ and boundary of $\mathcal{X}_k$ (right) for Example 7.2.

**Example 7.3** Let $\mathbf{A}$ be a $2 \times 2$ diagonal matrix, where $a_{11} = -0.4$, $a_{12} = a_{21} = 0$, and $a_{22} = 0.6$. Setting $\mathbf{f} = [0.3, -0.15]^T$, $\mathbf{B}^1 = \mathbf{B}^2 = \mathbf{I}$, $\mathbf{C}^1 = \mathbf{C}^2 = \mathbf{0}$, $d_1 = 2$, $d_2 = 1.2$, and $\eta = 1.7$, the graph of the objective function $P(x_1, x_2)$ is a saddle surface (Fig. 7.4 (left)), the constraint $g_1(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) \leq 2$ is a disk of radius 2, while $g_2(x_1, x_2) = \frac{1}{2}\left(\frac{1}{2}(x_1^2 + x_2^2) - 1.7\right)^2 \leq 1.2$ represents an annulus (see Fig. 7.4 (right)). Solving the dual problem, we get

$$\bar{\sigma}_1 = 0.5503198, \quad \bar{\sigma}_2 = 0, \quad \text{and} \quad \bar{\varsigma} = 0.3159349.$$

The primal solution is therefore

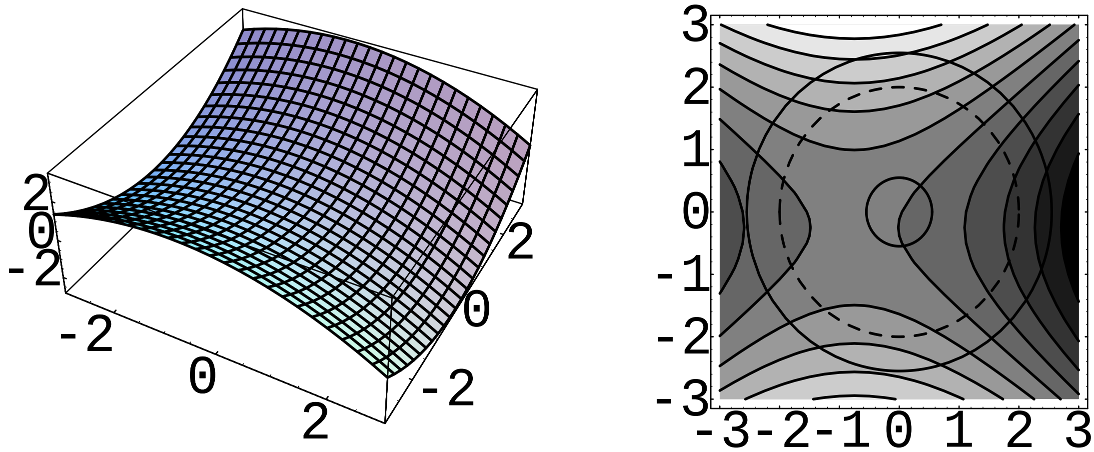
$$\bar{\mathbf{x}} = \begin{bmatrix} 1.9957445 \\ -0.1303985 \end{bmatrix},$$

which is located on the boundary $g_1(\bar{x}_1, \bar{x}_2) = 0$, and

$$P(\bar{\mathbf{x}}) = -1.4097812 = P^d(\bar{\sigma}_1, \bar{\sigma}_2, \bar{\varsigma}).$$

## 7.6 Conclusions

We have presented a detailed application of the canonical duality theory to the general differentiable nonconvex optimization problem. This problem arises in many real-world




(a) (b)

Figure 7.4: (a) Graph of $P(\mathbf{x})$; (b) Contours of $P(\mathbf{x})$, constraints $g_1(x_1, x_2) \leq d_1$ (disk with radius $R \leq 2$, dashed circle), and $g_2(x_1, x_2) \leq d_2$ (annulus with radius $0.55 \leq R \leq 2.55$) for Example 7.3.

applications. Using the canonical dual transformation, a unified canonical dual problem was formulated with zero duality gap, which can be solved by well-developed nonlinear optimization methods. The global optimizer can be identified by the triality theory. Insightful connections of this canonical duality with the classical Lagrangian duality have also been presented for two special cases.

Generally speaking, optimal solutions for constrained nonconvex minimization problems are usually KKT points located on the boundary of the feasible sets. Due to the lack of global optimality criteria, it is very difficult for direct methods and the classical Lagrangian relaxations to find global minimizers [82]. However, by the canonical duality theory, these KKT points can be determined by the critical points of the canonical dual problems. The triality theory can be used to develop effective algorithms for solving these problems.

# CHAPTER 8

# Nonlinear Systems of Equations

## 8.1 Problem Statement

We are interested in solving the following general nonlinear system of m quadratic equations:

$$(\mathcal{P}_0): \quad Q(\mathbf{x}) = \Lambda(\mathbf{x}) + B\mathbf{x} - \mathbf{c} = 0, \tag{8.1}$$

where $\mathbf{x} = \{x_i\} \in \mathbb{R}^n$ is an unknown vector, $\mathbf{c} = \{c^\alpha\} \in \mathbb{R}^m$ is a given data, $B = \{b_i^\alpha\} \in \mathbb{R}^{m\times n}$ is a matrix such that $B\mathbf{x} = \{\sum_i b_i^\alpha x_i\}$ is a vector in $\mathbb{R}^m$, and $\Lambda : \mathbb{R}^n \to \mathbb{R}^m$ is a quadratic operator defined by

$$\Lambda(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_i \sum_j x_i A_{ij}^\alpha x_j, \quad \alpha = 1, \dots, m, \tag{8.2}$$

in which, $\mathbf{A} = \{A_{ij}^\alpha\} \in \mathbb{R}^{n\times m\times n}$ is a given three-order tensor.

Problem $(\mathcal{P}_0)$ arises extensively in many complex systems of engineering science, data mining, chemistry, biomedicine, information theory, network communications, and ecology [32] [60] [61] [98]. The system is called *under determined* if $n > m$, and *over determined* if $n < m$. In either case, problem $(\mathcal{P}_0)$ possesses very high computation complexity, $O(n^m)$, due to several numerical issues [56] [23].

By using the least square method, Problem $(\mathcal{P}_0)$ can be relaxed as the following unconstrained optimization problem ($(\mathcal{P})$ for short):

$$(\mathcal{P}): \quad \min \left\{ P(\mathbf{x}) = \frac{1}{2}\|\Lambda(\mathbf{x}) + B\mathbf{x} - \mathbf{c}\|^2 : \ \mathbf{x} \in \mathbb{R}^n \right\}, \tag{8.3}$$

where $\|\mathbf{y}\|$ represents the Euclidian norm of $\mathbf{y}$.

**Lemma 8.1.** *If $\bar{\mathbf{x}}$ is a solution to $(\mathcal{P}_0)$, then $\bar{\mathbf{x}}$ must be a solution to $(\mathcal{P})$. On the other hand, if $(\mathcal{P}_0)$ has no solution, then Problem $(\mathcal{P})$ provides at least one optimal solution to Problem $(\mathcal{P}_0)$.*

*Proof.* The necessary condition $\nabla P(\bar{\mathbf{x}}) = 0$ for the unconstrained minimization problem

$(\mathcal{P})$ leads to the following equilibrium equation

$$(2\mathbf{A}\bar{\mathbf{x}} + \mathbf{B})^T(\Lambda(\bar{\mathbf{x}}) + \mathbf{B}\bar{\mathbf{x}} - \mathbf{c}) = 0. \tag{8.4}$$

Clearly, if $\bar{\mathbf{x}}$ is a solution to $(\mathcal{P}_0)$, i.e.,

$$\Lambda(\bar{\mathbf{x}}) + \mathbf{B}\bar{\mathbf{x}} - \mathbf{c} = 0, \tag{8.5}$$

then $\bar{\mathbf{x}}$ must be a critical point of $P(\mathbf{x})$. By the fact $P(\bar{\mathbf{x}}) = 0$, we know that $\bar{\mathbf{x}}$ is a (global) minimizer of $P(\mathbf{x})$. On the other hand, if the problem $(\mathcal{P}_0)$ has no solution, then $E(\mathbf{x}) = \Lambda(\mathbf{x}) + B\mathbf{x} - \mathbf{x} \neq 0$. Since the polynomial $P(\mathbf{x}) = \|E(\mathbf{x})\|^2 \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$ is bounded below and coercive, i.e., $\lim_{\mathbf{x}\to\infty} P(\mathbf{x}) = \infty$, the unconstrained minimization problem $(\mathcal{P})$ has at least one solution $\bar{\mathbf{x}}$. □

In the linear case that $\Lambda(\mathbf{x}) = 0$ and $n > m$, Problem $(\mathcal{P})$ is convex and the optimality condition leads to a linear equation

$$B^T B\mathbf{x} = B^T\mathbf{c}. \tag{8.6}$$

Clearly, this linear equation has at least one solution if $B^T\mathbf{c}$ is in the column space of $B^TB$. It has a unique solution if rank $B = m < n$.

Generally speaking, the target function $P: \mathbb{R}^n \to \mathbb{R}$ is a fourth-order polynomial:

$$P(\mathbf{x}) = W(\mathbf{x}) + \frac{1}{2}\mathbf{x}^T B^T B\mathbf{x} - \mathbf{c}^T\Lambda(\mathbf{x}) - \mathbf{c}^T(B\mathbf{x}) + d, \tag{8.7}$$

where $W(\mathbf{x}) = \frac{1}{2}\Lambda(\mathbf{x})^T\Lambda(\mathbf{x}) + (B\mathbf{x})^T\Lambda(\mathbf{x})$ and $d = \frac{1}{2}\mathbf{c}^T\mathbf{c}$, Problem $(\mathcal{P})$ may have multiple local extremal solutions. The standard techniques for solving nonconvex problems are mainly Newton type iteration methods [49] [50]. It was shown that the solutions to nonconvex minimization problems are difficult to be captured by Newton type direct approaches [42] [43]. From the criticality condition (8.4), we know that if $\bar{\mathbf{x}}$ solves the linear equation

$$2\mathbf{A}\bar{\mathbf{x}} + \mathbf{B} = 0, \tag{8.8}$$

it is also a critical point of $P(\mathbf{x})$. However, due to the fact that $P(\bar{\mathbf{x}}) = \frac{1}{2}\|\frac{1}{4}\mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} + \mathbf{c}\|^2 > 0$, this critical point is neither a solution to $(\mathcal{P})$, nor a solution to $(\mathcal{P}_0)$. Actually, due to the lack of global optimality conditions, many nonconvex minimization problems in global optimization are considered as NP-hard [31] [72] [92].

## 8.2 Canonical Dual Transformation

Following the standard procedure of the canonical dual transformation, we introduce a quadratic geometrical measure

$$\boldsymbol{\xi} = \Lambda(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} \in \mathbb{R}^m. \tag{8.9}$$

Thus, the nonconvex function $W(\mathbf{x}) = \frac{1}{2}\Lambda(\mathbf{x})^T\Lambda(\mathbf{x}) + (B\mathbf{x})^T\Lambda(\mathbf{x})$ can be written in the canonical form

$$W(\mathbf{x}) = V(\Lambda(\mathbf{x}), \mathbf{x}), \tag{8.10}$$

where $V(\boldsymbol{\xi}, \mathbf{x}) = \frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi} + (B\mathbf{x})^T\boldsymbol{\xi}$, i.e., the duality relation

$$\boldsymbol{\varsigma} = \frac{\partial V(\boldsymbol{\xi}, \mathbf{x})}{\partial \boldsymbol{\xi}} = \boldsymbol{\xi} + B\mathbf{x} \in \mathbb{R}^m \tag{8.11}$$

is invertible for any given $\boldsymbol{\xi} \in \mathbb{R}^m$. Thus, $(\boldsymbol{\xi}, \boldsymbol{\varsigma})$ forms a canonical duality pair on $\mathbb{R}^m \times \mathbb{R}^m$ (see [38]) and the Legendre conjugate $V^*$ can be uniquely defined by

$$V^*(\boldsymbol{\varsigma}, \mathbf{x}) = \mathrm{sta}\{\boldsymbol{\xi}^T\boldsymbol{\varsigma} - V(\boldsymbol{\xi}, \mathbf{x}) : \boldsymbol{\xi} \in \mathbb{R}^m\} = \frac{1}{2}(\boldsymbol{\varsigma} - B\mathbf{x})^2, \tag{8.12}$$

where sta{} denotes finding the stationary point of the statement in {}.

