

Introduction to Canonical Duality Theory

David Y. Gao

Department of Mathematics,
Virginia Tech, Blacksburg, VA 24061, USA. e-mail: gao@vt.edu

Summary. Canonical Duality Theory is a versatile and potentially powerful methodology which is composed mainly of a canonical dual transformation, a complementary-dual principle, and an associated triality theory. The canonical dual transformation can be used for modeling complex systems and to formulate perfect dual problems without duality gap; the complementary-dual principle presents a unified analytic solution form for general problems in continuous and discrete systems; the triality theory is comprised by a saddle min-max duality and two pairs of double-min, double-max dualities. This theory reveals an intrinsic duality pattern in complex phenomena and can be used to solve a very large class of challenging problems in complex systems.

This lecture presents, within a unified framework, a self-contained comprehensive introduction and some new developments on canonical duality theory for complex systems with emphasis on methods and applications in nonlinear analysis and optimization. Intrinsic relations among the popular semi-positive programming, semi-infinite programming, complementarity theory, variational inequality, penalty methods, and the Lagrangian duality theory are revealed within the unified framework of the canonical duality theory.

Applications are illustrated by a class of challenging (NP-hard) problems in global optimization and nonconvex analysis. It is shown that by the use of the canonical dual transformation, nonconvex constrained primal problems can be converted into certain simple canonical dual problems, which can be solved to obtain all extremal points, and NP-hard discrete optimization problems can be transformed to certain *minimal stationary problems* in continuous space. Optimality conditions (both local and global) for these extrema can be identified by the triality theory.

This lecture brings some fundamentally new insights into nonconvex analysis, global optimization, and computational science.

Key words: Duality, triality, Lagrangian duality, nonconvex mechanics, global optimization, nonconvex variations, canonical dual transformations, critical point theory, semi-linear equations, NP-hard problems, quadratic programming.

1.1 Introduction

Duality is a beautiful concept, which plays a fundamental role in multidisciplinary fields of mathematical science, especially in analysis and optimization.

The study of duality in mathematics and mechanics has had a long history since the well-known *Legendre transformation* was formally introduced in 1787. This elegant transformation plays a key role in complementary duality theory. In classical mechanical systems, each energy function defined in a configuration space is linked via the Legendre transformation with a complementary energy in the dual (source) space, through which the Lagrangian and Hamiltonian can be formulated. In static systems, the convex total potential energy leads to a saddle Lagrangian through which a beautiful saddle min-max duality theory can be constructed. This saddle Lagrangian plays a central role in classical duality theory in convex analysis and constrained optimization. In convex dynamic systems, however, the total action is usually a nonconvex d.c. function, that is the difference of convex kinetic energy and total potential functions. In this case, the classical Lagrangian is no longer a saddle function, but the Hamiltonian is convex in each of its variables. It turns out that instead of the Lagrangian, the Hamiltonian has been extensively used in convex dynamics. From a geometrical point of view, Lagrangian and Hamiltonian structures in convex systems and d.c. programming display an appealing symmetry, which was widely studied by their founders. Unfortunately, such a symmetry in nonconvex systems breaks down. It turns out that in recent times, tremendous effort and attention have been focused on the role of symmetry and symmetry-breaking in Hamiltonian mechanics in order to gain a deeper understanding into nonlinear and nonconvex phenomena (see Marsden and Ratiu, 1995).

The earliest examples of the Lagrangian duality in engineering mechanics are probably the complementary energy principles proposed by Haar and von Kármán in 1909 for elasto-perfectly plasticity and Hellinger in 1914 for continuum mechanics. Since the boundary conditions in Hellinger's principle were clarified by E. Reissner in 1953, the complementary-dual variational principles and methods have been studied extensively for more than 50 years by applied mathematicians and engineers.¹ The development of mathematical duality theory in convex variational analysis and optimization has had a similar history since W. Fenchel proposed the well-known Fenchel transformation in 1949. After the revolutionary concepts of super-potential and subdifferentials introduced by J. J. Moreau in 1966 in the study of frictional mechanics,

¹ Eric Reissner (PhD '1938) was a professor in the Department of Mathematics at MIT from 1949 to 1969. According to Gil Strang, since Reissner moved to the Department of Mechanical and Aerospace Engineering at University of California, San Diego in 1969, many applied mathematicians in the field of continuum mechanics, especially solid mechanics, switched from mathematical departments to engineering schools in the United States.

the modern mathematical theory of duality has been well-developed by celebrated mathematicians such as R. T. Rockafellar (1967, 1970), I. Ekeland, R. Temam (1976), F. H. Clarke (1983), and G. Strang (1986). Mathematically speaking, in linear elasticity where the total potential energy is convex, the Hellinger-Reissner complementary variational principle in engineering mechanics is equivalent to a Fenchel-Moreau-Rockafellar type dual variational problem. The so-called generalized complementary variational principle is actually the saddle Lagrangian duality theory, which serves as the foundation for hybrid/mixed finite element methods, and has been subjected to extensive study during the past 40 years (see Strang and Fix (1973), Oden and Lee (1977), Pian and Wu (2006), Han (2006), and the references cited therein).

Early in the beginning of the last century, Haar and von Karman [73] had already realized that in nonlinear variational problems of continuum mechanics, the direct approaches for solving minimum potential energy (primal problem) can only provide upper bounding solutions. However, the minimum complementary energy principle (i.e., the maximum Lagrangian dual problem) provides a lower bound (the mathematical proof of Haar-von Kármán's principle was given by Greenberg in 1949 [72]). In safety analysis of engineering structures, the upper and lower bounding approximations to the so-called collapse states of the elasto-plastic structures are equally important to engineers. Therefore, the primal-dual variational methods have been studied extensively by engineers for solving nonsmooth, nonlinear problems (see [30, 31, 90], and [58, 57, 66, 67, 62]). The article by Maier *et al.* (2000) serves as an excellent survey on the developments for applications of the Lagrangian duality in engineering structural mechanics. In mathematical programming and computational science, the so-called primal-dual interior point methods are also based on the Lagrangian duality theory, which has emerged as a revolutionary technique during the last 15 years. Complementary to the interior-point methods, the so-called *pan-penalty finite element programming* developed by Gao in 1988 [23, 24] is indeed a *primal-dual exterior-point method*. He proved that in rigid-perfectly plastic limit analysis, the exterior penalty functional and the associated perturbation method possess an elegant physical meaning, which led to an efficient dimension re-scaling technique in large-scale nonlinear mixed finite element programming problems [24].

In mathematical programming and analysis, the subject of complementarity is closely related to constrained optimization, variational inequality, and fixed point theory. Through the classical Lagrangian duality, the KKT conditions of constrained optimization problems lead to corresponding complementarity problems. The primal-dual schema has continued to evolve for linear and convex mathematical programming during the past 20 years. However, for nonconvex systems, it is well-known that the KKT conditions are only necessary under certain regularity conditions for global optimality. Moreover, the underlying nonlinear complementarity problems are fundamentally difficult due to the non-monotonicity of the nonlinear operators, and also, many problems in global optimization are NP-hard. The well-developed Fenchel-Moreau-

Rockafellar duality theory will produce a so-called duality gap between the primal problem and its Lagrangian dual. Therefore, how to formulate perfect dual problems (with a zero duality gap) is a challenging task in global optimization and nonconvex analysis. Extensions of the classical Lagrangian duality and the primal-dual schema to nonconvex systems are ongoing research endeavors. On the flip side, the Hellinger-Reissner complementary energy principle, emanating from large deformation mechanics, holds for both convex and nonconvex problems. It is very interesting to note that around the same time period of Reissner's work, the generalized potential variational principle in finite deformation elasto-plasticity was proposed independently by Hu Hai-chang (1955) and K. Washizu (1955). These two variational principles are *perfectly dual* to each other (i.e., with zero duality gap) and play important roles in large deformation mechanics and computational methods. The inner relations between the Hellinger-Reissner and Hu-Washizu principles were discovered by Wei-Zang Chien in 1964 when he proposed a systematic method to construct generalized variational principles in solid mechanics (see [10]).

Mechanics and mathematics have been complementary partners since Newton's time, and the history of science shows much evidence of the beneficial influence of these disciplines on each other. However, the independent developments of complementary-duality theory in mathematics and mechanics for more than a half century have generated a "duality gap" between the two partners. In modern analysis, the mathematical theory of duality was mainly based on the Fenchel transformation. During the last three decades, many modified versions of the Fenchel-Moreau-Rockafellar duality have been proposed. One, the so-called relaxation method in nonconvex mechanics, can be used to solve the relaxed convex problems (see, [2, 14, 142]). However, due to the duality gap, these relaxed solutions do not directly yield real solutions to the nonconvex primal problems. Thus, tremendous efforts have been focused recently on finding the so-called perfect duality theory in global optimization. On the other hand, it seems that most engineers and scientists prefer the classical Legendre transformation. It turns out that their attention has been mainly focused on how to use traditional Lagrange multiplier methods and complementary constitutive laws to correctly formulate complementary variational principles for numerical computational and application purposes. Although the generalized Hellinger-Reissner principle leads to a perfect duality between the nonconvex potential variational problem and its complementary-dual, and has many important consequences in large deformation theory and computational mechanics, the extremality property of this well-known principle, as well as the Hu-Washizu principle, remained an open problem for more than 40 years, and this raised many arguments in large deformation theory and nonconvex mechanics (see [88, 138, 82, 101, 102, 86, 87, 69]).

Actually, this open problem was partially solved in 1989 in the joint work of Gao and Strang [65] on nonconvex/nonsmooth variational problems. In order to recover the lost symmetry between the nonconvex primal problem

and its dual, they introduced a so-called *complementary gap function*, which leads to a nonlinear Lagrangian duality theory in fully nonlinear variational problems. They proved that if this gap function is positive on a dual feasible space, the generalized Hellinger-Reissner energy is a saddle-Lagrangian. Therefore, this gap function provides a sufficient condition in nonconvex variational problems. However, the extremality conditions for negative gap function were ignored until 1997 when Gao [34] got involved with a project on post-buckling problems in nonconvex mechanics. He discovered that if this gap function is negative, the generalized Hellinger-Reissner energy (the so-called super-Lagrangian) is concave in each of its variables, which led to a bi-duality theory. Therefore, a canonical duality theory has gradually developed, first in nonconvex mechanics, and then in global optimization. This new theory is composed mainly of a potentially useful canonical dual transformation and an associated *triviality theory*, whose components comprise a saddle min-max duality and two pairs of double-min, double-max dualities. The canonical dual transformation can be used to formulate perfect dual problems without a duality gap, whereas the triviality theory can be used to identify both global and local extrema.

The goal of this paper is to present a comprehensive review on the canonical duality theory within a unified framework, and to expose its role in establishing connections between nonconvex mechanics and global optimization. Applications to constrained nonconvex optimization problems are shown to reveal some important new results that are fundamental to global optimization theory. This paper should be of interest to both the operations research and applied mathematics communities. In order to make this presentation easy to follow by interdisciplinary readers, our attention here is mainly focused on smooth systems, although some concepts from nonsmooth analysis have been used in later sections.

1.2 Quadratic Minimization Problems

Let us begin with the simplest quadratic minimization problem (in short, the primal problem (\mathcal{P}_q)):

$$(\mathcal{P}_q) : \quad \min \left\{ P(u) = \frac{1}{2} \langle u, Au \rangle - \langle u, f \rangle \quad : \quad u \in \mathcal{U}_k \right\}, \quad (1.1)$$

where \mathcal{U}_k is an open subset of a linear space \mathcal{U} ; A is a linear symmetrical operator, which maps each $u \in \mathcal{U}$ into its dual space \mathcal{U}^* ; the bilinear form $\langle u, u^* \rangle : \mathcal{U} \times \mathcal{U}^* \rightarrow \mathbb{R}$ puts \mathcal{U} and \mathcal{U}^* in duality; $f \in \mathcal{U}^*$ is a given input, and $P : \mathcal{U} \rightarrow \mathbb{R}$ represents the total cost (action) of the system. The criticality condition $\delta P(u) = 0$ leads to a linear equation

$$Au = f, \quad (1.2)$$

which is called the *fundamental equation* (or equilibrium equation) in mathematical physics. By the fact that $A : \mathcal{U} \rightarrow \mathcal{U}^*$ is a symmetrical operator, we have the following canonical decomposition,

$$A = \Lambda^* D \Lambda, \quad (1.3)$$

where $\Lambda : \mathcal{U} \rightarrow \mathcal{V}$ is a so-called *geometrical operator*, which maps each $u \in \mathcal{U}$ into a so-called *intermediate space* \mathcal{V} , and the symmetrical operator D links \mathcal{V} with its dual space \mathcal{V}^* . The bilinear form $\langle v ; v^* \rangle : \mathcal{V} \times \mathcal{V}^* \rightarrow \mathbb{R}$ puts \mathcal{V} and \mathcal{V}^* in duality. We distinguish between the notations \langle , \rangle and $\langle ; \rangle$ according to the differences of the dual spaces $\mathcal{U} \times \mathcal{U}^*$ and $\mathcal{V} \times \mathcal{V}^*$ on which they are respectively defined. The mapping $v^* = Dv \in \mathcal{V}^*$ is called the *duality equation*. The adjoint operator $\Lambda^* : \mathcal{V}^* \rightarrow \mathcal{U}^*$, defined by

$$\langle \Lambda u ; v^* \rangle = \langle u, \Lambda^* v^* \rangle,$$

is also called the *balance operator*. Thus, by the use of the intermediate pair (v, v^*) , the fundamental equation (1.2) can be split into the so-called *tri-canonical form*

$$\left. \begin{array}{l} \text{(a) geometrical equation: } \Lambda u = v \\ \text{(b) duality equation: } Dv = v^* \\ \text{(c) balance equation: } \Lambda^* v^* = f \end{array} \right\} \Rightarrow \Lambda^* D \Lambda u = f. \quad (1.4)$$

In mathematical physics, the duality equation $v^* = Dv$ is also recognized as the *constitutive law* and the operator D depends on the physical properties of the system considered.