Replacing $W(\mathbf{x}) = V(\Lambda(\mathbf{x}), \mathbf{x})$ by $\Lambda(\mathbf{x})\boldsymbol{\varsigma} - V^*(\boldsymbol{\varsigma}, \mathbf{x})$, the *total complementary function* can be defined as:

$$\begin{aligned} \Xi(\mathbf{x}, \boldsymbol{\varsigma}) &= \Lambda(\mathbf{x})^T\boldsymbol{\varsigma} - V^*(\boldsymbol{\varsigma}, \mathbf{x}) + \frac{1}{2}\mathbf{x}^T B^T B\mathbf{x} - \mathbf{c}^T\Lambda(\mathbf{x}) - \mathbf{c}^T(B\mathbf{x}) + d \\ &= \frac{1}{2}\mathbf{x}^T G(\boldsymbol{\varsigma})\mathbf{x} - \mathbf{x}^T F(\boldsymbol{\varsigma}) - \frac{1}{2}\boldsymbol{\varsigma}^T\boldsymbol{\varsigma} + d, \end{aligned}$$

where

$$G(\boldsymbol{\varsigma}) = 2(\boldsymbol{\varsigma} - \mathbf{c})^T\mathbf{A} = \left\{ \sum_{\alpha=1}^{m} 2(\varsigma^\alpha - c^\alpha)\mathbf{A}_{ij}^\alpha \right\} \in \mathbb{R}^{n\times n}, \tag{8.13}$$

$$F(\boldsymbol{\varsigma}) = (\mathbf{c} - \boldsymbol{\varsigma})^T B = \left\{ \sum_{\alpha=1}^{m} (c^\alpha - \varsigma^\alpha)B_i^\alpha \right\} \in \mathbb{R}^n. \tag{8.14}$$

For a fixed $\boldsymbol{\varsigma}$, the criticality condition $\nabla_{\mathbf{x}}\Xi(\mathbf{x}, \boldsymbol{\varsigma}) = 0$ leads to the following canonical equilibrium equation:

$$G(\boldsymbol{\varsigma})\mathbf{x} = F(\boldsymbol{\varsigma}). \tag{8.15}$$

On the canonical dual feasible space $\mathcal{S}_a \subset \mathbb{R}^m$ defined by

$$\mathcal{S}_a = \{\boldsymbol{\varsigma} \in \mathbb{R}^m\}, \tag{8.16}$$

the solution of the canonical equilibrium equation can be uniquely determined as $\mathbf{x} = G^{+}(\boldsymbol{\varsigma})F(\boldsymbol{\varsigma})$. Substituting this result into the total complementary function $\Xi$, the canonical dual problem can be finally formulated as:

$$(\mathcal{P}^d): \quad \text{sta}\left\{ P^d(\boldsymbol{\varsigma}) = -\frac{1}{2}F(\boldsymbol{\varsigma})^T G^{+}(\boldsymbol{\varsigma})F(\boldsymbol{\varsigma}) - \frac{1}{2}\boldsymbol{\varsigma}^T\boldsymbol{\varsigma} + d: \ \boldsymbol{\varsigma} \in \mathcal{S}_a \right\}. \tag{8.17}$$

**Theorem 8.1** (Complementary-Dual Principle)**.** *If $\bar{\boldsymbol{\varsigma}}$ is a critical point of $(\mathcal{P}^d)$, then the vector*

$$\bar{\mathbf{x}} = G^{+}(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}}) \tag{8.18}$$

*is a critical point of $(\mathcal{P})$ and $P(\bar{\mathbf{x}}) = P^d(\bar{\boldsymbol{\varsigma}})$.*

*Proof.* Suppose that $\bar{\boldsymbol{\varsigma}}$ is a critical point of $(\mathcal{P}^d)$. Then,

$$\nabla P^d(\bar{\boldsymbol{\varsigma}}) = \bar{\mathbf{x}}(\bar{\boldsymbol{\varsigma}})^T \mathbf{A}\bar{\mathbf{x}}(\bar{\boldsymbol{\varsigma}}) + B\bar{\mathbf{x}}(\bar{\boldsymbol{\varsigma}}) - \bar{\boldsymbol{\varsigma}} = 0, \tag{8.19}$$

where $\bar{\mathbf{x}} = G^{+}(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}})$. Therefore, $\bar{\boldsymbol{\varsigma}} = \bar{\mathbf{x}}^T A\bar{\mathbf{x}} + B\bar{\mathbf{x}} = \Lambda(\bar{\mathbf{x}}) + B\bar{\mathbf{x}}$.

On the other hand,

$$\begin{aligned} P(\bar{\mathbf{x}}) &= \frac{1}{2}\Lambda(\bar{\mathbf{x}})^T\Lambda(\bar{\mathbf{x}}) + (B\bar{\mathbf{x}})^T\Lambda(\bar{\mathbf{x}}) + \frac{1}{2}\bar{\mathbf{x}}^T B^T B\bar{\mathbf{x}} - \mathbf{c}^T\Lambda(\bar{\mathbf{x}}) - \mathbf{c}^T(B\bar{\mathbf{x}}) + \frac{1}{2}\mathbf{c}^T\mathbf{c} \\ &= \frac{1}{2}(\Lambda(\bar{\mathbf{x}}) + B\bar{\mathbf{x}})^2 - \mathbf{c}^T\Lambda(\bar{\mathbf{x}}) - \mathbf{c}^T(B\bar{\mathbf{x}}) + \frac{1}{2}\mathbf{c}^T\mathbf{c}. \end{aligned}$$

Thus,

$$\bar{\mathbf{x}} \;=\; \frac{1}{2}((\bar{\boldsymbol{\varsigma}} - \mathbf{c})^T\mathbf{A})^{+}((\mathbf{c} - \bar{\boldsymbol{\varsigma}})^T B)$$

is a critical point of the primal problem $(\mathcal{P})$.

Moreover, in terms of $\bar{\mathbf{x}} = G^{+}(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}})$, we have

$$\begin{aligned} P^d(\bar{\boldsymbol{\varsigma}}) &= -\frac{1}{2}F(\bar{\boldsymbol{\varsigma}})^T G(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}}) - \frac{1}{2}\bar{\boldsymbol{\varsigma}}^T\bar{\boldsymbol{\varsigma}} + d \\ &= \frac{1}{2}\bar{\mathbf{x}}^T(2(\bar{\boldsymbol{\varsigma}} - \mathbf{c}))^T\mathbf{A}\bar{\mathbf{x}} - \bar{\mathbf{x}}^T((\mathbf{c} - \bar{\boldsymbol{\varsigma}})^T B) - \frac{1}{2}\bar{\boldsymbol{\varsigma}}^T\bar{\boldsymbol{\varsigma}} + \frac{1}{2}\mathbf{c}^T\mathbf{c} \\ &= \bar{\boldsymbol{\varsigma}}^T\Lambda(\bar{\mathbf{x}}) - \frac{1}{2}\bar{\boldsymbol{\varsigma}}^T\bar{\boldsymbol{\varsigma}} + (B\bar{\mathbf{x}})^T\bar{\boldsymbol{\varsigma}} - \mathbf{c}^T\Lambda(\bar{\mathbf{x}}) - \mathbf{c}^T(B\bar{\mathbf{x}}) + \frac{1}{2}\mathbf{c}^T\mathbf{c} \\ &= \frac{1}{2}\bar{\boldsymbol{\varsigma}}^T\bar{\boldsymbol{\varsigma}} - \mathbf{c}^T\Lambda(\bar{\mathbf{x}}) - \mathbf{c}^T(B\bar{\mathbf{x}}) + \frac{1}{2}\mathbf{c}^T\mathbf{c} \\ &= \frac{1}{2}\Lambda(\bar{\mathbf{x}})^T\Lambda(\bar{\mathbf{x}}) + (B\bar{\mathbf{x}})^T\Lambda(\bar{\mathbf{x}}) + \frac{1}{2}(B\bar{\mathbf{x}})^T(B\bar{\mathbf{x}}) - \mathbf{c}^T\Lambda(\bar{\mathbf{x}}) - \mathbf{c}^T(B\bar{\mathbf{x}}) + \frac{1}{2}\mathbf{c}^T\mathbf{c} \\ &= P(\bar{\mathbf{x}}) \end{aligned}$$

This proves the theorem. □

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

## 8.3 Global Optimality Criteria

It is known that the criticality condition is only necessary for local minimization of the nonconvex problem $(\mathcal{P})$. In order to identify global extrema among the critical points of Problem $(\mathcal{P})$, we need to introduce one useful feasible spaces

$$\mathcal{S}_a^+ = \{\boldsymbol{\varsigma} \in \mathcal{S}_a \mid \ G(\boldsymbol{\varsigma}) \ \succ 0\}. \tag{8.20}$$

Thus, we have the following result.

**Theorem 8.2.** *Suppose that the vector* $\bar{\boldsymbol{\varsigma}}$ *is a critical point of the canonical dual function* $P^d(\bar{\boldsymbol{\varsigma}})$. *Let* $\bar{\mathbf{x}} = G^+(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}})$. *If* $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a^+$, *then* $\bar{\boldsymbol{\varsigma}}$ *is a global maximizer of* $P^d(\boldsymbol{\varsigma})$ *on* $\mathcal{S}_a^+$. *The vector* $\bar{\mathbf{x}}$ *is a global minimizer of* $P(\mathbf{x})$ *on* $\mathbb{R}^n$, *and*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x}\in\mathbb{R}^n} P(\mathbf{x}) = \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}) = P^d(\bar{\boldsymbol{\varsigma}}). \tag{8.21}$$

*Proof.* By Theorem 8.1, we know that the vector $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a$ is a critical point of Problem $(\mathcal{P}^d)$ if and only if $\bar{\mathbf{x}} = G^+(\bar{\boldsymbol{\varsigma}})F(\bar{\boldsymbol{\varsigma}})$ is a critical point of Problem $(\mathcal{P})$, and

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

By the fact that the canonical dual function $P^d(\boldsymbol{\varsigma})$ is concave on $\mathcal{S}_a^+$, the critical point $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a^+$ is a global maximizer of $P^d(\boldsymbol{\varsigma})$ over $\mathcal{S}_a^+$, and $(\bar{\mathbf{x}}, \bar{\boldsymbol{\varsigma}})$ is a saddle point of the total complementary function $\Xi(\mathbf{x}, \boldsymbol{\varsigma})$ on $\mathbb{R}^n \times \mathcal{S}_a^+$, i.e., $\Xi$ is convex in $\mathbf{x} \in \mathbb{R}^n$ and concave in $\boldsymbol{\varsigma} \in \mathcal{S}_a^+$. Thus, by the saddle min-max duality theory (see [38]), we have

$$\begin{aligned}
P^d(\bar{\boldsymbol{\varsigma}}) &= \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}) = \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \min_{\mathbf{x}\in\mathbb{R}^n} \Xi(\mathbf{x}, \boldsymbol{\varsigma}) = \min_{\mathbf{x}\in\mathbb{R}^n} \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \Xi(\mathbf{x}, \boldsymbol{\varsigma}) \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} \{-\mathbf{c}^T\Lambda(\mathbf{x}) - \mathbf{c}^T(B\mathbf{x}) + \frac{1}{2}\mathbf{c}^T\mathbf{c} + \max_{\boldsymbol{\varsigma}\in\mathcal{S}_a^+} \{\Lambda(\mathbf{x})^T\boldsymbol{\varsigma} + (B\mathbf{x})^T\boldsymbol{\varsigma} - \frac{1}{2}\boldsymbol{\varsigma}^T\boldsymbol{\varsigma}\}\} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} \{-\mathbf{c}^T\Lambda(\mathbf{x}) - \mathbf{c}^T(B\mathbf{x}) + \frac{1}{2}\mathbf{c}^T\mathbf{c} + \frac{1}{2}((\Lambda(\bar{\mathbf{x}}) + B\bar{\mathbf{x}})^T(\Lambda(\bar{\mathbf{x}}) + B\bar{\mathbf{x}}))\} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} \{\frac{1}{2}\Lambda(\mathbf{x})^T\Lambda(\mathbf{x}) + (B\mathbf{x})^T\Lambda(\mathbf{x}) + \frac{1}{2}(B\mathbf{x})^T(B\mathbf{x}) - \mathbf{c}^T\Lambda(\mathbf{x}) - \mathbf{c}^T(B\mathbf{x}) + \frac{1}{2}\mathbf{c}^T\mathbf{c}\} \\
&= \min_{\mathbf{x}\in\mathbb{R}^n} P(\mathbf{x}) = P(\bar{\mathbf{x}}).
\end{aligned}$$

This proves the statement (8.21).

This theorem shows that the extremality criteria of the primal problem are controlled by the critical points of the canonical dual problem, i.e., if $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_a^+$, the vector $\bar{\mathbf{x}}(\bar{\boldsymbol{\varsigma}})$ is a global minimizer of $(\mathcal{P})$.

**Remark 8.1** (Perturbed Primal and Dual Problems)**.** Generally speaking, the solutions to the nonlinear problem $(\mathcal{P}_0)$ are not unique and the associated optimization problem $(\mathcal{P})$ may have multiple global minima. In order to solve Problem $(\mathcal{P})$ more efficiently by the canonical duality theory, we introduce the following perturbed problem

$$(\mathcal{P}_\epsilon): \quad \min\left\{P_\epsilon(\mathbf{x}) = P(\mathbf{x}) + \boldsymbol{\epsilon}^T\mathbf{x} \;:\; \mathbf{x} \in \mathbb{R}^n\right\}, \tag{8.22}$$

where $\boldsymbol{\epsilon} \geq 0 \in \mathbb{R}^n$ is a given perturbation vector. By the triality theory, the perturbed canonical dual problem is

$$(\mathcal{P}_\epsilon^d)_{\max}: \quad \max\left\{P_\epsilon^d(\boldsymbol{\varsigma}) = -\frac{1}{2}F_\epsilon(\boldsymbol{\varsigma})^T G^+(\boldsymbol{\varsigma}) F_\epsilon(\boldsymbol{\varsigma}) - \frac{1}{2}\boldsymbol{\varsigma}^T\boldsymbol{\varsigma} + d : \; \boldsymbol{\varsigma} \in \mathcal{S}_a^+\right\}, \tag{8.23}$$

where

$$F_\epsilon(\boldsymbol{\varsigma}) = (\mathbf{c} - \boldsymbol{\varsigma})^T B - \boldsymbol{\epsilon} = \left\{\sum_{\alpha=1}^{m}(c^\alpha - \varsigma^\alpha)B_i^\alpha - \epsilon_i\right\} \in \mathbb{R}^n. \tag{8.24}$$

It is easy to prove that the canonical dual function $P_\epsilon^d(\boldsymbol{\varsigma})$ is concave on $\mathcal{S}_a^+$. Therefore, for a given perturbation vector $\boldsymbol{\epsilon} \in \mathbb{R}^n$, if the canonical dual feasible space $\mathcal{S}_a^+$ is not empty, this perturbed canonical dual problem can be solved to yield a unique solution.

## 8.4 Numerical Examples

We now list a few examples to illustrate the applicability of the theory presented in this chapter. In order to find the global minimizer, we need to add a perturbation term.

**Example 8.1** We first consider a two dimension problem with only one equation $(m = 1)$:

$$Q(\mathbf{x}) = [x_1, x_2]^T \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [b_1, b_2]^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - c = 0.$$

Clearly, this equation has infinite number of solutions. The perturbed primal problem is a nonconvex minimization problem in $\mathbb{R}^2$:

$$\min\left\{P_\epsilon(\mathbf{x}) = \frac{1}{2}\left([x_1, x_2]^T \begin{bmatrix} A_{11} & A_{12} \\ A_{12} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [b_1, b_2]^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - c\right)^2 + \boldsymbol{\epsilon}^T\mathbf{x} : \; \mathbf{x} \in \mathbb{R}^2\right\}.$$

On the dual feasible space

$$\mathcal{S}_a = \{\varsigma \in \mathbb{R}\},$$

the canonical dual problem has the form of

$$P_\epsilon^d(\varsigma) = -\frac{1}{2}F_\epsilon(\varsigma)^T G^+(\varsigma) F_\epsilon(\varsigma) - \frac{1}{2}\varsigma^2 + \frac{1}{2}c^2,$$

where

$$G(\varsigma) = \begin{bmatrix} 2(\varsigma - c)A_{11} & 2(\varsigma - c)A_{12} \\ 2(\varsigma - c)A_{12} & 2(\varsigma - c)A_{22} \end{bmatrix}, \; F_\epsilon(\varsigma) = \begin{bmatrix} (c-\varsigma)b_1 \\ (c-\varsigma)b_2 \end{bmatrix} - \boldsymbol{\epsilon}.$$

If we choose $A_{11} = 0.5$, $A_{22} = 0.3$, $A_{12} = 0.2$, $b_1 = 2$, $b_2 = -1$, $c = 0.2$, $\boldsymbol{\epsilon} = [0.2, 0.2]^T$, the dual problem has three critical points (see Figure 8.2):

$$\bar{\varsigma}_3 = -5.68 < \bar{\varsigma}_2 = 0.12 < \bar{\varsigma}_1 = 0.28.$$

Since $\bar{\varsigma}_1 \in \mathcal{S}_a^+$, we know that $\bar{\mathbf{x}}_1 = [-4.80, 0.60]^T$ is a global minimizer; while $\bar{\mathbf{x}}_2 = [-2.49, 7.54]^T$ is a local minimizer, and $\bar{\mathbf{x}}_3 = [-3.62, 4.14]^T$ is a local maximizer (see Figure 8.1). We have

$$P_\epsilon(\bar{\mathbf{x}}_1) = -0.84 = P_\epsilon^d(\bar{\varsigma}_1) < P_\epsilon(\bar{\mathbf{x}}_2) = 1.01 = P_\epsilon^d(\bar{\varsigma}_2) < P_\epsilon(\bar{\mathbf{x}}_3) = 17.40 = P_\epsilon^d(\bar{\varsigma}_3).$$

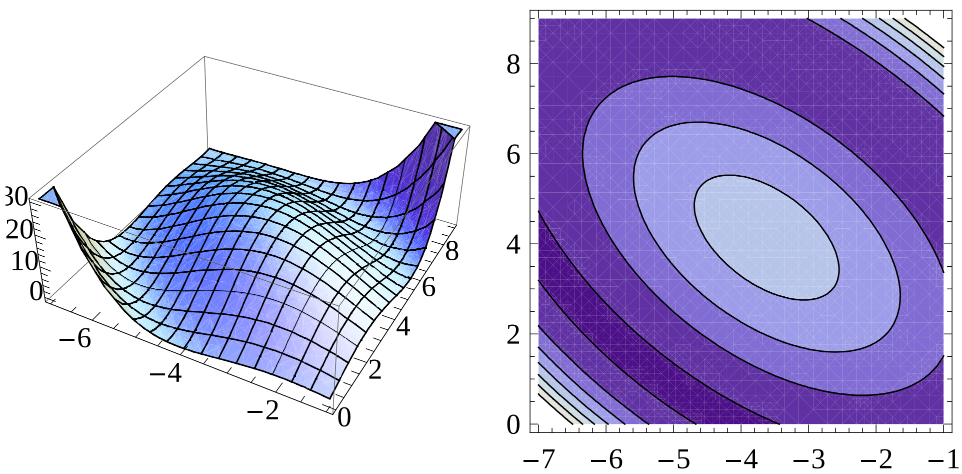



Figure 8.1: Graph of $P_\epsilon(\mathbf{x})$ (left); contours of $P_\epsilon(\mathbf{x})$ (right) for Example 8.1

By the fact that

$$P(\bar{\mathbf{x}}_1) = 0.00305261 < P(\bar{\mathbf{x}}_2) = 0.00313757,$$

both $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ are perturbed solutions to the original problem $(\mathcal{P}_0)$ and it is easy to

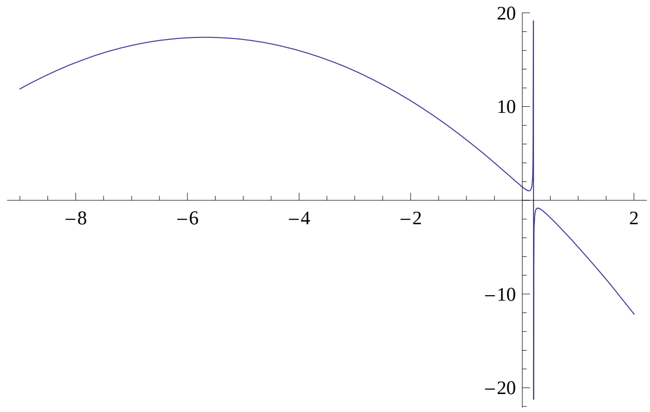



Figure 8.2: Graph of $P_\epsilon^d(\varsigma)$ for example 8.1.

verify that

$$Q(\bar{\mathbf{x}}_1) = 0.0781359, \quad Q(\bar{\mathbf{x}}_2) = -0.07921557.$$

Since $P(\bar{\mathbf{x}}_3) = 17.2917$, the local maximizer $\bar{\mathbf{x}}_3$ is a perturbed solution to the linear equation (8.8) and

$$2\mathbf{A}\bar{\mathbf{x}}_3 + B = \begin{bmatrix} 0.0340092 \\ 0.0340092 \end{bmatrix}.$$

**Example 8.2** Let $n = 3$, $m = 2$ such that the primal problem is

$$\min \left\{ P_\epsilon(\mathbf{x}) = \frac{1}{2}((\mathbf{x}^T A^1 \mathbf{x} + B^1 \mathbf{x} - c^1)^2 + (\mathbf{x}^T A^1 \mathbf{x} + B^1 \mathbf{x} - c^1)^2) + \boldsymbol{\epsilon}^T \mathbf{x} : \ \mathbf{x} \in \mathbb{R}^3 \right\},$$

where

$$A^1 = \begin{bmatrix} A_{11}^1 & A_{12}^1 & A_{13}^1 \\ A_{21}^1 & A_{22}^1 & A_{23}^1 \\ A_{31}^1 & A_{32}^1 & A_{33}^1 \end{bmatrix}, \ A^2 = \begin{bmatrix} A_{11}^2 & A_{12}^2 & A_{13}^2 \\ A_{12}^2 & A_{22}^2 & A_{23}^2 \\ A_{31}^2 & A_{32}^2 & A_{33}^2 \end{bmatrix},$$

$$B^1 = [b_1^1, \ b_2^1, \ b_3^1], \quad B^2 = [b_1^2, \ b_2^2, \ b_3^2].$$

Let

$$G(\varsigma_1, \varsigma_2) = 2((\varsigma_1 - c^1)A^1 + (\varsigma_2 - c^2)A^2) \in \mathbb{R}^{3\times 3},$$

$$F_\epsilon(\varsigma_1, \varsigma_2) = (c^1 - \varsigma_1)B^1 + (c^2 - \varsigma_2)B^2 - \boldsymbol{\epsilon} \in \mathbb{R}^3.$$

On the dual feasible space

$$\mathcal{S}_a = \{(\varsigma_1, \varsigma_2) \in \mathbb{R}^2 | \ \det G(\varsigma_1, \varsigma_2) \neq 0\},$$

the perturbed canonical dual function has the form of

$$P_\epsilon^d(\varsigma_1,\varsigma_2) = -\frac{1}{2}F_\epsilon(\varsigma_1,\varsigma_2)^T G^+(\varsigma_1,\varsigma_2)F_\epsilon(\varsigma_1,\varsigma_2) - \frac{1}{2}(\varsigma_1^2+\varsigma_2^2)+\frac{1}{2}(c_1^2+c_2^2).$$

Assume that

$$A^1 = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0 & 0.6 \end{bmatrix},\ A^2 = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.3 \end{bmatrix},$$

$$B^1 = [0.4, 0.9, 0.8]^T,\ \ B^2 = [0.9, 0.7, 0.6]^T,$$

$$c^1 = 0.8,\ c^2 = 0.6,\ \boldsymbol{\epsilon} = [0.05, 0.05, 0.05]^T.$$

The perturbed canonical dual problem has a unique solution $\bar{\boldsymbol{\varsigma}} = [0.8212, 0.6295]^T$ on $\mathcal{S}_a^+$, which leads to the global minimizer $\bar{\mathbf{x}} = [-1.65299, -1.68412, -1.962245]^T$. It is easy to check that $P_\epsilon(\bar{\mathbf{x}}) = -0.2642 = P_\epsilon^d(\bar{\boldsymbol{\varsigma}})$. The graph of $P_\epsilon^d(\mathbf{x})$ and its contours are shown in Figure 8.3.

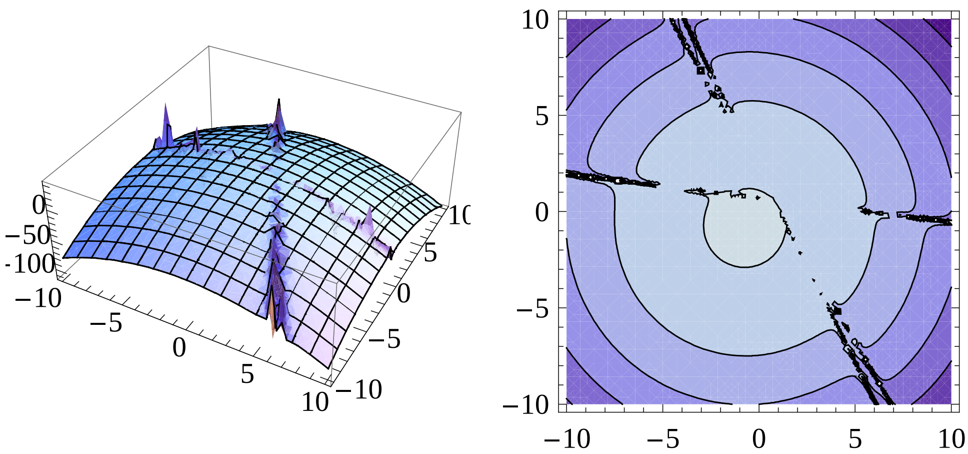



Figure 8.3: Graph of $P_\epsilon^d(\boldsymbol{\varsigma})$ (left); contours of $P^d(\boldsymbol{\varsigma})$ (right) for Example 8.2.

By the fact that

$$P(\bar{\mathbf{x}}) = 0.000756,$$

we know that $\bar{\mathbf{x}}$ is a perturbed solution to the original problem $(\mathcal{P}_0)$ and it is easy to verify that

$$Q(\bar{\mathbf{x}}) = \begin{bmatrix} 0.0224805 \\ 0.0317292 \end{bmatrix}.$$

## 8.5 Conclusions

We have presented a detailed application of the canonical duality theory to the general nonlinear systems of quadratic equations. Using the canonical dual transformation, the canonical dual problem was formulated with zero duality gap. Furthermore, the $n$-dimensional nonconvex problem $(\mathcal{P})$ can be reformulated as an $m(m < n)$-dimensional concave maximization dual problem $(\mathcal{P}^d)$ on $\mathcal{S}_a^+$, which can be solved by well-developed optimization techniques. Generally speaking, for given data and the perturbation vector $\boldsymbol{\epsilon} \in \mathbb{R}^n$, the perturbed canonical dual problem $(\mathcal{P}_\epsilon^d)$ has at most one solution in $\mathcal{S}_a^+$. Detailed study on the existence and uniqueness of the canonical dual solutions is an interesting future research topic.

For general higher order nonlinear systems, as long as the geometrical operator $\Lambda$ is chosen properly, the canonical dual transformation method can be used to establish useful theoretical results.

# CHAPTER 9

# Sensor Network Localization

## 9.1 Introduction

Sensor network localization [16], [24], [64], [79], [103] is an important problem in communication and information theory, and hence it has attracted an increasing attention. The information collected through a sensor network can be interpreted and relayed far more effectively if it is known where the information is coming from and where it needs to be sent. Therefore, it is often very useful to know the positions of the sensor nodes in a network. Wireless sensor network consists of a large number of wireless sensors located in a geographical area with the ability to communicate with their neighbors within a limited radio range. Sensors collect the local environmental information, such as temperature or humidity, and can communicate with each other. Wireless sensor network is applicable to a range of monitoring applications in civil and military scenarios, such as geographical monitoring, smart homes, industrial control and traffic monitoring. There is an urgent need to develop robust and efficient algorithms that can identify sensor positions in a network by using only the measurements of the mutual distances of the wireless sensors from their neighbors, which is called neighboring distance measurements. The advance of wireless communication technology has made the sensor network a low-cost and highly efficient method for environmental observations.

Sensor network localization can also be formulated as an optimization problem by least square method. However, this optimization is nonconvex, and hence its global optimal solutions are difficult to find. Several approximation methods have been developed for solving this difficult optimization problem [74] [78] [97], [105]. The semi-definite programming (SDP) and second-order cone programming (SOCP) relaxation are two of the most polular methods studied recently. The basic idea of SDP relaxation is to think of the quadratic terms as new variables subject to linear matrix inequality. The SOCP relaxation is developed in a similar way.