The pair (v, v^*) is said to be a *canonical dual pair* on $\mathcal{V}_a \times \mathcal{V}_a^* \subset \mathcal{V} \times \mathcal{V}^*$ if the duality mapping $D : \mathcal{V}_a \subset \mathcal{V} \rightarrow \mathcal{V}_a^* \subset \mathcal{V}^*$ is one-to-one and onto. Generally speaking, most physical variables appear in dual pairs, i.e., there exists a Gâteaux differentiable function $V : \mathcal{V}_a \rightarrow \mathbb{R}$ such that the duality relation $v^* = \delta V(v) : \mathcal{V}_a \rightarrow \mathcal{V}_a^*$ is revertible, where $\delta V(v)$ represents the Gâteaux derivative of V at v . In mathematical physics, such a function is called *free energy*. Its Legendre conjugate $V^*(v^*) : \mathcal{V}^* \rightarrow \mathbb{R}$, defined by the *Legendre transformation*

$$V^*(v^*) = \text{sta}\{\langle v ; v^* \rangle - V(v) : v \in \mathcal{V}_a\}, \quad (1.5)$$

is called *complementary energy*, where $\text{sta}\{ \}$ denotes finding stationary points of the statement in $\{ \}$. In order to study the canonical duality theory, consider the following definition.

Definition 1. A real-valued function $V : \mathcal{V}_a \subset \mathcal{V} \rightarrow \mathbb{R}$ is called a *canonical function* on \mathcal{V}_a if its Legendre conjugate $V^*(v^*)$ can be uniquely defined on $\mathcal{V}_a^* \subset \mathcal{V}^*$ such that the following relations hold on $\mathcal{V}_a \times \mathcal{V}_a^*$:

$$v^* = \delta V(v) \Leftrightarrow v = \delta V^*(v^*) \Leftrightarrow \langle v ; v^* \rangle = V(v) + V^*(v^*). \quad (1.6)$$

Clearly, if $D : \mathcal{V}_a \rightarrow \mathcal{V}_a^*$ is invertible, the quadratic function $V(v) = \frac{1}{2}\langle v; Dv \rangle$ is canonical on \mathcal{V}_a and its Legendre conjugate $V^*(v^*) = \frac{1}{2}\langle D^{-1}v^*; v^* \rangle$ is a canonical function on \mathcal{V}_a^* . Generally speaking, if $V : \mathcal{V}_a \rightarrow \mathbb{R}$ is a canonical function and $v^* = \delta V(v)$, then (v, v^*) is a canonical dual pair on $\mathcal{V}_a \times \mathcal{V}_a^*$. The one-to-one canonical duality relation serves as a foundation for the canonical dual transformation method reviewed in the following sections. The definition of the canonical pairs and functions can be generalized to nonsmooth systems where the Fenchel transformation and sub-differential have to be applied (see [40, 42]). This is discussed in the context of constrained global optimization problems in Section 8 of this paper.

In order to study general problems, we denote the linear function $\langle u, f \rangle$ by $U(u)$. If the feasible space \mathcal{U}_k can be written in the form of

$$\mathcal{U}_k = \{u \in \mathcal{U}_a \mid \Lambda u \in \mathcal{V}_a\}, \tag{1.7}$$

then the problem (\mathcal{P}_q) can be written in a general form

$$(\mathcal{P}) : \min\{P(u) = V(\Lambda u) - U(u) : u \in \mathcal{U}_k\}. \tag{1.8}$$

This general form covers many problems in applications. In continuum mechanics, the feasible set \mathcal{U}_k is usually called the *kinetically admissible space*. In statics, where the function $V(v)$ is viewed as an internal (or stored) energy and $U(u)$ is considered as an external energy, the cost function $P(u)$ is the so-called *total potential* and (\mathcal{P}) represents a minimal potential variational problem. In dynamical systems if $V(v)$ is considered as a kinetic energy and $U(u)$ is the total potential, then $P(u)$ is called the *total action* of the system. In this case, the variational problem associated with the general form (\mathcal{P}) is the well-known *least action principle*. A diagrammatic representation of this tri-canonical decomposition is shown in Figure 1.1.

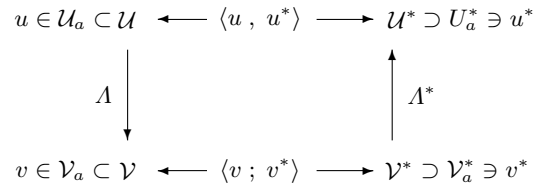


Fig. 1.1. Diagrammatic representation for quadratic systems

The development of the Λ^*DA -operator theory was apparently initiated by von Neumann in 1932, and was subsequently extended and put into a more general setting in the studies of complementary variational principles in continuum mechanics by Noble (1966), Rall (1969), Arthus (1970, 1980), Tonti

(1972), Oden and Reddy (1983), and Sewell (1987). In mathematical analysis, the tri-canonical form of $A = A^*DA$ has also been used to develop a mathematical theory of duality by Rockafellar (1970), Ekeland and Temam (1976), Toland (1978, 1979), Auchmuty (1983), Clark (1985), and many others. In the excellent textbook by Strang (1986), the tri-factorization $A = A^*DA$ for linear operators can be seen through an application of continuum theories to discrete systems. In what follows, we list some simple examples. More applications can be found in the monograph [40].

1.2.1 Quadratic Optimization Problems in \mathbb{R}^n

First, we consider \mathcal{U} as a finite-dimensional space such that $\mathcal{U} = \mathcal{U}^* = \mathbb{R}^n$. Thus $A : \mathcal{U} \rightarrow \mathcal{U}^*$ is a symmetric matrix in $\mathbb{R}^{n \times n}$ and the bilinear form $\langle u, u^* \rangle = u^T u^*$ is simply a dot-product in \mathbb{R}^n . By linear algebra, the canonical decomposition $A = A^*DA$ can be performed in many ways (see Strang, 1986), where $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a matrix, $D : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a symmetrical matrix, and $A^* = A^T$ maps $\mathcal{V}^* = \mathbb{R}^m$ back to $\mathcal{U}^* = \mathbb{R}^n$. The bilinear forms $\langle * , * \rangle$ and $\langle * ; * \rangle$ are simply dot products in \mathbb{R}^n and \mathbb{R}^m , respectively, that is

$$\langle Au; v^* \rangle = \sum_{i=1}^m \left(v_i^* \sum_{j=1}^n A_{ij} u_j \right) = \sum_{j=1}^n \left(u_j \sum_{i=1}^m A_{ij} v_i^* \right) = \langle u, A^T v^* \rangle.$$

If the matrix A is positive semi-definite, we can always choose a geometrical operator Λ to ensure that the matrix $D \in \mathbb{R}^{m \times m}$ is positive definite. In this case the problem (\mathcal{P}) is a convex program and any solution of the fundamental equation $Au = f$ also solves the minimization problem (\mathcal{P}) .

If the matrix A is indefinite, the quadratic function $\frac{1}{2}\langle u, Au \rangle$ is nonconvex. From linear algebra, it follows then that by choosing a particular linear operator $\Lambda : \mathbb{R}^n \rightarrow \mathbb{R}^m$, the matrix A can be written in the tri-canonical form:

$$A = (\Lambda^T, I) \begin{pmatrix} D & 0 \\ 0 & -C \end{pmatrix} \begin{pmatrix} \Lambda \\ I \end{pmatrix}, \quad (1.9)$$

where $D \in \mathbb{R}^{m \times m}$ is positive definite, $C \in \mathbb{R}^{n \times n}$ is positive semi-definite, and I is an identity in \mathbb{R}^n . In this case, both $V(v) = \frac{1}{2}\langle v; Dv \rangle$ and $U(u) = \frac{1}{2}\langle u, Cu \rangle + \langle u, f \rangle$ are convex quadratic functions, but

$$P(u) = V(Au) - U(u) = \frac{1}{2}\langle Au; DAu \rangle - \frac{1}{2}\langle u, Cu \rangle - \langle u, f \rangle$$

is a *nonconvex d.c. function*, that is a difference of convex functions. In this case, the problem (\mathcal{P}) is a nonconvex quadratic minimization and the solution of $Au = f$ is only a critical point of $P(u)$.

Nonconvex quadratic programming and d.c. programming are important from both the mathematical and application viewpoints. Sahni (1974) first showed that for a negative definite matrix A , the problem (\mathcal{P}) is NP-hard.

This result was also proved by Vavasis (1990, 1991) and by Pardalos (1991). During the last decade, several authors have shown that the general quadratic programming problem (\mathcal{P}) is an *NP-hard problem* in global optimization (cf. Murty and Kabadi, 1987; Horst *et al.*, 2000). It was shown by Pardalos and Vavasis (1990) that even when the matrix A is of rank one with exactly one negative eigenvalue, the problem is NP-hard. In order to solve this difficult problem, much effort has been devoted during the last decade. Comprehensive surveys have been given by Floudas and Visweswaran [21] for quadratic programming, and by Tuy [135] for d.c. optimization.

1.2.2 Variational Problems in Analysis

In continuous systems the linear space \mathcal{U} is usually a function space over a time-space domain, and the linear mapping A is a differential operator. In classical Newtonian dynamics, for example, the fundamental equation (1.2) is a second order differential equation

$$Au = -mu'' = f,$$

where f is an applied force field. In this case, $A = d/dt$ is a linear differential operator, $m > 0$ is a mass density, and $A^* = -d/dt$ can be defined by integrating by parts over a time domain $T \subset \mathbb{R}$ with boundary ∂T :

$$\langle Au; v^* \rangle = \int_T u'v^* dt = \int_T u(-v^*)' dt = \langle u, A^*v^* \rangle,$$

subject to the boundary conditions $u(t)v^*(t) = 0, \forall t \in \partial T$.

For Newton's law, $D = m$ is a constant and the tri-canonical form $Au = A^*DAu = -mu'' = f$ is Newton's equilibrium equation. The quadratic form

$$V(Au) = \frac{1}{2}\langle u, Au \rangle = \frac{1}{2}\langle Au; DAu \rangle = \frac{1}{2} \int_T mu'^2 dt$$

represents the internal (or kinetic) energy of the system, and the linear term

$$U(u) = \int_T uf dt$$

represents the external energy of the system. The function $P(u) = V(Au) - U(u)$ is called the total action, which is a convex functional.

For Einstein's law, however, $D = m(t) = m_o/\sqrt{1 - c^2/v^2}$ depends on the velocity $v = u'$, where $m_o > 0$ is a constant and c is the speed of light. In this case, the tri-canonical form $Au = f$ leads to Einstein's theory of special relativity:

$$-\frac{d}{dt} \left(\frac{m_o}{\sqrt{1 - u'^2/c^2}} \frac{d}{dt} u \right) = f.$$

The kinetic energy

$$V(v) = \int_T -m_o \sqrt{1 - v^2/c^2} dt$$

is no longer quadratic, but is still a convex functional on $\mathcal{V}_a = \{v \in \mathcal{L}^\infty(T) \mid v(t) < c, \forall t \in T\}$. By using the canonical dual transformation, the nonlinear minimization problem (\mathcal{P}) can be solved analytically (see [41]).

In mass-spring systems, $A = -(m\partial_{tt} + k)$ and the fundamental equation (1.2) has the form:

$$Au = -mu'' - ku = f.$$

The additional term ku represents the spring force and $k > 0$ is a spring constant. In this case, if we let $\Lambda = (\partial_t, 1)^T$ be a vector-valued operator, the second order linear differential operator A can still be written in the $\Lambda^*D\Lambda$ form as

$$A = -\left(m\frac{\partial^2}{\partial t^2} + k\right) = \begin{bmatrix} -\frac{\partial}{\partial t} & 1 \end{bmatrix} \begin{bmatrix} m & 0 \\ 0 & -k \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial t} \\ 1 \end{bmatrix}. \quad (1.10)$$

As evident here, if we let $\Lambda = (\partial_t, 1)^T$ be a vector-valued operator, the operator D is indefinite. However, if we let $\Lambda = \partial_t$, then similar to (1.9), we have $D = m$, which is positive definite. Thus in this dynamical system, we have.

$$V(v) = \int_T \frac{1}{2}mv^2 dt, \quad U(u) = \int_T \left(\frac{1}{2}ku^2 - uf\right) dt,$$

where the quadratic function $U(u)$ represents the total potential energy. The quadratic functional given by

$$P(u) = V(\Lambda u) - U(u) = \int_T \frac{1}{2}mu_{,t}^2 dt - \int_T \left[\frac{1}{2}ku^2 - uf\right] dt \quad (1.11)$$

is the well-known *total action*, which is again a d.c. functional.

Actually, every function $P(u) \in \mathcal{C}^2$ is d.c. on any compact convex set \mathcal{U}_k , and any d.c. optimization problem can be reduced to the canonical form (see Tuy, 1995):

$$\min\{V(\Lambda u) : U(u) \leq 0, G(u) \geq 0\}, \quad (1.12)$$

where V , U , and G are convex functions. In the next section, we demonstrate how the tri-canonical $\Lambda^*D\Lambda$ -operator theory serves as a framework for the Lagrangian duality theory.

1.3 Canonical Lagrangian Duality Theory

Classical Lagrangian duality was originally studied by Lagrange in analytical mechanics. In engineering mechanics it has been recognized as the *complementary variational principle*, and has been subjected to extensive study for more than several centuries. In this section, we show its connection to constrained optimization/variational problems. In addition to the well-known

saddle Lagrangian duality theory, a so-called super-Lagrangian duality is presented within a unified framework, which leads to a bi-duality theorem in d.c. programming and convex Hamiltonian systems.

Recall the general primal problem (1.8)

$$(\mathcal{P}) : \min\{P(u) = V(\Lambda u) - U(u) : u \in \mathcal{U}_k\}, \quad (1.13)$$

where $V : \mathcal{V}_a \subset \mathcal{V} \rightarrow \mathbb{R}$ is a canonical function, $U : \mathcal{U}_a \rightarrow \mathbb{R}$ is a Gâteaux differentiable function, either linear or canonical, and $\mathcal{U}_k = \{u \in \mathcal{U}_a \mid \Lambda u \in \mathcal{V}_a\}$ is a convex feasible set. Without loss of generality, we assume that the geometrical operator $\Lambda : \mathcal{U}_a \rightarrow \mathcal{V}$ can be chosen in a way such that the canonical function $V : \mathcal{V}_a \rightarrow \mathbb{R}$ is convex. By the definition of the canonical function, the duality relation $v^* = \delta V(v) : \mathcal{V}_a \rightarrow \mathcal{V}_a^*$ leads to the following Fenchel-Young equality on $\mathcal{V}_a \times \mathcal{V}_a^*$,

$$V(v) = \langle v; v^* \rangle - V^*(v^*).$$

Substituting this into equation (1.13), the Lagrangian $L(u, v^*) : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ associated with the canonical problem (\mathcal{P}) can be defined by

$$L(u, v^*) = \langle \Lambda u; v^* \rangle - V^*(v^*) - U(u). \quad (1.14)$$

Definition 2 (Canonical Lagrangian). *A function $L : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ associated with the problem (\mathcal{P}) is called a canonical Lagrangian if it is a canonical function on \mathcal{V}_a^* and a canonical or linear function on \mathcal{U}_a .*

The criticality condition $\delta L(\bar{u}, \bar{v}^*) = 0$ leads to the well-known *Lagrange equations*:

$$\begin{aligned} \Lambda \bar{u} &= \delta V^*(\bar{v}^*) \\ \Lambda^* \bar{v}^* &= \delta U(\bar{u}). \end{aligned} \quad (1.15)$$

By the fact that $V : \mathcal{V}_a \rightarrow \mathbb{R}$ is a canonical function, the Lagrange equations (1.15) are equivalent to $\Lambda^* \delta V(\Lambda \bar{u}) = \delta U(\bar{u})$. If (\bar{u}, \bar{v}^*) is a critical point of $L(u, v^*)$, then \bar{u} is a critical point of $P(u)$ on \mathcal{U}_k .

Because the canonical function V is assumed to be convex on \mathcal{V}_a , the canonical Lagrangian $L(u, v^*)$ is concave on \mathcal{V}_a^* . Thus, the extremality conditions of the critical point of $L(u, v^*)$ depends on the convexity of the function $U(u)$. Two important duality theories are associated with the canonical Lagrangian, as exposed in Sections 3.1 and 3.2 below.

1.3.1 Saddle-Lagrangian duality

First, we assume that $U(u)$ is a concave function on \mathcal{U}_a . In this case, $L(u, v^*)$ is a saddle-Lagrangian, that is $L(u, v^*)$ is convex on \mathcal{U}_a and concave on \mathcal{V}_a^* . By the traditional definition, a pair (\bar{u}, \bar{v}^*) is called a *saddle point* of $L(u, v^*)$ on $\mathcal{U}_a \times \mathcal{V}_a^*$ if

$$L(u, \bar{v}^*) \geq L(\bar{u}, \bar{v}^*) \geq L(\bar{u}, v^*), \quad \forall (u, v^*) \in \mathcal{U}_a \times \mathcal{V}_a^*. \quad (1.16)$$

The classical saddle-Lagrangian duality theory can be presented precisely by the following theorem.

Theorem 1 (Saddle-Minmax Theorem) *Suppose that the function $U : \mathcal{U}_a \rightarrow \mathbb{R}$ is concave and there exists a linear operator $\Lambda : \mathcal{U}_a \rightarrow \mathcal{V}_a$ such that the canonical Lagrangian $L : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ is a saddle function. If $(\bar{u}, \bar{v}^*) \in \mathcal{U}_a \times \mathcal{V}_a^*$ is a critical point of $L(u, v^*)$, then*

$$\min_{u \in \mathcal{U}_k} \max_{v^* \in \mathcal{V}_a^*} L(u, v^*) = L(\bar{u}, \bar{v}^*) = \max_{v^* \in \mathcal{V}_k^*} \min_{u \in \mathcal{U}_a} L(u, v^*). \quad (1.17)$$

By using this theorem, the dual function $P^d(v^*)$ can be defined as

$$P^d(v^*) = \min_{u \in \mathcal{U}_a} L(u, v^*) = U^b(\Lambda^* v^*) - V^*(v^*), \quad (1.18)$$

where $U^b : \mathcal{U}^* \rightarrow \mathbb{R}$ is a Fenchel conjugate function of U defined by the Fenchel transformation

$$U^b(u^*) = \min_{u \in \mathcal{U}_a} \{\langle u, u^* \rangle - U(u)\}. \quad (1.19)$$

Because $U(u)$ is a concave function on \mathcal{U}_a , the Fenchel conjugate U^b is also a concave function on $\mathcal{U}_a^* \subset \mathcal{U}^*$. Thus, on the dual feasible space \mathcal{V}_k^* defined by

$$\mathcal{V}_k^* = \{v^* \in \mathcal{V}_a^* \mid \Lambda^* v^* \in \mathcal{U}_a^*\}, \quad (1.20)$$

the problem, which is dual to (\mathcal{P}) , can be proposed as the following,

$$(\mathcal{P}^d) : \quad \max \{P^d(v^*) : v^* \in \mathcal{V}_k^*\}. \quad (1.21)$$

The saddle min-max duality theory leads to the following well-known result.

Theorem 2 (Saddle-Lagrangian Duality Theorem)

Suppose that $L(u, v^) : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ is a canonical saddle Lagrangian and (\bar{u}, \bar{v}^*) is a critical point of $L(u, v^*)$. Then \bar{u} is a global minimizer of $P(u)$, \bar{v}^* is a global maximizer of $P^d(v^*)$, and*

$$\min_{u \in \mathcal{U}_k} P(u) = P(\bar{u}) = L(\bar{u}, \bar{v}^*) = P^d(\bar{v}^*) = \max_{v^* \in \mathcal{V}_k^*} P^d(v^*). \quad (1.22)$$

Particularly, for a given $f \in \mathcal{U}_a^*$ such that $U(u) = \langle u, f \rangle$ is a linear function on \mathcal{U}_a , the Fenchel-conjugate $U^b(u^*)$ can be computed as

$$U^b(u^*) = \min_{u \in \mathcal{U}_a} \{\langle u, u^* \rangle - U(u)\} = \begin{cases} 0 & \text{if } u^* = f, \\ -\infty & \text{otherwise.} \end{cases} \quad (1.23)$$

Its effective domain is $\mathcal{U}_a^* = \{u^* \in \mathcal{U}^* \mid u^* = f\}$. Thus, the dual feasible space can be well-defined as $\mathcal{V}_k^* = \{v^* \in \mathcal{V}_a^* \mid \Lambda^* v^* = f\}$, and the dual problem is a concave maximization problem with a linear constraint:

$$(\mathcal{P}^d) : \max\{P^d(v^*) = -V^*(v^*) : \Lambda^*v^* = f, v^* \in \mathcal{V}_a^*\}. \quad (1.24)$$

By using the Lagrange multiplier $u \in \mathcal{U}_a$ to relax the linear constraint, we have

$$L(u, v^*) = -V^*(v^*) + \langle u, (\Lambda^*v^* - f) \rangle,$$

which is exactly the canonical Lagrangian (1.14) associated with the problem (\mathcal{P}) if the Lagrange multiplier u is in \mathcal{U}_a such that $V(\Lambda u)$ is a canonical function on \mathcal{V}_a . This shows that the classical Lagrangian can be obtained in two ways:

1. Legendre transformation method (by choosing a proper linear operator Λ in (\mathcal{P}))
2. Classical Lagrange multiplier method (by relaxing the constraint $\Lambda^*v^* = f$ in (\mathcal{P}^d))

In engineering mechanics, because V^* is called the complementary energy, the constrained problem

$$\min\{V^*(v^*) : \Lambda^*v^* = f, v^* \in \mathcal{V}_a^*\}$$

is also called the *complementary variational problem* and the Lagrangian $L(u, v^*)$ is called the *generalized complementary energy*. In computational mechanics, the saddle-Lagrangian duality theory serves as a foundation for mixed and hybrid finite element methods.

1.3.2 Super-Lagrangian Duality

If the function $U : \mathcal{U}_a \rightarrow \mathbb{R}$ is convex, the canonical Lagrangian $L(u, v^*)$ is concave in each of its variables $u \in \mathcal{U}_a$ and $v^* \in \mathcal{V}_a^*$. However, $L(u, v^*)$ may not be concave in $(u, v^*) \in \mathcal{U}_a \times \mathcal{V}_a^*$ (see examples in [40]). In this case, consider the following definition that was introduced in [40].

Definition 3. A point (\bar{u}, \bar{v}^*) is said to be a super-critical (or ∂^+ -critical) point of L on $\mathcal{U}_a \times \mathcal{V}_a^*$ if

$$L(\bar{u}, v^*) \leq L(\bar{u}, \bar{v}^*) \geq L(u, \bar{v}^*), \quad \forall (u, v^*) \in \mathcal{U}_a \times \mathcal{V}_a^*. \quad (1.25)$$

A function $L : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ is said to be a super-critical (or ∂^+) function on $\mathcal{U}_a \times \mathcal{V}_a^*$ if it is concave in each of its arguments, that is

$$\begin{aligned} L : \mathcal{U}_a &\rightarrow \mathbb{R} \text{ is concave, } \forall v^* \in \mathcal{V}_a^*, \\ L : \mathcal{V}_a^* &\rightarrow \mathbb{R} \text{ is concave, } \forall u \in \mathcal{U}_a. \end{aligned}$$

In particular, if the super-critical function $L : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ is a Lagrange form, it is called a super-Lagrangian.

From a duality viewpoint, a point (\bar{u}, \bar{v}^*) is said to be a *sub-critical* (or ∂^- -critical) point of L on $\mathcal{U}_a \times \mathcal{V}_a^*$ if

$$L(\bar{u}, v^*) \geq L(\bar{u}, \bar{v}^*) \leq L(u, \bar{v}^*), \quad \forall (u, v^*) \in \mathcal{U}_a \times \mathcal{V}_a^*. \quad (1.26)$$

This definition comes from the sub-differential (see [40]):

$$\bar{v}^* \in \partial^- V(v) = \{v^* \in \mathcal{V}_a^* \mid V(v) - V(\bar{v}) \geq \langle v - \bar{v}; \bar{v}^* \rangle, \quad \forall v \in \mathcal{V}_a\}.$$

Clearly, (\bar{u}, \bar{v}^*) is a super-critical point of L on $\mathcal{U}_a \times \mathcal{V}_a^*$ if and only if it is a sub-critical point of $-L$ on $\mathcal{U}_a \times \mathcal{V}_a^*$.

Theorem 3 (Super-Lagrangian Duality Theorem [40])

Suppose that there exists a linear operator $\Lambda : \mathcal{U}_a \rightarrow \mathcal{V}_a$ such that $L : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ is a super-Lagrangian. If $(\bar{u}, \bar{v}^*) \in \mathcal{U}_a \times \mathcal{V}_a^*$ is a super-critical point of $L(u, v^*)$ on $\mathcal{U}_a \times \mathcal{V}_a^*$, then either the super-maximum theorem in the form

$$\max_{u \in \mathcal{U}_k} \max_{v^* \in \mathcal{V}_a^*} L(u, v^*) = L(\bar{u}, \bar{v}^*) = \max_{v^* \in \mathcal{V}_k^*} \max_{u \in \mathcal{U}_a} L(u, v^*) \quad (1.27)$$

holds, or the super-minimax theorem in the form

$$\min_{u \in \mathcal{U}_k} \max_{v^* \in \mathcal{V}_a^*} L(u, v^*) = L(\bar{u}, \bar{v}^*) = \min_{v^* \in \mathcal{V}_k^*} \max_{u \in \mathcal{U}_a} L(u, v^*) \quad (1.28)$$

holds.

Based on this super-Lagrangian duality theorem, a dual function to the nonconvex d.c. function $P(u) = V(\Lambda u) - U(u)$ can be formulated as

$$P^d(v^*) = \max_{u \in \mathcal{U}_a} L(u, v^*) = U^\sharp(\Lambda^* v^*) - V^*(v^*), \quad (1.29)$$

where $U^\sharp : \mathcal{V}^* \rightarrow \mathbb{R}$ is defined by the super-Fenchel transformation

$$U^\sharp(u^*) = \max\{\langle u, u^* \rangle - U(u) \quad : \quad u \in \mathcal{U}_a\}. \quad (1.30)$$

Suppose that $\mathcal{U}_a^* \subset \mathcal{U}^*$ is an effective domain of U^\sharp . Then on the dual feasible space $\mathcal{V}_k^* = \{v^* \in \mathcal{V}_a^* \mid \Lambda^* v^* \in \mathcal{U}_a^*\}$, we have the following result.

Theorem 4 (Bi-Duality Theory [40]) If (\bar{u}, \bar{v}^*) is a super-critical point of $L(u, v^*)$, then either the double-min theorem in the form

$$\min_{u \in \mathcal{U}_k} P(u) = P(\bar{u}) = L(\bar{u}, \bar{v}^*) = P^d(\bar{v}^*) = \min_{v^* \in \mathcal{V}_k^*} P^d(v^*) \quad (1.31)$$

holds, or the double-max theorem in the form

$$\max_{u \in \mathcal{U}_k} P(u) = P(\bar{u}) = L(\bar{u}, \bar{v}^*) = P^d(\bar{v}^*) = \max_{v^* \in \mathcal{V}_k^*} P^d(v^*) \quad (1.32)$$

holds.

The Hamiltonian $H : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ associated with the Lagrangian is defined by

$$H(u, v^*) = \langle \Lambda u; v^* \rangle - L(u, v^*) = V^*(v^*) + U(u). \quad (1.33)$$

Clearly, if $L(u, v^*)$ is a super-Lagrangian, the Hamiltonian $H(u, v^*)$ is convex in each of its variables and in terms of $H(u, v^*)$, the Lagrange equations (1.15) can be written in the so-called *Hamiltonian canonical form*:

$$\Lambda u = \delta_{v^*} H(u, v^*), \quad \Lambda^* v^* = \delta_u H(u, v^*). \quad (1.34)$$

However, this nice symmetrical form and the convexity of the Hamiltonian does not afford new insights into understanding the extremality conditions of the nonconvex problem. The super-Lagrangian duality theory plays an important role in d.c. programming, convex Hamilton systems, and global optimization.