## 9.2 Problem Statement

We use the least square method for the formulation of a new optimization problem. Consider a general sensor network localization problem, where the sensor locations are to be determined by solving the system of polynomial equations [6] [22]:

$$(\mathcal{P}_0): \qquad \|\mathbf{x}_i - \mathbf{x}_j\| = d_{ij}, \quad (i,j) \in \mathcal{A}_d, \tag{9.1}$$

$$\|\mathbf{x}_i - \mathbf{a}_k\| = e_{ik}, \quad (i,k) \in \mathcal{A}_e. \tag{9.2}$$

Here, the vectors $\mathbf{a}_k,\ k = 1, \cdots, m$, are specified anchors, where

$$\|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{\sum_{\alpha=1}^{d} (x_i^\alpha - x_j^\alpha)^2}$$

denotes the Euclidian distance between locations $\mathbf{x}_i$ and $\mathbf{x}_j \in \mathbb{R}^d,\ i = 1, \cdots, n; j = 1, \cdots, n$, and

$$\begin{aligned} \mathcal{A}_d &= \{(i,j) \in [n] \times [n] \mid \ \|\mathbf{x}_i - \mathbf{x}_j\| = d_{ij}, \quad i < j, \ \ d_{ij} \text{ are given distances}\}, \\ \mathcal{A}_e &= \{(i,k) \in [n] \times [m] \mid \ \|\mathbf{x}_i - \mathbf{a}_k\| = e_{ik}, \quad e_{ik} \text{ are given distances}\}, \end{aligned}$$

where $[N] = \{1, \cdots, N\}$ for any integer $N$.

For a small number of sensors, it might be possible to compute sensor locations by solving equations (9.1)-(9.2). However, solving this algebraic system can be very expensive computationally when the number of sensors is large.

By the least squares method, the sensor network localization problem $(\mathcal{P}_0)$ can be reformulated as a fourth-order polynomial optimization problem stated below:

$$(\mathcal{P}_1): \quad \min \quad \left\{ P(\mathbf{X}) = \sum_{(i,j)\in\mathcal{A}_d} \frac{1}{2} w_{ij} (\|\mathbf{x}_i - \mathbf{x}_j\|^2 - d_{ij}^2)^2 + \sum_{(i,k)\in\mathcal{A}_e} \frac{1}{2} q_{ik} (\|\mathbf{x}_i - \mathbf{a}_k\|^2 - e_{ik}^2)^2 \right\},$$

where $\mathbf{X} = [x_1, x_2, \cdots, x_n] = \{x_i^\alpha\} \in \mathbb{R}^{d\times n}$ is a matrix with each column $x_i$ being a position in $\mathbb{R}^d$, $w_{ij}, q_{ik} > 0$ are given weights. Obviously, $\mathbf{X}$ are true sensor locations if and only if the optimal value is zero. This nonconvex optimization problem appears extensively in mathematical physics [45], computational biology [102], numerical algebra [81] as well as finite element analysis of structural mechanics [84].

The sensor network localization problem can also be viewed as a variant of Graph Realization problem, or a distance geometry problem [6], which has been studied extensively in computational biology, Euclidean ball packing, molecular confirmation and wireless network communication. In general, the sensor network localization problem is considered to be NP-hard even for the simplest case $d = 1$ [70] [85]. Therefore, many approximation method have been proposed for solving this nonconvex global optimization problem approximately. The semi-definite programming (SD)) and second order cone programming (SOCP) relaxation are two of the popular methods studied recently [5] [6] [13] [97] [98].

In the following, we shall see that by using the canonical duality theory, this nonconvex minimization problem is equivalent to a concave maximization dual problem over a convex set, which can be solved by well-developed optimization techniques.

## 9.3 Canonical Dual Transformation

In order to use the canonical dual transformation, we transfer variables from matrix to vector, and let

$$\begin{aligned}
\mathbf{y} &= [x_1^1 \cdots x_1^d \cdots x_n^1 \cdots x_n^d]^T \in \mathbb{R}^{nd} : \text{Locations of sensors (variables)},\\
\mathbf{W} &= [w_{11} \cdots w_{1n} \cdots w_{n1} \cdots w_{nn}]^T \in \mathbb{R}^{nn} : \text{Weights for the optimization problem } (\mathcal{P}_1),\\
\mathbf{Q} &= [q_{11} \cdots q_{1m} \cdots q_{n1} \cdots q_{nm}]^T \in \mathbb{R}^{nm} : \text{Weights for the optimization problem } (\mathcal{P}_1),\\
\mathbf{a} &= [\sum_{\alpha=1}^{d} (a_1^\alpha)^2, \cdots, \sum_{\alpha=1}^{d} (a_m^\alpha)^2]^T : \text{Sums of squares of anchors},\\
\mathbf{d} &= [d_{11}^2 \cdots d_{1n}^2 \cdots d_{n1}^2 \cdots d_{nn}^2]^T \in \mathbb{R}^{nn} : \text{Squares of distances between sensors},\\
\mathbf{e} &= [e_{11}^2 \cdots e_{1m}^2 \cdots e_{n1}^2 \cdots e_{nm}^2]^T \in \mathbb{R}^{nm} : \text{Squares of distances between sensors and anchors.}
\end{aligned}$$

Then, Problem $(\mathcal{P}_1)$ can be written in a vector form given below.

$$(\mathcal{P}): \quad \min \quad \Bigg\{ P(\mathbf{y}) = \sum_{(i,j)\in A_d} \frac{1}{2} w_{ij} \left(\mathbf{y}^T D_{ij} \mathbf{y} - d_{ij}^2\right)^2 + \sum_{(i,k)\in A_e} \frac{1}{2} q_{ik} \left(\mathbf{y}^T E_{ik} \mathbf{y} - 2A_{ik}^T \mathbf{y} + \sum_{\alpha=1}^{d} (a_{ik}^\alpha)^2 - e_{ik}^2\right)^2 \Bigg\},$$

where $\mathbf{y} \in \mathbb{R}^{nd}$, $E_{ik} \in \mathbb{R}^{nd\times nd}$ is a diagonal matrix defined by

$$E_{ik} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_{ik} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

with $I_{ik} \in \mathbb{R}^{d\times d}$ being the identity matrix corresponding to sensor $i$ and anchor $k$, so that the (1,1) entry of $I_{ik}$ coincides with the $(i, k)$ entry of $E_{ik}$. Similarly, $D_{ij}$ is an $nd \times nd$ matrix defined by

$$D_{ij} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & I_{ii} & 0 & -I_{ij} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -I_{ji} & 0 & I_{jj} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

with $I_{ii}$,$I_{jj}$, $I_{ij}$,$I_{ji} \in \mathbb{R}^{d\times d}$ being the identity matrices, so that the (1,1) entry of $I_{ii}$ coincides with the $(i,i)$ entry of the matrix $D_{ij}$. For $I_{jj}$, $I_{ij}$, $I_{ji}$, they are defined similarly. Let

$$\xi_{ij} = \Lambda_{ij}(\mathbf{y}) \;=\; \mathbf{y}^T D_{ij}\mathbf{y}, \tag{9.3}$$

$$\epsilon_{ik} = \Lambda_{ik}(\mathbf{y}) \;=\; \mathbf{y}^T E_{ik}\mathbf{y} - 2A_{ik}^T\mathbf{y}, \tag{9.4}$$

where $\Lambda_{ij}$ and $\Lambda_{ik}$ are, respectively, geometrical operators from $\mathbb{R}^{nd}$ into

$$\mathcal{E}_d \;=\; \{\boldsymbol{\xi} \in \mathbb{R}^{nn} |\ \xi_{ij} \geq 0,\ \xi_{ij} = 0 \text{ if } i = j\}$$

and

$$\mathcal{E}_e \;=\; \{\boldsymbol{\epsilon} \in \mathbb{R}^{mn} |\ \epsilon_{ik} \geq 0\}.$$

By introducing quadratic functions $V_{ij} : \mathcal{E}_d \to \mathbb{R}$ and $V_{ik} : \mathcal{E}_e \to \mathbb{R}$ such that

$$V_{ij}(\xi_{ij}) \;=\; \frac{1}{2}w_{ij}(\xi_{ij} - d_{ij}^2)^2 \tag{9.5}$$

and

$$V_{ik}(\epsilon_{ik}) \;=\; \frac{1}{2}q_{ik}(\epsilon_{ik} + \sum_{\alpha=1}^{d}(a_{ik}^{\alpha})^2 - e_{ik}^2)^2. \tag{9.6}$$

Problem $(\mathcal{P})$ can then be reformulated in the canonical form given below:

$$(\mathcal{P}):\ \ \min\left\{\Pi(\mathbf{y}) = \sum_{(i,j)\in\mathcal{A}_d} V_{ij}(\Lambda_{ij}(\mathbf{y})) + \sum_{(i,k)\in\mathcal{A}_e} V_{ik}(\Lambda_{ik}(\mathbf{y}))|\ \ \mathbf{y} \in \mathbb{R}^{nd}\right\}.$$

Note that the function $V_{ij}(\xi_{ij})$ and $V_{ik}(\epsilon_{ik})$ are both convex. Their duality relations are given, respectively, by

$$\varsigma_{ij} = \nabla V_{ij}(\xi_{ij}) = w_{ij}(\xi_{ij} - d_{ij}^2),\ \ (i,j) \in \mathcal{A}_d, \tag{9.7}$$

and

$$\sigma_{ik} = \nabla V_{ik}(\epsilon_{ik}) = q_{ik}(\epsilon_{ik} + \sum_{\alpha=1}^{d}(a_{ik}^{\alpha})^2 - e_{ik}^2), \quad (i,k) \in \mathcal{A}_e, \tag{9.8}$$

where $\varsigma_{ij}$ and $\sigma_{ik}$ are dual variables. Let $\mathcal{S}_d$ be the range of the duality mapping $\boldsymbol{\varsigma}_{ij} = \nabla V_{ij}(\xi_{ij})$, and let $\mathcal{S}_e$ be the range of the duality mapping $\boldsymbol{\sigma}_{ik} = \nabla V_{ik}(\delta_{ik})$. Then, for any given $\boldsymbol{\varsigma} \in \mathcal{S}_d$ and $\boldsymbol{\sigma} \in \mathcal{S}_e$, the Legendre conjugate $V_{ij}^*$ and $V_{ik}^*$ can be uniquely defined by

$$V_{ij}^*(\varsigma_{ij}) \;=\; \text{sta}\{\xi_{ij}^T\varsigma_{ij} - V_{ij}(\xi_{ij}) \mid \xi_{ij} \in \mathcal{V}_d\} = \frac{1}{2}w_{ij}^{-1}\varsigma_{ij}^2 + d_{ij}^2\varsigma_{ij}, (ij) \in \mathcal{A}_d$$

and

$$V_{ik}^*(\sigma_{ik}) \;=\; \text{sta}\{\delta_{ik}^T\sigma_{ik} - V_{ik}(\delta_{ik}) \mid \xi_{ik} \in \mathcal{V}_e\} = \frac{1}{2}q_{ik}^{-1}\sigma_{ik}^2 + (e_{ik}^2 - \sum_{\alpha=1}^{d}(a_{ik}^{\alpha})^2)\sigma_{ik}, (i,k) \in \mathcal{A}_e,$$

where sta{} denotes finding the stationary point of the statement within {}. Clearly, $(\boldsymbol{\xi}, \boldsymbol{\varsigma})$ and $(\boldsymbol{\epsilon}, \boldsymbol{\sigma})$ form a *canonical duality pair* (see [38]). The following canonical duality relations hold on both $\mathcal{E}_d \times \mathcal{S}_d$ and $\mathcal{E}_e \times \mathcal{S}_e$

$$\varsigma_{ij} \;=\; \nabla V_{ij}(\xi_{ij}) \;\Leftrightarrow\; \xi_{ij} = \nabla V_{ij}^*(\varsigma_{ij}) \Leftrightarrow \xi_{ij}^T\varsigma_{ij} = V_{ij}(\xi_{ij}) + V_{ij}^*(\varsigma_{ij}), \tag{9.9}$$

$$\sigma_{ik} \;=\; \nabla V_{ik}(\epsilon_{ik}) \;\Leftrightarrow\; \epsilon_{ik} = \nabla V_{ik}^*(\sigma_{ik}) \Leftrightarrow \epsilon_{ik}^T\sigma_{ik} = V_{ik}(\epsilon_{ik}) + V_{ik}^*(\sigma_{ik}), \tag{9.10}$$

respectively.