1.3.3 Applications in Quadratic Programming and Commentary

Now, let us consider the nonconvex quadratic programming problem (\mathcal{P}_q) where the cost function is a d.c. function

$$P(u) = \frac{1}{2} \langle \Lambda u; D \Lambda u \rangle - \frac{1}{2} \langle u, C u \rangle - \langle u, f \rangle$$

as discussed in (1.2.1), where D is a positive definite matrix in $\mathbb{R}^{m \times m}$, and $C \in \mathbb{R}^{n \times n}$ is positive semi-definite. Because $U(u) = \frac{1}{2} \langle u, C u \rangle + \langle u, f \rangle$ in this case is convex, the Lagrangian

$$L(u, v^*) = \langle \Lambda u; v^* \rangle - \frac{1}{2} \langle D^{-1} v^*; v^* \rangle - \frac{1}{2} \langle u, C u \rangle - \langle u, f \rangle$$

is a super-Lagrangian. By using the super-Fenchel transformation, we have

$$\begin{aligned} U^\sharp(u) &= \max_{u \in \mathcal{U}_a} \{ \langle u, u^* - f \rangle - \frac{1}{2} \langle u, C u \rangle \} \\ &= \frac{1}{2} \langle C^+(u^* - f), (u^* - f) \rangle, \end{aligned}$$

subject to $u^* - f \in \mathcal{C}(C)$, where C^+ is a pseudo-inverse of C and $\mathcal{C}(C)$ represents the column space of C . Thus, on the dual feasible space

$$\mathcal{V}_k^* = \{ v^* \in \mathcal{V}_a \subset \mathbb{R}^m \mid \Lambda^T v^* - f \in \mathcal{C}(C) \}, \quad (1.35)$$

the dual function

$$P^d(v^*) = \frac{1}{2} \langle C^+(\Lambda^* v^* - f), \Lambda^* v^* - f \rangle - \frac{1}{2} \langle D^{-1} v^*; v^* \rangle \quad (1.36)$$

is also a d.c. function. The bi-duality theorem shows that the optimal values of the primal and dual problems are equal. If \bar{u} solves the primal (either minimization or maximization) and $\Lambda^* \bar{v}^* - f \in \partial^- U(\bar{u})$, then \bar{v}^* solves the dual.

One of the earliest and best known double-min duality schemes was formulated by Toland [131] for the d.c. minimization problem

$$\min\{W(u) - U(u) : u \in \text{dom } W\}, \quad (1.37)$$

where $W(u)$ is an arbitrary function, $U(u)$ is a convex proper lsc function on \mathbb{R}^n , and $\text{dom } W$ represents effective domain of W . The dual problem is

$$\min\{U^\sharp(u^*) - W^\sharp(u^*) : u^* \in \text{dom } U^\sharp\}, \quad (1.38)$$

which is also a d.c. minimization problem in \mathbb{R}^n . The generalizations were made by Auchmuty (1983) to general nonconvex functionals with a linear operator Λ . Since then, several important duality concepts have been developed and studied for nonconvex optimization and d.c. programming by Crouzeix (1981), Hiriart-Urruty (1985), Singer (1988), Penot & Volle (1990), Tuy (1990, 1991, 1992), Tuach (1993, 1995), and many others. A detailed review on duality in d.c. programming appears in Tuy [135]. Much of the foregoing discussion is based on generalized nonconvex functionals, which are allowed to be extended-real-valued. In order to avoid difficulties such as $\infty - \infty$, a modified version of the double-min duality in optimization was presented in Rockafellar and Wets [110]. It is traditional in the calculus of variations and optimization that the primal problem is always taken to be a minimization problem. However, this tradition somewhat obscures our view of more general problems. In convex Hamiltonian systems where $V(v) = \frac{1}{2}\langle \Lambda v, D\Lambda v \rangle$ is a kinetic energy function and $U(u) = \frac{1}{2}\langle u, Cu \rangle + \langle u, f \rangle$ is a total potential energy function, the d.c. function $P(u) = V(\Lambda u) - U(u)$ represents a total action of the system. As pointed out in [16, 40], in the context of dynamical systems, the least action principle is somehow misleading because the action is a d.c. function that takes minimum and maximum values periodically over the time domain. Both the min- and the max-primal problems have to be considered simultaneously in a period. The bi-duality theorem reveals a periodic behavior of dynamical systems.

In two-person game theory, the bi-duality theory shows that the d.c. programming problem has two Nash equilibrium points.

The super-Lagrangian duality and the associated bi-duality theory were first proposed in the monograph [40]. Based on this theory and the tri-canonical form $\Lambda^* D\Lambda$, we reformulated the nonconvex quadratic programming problem in a dual form of (1.36), which is well-defined on the dual feasible space $\mathcal{V}_k^* \subset \mathbb{R}^m$ (1.35). Because $m \leq n$, we believe this new dual form will play an important role in nonconvex quadratic programming theory.

1.4 Complementary Variational Principles in Continuum Mechanics

This section presents two simple applications of the canonical Lagrange duality theory in continuum mechanics. The first application shows the connection between the mathematical theory of saddle-Lagrangian duality and the complementary energy variational principles in static linear elasticity, which are well-known in solid mechanics and computational mechanics. Indeed, the application of the super-Lagrangian duality theory to convex Hamiltonian systems may bring some important insights into extremality conditions in dynamic systems.

1.4.1 Linear Elasticity

Let us consider an elastic material in \mathbb{R}^3 occupying a simple connected domain $\Omega \subset \mathbb{R}^3$ with boundary $\Gamma = \partial\Omega = \Gamma_u \cup \Gamma_t$ such that $\Gamma_u \cap \Gamma_t = \emptyset$. On Γ_u , the boundary displacement \bar{u} is given, whereas on Γ_t , a surface traction $\bar{\mathbf{t}}$ is prescribed. Suppose that the elastic body is subjected to a distributed force field f . The equilibrium equation $Au = f$ has the following form,

$$-\frac{\partial}{\partial x_j} \left(D_{ijkl} \frac{\partial u_k(x)}{\partial x_l} \right) = f_i(x), \quad \forall x \in \Omega, \quad (1.39)$$

where $D = \{D_{ijkl}\}$ ($i, j, k, l = 1, 2, 3$) is a positive definite fourth-order elastic tensor, satisfying $D_{ijkl} = D_{jikl} = D_{klij}$, and Einstein's summation convention over the repeated sub-indices is used here. In this problem, $A = -\text{div } D \text{ grad}$ is an elliptic operator, $A = \text{grad}$ is a gradient, and $v = \text{grad}u$ is called the deformation gradient. Its symmetrical part is an infinitesimal strain tensor, denoted as $\epsilon = \frac{1}{2}(\nabla u + (\nabla u)^T)$. The dual variable $v^* = D\epsilon$ is a stress tensor, usually denoted by σ . In this infinite dimensional system $\mathcal{U} = \mathcal{L}^2(\Omega; \mathbb{R}^3) = \mathcal{U}^*$ and $\mathcal{V} = \mathcal{L}^2(\Omega; \mathbb{R}^{3 \times 3}) = \mathcal{V}^*$. The bilinear forms are defined by

$$\langle u, f \rangle = \int_{\Omega} u \cdot f \, d\Omega, \quad \langle \epsilon, \sigma \rangle = \int_{\Omega} \epsilon : \sigma \, d\Omega,$$

where $\epsilon : \sigma = \text{tr}(\epsilon \cdot \sigma) = \epsilon_{ij}\sigma_{ij}$. The adjoint operator A^* in this case is $A^* = \{-\text{div in } \Omega, \mathbf{n} \cdot \text{ on } \Gamma\}$, and *dive* is also called the formal adjoint of $A = \text{grad}$. Let

$$\begin{aligned} \mathcal{U}_a &= \{u \in \mathcal{U} \mid u(x) = \bar{u}(x), \quad \forall x \in \Gamma_u\} \\ \mathcal{V}_a &= \{\epsilon \in \mathcal{V} \mid \epsilon(x) = \epsilon^T(x), \quad \forall x \in \Omega\}. \end{aligned}$$

Thus on the feasible space, i.e., the so-called *statically admissible space* $\mathcal{U}_k = \{u \in \mathcal{U}_a \mid Au \in \mathcal{V}_a\}$, the quadratic form

$$P(u) = \int_{\Omega} \frac{1}{2} (\nabla u) : D : (\nabla u) \, d\Omega - \int_{\Gamma_t} u \cdot f \, d\Gamma \quad (1.40)$$

is the so-called *total potential* of the deformed elastic body. The *minimal potential principle* leads to the convex variational problem

$$\min \{P(u) : u \in \mathcal{U}_k\}. \quad (1.41)$$

The functional $V(\boldsymbol{\epsilon}) = \frac{1}{2}\langle \boldsymbol{\epsilon}; D\boldsymbol{\epsilon} \rangle$ is call the *internal (or stored) potential*. Its Legendre conjugate

$$V^*(\boldsymbol{\sigma}) = \{ \langle \boldsymbol{\epsilon}, \boldsymbol{\sigma} \rangle - U(\boldsymbol{\epsilon}) \mid \boldsymbol{\sigma} = D : \boldsymbol{\epsilon} \} = \int_{\Omega} \frac{1}{2} \boldsymbol{\sigma} : D^{-1} : \boldsymbol{\sigma} \, d\Omega$$

is known as the *complementary energy* in solid mechanics. Because

$$U(u) = \int_{\Omega} u \cdot f \, d\Omega + \int_{\Gamma_u} u \cdot \bar{\mathbf{t}} \, d\Gamma$$

is linear, which is also called the *external potential*, the Lagrangian associated with the total potential $P(u)$, as given by

$$L(u, \boldsymbol{\sigma}) = \int_{\Omega} [(\nabla u) : \boldsymbol{\sigma} - \frac{1}{2} \boldsymbol{\sigma} : D^{-1} : \boldsymbol{\sigma}] \, d\Omega - \int_{\Gamma_u} u \cdot \bar{\mathbf{t}} \, d\Gamma, \quad (1.42)$$

can be considered as a saddle Lagrangian, which is the well-known generalized *Hellinger-Reissner complementary energy*. Thus, by the saddle Lagrangian duality, the dual functional $P^d(\boldsymbol{\sigma})$ is defined by

$$P^d(\boldsymbol{\sigma}) = \min_{u \in \mathcal{U}_a} L(u, \boldsymbol{\sigma}) = U^b(\Lambda^* \boldsymbol{\sigma}) - V^*(\boldsymbol{\sigma}),$$

where

$$\begin{aligned} U^b(\Lambda^* \boldsymbol{\sigma}) &= \min_u \left\{ \int_{\Omega} (\nabla u) : \boldsymbol{\sigma} \, d\Omega - \int_{\Omega} u \cdot f \, d\Omega - \int_{\Gamma_t} u \cdot \bar{\mathbf{t}} \, d\Gamma \right\} \\ &= \begin{cases} \int_{\Gamma_u} \bar{\mathbf{u}} \cdot \boldsymbol{\sigma} \cdot \mathbf{n} \, d\Gamma & \text{if } -\operatorname{div} \boldsymbol{\sigma} = 0 \text{ in } \Omega, \boldsymbol{\sigma} \cdot \mathbf{n} = \bar{\mathbf{t}} \text{ on } \Gamma_t, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, on the dual feasible space, that is the so-called *statically admissible space* defined by

$$\mathcal{V}_k^* = \{ \boldsymbol{\sigma} \in \mathcal{V}_a^* \mid -\operatorname{div} \boldsymbol{\sigma} = 0 \text{ in } \Omega, \boldsymbol{\sigma} \cdot \mathbf{n} = \bar{\mathbf{t}} \text{ on } \Gamma_t \},$$

the dual problem for this linear elasticity case is given by

$$\max \left\{ P^d(\boldsymbol{\sigma}) = \int_{\Gamma_u} \bar{\mathbf{u}} \cdot \boldsymbol{\sigma} \cdot \mathbf{n} \, d\Gamma - \int_{\Omega} \frac{1}{2} \boldsymbol{\sigma} : D^{-1} : \boldsymbol{\sigma} \, d\Omega : \boldsymbol{\sigma} \in \mathcal{V}_k^* \right\}. \quad (1.43)$$

This is a concave maximization problem with linear constraints. The Lagrange multiplier u for the equilibrium constraints is the solution of the primal problem.

In continuum mechanics, the functional $-P^d$, denoted by

$$P^c(\boldsymbol{\sigma}) = \int_{\Omega} \frac{1}{2} \boldsymbol{\sigma} : D^{-1} : \boldsymbol{\sigma} \, d\Omega - \int_{\Gamma_u} \bar{u} \cdot \boldsymbol{\sigma} \cdot \mathbf{n} \, d\Gamma,$$

is called the total complementary energy. Thus, instead of the dual problem (1.43), the minimum complementary variational problem

$$\min \{P^c(\boldsymbol{\sigma}) : \boldsymbol{\sigma} \in \mathcal{V}_k^*\}$$

has been extensively studied by engineers, which serves as a foundation for the so-called *stress, or equilibrium, finite element methods*.

1.4.2 Convex Hamiltonian Systems

Recall the mass-spring dynamical system discussed in Section 2, where the total action is a d.c. function of the form

$$P(u) = V(\Lambda u) - U(u) = \int_T \frac{1}{2} m(u,t)^2 \, dt - \int_T [\frac{1}{2} k u^2 - u f] \, dt. \quad (1.44)$$

Because the Lagrangian

$$L(u, p) = \int_T [u,t p - \frac{1}{2} m^{-1} p^2 - \frac{1}{2} k u^2] \, dt - \int_T u f \, dt$$

is not a saddle function, thus the Hamiltonian

$$\begin{aligned} H(u, p) &= \langle \Lambda u, p \rangle - L(u, p) \\ &= \int_T [\frac{1}{2} m^{-1} p^2 + \frac{1}{2} k u^2] \, dt + \int_T u f \, dt \end{aligned} \quad (1.45)$$

was extensively used in classical dynamical systems. One of the main reasons for this could be that $H(u, p)$ is convex. Thus, the original differential equation $Au = -m u_{,tt} - k u = f$ can be written in the well-known *Hamiltonian canonical form*:

$$\Lambda u = \delta_p H(u, p), \quad \Lambda^* p = \delta_u H(u, p). \quad (1.46)$$

However, an important phenomenon has been hiding in the shadow of this convex Hamiltonian for centuries. Because $L(u, p)$ is a super-Lagrangian, the dual action can be formulated as

$$P^d(p) = \max_u L(u, p) = \frac{1}{2} \int_T k^{-1} (p_{,t} - f)^2 \, dt - \frac{1}{2} \int_T m^{-1} p^2 \, dt,$$

which is also a d.c. functional. The bi-duality theory

$$\begin{aligned} \min P(u) &= \min P^d(p), \\ \max P(u) &= \max P^d(p) \end{aligned}$$

shows that the well-known *least action* principle in periodic dynamical systems is actually a misnomer, that is the periodic solution $u(t)$ does not minimize the total action $P(u)$, which could be either a minimizer or a maximizer, depending on the time period (see [40]).