Replacing $V_{ij}(\Lambda_{ij}(\mathbf{y}))$ by $\Lambda_{ij}(\mathbf{y})^T\varsigma_{ij} - V_{ij}^*(\varsigma_{ij})$ and $V_{ik}(\Lambda_{ik}(\mathbf{y}))$ by $\Lambda_{ik}(\mathbf{y})^T\sigma_{ik} - V_{ik}^*(\sigma_{ik})$, the generalized complementary function is given by

$$\begin{aligned}
\Xi(\mathbf{y}, \boldsymbol{\varsigma}, \boldsymbol{\sigma}) &= \sum_{(i,j)\in A_d}(\Lambda_{ij}(\mathbf{y})\varsigma_{ij} - V_{ij}^*(\varsigma_{ij})) + \sum_{(i,k)\in A_e}(\Lambda_{ik}(\mathbf{y})\varsigma_{ik} - V_{ik}^*(\varsigma_{ik})) \\
&= \sum_{(i,j)\in A_d}((\mathbf{y}^T D_{ij}\mathbf{y})\varsigma_{ij} - (\frac{1}{2}w_{ij}^{-1}\varsigma_{ij}^2 + d_{ij}^2\varsigma_{ij})) \\
&\quad + \sum_{(i,k)\in A_e}((\mathbf{y}^T E_{ik}\mathbf{y} - 2A_{ik}^T\mathbf{y})\sigma_{ik} - (\frac{1}{2}q_{ik}^{-1}\sigma_{ik}^2 + (e_{ik}^2 - \sum_{\alpha=1}^{d}(a_{ik}^{\alpha})^2)\sigma_{ik})) \\
&= \frac{1}{2}\mathbf{y}^T\mathbf{G}(\boldsymbol{\varsigma}, \boldsymbol{\sigma})\mathbf{y} - \mathbf{F}^T(\boldsymbol{\sigma})\mathbf{y} - \frac{1}{2}(\mathbf{W}^{-1})^T(\boldsymbol{\varsigma} \circ \boldsymbol{\varsigma}) - \frac{1}{2}(\mathbf{Q}^{-1})^T(\boldsymbol{\sigma} \circ \boldsymbol{\sigma}) \\
&\quad - \mathbf{d}^T\boldsymbol{\varsigma} + \mathbf{a}^T\boldsymbol{\sigma} - \mathbf{e}^T\boldsymbol{\sigma},
\end{aligned}$$

where $\mathbf{s} \circ \mathbf{t} := [s_1t_1, s_2t_2, \cdots, s_nt_n]^T$ denotes the Hadamard product of any two vectors $\mathbf{s}$,

$\mathbf{t} \in \mathbb{R}^n$,

$$\mathbf{F}(\boldsymbol{\sigma}) = \left[ \sum_{k=1}^{m} 2a_k^1 \sigma_{1k} \cdots \sum_{k=1}^{m} 2a_k^d \sigma_{1k} \cdots \sum_{k=1}^{m} 2a_k^1 \sigma_{nk} \cdots \sum_{k=1}^{m} 2a_k^d \sigma_{nk} \right]^T, \tag{9.11}$$

$$\mathbf{G}(\boldsymbol{\varsigma}, \boldsymbol{\sigma}) = 2(\text{Diag } (F_1(\boldsymbol{\varsigma})) + \text{Diag } (F_2(\boldsymbol{\sigma})) + G_3(\boldsymbol{\varsigma})), \tag{9.12}$$

with

$$F_1(\boldsymbol{\varsigma}) = \begin{bmatrix} \sum_{i=1}^{n} \varsigma_{1i} + \sum_{i=n}^{1} \varsigma_{i1} \\ \vdots \\ \sum_{i=1}^{n} \varsigma_{1i} + \sum_{i=n}^{1} \varsigma_{i1} \\ \vdots \\ \sum_{i=1}^{n} \varsigma_{ni} + \sum_{i=n}^{1} \varsigma_{in} \\ \vdots \\ \sum_{i=1}^{n} \varsigma_{ni} + \sum_{i=n}^{1} \varsigma_{in} \end{bmatrix},$$

$$F_2(\boldsymbol{\sigma}) = \begin{bmatrix} \sum_{k=1}^{m} \sigma_{1k} \\ \vdots \\ \sum_{k=1}^{m} \sigma_{1k} \\ \vdots \\ \sum_{k=1}^{m} \sigma_{nk} \\ \vdots \\ \sum_{k=1}^{m} \sigma_{nk} \end{bmatrix},$$

$$G_3(\boldsymbol{\varsigma}) = \begin{bmatrix} -\varsigma_{11} I_{11} & \cdots & -\varsigma_{1n} I_{1n} \\ \vdots & \vdots & \vdots \\ -\varsigma_{n1} I_{n1} & \cdots & -\varsigma_{nn} I_{nn} \end{bmatrix}.$$

For a fixed $\boldsymbol{\varsigma} \in \mathcal{S}_d$ and $\boldsymbol{\sigma} \in \mathcal{S}_e$, the criticality condition $\nabla_{\mathbf{y}} \Xi(\mathbf{y}, \boldsymbol{\varsigma}, \boldsymbol{\sigma}) = 0$ leads to the following *canonical equilibrium equation*:

$$\mathbf{G}(\boldsymbol{\varsigma}, \boldsymbol{\sigma})\mathbf{y} - \mathbf{F}(\boldsymbol{\sigma}) = 0. \tag{9.13}$$

Thus, on the dual feasible space defined by $\mathcal{S}_d \times \mathcal{S}_e$, the canonical dual function can be formulated as:

$$\begin{aligned} P^d(\boldsymbol{\varsigma},\boldsymbol{\sigma}) &= \text{sta}\{\Xi(\mathbf{y},\boldsymbol{\varsigma},\boldsymbol{\sigma})|\ \mathbf{y}\in\mathcal{Y}_a\} \\ &= -\frac{1}{2}\mathbf{F}(\boldsymbol{\sigma})^T\mathbf{G}^+(\boldsymbol{\varsigma},\boldsymbol{\sigma})\mathbf{F}(\boldsymbol{\sigma}) - \frac{1}{2}(\mathbf{W}^{-1})^T(\boldsymbol{\varsigma}\circ\boldsymbol{\varsigma}) \\ & \quad -\frac{1}{2}(\mathbf{Q}^{-1})^T(\boldsymbol{\sigma}\circ\boldsymbol{\sigma}) - \mathbf{d}^T\boldsymbol{\varsigma} + \mathbf{a}^T\boldsymbol{\sigma} - \mathbf{e}^T\boldsymbol{\sigma}, \end{aligned}$$

where

$$\mathbf{F}(\boldsymbol{\sigma}) = \left[\sum_{k=1}^{m} 2a_k^1\sigma_{1k}\cdots\sum_{k=1}^{m} 2a_k^d\sigma_{1k}\cdots\sum_{k=1}^{m} 2a_k^1\sigma_{nk}\cdots\sum_{k=1}^{m} 2a_k^d\sigma_{nk}\right]^T, \tag{9.14}$$

$\mathbf{G}^+(\boldsymbol{\varsigma},\boldsymbol{\sigma})$ denotes the generalized inverse of $\mathbf{G}(\boldsymbol{\varsigma},\boldsymbol{\sigma})$, and

$$\begin{aligned} \mathbf{W}^{-1} &= \left[\frac{1}{w_{11}}\cdots\frac{1}{w_{1n}}\cdots\frac{1}{w_{n1}}\cdots\frac{1}{w_{nn}}\right]^T, \\ \mathbf{Q}^{-1} &= \left[\frac{1}{q_{11}}\cdots\frac{1}{q_{1m}}\cdots\frac{1}{q_{n1}}\cdots\frac{1}{q_{nm}}\right]^T. \end{aligned}$$

Therefore, the canonical dual problem can be written in the form given below:

$$(\mathcal{P}^d):\ \ \text{sta}\left\{P^d(\boldsymbol{\varsigma},\boldsymbol{\sigma})|\ \ \boldsymbol{\varsigma}\in\mathcal{S}_d, \boldsymbol{\sigma}\in\mathcal{S}_e\right\}.$$

We have following theorems:

**Theorem 9.1.** *Problem* $(\mathcal{P}^d)$ *is a canonical dual of the primal problem* $(\mathcal{P})$ *in the sense that if* $(\bar{\boldsymbol{\varsigma}},\bar{\boldsymbol{\sigma}})$ *is a critical point of* $(\mathcal{P}^d)$*, then*

$$\bar{\mathbf{y}} = \mathbf{G}^+(\bar{\boldsymbol{\varsigma}},\bar{\boldsymbol{\sigma}})\mathbf{F}(\bar{\boldsymbol{\sigma}}) \tag{9.15}$$

*is a critical point of* $(\mathcal{P})$ *and*

$$P(\bar{\mathbf{y}}) = P^d(\bar{\boldsymbol{\varsigma}},\bar{\boldsymbol{\sigma}}). \tag{9.16}$$

Theorem 9.1 shows that the nonconvex primal problem $(\mathcal{P})$ is equivalent to its canonical dual problem $(\mathcal{P}^d)$ with zero duality gap, and the solution of $(\mathcal{P})$ can be analytically expressed in the form of (9.15). The extremality of this analytical solution will be discussed in the next section.

For further discussion on extremality properties of the analytical solution (9.15), we introduce the following feasible space

$$\mathcal{S}_a^+ = \{(\boldsymbol{\varsigma},\boldsymbol{\sigma})\in\mathcal{S}_d\times\mathcal{S}_e\ |\ \ \mathbf{G}(\boldsymbol{\varsigma},\boldsymbol{\sigma})\succ 0\}. \tag{9.17}$$

We have the following results.

**Theorem 9.2.** *Suppose that* $(\bar{\varsigma}, \bar{\boldsymbol{\sigma}})$ *is a critical point of the canonical dual function* $P^d(\bar{\varsigma}, \bar{\boldsymbol{\sigma}})$ *and* $\bar{\mathbf{y}} = \mathbf{G}^+(\bar{\varsigma}, \bar{\boldsymbol{\sigma}})F(\bar{\boldsymbol{\sigma}})$*. Let* $(\bar{\varsigma}, \bar{\boldsymbol{\sigma}}) \in \mathcal{S}_a^+$*. Then,* $\bar{\mathbf{y}}$ *is a global minimizer of* $P(\mathbf{y})$ *on* $\mathbb{R}^{nd}$ *if and only if* $(\bar{\varsigma}, \bar{\boldsymbol{\sigma}})$ *is a global maximizer of* $P^d(\bar{\varsigma}, \bar{\boldsymbol{\sigma}})$ *on* $\mathcal{S}_a^+$*, i.e.,*

$$P(\bar{\mathbf{y}}) = \min_{\mathbf{y}\in\mathbb{R}^{nd}} P(\mathbf{y}) \Leftrightarrow \max_{(\boldsymbol{\varsigma},\boldsymbol{\sigma})\in\mathcal{S}_a^+} P^d(\boldsymbol{\varsigma}, \boldsymbol{\sigma}) = P^d(\bar{\varsigma}, \bar{\boldsymbol{\sigma}}). \tag{9.18}$$

Theorem 9.2 shows that the extremality condition of the analytical solution (9.15) is controlled by the critical point of the canonical dual function. If the primal problem has a global minimal solution, then $\mathcal{S}_a^+ \neq \emptyset$ and the primal problem is equivalent to the canonical dual problem

$$(\mathcal{P}^d_{\max}) : \quad \max\{P^d(\boldsymbol{\varsigma}, \boldsymbol{\sigma})| \ (\boldsymbol{\varsigma}, \boldsymbol{\sigma}) \in \mathcal{S}_a^+\}. \tag{9.19}$$

## 9.4 Numerical simulations

### 9.4.1 18 sensors network localization problem

We now consider sensor network localization problem with 18 sensors. In this case, we have Problem $(\mathcal{P}_1)$ with $d = 2$. Define $\mathbf{y} = [x_1^1, x_1^2, \cdots, x_n^1, x_n^2]^T \in \mathbb{R}^{2n}$, and let $w_{ij} = q_{ik} = 1$ in Problem $(\mathcal{P}_1)$. Here, we do not consider noise.

The 18 sensors $\{\hat{\mathbf{x}}_i = [\hat{x}_i^1, \hat{x}_i^2] : i = 1, \cdots, 18\}$ are randomly generated in the unit square [-0.5, 0.5] $\times$ [-0.5, 0.5]. The four anchors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$ are placed at the positions $(\pm 0.45, \pm 0.45)$. The distances $\mathbf{d} = \{d_{ij}\}$, $i = 1, \cdots, 18$; $j = 1, \cdots, 18$, and $\mathbf{e} = \{e_{ik}\}$, $i = 1, \cdots, 18$; $k = 1, \cdots, 4$, are computed as follows:

$$d_{ij} = \|x_i^* - x_j^*\|, \ e_{ik} = \|x_i^* - a_k\|$$

We now assume that the locations of the 18 sensors are unknown. They are to be determined by the approach proposed in the chapter. The sequential quadratic programming approximation with active set strategy in the optimization toolbox within the Matlab environment is used to solve the canonical dual problem.

By Theorem 9.2, we obtain $\bar{\mathbf{y}} = [\bar{\mathbf{x}}_1, \cdots, \bar{\mathbf{x}}_{18}]^T$ with $\bar{\mathbf{x}}_i = [\bar{\mathbf{x}}_i^1, \bar{\mathbf{x}}_i^2]^T, i = 1, \cdots, 18$, which is a global minimizer of $P(\mathbf{y})$,

Furthermore, we have

$$\Pi(\bar{\mathbf{y}}) = 1.30 \times 10^{-8} \simeq 3.03 \times 10^{-8} = \Pi^d(\bar{\varsigma}, \bar{\boldsymbol{\sigma}}).$$

This problem is also solved by the standard semi-definite programming (SDP) method.

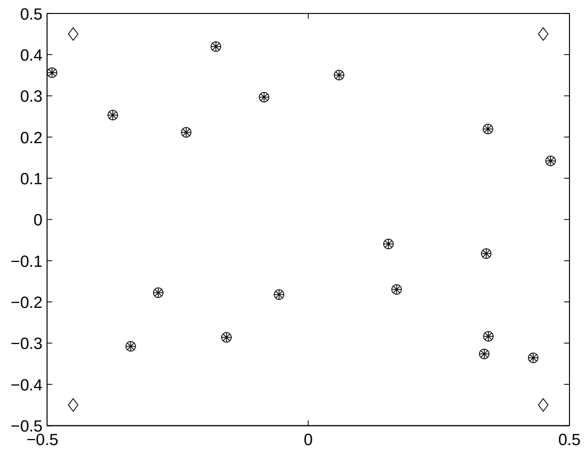



Figure 9.1: Sensor network with 18 sensors by the canonical dual method.

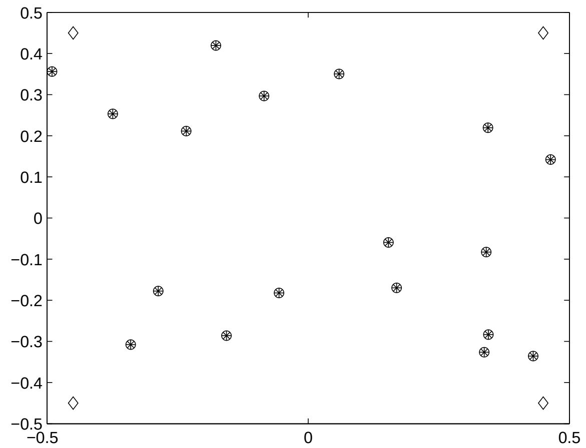



Figure 9.2: Sensor network with 18 sensors by the standard SDP method.

The RMSD obtained using the canonical dual method is $4.61 \times 10^{-7}$, while the RMSD obtained using the standard SDP method is $4.45 \times 10^{-5}$, where RMSD is the Root Mean Square Distance defined by

$$\text{RMSD} = \Big(\frac{1}{n}\sum_{i=1}^{n} \|\hat{\mathbf{x}}_i - \bar{\mathbf{x}}_i\|^2\Big)^{\frac{1}{2}},$$

which is to measure the accuracy of the computed locations.

The computed results by the canonical dual method and the standard SDP method are plotted in Fig. 9.1 and Fig. 9.2, respectively. The true sensor locations (denoted by circles) and the computed locations (denoted by stars) are connected by solid lines. Our program is implemented in the MATLAB environment, where SEDUMI [67] is used as the SDP solver.

From the results obtained, we see that, when there is no noise and the sensor size is not too large, both the canonical dual method and SDP method are very effective method for finding sensor locations. In particular, for the canonical dual method, all the stars are exactly located inside circles.

### 9.4.2 A 20-sensor-network localization problem with distance errors

A network of 20 uniform randomly distributed unknown points is generated in the square area $[0,1] \times [0,1]$. We assume

$$\begin{aligned} \text{If } \|x_i - x_j\| &\leq \text{ radio range, a distance (with noise) is given between } x_i \text{ and } x_j, \\ \text{If } \|x_i - x_j\| &> \text{ radio range, no distance is given between } x_i \text{ and } x_j. \end{aligned}$$

4 anchors are located in (0,0), (0,1), (1,0) and (1,1). The distances between the nodes are calculated. If the distance between two nodes is within the specified radio range of 0.4, the distance is included in the edge set for solving the problem after adding a random error to it in the following manner:

$$d_{ij} = \hat{d}_{ij} |1 + N(0, \sqrt{0.001})|$$

where $\hat{d}_{ij}$ is the actual distance between the 2 nodes, and $N(0, \sqrt{0.001})$ is a random variable.

The computed results obtained by the canonical dual method and the standard SDP method are plotted in Fig. 9.3 and Fig. 9.4, respectively. The true sensor locations (denoted by circles) and the computed locations (denoted by stars) are connected by solid lines.

### 9.4.3 A 200-sensor-network localization problem with distance errors

A network of 200 uniform randomly distributed unknown points is generated in the square area $[0,1] \times [0,1]$. 4 anchors are located in (0,0), (0,1), (1,0) and (1,1). For all sensors, the radio range = 0.2. The distance, including a random error, is generated in the following manner:

$$d_{ij} = \hat{d}_{ij} |1 + N(0, 0.01)|$$

The computed results obtained by the canonical dual method and the standard SDP

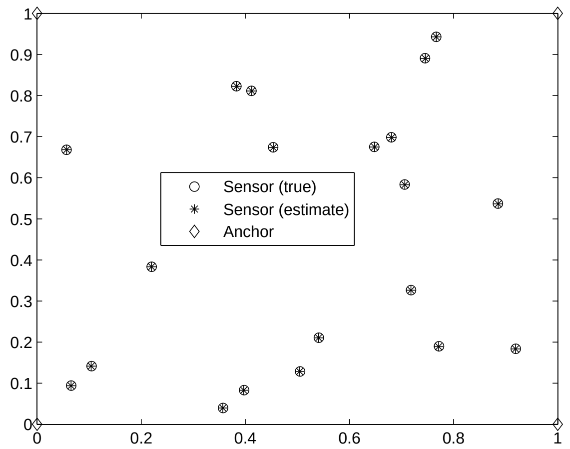



Figure 9.3: Sensor network with 20 sensors solved by the canonical dual method.

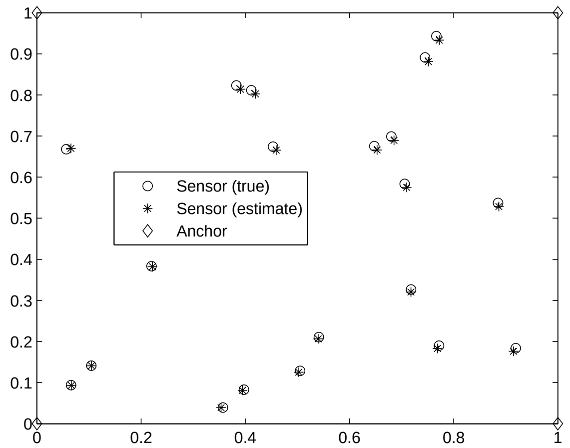



Figure 9.4: Sensor network with 20 sensors solved by the standard SDP method.

method are plotted in Fig. 9.5 and Fig. 9.6, respectively.

Careful examination of the results obtained for the cases involving 20 sensors and 200 sensors, we observe that when noise is taken into consideration, the canonical dual method gives rise to much better solutions. In particular, if the level of noise or the sensor size is large, the standard SDP is usually having difficulty to find the exact sensor positions. See Figure 9.3 and Figure 9.5.

## 9.5 Conclusions

We have presented an effective computational method based on the canonical duality theory for solving large scale sensor network localization problems. The form of the analytical solution is obtained by using the complementary-duality principle, yielding a

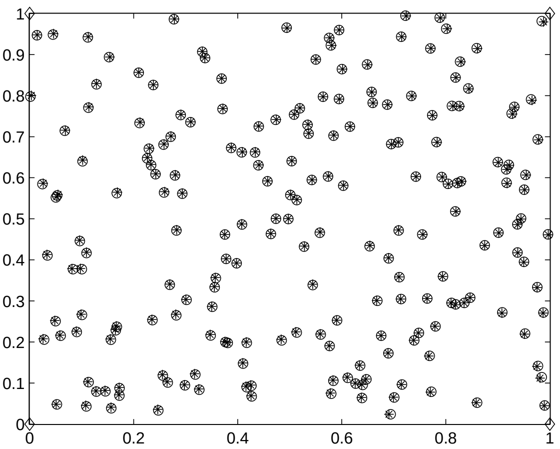



Figure 9.5: Sensor network with 200 sensors solved by the canonical dual method.

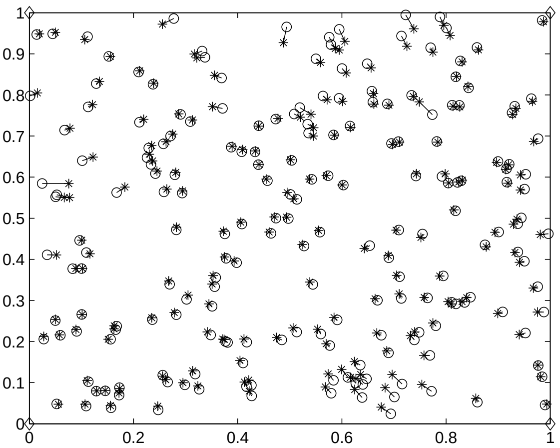



Figure 9.6: Sensor network with 200 sensors solved by the standard SDP method.

concave maximization dual problem, which can be solved by any nonlinear optimization technique. From the numerical studies, it is seen that large scale sensor network localization problems can be solved by the method proposed, yielding global solutions.

# CHAPTER 10

# Summary and future research directions

## 10.1 Main contributions of this thesis

We have presented a detailed application of the canonical duality theory for solving general sum of fourth-order polynomial optimization problem. An analytical solution is obtained by the complementary-dual principle and its extremality property is classified by the triality theory. Results show that by using the canonical dual transformation, the nonconvex primal problem in $\mathbb{R}^n$ can be converted into a concave maximization dual problem $(\mathcal{P}^d_{\max})$ in $\mathbb{R}^m$, which can be solved by well-developed convex minimization techniques .

Generally speaking, the nonconvex quadratic form with an exponential objective function can be used to model many nonconvex systems. By using the canonical dual transformation, the nonconvex primal problem in n-dimensional space can be converted into a one-dimensional canonical dual problem. As indicated in [38], for any given nonconvex problem, as long as the geometrical operator $\Lambda(x)$ can be chosen properly and the canonical duality pairs can be identified correctly, the canonical dual transformation can be used to formulate perfect dual problems. In global optimization, extensive applications of the canonical duality theory have been given to the problems including concave minimization with inequality constraints, polynomial minimization, nonconvex minimization with box constraints, quadratic minimization with general nonconvex constraints, nonconvex fractional programming, and integer programming.

We have presented a detailed application of the canonical duality theory for solving box and integer constrained quadratic optimization problems $(\mathcal{P})$ and $(\mathcal{P}_{ip})$. By using the canonical dual transformation, several canonical dual problems and their perturbations are proposed.

For any given $\mathbf{Q}$ and $\mathbf{f}$, the discrete integer constrained problem $(\mathcal{P}_{ip})$ is equivalent to the continuous unconstrained canonical dual problem $(\mathcal{P}^d_{\alpha})$. For convex-perturbation $\mathbf{Q} + \text{Diag}\ (\boldsymbol{\alpha}) \succ 0$, if the concave maximization problem

$$(\mathcal{P}^\sharp_{\alpha}): \quad \max\{P^d_{\alpha}(\boldsymbol{\sigma})|\ \boldsymbol{\sigma} \in \mathbb{R}^n\}. \qquad (10.1)$$

has a critical solution, the discrete problem $(\mathcal{P}_{ip})$ can be solved uniquely. Otherwise, the nonsmooth problem $(\mathcal{P}^{\sharp}_{\alpha})$ provides a lower bound for box constrained problem $(\mathcal{P})$.

The canonical duality theory was originally developed for general complex systems [38] [44], which has been successfully applied for solving a class of nonconvex/nonsmooth variational/boundary value problems [36]. Complete sets of solutions to a class of well-known problems in finite deformation mechanics and phase transitions of solids have been obtained [42]. Recent applications in finite dimensional systems have shown that this theory is potentially useful for solving both continuous and discrete global optimization problems [29] [40] [41].

As indicated in [38], the key step in the canonical dual transformation is to choose the (nonlinear) geometrical operator $\Lambda(\mathbf{x})$. Different forms of $\Lambda(\mathbf{x})$ may lead to different (but equivalent) canonical dual problems.

To see this, instead of the vector-valued (pure) quadratic geometrical operator $\Lambda(\mathbf{x}) = \frac{1}{2}\{\mathbf{x}^T\mathbf{Q}\mathbf{x}, \mathbf{x}\circ\mathbf{x}\}$ for integer programming, we simply let $\Lambda(\mathbf{x})$ be a matrix-valued geometrical operator:

$$\boldsymbol{\xi} = \Lambda(\mathbf{x}) = \frac{1}{2}\mathbf{x}\mathbf{x}^T : \mathbb{R}^n \to \mathcal{E} = \mathbb{R}^{n\times n}. \tag{10.2}$$

Then, both the primal problems $(\mathcal{P})$ and $(\mathcal{P}_{ip})$ can be written in the following unified canonical form

$$\min\{\Pi(\mathbf{x}) = V(\Lambda(\mathbf{x})) - \langle \mathbf{x}, \mathbf{f}\rangle |\ \mathbf{x} \in \mathbb{R}^n\}, \tag{10.3}$$

where the canonical function $V : \mathcal{E} \to \mathbb{R}\cup\{+\infty\}$ is defined by

$$V(\boldsymbol{\xi}) = \langle \mathbf{Q}; \boldsymbol{\xi}\rangle + \begin{cases} 0 & \text{if } \boldsymbol{\xi} \in \mathcal{E}_a \\ \infty & \text{if } \boldsymbol{\xi} \notin \mathcal{E}_a. \end{cases} \tag{10.4}$$

For box constrained problem, the effective domain $\mathcal{E}_a$ of $V(\boldsymbol{\xi})$ is defined by

$$\mathcal{E}_a = \{\boldsymbol{\xi} \in \mathcal{E} |\ \boldsymbol{\xi} = \boldsymbol{\xi}^T,\ \boldsymbol{\xi} \succeq 0,\ 2\xi_{ii} \le 1\ \ \forall i \in \{1, \ldots, n\},\ \boldsymbol{\xi} \text{ rank-one }\}. \tag{10.5}$$

While for integer constrained problem, the inequality $2\xi_{ii} \le 1$ in $\mathcal{E}_a$ should be replaced by equality. The bilinear form $\langle \boldsymbol{\xi}; \boldsymbol{\xi}^*\rangle : \mathcal{E}\times\mathcal{E}^* \to \mathbb{R}$ is defined by

$$\langle \boldsymbol{\xi}; \boldsymbol{\xi}^*\rangle = \text{trace}(\boldsymbol{\xi}^T\boldsymbol{\xi}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \xi_{ij}\xi^*_{ij}. \tag{10.6}$$