1.5 Nonconvex Problems with Double-Well Energy

We now turn our attention to duality theory in nonconvex systems by considering a very simple problem in \mathbb{R}^n :

$$(\mathcal{P}_w) : \min \left\{ P(u) = \frac{1}{2}\alpha \left(\frac{1}{2}|Bu|^2 - \lambda \right)^2 - \langle u, f \rangle : u \in \mathbb{R}^n \right\}, \quad (1.47)$$

where $B \in \mathbb{R}^{m \times n}$ is a matrix, $\alpha, \lambda > 0$ are positive constants, and $|v|$ denotes the Euclidean norm of v . The criticality condition $\delta P(u) = 0$ leads to a coupled nonlinear algebraic system in \mathbb{R}^n :

$$\alpha \left(\frac{1}{2}|Bu|^2 - \lambda \right) B^T B u = f. \quad (1.48)$$

Clearly, it is difficult to solve this nonlinear system by direct methods. Also, due to the nonconvexity of $P(u)$, any solution to this nonlinear system satisfies only a necessary condition. The nonconvex function $W(v) = \frac{1}{2}\alpha \left(\frac{1}{2}|v|^2 - \lambda \right)^2$

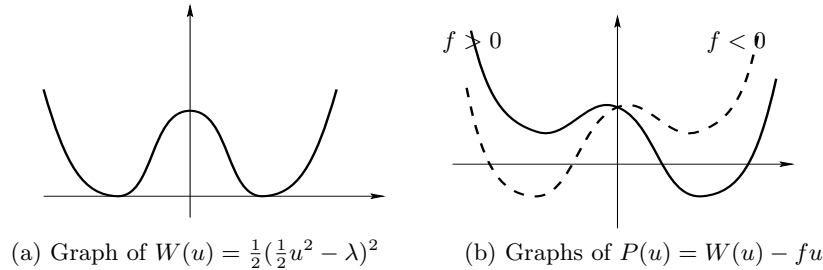


Fig. 1.2. Double-well energy and nonconvex potential functions

is a so-called *double-well energy*, which was first studied by van der Waals in fluid mechanics in 1985. For each given parameter $\lambda > 0$, $W(v)$ has two minimizers and one local maximizer (see Fig. 1.2(a)). The global and local minimizers depend on the input f (see Fig. 1.2(b)). This double-well function has extensive applications in mathematical physics. In phase transitions of shape memory alloys, or in the mathematical theory of super-conductivity, $W(v)$ is the well-known Landau second order free energy, and each of its local minimizers represents a possible phase state of the material. In quantum mechanics, if v represents the Higgs' field strength, then $W(v)$ is the energy. It was discovered in the context of post-buckling analysis of large deformed beam models, that the total potential is also a double-well energy (see Gao, [43]), and each potential well represents a possible buckled beam state. More examples can be found in a recent review article [49].

1.5.1 Classical Lagrangian and Duality Gap

If we choose $A = B$ as a linear operator, the primal function can be written in the traditional form $P(u) = W(Bu) - U(u)$, where $U(u) = \langle u, f \rangle$ is a linear function. Because the duality relation $v^* = \delta W(v) = \alpha(\frac{1}{2}|v|^2 - \lambda)v$ is not one-to-one, the Legendre conjugate

$$W^*(v^*) = \text{sta}\{\langle v, v^* \rangle - W(v) \quad : \quad v \in \mathbb{R}^m\}$$

is not uniquely defined. Thus, the entity (v, v^*) associated with the nonconvex function $W(v)$ is not a canonical dual pair. By using the Fenchel transformation

$$W^\sharp(v^*) = \max\{\langle v, v^* \rangle - W(v) \quad : \quad v \in \mathbb{R}^m\},$$

the traditional Lagrangian (associated with the linear operator $A = B$) can still be defined as

$$L(u, v^*) = \langle Bu, v^* \rangle - W^\sharp(v^*) - \langle u, f \rangle. \tag{1.49}$$

Thus, the classical Lagrangian duality theory $P^\sharp(v^*) = \max_u L(u, v^*)$ leads to the well-known Fenchel-Rockafellar dual problem

$$(\mathcal{P}^\sharp) : \max_{v^* \in \mathbb{R}^m} \{P^\sharp(v^*) = -W^\sharp(v^*) \quad : \quad B^T v^* = f\}. \tag{1.50}$$

This is a linearly constrained concave maximization problem. The Lagrange multiplier for the linear constraint set is u . However, due to the nonconvexity of $W(v)$, the Fenchel-Young inequality

$$W(v) + W^\sharp(v^*) \leq \langle v, v^* \rangle$$

leads to a weak duality relation

$$\min P \geq \max P^\sharp.$$

The non-zero value $\theta \equiv \min P(u) - \max P^\sharp(v^*)$ is called the *duality gap*. This duality gap shows that the classical Lagrange multiplier u may not be a solution to the primal problem. Thus, the Fenchel-Rockafellar duality theory can be used mainly for solving convex problems. In order to eliminate this duality gap, many modified Lagrangian dualities have been proposed during recent years (see, for examples, Rubinov *et al.* (2001 - 2005), Goh and Yang (2002), Huang and Yang (2003), Zhou and Yang (2004)). Most of these mathematical approaches are based on penalization of a class of augmented Lagrangian functions. On the other hand, the canonical duality theory addressed in the next section is based on a fundamental truth in physics, that is physical variables appear in (canonical) pairs. The one-to-one canonical duality relation leads to a perfect duality theory in mathematical physics and global optimization.

1.5.2 Canonical Dual Transformation and Triality Theory

In order to recover the duality gap, a *canonical duality theory* was developed during the last 15 years: first in nonconvex mechanics and analysis (see [65, 66, 34, 35, 40]), then in global optimization (see [40, 42, 48, 51]). The key idea of this theory is to choose a right operator (usually nonlinear) $\xi = \Lambda(u)$ such that the nonconvex function $W(u)$ can be written in the canonical form

$$W(u) = V(\Lambda(u)),$$

where $V(\xi)$ is a canonical function of $\xi = \Lambda(u)$. For the present nonconvex problem (1.47), instead of $\Lambda = B$, we choose

$$\xi = \Lambda(u) = \frac{1}{2}|Bu|^2, \quad (1.51)$$

which is a quadratic map from $\mathcal{U} = \mathbb{R}^n$ into $\mathcal{V}_a = \{\xi \in \mathbb{R} \mid \xi \geq 0\}$. Thus, the canonical function

$$V(\xi) = \frac{1}{2}\alpha(\xi - \lambda)^2$$

is simply a scale-valued quadratic function well-defined on \mathcal{V}_a , which leads to a linear duality relation

$$\varsigma = \delta V(\xi) = \alpha(\xi - \lambda).$$

Let $\mathcal{V}_a^* = \{\varsigma \in \mathbb{R} \mid \varsigma \geq -\alpha\lambda\}$ be the range of this duality mapping. So (ξ, ς) forms a canonical duality pair on $\mathcal{V}_a \times \mathcal{V}_a^*$, and the Legendre conjugate V^* is also a quadratic function:

$$V^*(\varsigma) = \text{sta} \left\{ \langle \xi; \varsigma \rangle - \frac{1}{2}\alpha(\xi - \lambda)^2 \quad : \quad \xi \in \mathcal{V}_a \right\} = \frac{1}{2}\alpha^{-1}\varsigma^2 + \lambda\varsigma.$$

Thus, replacing $W(u) = V(\Lambda(u)) = \langle \Lambda(u); \varsigma \rangle - V^*(\varsigma)$ in $P(u) = W(u) - U(u)$, the so-called *total complementary function* [65, 40] can be defined by

$$\begin{aligned} \Xi(u, \varsigma) &= \langle \Lambda(u); \varsigma \rangle - V^*(\varsigma) - U(u) \\ &= \frac{1}{2}|Bu|^2\varsigma - \frac{1}{2}\alpha^{-1}\varsigma^2 - \lambda\varsigma - u^T f. \end{aligned} \quad (1.52)$$

The criticality condition $\delta\Xi(u, \varsigma) = 0$ leads to the following canonical equilibrium equations

$$\left(\frac{1}{2}|Bu|^2 - \lambda\right) = \alpha^{-1}\varsigma, \quad (1.53)$$

$$\varsigma B^T Bu = f. \quad (1.54)$$

Equation (1.53) is actually the inverse duality relation $\xi = \delta V^*(\varsigma)$, which is equivalent to $\varsigma = \alpha(\frac{1}{2}|Bu|^2 - \lambda)$. Thus, equation (1.54) is identical to the Euler

equation (1.48). This shows that the critical point of the total complementary function is also a critical point of the primal problem. For a fixed $\varsigma \neq 0$, solving (1.54) for u gives

$$u = \frac{1}{\varsigma}(B^T B)^{-1} f. \quad (1.55)$$

Substituting this result into the total complementary function leads to the canonical dual function

$$P^d(\varsigma) = -\frac{1}{2\varsigma} f^T (B^T B)^{-1} f - \lambda\varsigma - \frac{1}{2}\alpha^{-1}\varsigma^2, \quad (1.56)$$

which is well-defined on the dual feasible space given by

$$\mathcal{V}_k^* = \{\varsigma \in \mathcal{V}_a^* \mid \varsigma \neq 0\} = \{\varsigma \in \mathbb{R} \mid \varsigma \geq -\alpha\lambda, \varsigma \neq 0\}.$$

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

$$2\varsigma^2(\alpha^{-1}\varsigma + \lambda) = f^T (B^T B)^{-1} f. \quad (1.57)$$

Theorem 5 (Gao [42]) *For any given parameters $\alpha, \lambda > 0$, and vector $f \in \mathbb{R}^n$, the canonical dual function (1.56) has at most three critical points $\bar{\varsigma}_i, (i = 1, 2, 3)$ satisfying*

$$\bar{\varsigma}_1 > 0 > \bar{\varsigma}_2 \geq \bar{\varsigma}_3. \quad (1.58)$$

For each of these roots, the vector

$$\bar{u}_i = (B^T B)^{-1} f / \bar{\varsigma}_i, \quad \text{for } i = 1, 2, 3, \quad (1.59)$$

is a critical point of the nonconvex function $P(u)$ in Problem (1.47), and we have

$$P(\bar{u}_i) = P^d(\bar{\varsigma}_i), \quad \forall i = 1, 2, 3. \quad (1.60)$$

The original version of this theorem was first discovered in a post-bifurcation problem of a large deformed beam model in 1997 [34], which shows that there is no duality gap between the nonconvex function $P(u)$ and its canonical dual $P^d(\varsigma)$. The dual algebraic equation (1.57) can be solved exactly to obtain all critical points, therefore the vector $\{\bar{u}_i\}$ defined by (1.59) yields a complete set of solutions to the nonlinear algebraic system (1.48).

Let $\tau^2 = f^T (B^T B)^{-1} f$. In algebraic geometry, the graph of the algebraic equation $\tau^2 = 2\varsigma^2(\alpha^{-1}\varsigma + \lambda)$ is the so-called *singular algebraic curve* in (ς, τ) -space (i.e., the point $\varsigma = 0$ is on the curve; cf. [119]). From this algebraic curve, we can see that there exists a constant τ_c such that if $\tau^2 > \tau_c^2$, the dual algebraic equation (1.57) has a unique solution $\varsigma > 0$. It has three real solutions if and only if $\tau^2 < \tau_c^2$.

It is interesting to note that for $\varsigma > 0$, the total complementary function $\Xi(u, \varsigma)$ is a saddle function and the well-known saddle min-max theory leads to

$$\min_u \max_{\varsigma > 0} \Xi(u, \varsigma) = \Xi(\bar{u}, \bar{\varsigma}) = \max_{\varsigma > 0} \min_u \Xi(u, \varsigma). \quad (1.61)$$

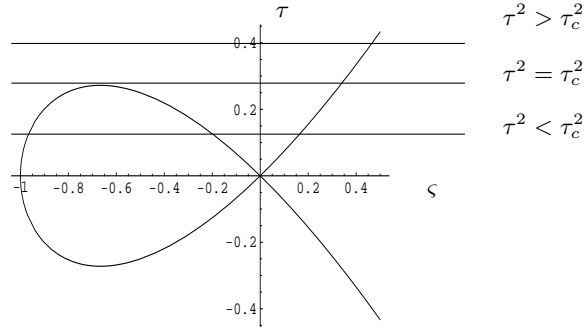


Fig. 1.3. Graph of the dual algebraic equation (1.57) and a geometrical proof of the triality theorem.

This means that \bar{u}_1 is a global minimizer of $P(u)$ and $\bar{\varsigma}_1$ is a global maximizer on the open domain $\varsigma > 0$. However, for $\varsigma < 0$, the total complementary function $\Xi(u, \varsigma)$ is concave in both u and $\varsigma < 0$, i.e., it is a super-critical function. Thus, by the bi-duality theory, we have that either

$$\min_u \max_{\varsigma < 0} \Xi(u, \varsigma) = \Xi(\bar{u}, \bar{\varsigma}) = \min_{\varsigma < 0} \max_u \Xi(u, \varsigma) \tag{1.62}$$

holds on a neighborhood of $(\bar{u}, \bar{\varsigma})$, or

$$\max_u \max_{\varsigma < 0} \Xi(u, \varsigma) = \Xi(\bar{u}, \bar{\varsigma}) = \max_{\varsigma < 0} \max_u \Xi(u, \varsigma). \tag{1.63}$$

Actually, the extremality conditions can be easily viewed through the graph of $P^d(\varsigma)$ (see Fig. 1.4). To compare with this canonical dual function, the graph of $P(u)$ for $n = 1$ is also shown in Fig. 1.4. Precisely, we have the following result (see [40, 41]).