Using $\boldsymbol{\sigma} \in \mathbb{R}^n$ to relax the inequality condition $\xi_{ii} \le \frac{1}{2}$ and let $\boldsymbol{\xi} = \frac{1}{2}\mathbf{x}\mathbf{x}^T$ to relax $\boldsymbol{\xi} = \boldsymbol{\xi}^T$, $\boldsymbol{\xi} \succeq 0$, and rank-one conditions in $\mathcal{E}_a$, the Fenchel sup-conjugate of the canonical function $V(\boldsymbol{\xi})$ can be obtained as:

$$\begin{aligned} V^\sharp(\boldsymbol{\xi}^*) &= \sup_{\boldsymbol{\xi}\in\mathcal{E}}\{\langle\boldsymbol{\xi};\boldsymbol{\xi}^*\rangle - V(\boldsymbol{\xi})\} = \sup_{\boldsymbol{\xi}\in\mathcal{E}_a}\{\langle\boldsymbol{\xi};\boldsymbol{\xi}^* - \mathbf{Q}\rangle\} \\ &= \sup_{\mathbf{x}\in\mathbb{R}^n}\left\{\frac{1}{2}\langle\mathbf{x},(\boldsymbol{\xi}^*-\mathbf{Q})\mathbf{x}\rangle - \sum_{i=1}^n \frac{1}{2}\sigma_i(\mathbf{x}_i^2-1)\right\} \\ &= \begin{cases} \frac{1}{2}\langle\mathbf{e},\boldsymbol{\sigma}\rangle & \text{if } \boldsymbol{\xi}^*\in\mathcal{E}_a^* \\ +\infty & \text{otherwise,}\end{cases} \end{aligned}$$

where

$$\mathcal{E}_a^* = \{\boldsymbol{\xi}^*\in\mathcal{E}_a^* = \mathbb{R}^{n\times n}|\ \boldsymbol{\xi}^* = \mathbf{Q} + \text{Diag}\ (\boldsymbol{\sigma}),\ \ \boldsymbol{\sigma}\in\mathbb{R}_+^n\}. \tag{10.7}$$

Therefore, in term of $\boldsymbol{\sigma}$, the standard total complementary function is

$$\Xi(\mathbf{x},\boldsymbol{\xi}^*) = \langle\Lambda(\mathbf{x});\boldsymbol{\xi}^*\rangle - V^*(\boldsymbol{\xi}^*) - \langle\mathbf{x},\mathbf{f}\rangle.$$

Since the geometrical operator $\Lambda(\mathbf{x}) = \frac{1}{2}\mathbf{x}\mathbf{x}^T$ is a pure quadratic function of $\mathbf{x}$, its variation at $\bar{\mathbf{x}}$ in the direction of $\mathbf{x}$ is $\delta\Lambda(\bar{\mathbf{x}},\mathbf{x}) = \Lambda_t(\bar{\mathbf{x}})\mathbf{x} = \mathbf{x}\bar{\mathbf{x}}^T$, where $\Lambda_t(\bar{\mathbf{x}}) = \nabla\Lambda(\bar{\mathbf{x}})$ denotes the derivative of $\Lambda(\mathbf{x})$ at $\bar{\mathbf{x}}$. Its complementary operator is defined as $\Lambda_c(\mathbf{x}) = \Lambda(\mathbf{x}) - \Lambda_t(\mathbf{x}) = -\frac{1}{2}\mathbf{x}\mathbf{x}^T$ , where $\Lambda, \Lambda_t, \Lambda_c$ are denoted as $A$, $T$, $N$, respectively. By $\Lambda_t$, the canonical equilibrium condition

$$\langle\Lambda_t(\bar{\mathbf{x}})\mathbf{x};\boldsymbol{\xi}^*\rangle = \langle\mathbf{x},\mathbf{f}\rangle\ \ \forall\mathbf{x}\in\mathbb{R}^n$$

leads to the analytical solution form $\bar{\mathbf{x}} = G^{-1}(\boldsymbol{\sigma})\mathbf{f}$. By $\Lambda_c$, the complementary gap function is given by

$$G_{ap}(\mathbf{x},\boldsymbol{\xi}^*) = \langle-\Lambda_c(\mathbf{x});\boldsymbol{\xi}^*\rangle = \frac{1}{2}\langle\mathbf{x}\mathbf{x}^T;\mathbf{Q}+\text{Diag}\ (\boldsymbol{\sigma})\rangle = \frac{1}{2}\langle\mathbf{x},\mathbf{G}(\boldsymbol{\sigma})\mathbf{x}\rangle.$$

Clearly, the sufficient condition $G_{ap}(\mathbf{x},\boldsymbol{\sigma}) \geq 0\ \ \forall\mathbf{x}\in\mathbb{R}^n$ for global minimizer of the primal problem leads to the semi-positive definite condition $\mathbf{G}(\boldsymbol{\sigma}) \succeq 0$. Therefore, we have

$$\min_{\mathbf{x}\in\mathcal{X}_a} P(\mathbf{x}) = \max_{\boldsymbol{\sigma}\in\mathcal{S}_g^+} P^d(\boldsymbol{\sigma}).$$

Thus, we have shown again that the equivalent (or the same) canonical dual problem can be obtained by using different quadratic geometrical operator $\Lambda(\mathbf{x})$.

In finite deformation theory and differential geometry, the pure quadratic geometrical measure $\boldsymbol{\xi} = \Lambda(\mathbf{x})$ is similar to the Cauchy-Riemann type metric tensor, which has been used extensively in the canonical duality theory [38]. In semi-definite optimization, the bilinear form $\langle\boldsymbol{\xi};\boldsymbol{\xi}^*\rangle$ is denoted by $\boldsymbol{\xi}\bullet\boldsymbol{\xi}^*$. Therefore, in a very special case of $\mathbf{f} = 0$, the canonical primal problem (10.3) for integer programming can be written (in term of

$\mathbf{X} = 2\boldsymbol{\xi} = \mathbf{x}\mathbf{x}^T$) as :

$$(\mathcal{P}_{mc}): \ \ \min \frac{1}{2}\mathbf{Q} \bullet \mathbf{X}, \ \ s.t. \ \ \mathbf{X} \succeq 0, \ \ X_{ii} = 1, \ \ \mathbf{X} = \mathbf{X}^T, \ \mathbf{X} \text{ rank-one.} \tag{10.8}$$

If both the symmetrical and rank-one constraints are ignored, this problem is exactly a semi-definite programming problem. However, we must emphasize that for quadratic integer programming problems, these two conditions imply that $\mathbf{X} = \mathbf{x}\mathbf{x}^T$ and hence can not be ignored. Otherwise, Problem (10.8) will have $n \times n$ unknowns, and the dual variable of $\mathbf{X}$ should be also a tensor $\mathbf{X}^* \in \mathbb{R}^{n \times n}$, instead of a vector in $\mathbb{R}^n$. Therefore, by introducing Lagrange multiplier $\boldsymbol{\sigma} \in \mathbb{R}^n$ to relax the equality and rank-one conditions, we have

$$(\mathcal{P}^d_{mc}): \ \ \max \left\{ -\frac{1}{2}\langle \mathbf{e}, \boldsymbol{\sigma} \rangle \ | \ \mathbf{G}(\boldsymbol{\sigma}) \succeq 0, \ \ \boldsymbol{\sigma} \neq 0 \right\}, \tag{10.9}$$

which is clearly a special case of the canonical dual problem $(\mathcal{P}^g_{\max})$.

we know that the vector $\mathbf{f}$ plays a fundamental role for ensuring unique solution of the nonconvex quadratic programming problems. From the view point of systems theory, $\mathbf{f}$ represents input (or source) and $\mathbf{x}$ denotes the output (or state). If there is no input, the system either has trivial solution ($\mathbf{x} = 0$) or more than one solution. The reason for multi-solutions is due to the symmetry of the systems. The input usually destroys certain symmetry and leads to the possibility of unique solution.

## 10.2 Future research directions

The canonical duality theory was originally developed for general complex systems [38]. It has been successfully applied to some nonconvex optimization problems. It is mathematically challenging and practically significant to develop efficient and effective numerical methods based on the use of canonical theory, to finding global solutions of discrete optimization problems and nonconvex optimization problems.

The $\alpha$-perturbed problem $(\mathcal{P}_\alpha)$ is actually a quadratic perturbation for solving general Euclidean distance geometry problems in network optimization. However, how to choose the perturbation vector $\boldsymbol{\alpha}$ is fundamentally important and deserves further investigation.

Also if the dual feasible space $\mathcal{S}^+_a$ contains no KKT point, the primal problem could be NP-hard. In this case, the canonical dual problem $(\mathcal{P}^\sharp)$ can be used to provide an optimal lower bound approach to the NP-hard primal problems. Finding conditions for $Q, \mathbf{c}, \mathbf{A}, \mathbf{b}$ such that $\mathcal{S}^+_a$ has no KKT point is an open problem which is fundamentally important for understanding NP-hard problems.

# Bibliography


[1] "Introduction to global opotimization," in *Nonconvex Optimization and its Applications*, R. Horst, P. Pardalos, and N. Thoai, Eds. Dordrecht: Kluwer Academic Publishers, 2000.

[2] K. Aardal, "Capacitated facility location: separation algorithms and computational experience," *Math. Program.(ser. B)*, vol. 81, pp. 149–175, 2000.

[3] I. G. Akrotirianakis and C. Floudas, "Computational experience with a new class of convex underestimators: Box-constrained nlp problems," *J. Global Optim.*, vol. 29, pp. 249–264, 2004.

[4] ——, "A new class of improved convex underestimators for twice continuously differentiable constrained nlps," *J. Global Optim.*, vol. 30, pp. 367–390, 2004.

[5] A. Alfakih, A. Khandani, and H. Wolkowicz, "Solving euclidean distance matris completion problems via semidefinite programming," *Comput. Opt. and Appl.*, vol. 12, pp. 13–30, 1999.

[6] A. Alfakin, "Graph rigidity via euclidean distance matrices," *Linear Algebra Appl.*, vol. 310, pp. 149–165, 2000.

[7] B. Alidaee and G. Kochenberger, "A note on a simple synamic programming approach to the single-sink fixed-charge transportation problem," *Transport. Sci.*, vol. 39, pp. 140–143, 2005.

[8] J. Arora, M. Huang, and C. Hsieh, "Methods for optimization of nonlinear problems with discrete variables: a review," *Struct. Optim.*, vol. 8, pp. 69–85, 1994.

[9] A. Atamtürk, "Flow pack facets of the single node fixed-charge flow polytope," *Oper. Res. Lett.*, vol. 29, pp. 107–114, 2001.

[10] G. Auchmuty, "Duality for non-convex variational principles," *J. Differential Equations*, vol. 1, pp. 83–145, 1983.

[11] E. Balas, S.Ceria, and G. Cornuéjols, "A lift-and-project cutting plane algorithm for mixed 0-1 programs," *Math. Program.*, vol. 58, pp. 295–324, 1993.

[12] I. Barany, T. V. Roy, and L. Wolsey, "Strong formulations for multi-item capacitated lot sizing," *Management Science*, vol. 30, pp. 1255–1261, 1984.

[13] A. Barvinok, "Problems of distance geometry and convex properties of quadratic maps," *Disc Comp. Geom.*, vol. 13, pp. 189–202, 1995.

[14] M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear programming: Theory and algorithm.* New York: Wiley, 2006.

[15] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Waechter, "Branching and bounds tightening techniques for non-convex minlp," *Optim. Method Softw.*, vol. 24, pp. 597–634, 2009.

[16] A. Bishop, B. Fidan, B. O. Anderson, K. Dogcancay, and P. Pathirana, "Optimality analysis of sensor-target localization geometries," *Automatica*, vol. 46, pp. 479–492, 2010.

[17] P. Biswas, T. Liang, K. Toh, Y. Ye, and T. Wang, "Semidefinite programming approaches for sensor network localization with noisy distance measurement," *IEEE Trans. Autom. Sci. Eng.*, vol. 3, no. 4, pp. 360–371, 2006.

[18] B. Borchers and J. Mitchell, "An improved branch and bound algorithm for mixed integer nonlinear programs," *Comput. Oper. Res.*, vol. 21, pp. 359–367, 1994.

[19] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[20] G. Briggs, F. Kidd, and C. West, "A quantitative analysis of plant growth," *Ann. Appl. Biol.*, vol. 7, pp. 202–223, 1920.

[21] D. S. Chen, R. Batson, and Y. Dang, *Applied Integer Programming: Modeling and Solution.* John Wiley and Sons, 2010.

[22] M. Chiang, *Nonconvex optimization for communication systems*, 1st ed., D. Gao and H. Sherali, Eds. Berlin, Germany: Springer-Verlag, 2007.

[23] N. Courtois, L. Goubin, W. Meier, and J. D. Tacier, "Solving underdefined systems of multivariate quadratic equations," in *Proceedings of Public Key Cryptography 2002, LNCS 2274.* Springer-Verlag, 2002, pp. 211–227.

[24] M. Deghat, I. Shames, B. Anderson, and J. Moura, "Distributed localization via barycentric coordinates: Finite-timeconvergence," in *18th World Congress of the International Federation of Automatic Control,(IFAC 2011)*, Milan, Italy, 2011, pp. 7824–7829.

[25] D.Li and X. Sun, *Nonlinear Integer Programming.* New York: Springer, 2006.

[26] M. A. Duran and L. E. Grossmann, "An outer-approximation algorithm for a class of mixed integer nonlinear programs," *Math. Program.*, vol. 36, pp. 307–339, 1986.

[27] I. Ekeland, "Legendre duality in nonconvex optimization and calculus variations," *SIAM J. Control Optimization*, vol. 6, pp. 905–934, 1977.

[28] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems.* North-Holland, 1976.

[29] S. Fang, D. Gao, R. Sheu, and W. Xing, "Global optimization for a class of fractional programming problem," *J. Glob. Optim.*, vol. 45, pp. 337–353, 2009.

[30] R. Fletcher and S. Leyffer, "Solving mixed integer nonlinear programs by outer approximation," *Math. Program.*, vol. 66, pp. 327–349, 1994.

[31] C. Floudas, *Nonlinear and Mixed-Integer Optimization: Theory, Methods and Applications.* Dordrecht/Boston/London: Kluwer Academic Publishers, 2000.