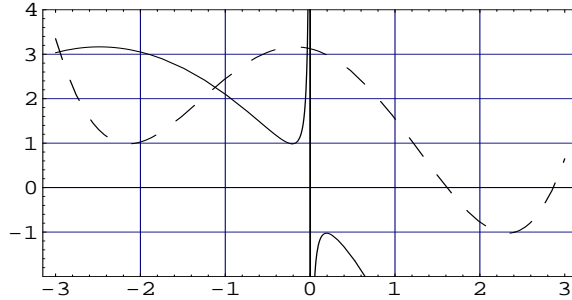


Fig. 1.4. Graphs of $P(x)$ (dashed) and $P^d(\varsigma)$ (solid) for $n = 1$.

Theorem 6 (Complete Solutions for Problem (\mathcal{P}_w) [35, 40]) For certain given parameters α , $\lambda > 0$, and the vector $f \in \mathbb{R}^n$, if $\tau^2 > \tau_c^2 = 8\alpha^2\lambda^3/27$, then the canonical dual function $P^d(\varsigma)$ has only one critical point $\bar{\varsigma} > 0$, which is a global maximizer of $P^d(\varsigma)$, and $\bar{u} = (B^T B)^{-1} f/\bar{\varsigma}$ is a global minimizer of $P(u)$.

If $\tau^2 < \tau_c^2$, the canonical dual function $P^d(\varsigma)$ has three critical points $\bar{\varsigma}_1 > 0 > \bar{\varsigma}_2 > \bar{\varsigma}_3$ such that \bar{u}_1 is a global minimizer, \bar{u}_2 is a local minimizer, and \bar{u}_3 is a local maximizer of $P(u)$.

1.5.3 Canonical Dual Solutions to Nonconvex Variational Problems

Similar to the nonconvex optimization problem (1.47) with the double-well function, let us now consider the following typical nonconvex variational problem:

$$(\mathcal{P}) : \min_{u \in \mathcal{U}_k} \left\{ P(u) = \int_0^1 \frac{1}{2} \alpha \left(\frac{1}{2} u'^2 - \lambda \right)^2 dx - \int_0^1 u f dx \right\}, \quad (1.64)$$

where $f(x)$ is a given function, $\lambda > 0$ is a parameter, and

$$\mathcal{U}_k = \{u \in \mathcal{L}^2[0, 1] \mid u' \in \mathcal{L}^4[0, 1], u(0) = 0\}$$

is an admissible space. Compared with Problem (1.47), we see that the linear operator B in this case is a differential operator $\frac{d}{dx}$. This variational problem appears frequently in association with phase transitions in fluids and solids, and in post-buckling analysis of large deformed structures. The criticality condition $\delta P(u) = 0$ leads to a nonlinear differential equation in the domain $(0, 1)$ with the natural boundary condition at $x = 1$, that is

$$\left[\alpha u' \left(\frac{1}{2} u'^2 - \lambda \right) \right]' + f(x) = 0, \quad \forall x \in (0, 1), \quad (1.65)$$

$$\alpha u' \left(\frac{1}{2} u'^2 - \lambda \right) = 0 \quad \text{at } x = 1. \quad (1.66)$$

Due to its nonlinearity, a solution to this boundary value problem is not unique. Particularly, if we let $f(x) = 0$, the equation (1.65) could have three real roots $u'(x) = \{0, \pm\sqrt{2\lambda}\}$. Thus, any zig-zag curve $u(x)$ with slope $\{0, \pm\sqrt{2\lambda}\}$ solves the boundary value problem, but may not be a global minimizer of the total energy $P(u)$. This problem shows an important fact that in nonconvex analysis: the criticality condition is only necessary, but not sufficient for solving variational problems. Traditional direct approaches for solving nonconvex variational problems are very difficult, or impossible. However, by using the canonical dual transformation, this problem can be solved completely. To see this, we introduce a new “strain measure”

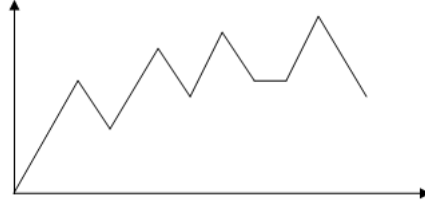


Fig. 1.5. Zig-zag function: Solution to the nonlinear boundary value problem (1.65)

$$\xi = \Lambda(u) = \frac{1}{2}u'^2,$$

such that the canonical functional

$$V(\xi) = \int_0^1 \frac{1}{2}\alpha(\xi - \lambda)^2 dx$$

is convex on $\mathcal{V}_a = \{\xi \in \mathcal{L}^2[0, 1] \mid \xi(x) \geq 0 \ \forall x \in (0, 1)\}$, and the duality relation $\varsigma = \delta V(\xi) = \alpha(\xi - \lambda)$ is one-to-one. Thus, its Legendre conjugate can be simply obtained as

$$\begin{aligned} V^*(\varsigma) &= \text{sta} \left\{ \int_0^1 \xi \varsigma dx - V(\xi) \ : \ \xi \in \mathcal{V}_a \right\} \\ &= \int_0^1 \left(\frac{1}{2}\alpha^{-1}\varsigma^2 + \lambda\varsigma \right) dx. \end{aligned}$$

Similar to (1.52), the total complementary function is

$$\Xi(u, \varsigma) = \int_0^1 \left(\frac{1}{2}u'^2\varsigma - \frac{1}{2}\alpha^{-1}\varsigma^2 - \lambda\varsigma \right) dx - \int_0^1 u f dx. \quad (1.67)$$

For a given $\varsigma \neq 0$, the canonical dual functional can be obtained as

$$P^d(\varsigma) = \text{sta}\{\Xi(u, \varsigma) \ : \ u \in \mathcal{U}_k\} = - \int_0^1 \left(\frac{\tau^2}{2\varsigma} + \lambda\varsigma + \frac{1}{2}\alpha^{-1}\varsigma^2 \right) dx, \quad (1.68)$$

where $\tau(x)$ is defined by

$$\tau = - \int_0^x f(x) dx + c, \quad (1.69)$$

and the integral constant c depends on the boundary condition. The criticality condition $\delta P^d(\varsigma) = 0$ leads to the dual equilibrium equation

$$2\varsigma^2(\alpha^{-1}\varsigma + \lambda) = \tau^2. \quad (1.70)$$

This algebraic equation is the same as (1.57), which can be solved analytically as stated below.

Theorem 7 (Analytical Solutions and Triality Theorem [35, 41])

For any given input function $f(x)$ such that $\tau(x)$ is defined by (1.69), the dual algebraic equation (1.70) has at most three real roots ς_i ($i = 1, 2, 3$) satisfying

$$\bar{\varsigma}_1(x) > 0 > \bar{\varsigma}_2(x) \geq \bar{\varsigma}_3(x).$$

For each $\bar{\varsigma}_i$, the function

$$\bar{u}_i(x) = \int_0^x \frac{\tau}{\bar{\varsigma}_i} dx \tag{1.71}$$

is a critical point of the variational problem (1.64). Moreover, $\bar{u}_1(x)$ is a global minimizer, $\bar{u}_2(x)$ is a local minimizer, and $\bar{u}_3(x)$ is a local maximizer; that is

$$P(\bar{u}_1) = \min_u \max_{\varsigma > 0} \Xi(u, \varsigma) = \max_{\varsigma > 0} \min_u \Xi(u, \varsigma) = P^d(\bar{\varsigma}_1); \tag{1.72}$$

$$P(\bar{u}_2) = \min_u \max_{\varsigma \in (\bar{\varsigma}_3, 0)} \Xi(u, \varsigma) = \min_{\varsigma \in (\bar{\varsigma}_3, 0)} \max_u \Xi(u, \varsigma) = P^d(\bar{\varsigma}_2); \tag{1.73}$$

$$P(\bar{u}_3) = \max_u \max_{\varsigma < \bar{\varsigma}_2} \Xi(u, \varsigma) = \max_{\varsigma < \bar{\varsigma}_2} \max_u \Xi(u, \varsigma) = P^d(\bar{\varsigma}_3). \tag{1.74}$$

As a complete theory, the triality theorem was first discovered in post-buckling analysis of large deformed elastic beam models [34]. The bi-duality theory was developed two years later during the writing of the monograph [40]. However, the original idea of the canonical dual transformation and the saddle-minmax theorem (1.72) were from the joint work by Gao and Strang in the study of complementary variational problems in nonconvex/nonsmooth boundary value problems [65, 66]. Theorems 5 and 7 were also first proposed in the context of continuum mechanics (see Section 7, and [37, 39, 41, 89]).

1.6 Canonical Duality Theory in General Nonconvex Systems

In this section, we discuss the canonical dual transformation and its associated triality theory for solving the following general nonconvex problem

$$(\mathcal{P}) : \min \{P(u) = W(u) - U(u) : \forall u \in \mathcal{U}_k\}, \tag{1.75}$$

where $W(u)$ is a general nonconvex function on an open set $\mathcal{U}_a \subset \mathcal{U}$ and $U : \mathcal{U}_a \rightarrow \mathbb{R}$ is a Gâteaux differentiable function, either linear or canonical, and $\mathcal{U}_k \subset \mathcal{U}_a$ is a feasible space. The canonical dual transformation for solving more general problems can be found in [35, 40, 42].

1.6.1 Canonical Dual Transformation and Framework

The key idea of the canonical dual transformation is to choose a Gâteaux differentiable *geometrical operator* $\xi = \Lambda(u) : \mathcal{U}_a \rightarrow \mathcal{V}_a$ and a canonical

function $V(\xi) : \mathcal{V}_a \rightarrow \mathbb{R}$ such that the nonconvex function $W(u)$ can be written as

$$W(u) = V(\Lambda(u)). \quad (1.76)$$

Because $V(\xi)$ is a canonical function on \mathcal{V}_a , its Legendre conjugate can be defined uniquely on $\mathcal{V}_a^* \subset \mathcal{V}^*$ by

$$V^*(\varsigma) = \text{sta}\{\langle \xi, \varsigma \rangle - V(\xi) : \forall \xi \in \mathcal{V}_a\}, \quad (1.77)$$

and on $\mathcal{V}_a \times \mathcal{V}_a^*$, we have

$$\varsigma = \delta V(\xi) \Leftrightarrow \xi = \delta V^*(\varsigma) \Leftrightarrow \langle \xi; \varsigma \rangle = V(\xi) + V^*(\varsigma). \quad (1.78)$$

Replacing $W(u)$ by $V(\Lambda(u))$ and letting $\mathcal{U}_k = \{u \in \mathcal{U}_a \mid \Lambda(u) \in \mathcal{V}_a\}$, the primal problem (\mathcal{P}) can be written in the canonical form:

$$(\mathcal{P}) : \min\{P(u) = V(\Lambda(u)) - U(u) : \forall u \in \mathcal{U}_k\}. \quad (1.79)$$

Because $\Lambda(u)$ is Gâteaux differentiable, by the chain rule we have $\delta V(\Lambda(u)) = \Lambda_t(u)\delta_\xi V(\Lambda(u))$, where $\Lambda_t(u)$ is the Gâteaux derivative of $\Lambda(u)$ and $\delta_\xi V(\Lambda(u))$ represents the Gâteaux derivative of V with respect to $\xi = \Lambda(u)$. Its adjoint $\Lambda_t^*(u)$ is defined by

$$\langle \Lambda_t(u)u; \varsigma \rangle = \langle u, \Lambda_t^*(u)\varsigma \rangle.$$

Thus, the criticality condition $\delta P(u) = 0$ leads to the *canonical equilibrium equation*

$$\Lambda_t^*(u)\delta_\xi V(\Lambda(u)) - \delta U(u) = 0. \quad (1.80)$$

In terms of the canonical duality pair (ξ, ς) , the canonical equilibrium equation (1.80) can be written in the tri-canonical forms:

$$\begin{aligned} \text{(a) Geometrical equation: } & \Lambda(u) = \xi \\ \text{(b) Constitutive equation: } & \delta V(\xi) = \varsigma \\ \text{(c) Balance equation: } & \Lambda_t^*(u)\varsigma = \delta U(u). \end{aligned} \quad (1.81)$$

In many applications, where the function $U(u)$ is usually linear on \mathcal{U}_a , the nonlinearity of the problem (\mathcal{P}) mainly depends on Λ and V . In this case, the nonlinearities of the general nonconvex problem can be classified by the following definition [40].

Definition 4 (Nonlinearity Classification). *The problem (\mathcal{P}) is said to be geometrically nonlinear if the operator $\Lambda(u)$ is nonlinear, physically nonlinear if the constitutive relation $\varsigma = \delta V(\xi)$ is nonlinear, and fully nonlinear if it is both geometrically and physically nonlinear.*

Generally speaking, the nonconvexity of $P(u)$ is mainly due to the geometrical nonlinearity. For a nonlinear operator $\Lambda(u)$, the following operator decomposition introduced by Gao and Strang [65] plays an important role in canonical duality theory

$$\Lambda(u) = \Lambda_t(u)u + \Lambda_c(u), \quad (1.82)$$

where $\Lambda_c = \Lambda(u) - \Lambda_t(u)u$ is the so-called *complementary operator* of Λ_t . By this decomposition (1.82), Gao and Strang discovered in the case where $U(u)$ is a linear function, that the duality gap existing in classical Lagrangian duality theory can be naturally recovered by the so-called *complementary gap function* defined by

$$G_c(u, \varsigma) = -\langle \Lambda_c(u); \varsigma \rangle. \quad (1.83)$$

The diagrammatic representation for a *fully nonlinear canonical system* is given in Fig. 1.6.

$$\begin{array}{ccc} u \in \mathcal{U}_a \subset \mathcal{U} & \longleftarrow \langle u, u^* \rangle & \longrightarrow \mathcal{U}^* \supset \mathcal{U}_a^* \ni u^* \\ \downarrow \Lambda_t + \Lambda_c = \Lambda & & \uparrow \Lambda_t^* = (\Lambda - \Lambda_c)^* \\ \xi \in \mathcal{V}_a \subset \mathcal{V} & \longleftarrow \langle \xi; \varsigma \rangle & \longrightarrow \mathcal{V}^* \supset \mathcal{V}_a^* \ni \varsigma \end{array}$$

Fig. 1.6. Diagrammatic representation in fully nonlinear systems.

Based on the canonical form of the primal problem (1.79), the total complementary function $\Xi : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ can be formulated as

$$\Xi(u, \varsigma) = \langle \Lambda(u); \varsigma \rangle - V^*(\varsigma) - U(u), \quad (1.84)$$

which is also called the *generalized complementary energy* in nonconvex variational problems and continuum mechanics [65, 40], or the *nonlinear Lagrangian* in global optimization [42]. For each fixed $u \in \mathcal{U}_a$, the mapping $\Xi(u, \cdot) : \mathcal{V}_a^* \rightarrow \mathbb{R}$ is a canonical function. However, the property of the mapping $\Xi(\cdot, \varsigma) : \mathcal{U}_a \rightarrow \mathbb{R}$ will depend on the geometrical operator $\Lambda(u)$. Therefore, for a given $\varsigma \in \mathcal{V}_a^*$, we introduce a new (parametric) function

$$G_\varsigma(u) := \langle \Lambda(u); \varsigma \rangle - U(u), \quad \forall u \in \mathcal{U}_a. \quad (1.85)$$

Clearly, for a fixed $\varsigma \in \mathcal{V}_a^*$, the criticality condition

$$\delta G_\varsigma(\bar{u}; u) = \langle \Lambda_t(\bar{u})u; \varsigma \rangle - \delta U(\bar{u}; u) = 0, \quad \forall u \in \mathcal{U}_a$$

leads to the balance equation $\Lambda_t^*(\bar{u})\varsigma - \delta U(\bar{u}) = 0$. This function plays an important role in canonical duality theory. By introducing the so-called *canonical dual feasible space* \mathcal{V}_k^* defined by

$$\mathcal{V}_k^* = \{\varsigma \in \mathcal{V}_a^* \mid \Lambda_t^*(u)\varsigma = \delta U(u), \quad \forall u \in \mathcal{U}_a\}, \quad (1.86)$$

the canonical dual function $P^d : \mathcal{V}_k^* \rightarrow \mathbb{R}$ can be formulated via $\Xi(u, \varsigma)$ as

$$P^d(\varsigma) = \text{sta}\{\Xi(u, \varsigma) : u \in \mathcal{U}_a\} = U^A(\varsigma) - V^*(\varsigma), \quad (1.87)$$

where $U^A : \mathcal{V}_k^* \rightarrow \mathbb{R}$ is called A -conjugate transformation of U , defined by (see [40])

$$U^A(\varsigma) = \text{sta}\{\langle A(u) ; \varsigma \rangle - U(u) : u \in \mathcal{U}_a\}. \quad (1.88)$$

Theorem 8 (Canonical Dual Transformation [40]) *The function*

$$P^d(\varsigma) = U^A(\varsigma) - V^*(\varsigma) : \mathcal{V}_k^* \rightarrow \mathbb{R}$$

is canonically dual to $P(u) = V(A(u)) - U(u) : \mathcal{U}_k \rightarrow \mathbb{R}$ in the sense that if $(\bar{u}, \bar{\varsigma})$ is a critical point of $\Xi(u, \varsigma)$, then \bar{u} is a critical point of $P(u)$, $\bar{\varsigma}$ is a critical point of $P^d(\varsigma)$, and

$$P(\bar{u}) = \Xi(\bar{u}, \bar{\varsigma}) = P^d(\bar{\varsigma}). \quad (1.89)$$

This theorem can be easily proved by examining the criticality condition $\delta\Xi(\bar{u}, \bar{\varsigma}) = 0$, which leads to the following canonical Lagrangian equations,

$$A(\bar{u}) = \delta V^*(\bar{\varsigma}), \quad A_t^*(\bar{u})\bar{\varsigma} = \delta U(\bar{u}), \quad (1.90)$$

which are equivalent to the tri-canonical forms (1.81) because $V^*(\varsigma)$ is a canonical function on \mathcal{V}_a^* . Thus, \bar{u} is a critical point of $P(u)$. By the definition of the canonical dual function, $\bar{\varsigma}$ is also a critical point of $P^d(\varsigma)$. \square

Theorem 8 shows that there is no duality gap between the primal function and its canonical dual. Actually, in the case that $U(u) = \langle u, f \rangle$ is a linear function, we have

$$U^A(\varsigma) = G_\varsigma(u) = \langle A_c(u) ; \varsigma \rangle = -G_c(u, \varsigma) \quad \text{s.t.} \quad A_t^*(u)\varsigma = f;$$

that is the duality gap is recovered by the complementary gap function $G_c(u, \varsigma)$. In this case, the function $P^c(u, \varsigma) = -\Xi(u, \varsigma)$ defined by

$$P^c(u, \varsigma) = G_c(u, \varsigma) + V^*(\varsigma) \quad (1.91)$$

is the total complementary energy introduced by Gao and Strang in 1989 [65]. They proved that if $(\bar{u}, \bar{\varsigma})$ is a critical point of $P^c(u, \varsigma)$, then \bar{u} is a critical point of $P(u)$, and $P(\bar{u}) + P^c(\bar{u}, \bar{\varsigma}) = 0$. Because the operator $A(u)$ is usually nonlinear in nonconvex problems, therefore the explicit format of the canonical dual function $P^d(\varsigma)$ will depend on the properties of the function $G_\varsigma(u)$. By the implicit function theory, if $A(u)$ is twice Gâteaux differentiable and the second Gâteaux differential

$$\delta_u^2 G_\varsigma(u; \delta u^2) \neq 0 \quad \forall \delta u \neq 0, \quad (1.92)$$

then $U^A(\varsigma)$ can be formulated explicitly by the Λ -conjugate transformation (1.88). Some simple illustrative examples are given below.

Example 1. Recall the nonconvex optimization problem with the double well function (1.47):

$$\min\{P(u) = \frac{1}{2}\alpha(\frac{1}{2}|Bu|^2 - \lambda)^2 - \langle u, f \rangle : u \in \mathbb{R}^n\},$$

where $W(u)$ is a double-well function and $U(u)$ is a linear function. If we choose $\xi = \Lambda(u) = \frac{1}{2}|Bu|^2$ as a quadratic operator, then we have $\Lambda_t(u) = (Bu)^T B$ and $\Lambda_c(u) = \Lambda(u) - \Lambda_t(u) = -\frac{1}{2}|Bu|^2$. Because for each $\varsigma \neq 0$, $G_\varsigma(u) = \frac{1}{2}|Bu|^2\varsigma - \langle u, f \rangle$ is a quadratic function and $\delta^2 G_\varsigma(\bar{u}) = \varsigma$, the Λ -conjugate U^A is well-defined by

$$U^A(\varsigma) = \text{sta} \left\{ \left\langle \frac{1}{2}|Bu|^2; \varsigma \right\rangle - \langle u, f \rangle : u \in \mathbb{R}^n \right\} = -\frac{1}{2\varsigma} f^T (B^T B)^{-1} f.$$

The complementary gap function in this case is $G_c(u, \varsigma) = \frac{1}{2}|Bu|^2\varsigma$. Clearly, for any $u \in \mathbb{R}^n$ and $u \neq 0$, $G_c(u, \varsigma) > 0$ if and only if $\varsigma > 0$. Thus, the total complementary function $\Xi(u, \varsigma)$ given by (1.52) is a saddle function for $\varsigma > 0$. This leads to the saddle min-max duality (1.61) in the triality theory.

Example 2. In the nonconvex variational problem (1.64), the quadratic differential operator $\xi = \Lambda(u) = \frac{1}{2}u'^2$ has a physical meaning. In finite deformation theory, if u is considered as the displacement of a deformed body, then ξ can be considered as a *Cauchy-Green strain measure* (see the following section). The Gâteaux derivative of the quadratic differential operator $\Lambda(u)$ is $\Lambda_t(u) = u'd/ dx$. For any given $u \in \mathcal{U}_a$, using integration by parts, we get

$$\langle \Lambda_t(u)u; \varsigma \rangle = \int_0^1 u u' \varsigma \, dx = u u' \varsigma \Big|_{x=0}^{x=1} - \int_0^1 u [u' \varsigma]' \, dx = \langle u, \Lambda_t^*(u) \varsigma \rangle,$$

which gives the adjoint operator Λ_t^* via

$$\Lambda_t^*(u) \varsigma = \begin{cases} u' \varsigma & \text{on } x = 1 \\ [u' \varsigma]', & \forall x \in (0, 1). \end{cases}$$

For any given $\varsigma \in \mathcal{V}_a$, the Λ -conjugate transformation

$$U^A(\varsigma) = \text{sta} \{ \langle \Lambda(u), \varsigma \rangle - U(u) : u \in \mathcal{U}_k \} = - \int_0^1 \tau^2 \varsigma^{-1} \, dx.$$

The complementary operator in this problem is $\Lambda_c(u) = \Lambda(u) - \Lambda_t(u)u = -\frac{1}{2}u'^2$, which leads to the complementary gap function

$$G_c(u, \varsigma) = \int_0^1 \frac{1}{2} u'^2 \varsigma \, dx.$$

Clearly, this is positive if $\varsigma \geq 0$.

1.6.2 Extremality Conditions: Triality Theory

In order to study the extremality conditions of the nonconvex problem, we need to clarify the convexity of the canonical function $V(\xi)$. Without loss of generality, we assume that $V : \mathcal{V}_a \rightarrow \mathbb{R}$ is convex. Thus, for each $u \in \mathcal{U}_a$, the total complementary function

$$\Xi(u, \varsigma) = \langle \Lambda(u) ; \varsigma \rangle - V^*(\varsigma) - U(u) : \mathcal{V}_a^* \rightarrow \mathbb{R}$$

is concave in $\varsigma \in \mathcal{V}_a^*$. The convexity of $\Xi(\cdot, \varsigma) : \mathcal{U}_a \rightarrow \mathbb{R}$ will depend on the geometrical operator $\Lambda(u)$ and the function $U(u)$. We furthermore assume that the function $G_\varsigma(u) = \langle \Lambda(u) ; \varsigma \rangle - U(u) : \mathcal{U}_a \rightarrow \mathbb{R}$ is twice Gâteaux differentiable on \mathcal{U}_a and let

$$\mathcal{G} := \{(u, \varsigma) \in \mathcal{U}_a \times \mathcal{V}_a^* \mid \delta^2 G_\varsigma(u; \delta u^2) \neq 0, \forall \delta u \neq 0\}, \quad (1.93)$$

$$\mathcal{G}^+ := \{(u, \varsigma) \in \mathcal{U}_a \times \mathcal{V}_a^* \mid \delta^2 G_\varsigma(u; \delta u^2) > 0, \forall \delta u \neq 0\}, \quad (1.94)$$

$$\mathcal{G}^- := \{(u, \varsigma) \in \mathcal{U}_a \times \mathcal{V}_a^* \mid \delta^2 G_\varsigma(u; \delta u^2) < 0, \forall \delta u \neq 0\}. \quad (1.95)$$

Theorem 9 (Triality Theorem) *Suppose that $(\bar{u}, \bar{\varsigma}) \in \mathcal{G}$ is a critical point of $\Xi(u, \varsigma)$ and $\mathcal{U}_o \times \mathcal{V}_o^* \subset \mathcal{U}_k \times \mathcal{V}_k^*$ is a neighborhood of $(\bar{u}, \bar{\varsigma})$.*

If $(\bar{u}, \bar{\varsigma}) \in \mathcal{G}^+$, then $(\bar{u}, \bar{\varsigma})$ is a saddle point of $\Xi(u, \varsigma)$; that is

$$\min_{u \in \mathcal{U}_o} \max_{\varsigma \in \mathcal{V}_o^*} \Xi(u, \varsigma) = \Xi(\bar{u}, \bar{\varsigma}) = \max_{\varsigma \in \mathcal{V}_o^*} \min_{u \in \mathcal{U}_o} \Xi(u, \varsigma). \quad (1.96)$$

If $(\bar{u}, \bar{\varsigma}) \in \mathcal{G}^-$, then $(\bar{u}, \bar{\varsigma})$ is a super-critical point of $\Xi(u, \varsigma)$, and we have that either

$$\min_{u \in \mathcal{U}_o} \max_{\varsigma \in \mathcal{V}_o^*} \Xi(u, \varsigma) = \Xi(\bar{u}, \bar{\varsigma}) = \min_{\varsigma \in \mathcal{V}_o^*} \max_{u \in \mathcal{U}_o} \Xi(u, \varsigma) \quad (1.97)$$

holds, or

$$\max_{u \in \mathcal{U}_o} \max_{\varsigma \in \mathcal{V}_o^*} \Xi(u, \varsigma) = \Xi(\bar{u}, \bar{\varsigma}) = \max_{\varsigma \in \mathcal{V}_o^*} \max_{u \in \mathcal{U}_o} \Xi(u, \varsigma). \quad (1.98)$$

Proof. By the assumption on the canonical function $V(\xi)$, we know that $\Xi(u, \varsigma)$ is concave on \mathcal{V}_a^* . Because $G_\varsigma(u)$ is twice Gâteaux differentiable on \mathcal{U}_a , the theory of implicit functions tells us that if $(\bar{u}, \bar{\varsigma}) \in \mathcal{G}$, then there exists a unique $u \in \mathcal{U}_o \subset \mathcal{U}_k$ such that the dual feasible set \mathcal{V}_k^* is non-empty. If such a point $(\bar{u}, \bar{\varsigma}) \in \mathcal{G}^+$, then $G_\varsigma(u)$ is convex in u and $(\bar{u}, \bar{\varsigma})$ is a saddle point of Ξ on $\mathcal{U}_o \times \mathcal{V}_o^*$. The saddle-Lagrangian duality leads to (1.96). If $(\bar{u}, \bar{\varsigma}) \in \mathcal{G}^-$, then $G_\varsigma(u)$ is locally concave in u and $(\bar{u}, \bar{\varsigma})$ is a super-critical point of $\Xi(u, \varsigma)$ on $\mathcal{U}_o \times \mathcal{V}_o^*$. In this case the bi-duality theory leads to (1.97) and (1.98). \square

If the geometrical operator $\Lambda(u)$ is a quadratic function and $U(u)$ is either quadratic or linear, then the second order Gâteaux derivative $\delta^2 G_\varsigma(u)$ does not depend on u . In this case, we let

$$\mathcal{V}_+^* := \{\varsigma \in \mathcal{V}_a^* \mid \delta^2 G_\varsigma(u) \text{ is positive definite}\}, \quad (1.99)$$

$$\mathcal{V}_-^* := \{\varsigma \in \mathcal{V}_a^* \mid \delta^2 G_\varsigma(u) \text{ is negative definite}\}. \quad (1.100)$$

The following theorem provides extremality criteria for critical points of $\Xi(u, \varsigma)$.