[32] C. Floudas, I. Akrotirianakis, S. Caratzoulas, C. Meyer, and J. Kallrath, "Global optimization in the 21th century: advances and challenges," *Comput. Chem. Eng.*, vol. 29, pp. 1185–1202, 2005.

[33] C. Floudas and V. Visweswaran, "A primal-relaxed dual global optimization approach," *J. Optimiz. Theory App.*, vol. 78, no. 2, pp. 187–225, 1993.

[34] ——, *Quadratic optimizations*, R. Horst and P. Pardalos, Eds. Dordrecht/Boston/London: Kluwer Academic Publishers, 1995.

[35] D. Gao, "Complementary finite element method for finite deformation nonsmooth mechanics," *J. Eng. Math.*, vol. 30, pp. 339–353, 1996.

[36] ——, "Duality, triality and complementary extremun principles in nonconvex parametric variational problems with applications," *IMA J. Appl. Math.*, vol. 61, pp. 199–235, 1998.

[37] ——, "Canonical dual transformation method and generalized triality theory in nonsmooth global optimization," *J. Glob. Optim.*, vol. 17, pp. 127–160, 2000.

[38] ——, *Duality Principles in Nonconvex Systems: Theory, Methods and Applications.* Dordrecht/Boston/London: Kluwer Academic Publishers, 2000.

[39] ——, “Nonconvex semi-linear problems and canonical dual solutions,” in *Advances in Mechanics and Mathematics*. Kluwer Academic Publishers, 2003, vol. II, pp. 261–312.

[40] ——, “Perfect duality theory and complete solutions to a class of global optimization problems,” *Optimization*, vol. 52, pp. 467–493, 2003.

[41] ——, “Solutions and optimality to box constrained nonconvex minimization problems,” *J. Ind. Manag. Optim.*, vol. 3, no. 2, pp. 293–304, 2007.

[42] D. Gao and R. Ogden, “Closed-form solutions, extremality and nonsmoothness criteria to nonconvex variational problem in finite elasticity,” *Z. angew. Math. Phys. (ZAMP)*, vol. 59, p. 498517, 2008.

[43] ——, “Multiple solutions to nonconvex variational problems with implications for phase transitions and numerical computation,” *Q.J. Mech. Appl. Math.*, vol. 61, pp. 497–522, 2008.

[44] D. Gao and G. Strang, “Geometric nonlinearity: Potential energy, complementary energy, and the gap function,” *Quart. Appl. Math.*, vol. 47, no. 3, pp. 487–504, 1989.

[45] D. Gao and H. Yu, “Multi-scale modelling and canonical dual finite element method in phase transitions of solids,” *Int. J. Solids Struct.*, vol. 45, pp. 3660–3673, 2008.

[46] R. Gasimov, “Augmented lagrangian duality and nondifferentiable optimization methods in nonconvex programming,” *J. Global Optim.*, vol. 24, pp. 187–203, 2002.

[47] B. Ghaddar, J. Vera, and M. Anjos, “Second-order cone relaxations for binary quadratic polynomial programs,” *SIAM J. Optim.*, vol. 21, pp. 391–414, 2011.

[48] F. Glover and H. Sherali, “Some classes of valid inequalities and convex hull characterizations for dynamic fixed-charge problems under nested constraints,” *Ann. Oper. Res.*, vol. 40, pp. 215–234, 2005.

[49] L. Grippo and S. Lucidi, “A differentiable exact penalty function for bound constrained quadratic programming problems,” *Optimization*, vol. 22, pp. 557–578, 1991.

[50] C. Grosan and A. Abraham, “A new approach for solving nonlinear equations systems,” *IEEE Trans. Syst. Man Cybern.*, vol. 38, pp. 698–714, 2008.

[51] Z. Gu, G. Nemhauser, and M. Savelsbergh, “Lifted flow cover inequalities for mixed 0-1 integer programs,” *Math. Program. (Ser.A )*, vol. 85, pp. 439–467, 1999.

[52] O. Gupta and A. Ravindran, "Branch and bound experiments in convex nonlinear integer programming," *Manag. Sci.*, vol. 31, pp. 1553–1546, 1985.

[53] J. Haberl, "Exact algorithm for solving a special fixed-charge linear programming problem," *J. Optimiz. Theory App.*

[54] B. Hajek and G. Sasaki, "Simulated annealing-to coll or not," *Systems Control Lett.*, vol. 12, pp. 443–447, 1989.

[55] P. Hansen, B. Jaumard, M. Ruiz, and J. Xiong, "Global minimization of indefinite quadratic functions subject to box constraints," *Nav. Res. Logist.*

[56] P. Hentenryck, D. Mcallester, and D. Kapur, "Solving polynomial systems using a branch and prune approach," *Math. Program. (Ser.A )*, vol. 34, pp. 797–827, 1997.

[57] J. Holland, Ed., *Adaption in Natural and Artificial Systems.* Ann Arbor, Michigan: The University of Michigan Press, 1975.

[58] G. Holzapfel and R. Ogden, Eds., *Mechanics of biological Tissue.* Berlin, Germany: Springer-Verlag, 2006.

[59] F. Jeng and J. Woods, "Simulated annealing in compound gaussian random fields," *IEEE Trans. Inform. Theory*, vol. 36, pp. 94–107, 1990.

[60] Y. Jiao and Y. Chen, "An application of generalized linear model in production model and sequential population analysis," *Fish. Res.*, vol. 70, pp. 367–376, 2004.

[61] Y. Jiao, R. Neves, and J. Jones, "Models and model selection uncertainty in estimating growth rates of endangered freshwater mussel populations," *Can. J. Fish. Aquat. Sci*, vol. 65, pp. 2389–2398, 2008.

[62] J. Karlof, Ed., *Integer Programming: Theory and Practice.* CRC Press, 2006.

[63] P. Kesavan and P. I. Barton, "Generalized branch-and-cut framework for mixed-integer nonlinear optimization problems." *Comput. Chem. Eng.*, vol. 24, pp. 1361–1366, 2000.

[64] U. A. Khan, S. Kar, and J. Moura, "Distributed sensor localization in random environments using minimal number of anchor nodes," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 2000–2016, 2009.

[65] A. Klose, "Algorithms for solving the single-sink fixed-charge transportation problem," *Comput. Oper. Res.*, vol. 35, no. 6, pp. 2079–2092, 2008.

[66] H. Loh and P. Papalambros, "Computational implementation and tests of a sequential linearization approach for solving mixed-discrete nonlinear design optimization," *J.Mech. Des. ASME*, vol. 113, pp. 335–345, 1991.

[67] W. Lu, *Use SeDuMi to solve LP, SDP and SCOP problem: Remarks and examples.* [Online]. Available: http://www.ece.uvic.ca/~wslu/Talk/SeDuMi-Remarks.pdf

[68] R. Marsten and T. Morin, "A hybrid approach to discrete mathematical programming," *Math. Program.*, vol. 14, pp. 21–40, 1978.

[69] R. Meyer and C. Nachtsheim, "Constructing exact d-optimal experimental designs by simulated annealing," *Amer. J. Math. Management Sci.*, vol. 8, pp. 329–359, 1988.

[70] J. J. Moré and Z. J. Wu, "Distance geometry optimization for protein structures," *J. Global Optimization*, vol. 15, pp. 297–312, 1999.

[71] J. Mosler and M. Ortiz, "Variational h-adaption in finite deformatin elasticity and plasticity," *Int. J. Numer. Meth. Engng.*, vol. 72, pp. 505–523, 2007.

[72] K. Murty and S. Kabadi, "Some NP-hard problems in quadratic and nonlinear programming," *Math. Program.*, vol. 39, pp. 117–129, 1987.

[73] M. Padberg, T. V. Roy, and L. Wolsey, "Valid linear inequalities for fixed charge problems," *Oper. Res.*, vol. 33, pp. 842–861, 1985.

[74] A. Pal, "Localization algorithms in wireless sensor networks:current approaches and future challenges," *Network Protocols and Algorithms*, vol. 2, pp. 45–74, 2010.

[75] P. Pardalos and G. Schnitger, "Checking local optimality in constrained quadratic and nonlinear programming," *Oper. Res. Lett.*, vol. 7, pp. 33–35, 1988.

[76] P. Pardalos and S. Vavasis, "Quadratic programming with one negative eigenvalue is NP-hard," *J. Glob. Optim.*, vol. 1, pp. 15–23, 1991.

[77] R. Pŏrn and T. Westerlund, "A cutting plane method for minimizing pesudo-convex functions in the mixed integer case," *Comput. Chem. Eng.*, vol. 24, pp. 2655–2665, 2000.

[78] T. Pong and P. Tseng, "(robust) edge-based semidefinite programming relaxation of sensor network localization," *Math. Program.*, vol. 130, pp. 321–358, 2011, ser. A.

[79] R. Rangarajan, R. Raich, and A. Hero, "Euclidean matrix completion problems in tracking and geo-localization," in *IEEE Intl Conf. on Acoustics, Speech and Signal Processing*, Apple 2008.

[80] R. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton University Press, 1970.

[81] N. Ruan, D. Y. Gao, and Y. Jiao, "Canonical dual least square method for solving general nonlinear systems of quadratic equations," *Comput. Optim. Appl.*, vol. 47, pp. 335–347, 2010.

[82] S. Sahni, "Computationally related problems," *SIAM J. Comp*, vol. 3, p. 26279, 1974.

[83] E. Sandgren, "Nonlinear integer and discreter programming in mechanical design optimizatin," *J.Mech. Des. ASME.*, vol. 112, pp. 223–229, 1990.

[84] H. Santosa and D. Gao, "Canonical dual finite element method for solving post-buckling problems of a large deformation elastic beam," 2011, doi:10.1016/j.ijnonlinmec.2011.05.012,.

[85] J. Saxe, "Embeddability of weighted graphs in k-space is strongly np-hard," in *Proc. 17th Allerton Conference in Communications, Control, and Computing*, Monticello, IL, 1979, pp. 480–489.

[86] Y. Shang, W. Rumi, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *IEEE Trans. Parall. Distr.*, vol. 15, no. 11, pp. 961–974, 2004.

[87] H. Sherali and J. Smith, "An improved linearization stategy for zero-one quadratic programming problems." *Optim. Lett.*, vol. 1, pp. 33–47, 2007.

[88] H. Sherali and C. Tuncbileck, "A global optimization algorithm for polynomial programming problem using a reformulation-linearization technique." *J. Global Optim.*, vol. 2, pp. 101–112, 1992.

[89] ——, "A reformulation-convexification approach for solving nonconvex quadratic programming problems." *J. Global Optim.*, vol. 7, pp. 1–31, 1995.

[90] ——, "New reformulation-linearization technique based relaxation for univariate and multivariate polynominal programming problems." *Oper. Res. Lett.*, vol. 21, pp. 1–10, 1997.

[91] I. Singer, "Duality for optimization and best approximation over finite intersections," *Numer. Funct. Anal. Optim.*, vol. 19, pp. 903–915, 1998.

[92] R. Strongin and Y. D. Sergeyev, Eds., *Global Optimization with Non-Convex Constraints: Squential and Parallel Algorithms.* Dordrecht: Kluwer Academic, 2000.

[93] R. Stubbs and S. Mehrotra, "A branch-and-cut method for 0-1 mixed convex programming," *Math. Program.*, vol. 86, pp. 515–532, 1999.

[94] P. T. Thach, "Global optimality criterion and a duality with a zero gap in nonconvex optimization," *SIAM J. Math. Anal.*, vol. 24, pp. 1537–1556, 1993.

[95] P. T. Thanedar and G. Vanderplaats, "A survey of discrete variable optimization for structural design," *J. Struct. Eng. ASCE*, vol. 121, pp. 301–306, 1994.

[96] J. F. Toland, "Duality in non-convex optimization," *J. math. anal. appl.*, vol. 66, pp. 399–415, 1978.

[97] P. Tseng, "Second-order cone programming relaxation of sensor network localization," *SIAM J. Optimiz.*, vol. 18, no. 1, pp. 156–185, 2007.

[98] H. Waki, S. Kim, M. Kojima, and M. Muramatsu, "Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity," *SIAM J. Optimiz.*, vol. 17, no. 1, pp. 218–242, 2006.

[99] S. Wang, K. Teo, and H. Lee, "A new approach to nonlinear mixed discrete programming problems," *Eng. Optimiz.*, vol. 30, no. 3, pp. 249–262, 1998.

[100] D. Xie, S. Singh, E. Fluder, and T. Schlick, "Principal component analysis combined with truncated-newton minimization for dimensionality reduction of chemical databases," *Math. Program. (Series B)*, vol. 95, p. 161185, 2003.

[101] A. Zecchin, A. Simpson, H. Maier, and J. Nixon, "Parametric study for an ant algorithm applied to water distribution system optimization," *IEEE Transactions on Evolutionary Computation*, vol. 9, pp. 175–191, 2005.

[102] J. Zhang, D. Gao, and J. Yearwood, "A novel canonical dual computational approach for prion agaaaaga amyloid fibril molecular modeling," *J. Theoretical Biology*, vol. 284, pp. 149–157, 2011.

[103] P. Zhang and M. Martonosi, "Locate: Collabotive localization estimation for sparse mobile sensor networks," in *The International Conference on Information Processing in Sensor Networks*, Milan, Italy, Apple 2008, (IPSN 2008).

[104] H. Zhu, Z. Zhou, R. Yang, and A. Yu, "Discrete particle simulation of particulate systems: Theoretical developments," *Chem. Eng. Sci.*, vol. 62, pp. 3378–3396, 2007.

[105] Z.Wang, S. Zheng, Y. Ye, and S. Boyd, "Further relaxation of the semidefinite programming approach to sensor network localization," *SIAM J. Optimiz.*, vol. 19, no. 2, pp. 655–673, 2008.