Theorem 10 (Tri-Duality Theorem [35, 40]) *Suppose that $G_\varsigma(u) = \langle \Lambda(u); \varsigma \rangle - U(u)$ is a quadratic function of $u \in \mathcal{U}_a$ and $(\bar{u}, \bar{\varsigma})$ is a critical point of $\Xi(u, \varsigma)$.*

If $\bar{\varsigma} \in \mathcal{V}_+^$, then \bar{u} is a global minimizer of $P(u)$ on \mathcal{U}_k if and only if $\bar{\varsigma}$ is a global maximizer of $P^d(\varsigma)$ on \mathcal{V}_+^* , and*

$$P(\bar{u}) = \min_{u \in \mathcal{U}_k} P(u) = \max_{\varsigma \in \mathcal{V}_+^*} P^d(\varsigma) = P^d(\bar{\varsigma}). \quad (1.101)$$

If $\bar{\varsigma} \in \mathcal{V}_-^$, then on the neighborhood $\mathcal{U}_o \times \mathcal{V}_o^* \subset \mathcal{U}_a \times \mathcal{V}_a^*$ of $(\bar{u}, \bar{\varsigma})$, we have that either*

$$P(\bar{u}) = \min_{u \in \mathcal{U}_o} P(u) = \min_{\varsigma \in \mathcal{V}_o^*} P^d(\varsigma) = P^d(\bar{\varsigma}) \quad (1.102)$$

holds, or

$$P(\bar{u}) = \max_{u \in \mathcal{U}_o} P(u) = \max_{\varsigma \in \mathcal{V}_o^*} P^d(\varsigma) = P^d(\bar{\varsigma}). \quad (1.103)$$

This theorem shows that the canonical dual solution $\bar{\varsigma} \in \mathcal{V}_+^*$ provides a global optimality condition for the nonconvex primal problem, whereas the condition $\bar{\varsigma} \in \mathcal{V}_-^*$ provides local extremality conditions.

The triality theory was originally discovered in nonconvex mechanics [34, 39]. Since then, several modified versions have been proposed in nonconvex parametrical variational problems (for quadratic $\Lambda(u)$ and linear $U(u)$ [35]), general nonconvex systems (for nonlinear $\Lambda(u)$ and linear $U(u)$ [40]), global optimization (for general nonconvex functions of type $\Phi(u, \Lambda(u))$ [42], quadratic $U(u)$ [48, 49]), and dissipative Hamiltonian system (for nonconvex/nonsmooth functions of type $\Phi(u, u_t, \Lambda(u))$ [46]). In terms of the parametrical function $G_\varsigma(u) = \langle \Lambda(u); \varsigma \rangle - U(u)$, the current version (Theorems 9 and 10) can be used for solving general nonconvex problem (1.75) with the canonical function $U(u)$.

1.6.3 Complementary Variational Principles in Finite Deformation Theory

In finite deformation theory, the deformation $\mathbf{u}(\mathbf{x})$ is a smooth, vector-valued mapping from an open, simply connected, and bounded domain $\Omega \subset \mathbb{R}^n$ into a deformed domain² $\omega \subset \mathbb{R}^m$. Let $\Gamma = \partial\Omega = \Gamma_u \cup \Gamma_t$ be the boundary of Ω such that on Γ_u , the boundary condition $\mathbf{u}(\mathbf{x}) = \bar{\mathbf{u}}$ is prescribed, whereas on the remaining boundary Γ_t , the surface traction (external force) $\bar{\mathbf{t}}(\mathbf{x})$ is applied. Similar to the nonconvex optimization problem (1.48), the primal problem is to minimize the total potential energy functional:

² If $m = n + 1$, then the deformation $\mathbf{u}(\mathbf{x})$ represents a hyper-surface in m -dimensional space. Applications of the canonical duality theory in differential geometry were discussed in [68].

$$\min \left\{ P(\mathbf{u}) = \int_{\Omega} [W(\nabla \mathbf{u}) - \mathbf{u} \cdot \mathbf{f}] \, d\Omega - \int_{\Gamma_t} \mathbf{u} \cdot \bar{\mathbf{t}} \, d\Gamma : \mathbf{u} = \bar{\mathbf{u}} \text{ on } \Gamma_u \right\} \quad (1.104)$$

where the stored energy $W(\mathbf{F})$ is a Gâteaux differentiable function of $\mathbf{F} = \nabla \mathbf{u}$, and $\mathbf{f}(\mathbf{x})$ is a given force field. Because the deformation gradient $\mathbf{F} = \nabla \mathbf{u} \in \mathbb{R}^{n \times m}$ is a so-called two-point tensor, which is no longer a strain measure in finite deformation theory, the stored energy $W(\mathbf{F})$ is usually nonconvex. Particularly, for *St. Venant-Kirchhoff material* (see [40]), we have

$$W(\boldsymbol{\epsilon}) = \frac{1}{2} \left[\frac{1}{2}(\mathbf{F}^T \mathbf{F} - I) \right] : D : \left[\frac{1}{2}(\mathbf{F}^T \mathbf{F} - I) \right], \quad (1.105)$$

where I is an identity tensor in $\mathbb{R}^{n \times n}$. Due to nonconvexity, the duality relation

$$\boldsymbol{\tau} = \delta W(\mathbf{F})$$

is not one-to-one. Although the two-point tensor $\boldsymbol{\tau} \in \mathbb{R}^{m \times n}$ is called the *first Piola-Kirchhoff stress*, according to Hill's constitutive theory, $(\mathbf{F}, \boldsymbol{\tau})$ is not considered as a work-conjugate (canonical) strain-stress pair (see [40]). The Fenchel-Rockafellar type dual variational problem is

$$\max \left\{ P^\sharp(\boldsymbol{\tau}) = \int_{\Gamma_u} \bar{\mathbf{u}} \cdot \boldsymbol{\tau} \cdot \mathbf{n} \, d\Gamma - \int_{\Omega} W^\sharp(\boldsymbol{\tau}) \, d\Omega \right\} \quad (1.106)$$

$$s.t. \quad -\nabla \cdot \boldsymbol{\tau}^T = \mathbf{f} \text{ in } \Omega, \quad \mathbf{n} \cdot \boldsymbol{\tau}^T = \bar{\mathbf{t}} \text{ on } \Gamma_t. \quad (1.107)$$

In the case where the stored energy $W(\mathbf{F})$ is convex, then $W^\sharp(\boldsymbol{\tau}) = W^*(\boldsymbol{\tau})$ which is called the *complementary energy* in elasticity. In this case, the functional

$$\Pi^c(\boldsymbol{\tau}) = \int_{\Omega} W^*(\boldsymbol{\tau}) \, d\Omega - \int_{\Gamma_u} \bar{\mathbf{u}} \cdot \boldsymbol{\tau} \cdot \mathbf{n} \, d\Gamma$$

is the well-known *Levinson-Zubov complementary energy*. As discussed before, if the stored energy $W(\mathbf{F})$ is nonconvex, the Legendre conjugate W^* is not uniquely defined. It turns out that the Levinson-Zubov complementary variational principle can be used only for solving convex problems (see [31]). Although the Fenchel conjugate $W^\sharp(\boldsymbol{\tau})$ can be uniquely defined, the Fenchel-Young inequality $W(\mathbf{F}) + W^\sharp(\boldsymbol{\tau}) \geq \langle \mathbf{F}; \boldsymbol{\tau} \rangle$ leads to a duality gap between the minimal potential variational problem (1.104) and its Fenchel-Rockafellar dual (see [31]), i.e., in general,

$$\min P(\mathbf{u}) \geq \max P^\sharp(\boldsymbol{\tau}). \quad (1.108)$$

By the fact that the criticality condition $\delta P^\sharp(\boldsymbol{\tau}) = 0$ is not equivalent to the primal variational problem and the weak duality is not appreciated in the field of continuum mechanics, the existence of a *perfect* (i.e., without a duality gap), *pure* (i.e., involving only stress tensor as variational argument) complementary variational principle in finite elasticity has been argued among

well-known scientists for more than three decades (see [75, 76, 81, 82, 86, 87, 88, 101, 102, 145]). This problem was finally solved by the canonical dual transformation and triality theory in [31, 39].

Similar to the quadratic operator $\Lambda(\mathbf{u}) = \frac{1}{2}|B\mathbf{u}|^2$ (see equation (1.51)) chosen for the nonconvex optimization problem (1.48), we let

$$\mathbf{E} = \Lambda(\mathbf{u}) = \frac{1}{2}[(\nabla\mathbf{u})^T(\nabla\mathbf{u}) - I], \quad (1.109)$$

which is a symmetrical tensor field in $\mathbb{R}^{n \times n}$. In finite deformation theory, \mathbf{E} is the well-known *Green-St. Venant strain tensor*. Thus, in terms of \mathbf{E} , the stored energy for St. Venant-Kirchhoff material can be written in the canonical form $W(\nabla\mathbf{u}) = V(\Lambda(\nabla\mathbf{u}))$, and

$$V(\mathbf{E}) = \frac{1}{2}\mathbf{E} : D : \mathbf{E}$$

is a (quadratic) convex function of the symmetrical tensor $\mathbf{E} \in \mathbb{R}^{n \times n}$. The canonical dual variable $\mathbf{E}^* = \delta V(\mathbf{E}) = D \cdot \mathbf{E}$ is called the *second Piola-Kirchhoff stress tensor*, denoted as \mathbf{T} . The Legendre conjugate

$$V^*(\mathbf{T}) = \frac{1}{2}\mathbf{T} : D^{-1} : \mathbf{T} \quad (1.110)$$

is also a quadratic function. Let $\mathcal{U}_a = \{\mathbf{u} \in \mathcal{W}^{1,p}(\Omega; \mathbb{R}^3) \mid \mathbf{u} = \bar{\mathbf{u}} \text{ on } \Gamma_u\}$ (where $\mathcal{W}^{1,p}$ is a standard Sobolev space with $p \in (1, \infty)$) and $\mathcal{V}_a^* = \mathcal{C}(\Omega; \mathbb{R}^{n \times n})$. Replacing $W(\nabla\mathbf{u})$ by its canonical dual transformation $V(\Lambda(\mathbf{u})) = \mathbf{E}(\mathbf{u}) : \mathbf{T} - V^*(\mathbf{T})$, the generalized complementary energy $\Xi : \mathcal{U}_a \times \mathcal{V}_a^* \rightarrow \mathbb{R}$ has the following format,

$$\Xi(\mathbf{u}, \mathbf{T}) = \int_{\Omega} [\mathbf{E}(\mathbf{u}) : \mathbf{T} - V^*(\mathbf{T}) - \mathbf{u} \cdot \mathbf{f}] \, d\Omega - \int_{\Gamma_t} \mathbf{u} \cdot \bar{\mathbf{t}} \, d\Gamma, \quad (1.111)$$

which is the well-known *Hellinger-Reissner generalized complementary energy* in continuum mechanics.

Furthermore, if we replace $V^*(\mathbf{T})$ by its bi-Legendre transformation $\mathbf{E} : \mathbf{T} - V(\mathbf{E})$, then $\Xi(\mathbf{u}, \mathbf{T})$ can be written as

$$\Xi_{hw}(\mathbf{u}, \mathbf{T}, \mathbf{E}) = \int_{\Omega} [\Lambda(\nabla\mathbf{u}) - \mathbf{E}] : \mathbf{T} + V(\mathbf{E}) - \mathbf{u} \cdot \mathbf{f}] \, d\Omega - \int_{\Gamma_t} \mathbf{u} \cdot \bar{\mathbf{t}} \, d\Gamma. \quad (1.112)$$

This is the well-known *Hu-Washizu generalized potential energy* in nonlinear elasticity. The Hu-Washizu variational principle has important applications in computational analysis of thin-walled structures, where the geometrical equation $\mathbf{E} = \Lambda(u)$ is usually proposed by certain geometrical hypothesis.

Because $\Lambda(\mathbf{u})$ is a quadratic operator, its Gâteaux differential at $\bar{\mathbf{u}}$ in the direction \mathbf{u} is $\delta\Lambda(\bar{\mathbf{u}}; \mathbf{u}) = \Lambda_t(\bar{\mathbf{u}})\mathbf{u} = (\nabla\bar{\mathbf{u}})^T(\nabla\mathbf{u})$ and

$$\Lambda_c(\mathbf{u}) = \Lambda(\mathbf{u}) - \Lambda_t(\mathbf{u})\mathbf{u} = -\frac{1}{2}[(\nabla\mathbf{u})^T(\nabla\mathbf{u}) + I].$$

By using the Gauss-Green theorem, the balance operator $A_t^*(\mathbf{u})$ can be defined as

$$A_t^*(\mathbf{u})\mathbf{T} = \begin{cases} -\nabla \cdot [(\nabla \mathbf{u})^T \cdot \mathbf{T}]^T & \text{in } \Omega, \\ \mathbf{n} \cdot [(\nabla \mathbf{u})^T \cdot \mathbf{T}]^T & \text{on } \Gamma. \end{cases}$$

The complementary gap function in this problem is a quadratic functional:

$$G_c(\mathbf{u}, \mathbf{T}) = \langle -A_c(\mathbf{u}); \mathbf{T} \rangle = \int_{\Omega} \frac{1}{2} \text{tr}[(\nabla \mathbf{u})^T \cdot \mathbf{T} \cdot (\nabla \mathbf{u}) + \mathbf{T}] \, d\Omega. \quad (1.113)$$

Thus, the complementary variational problem is to find critical (stationary) points $(\bar{\mathbf{u}}, \bar{\mathbf{T}})$ such that

$$P^c(\bar{\mathbf{u}}, \bar{\mathbf{T}}) = \text{sta} \left\{ \int_{\Omega} \frac{1}{2} \text{tr}[(\nabla \mathbf{u})^T \cdot \mathbf{T} \cdot (\nabla \mathbf{u}) + \mathbf{T}] \, d\Omega + \int_{\Omega} V^*(\mathbf{T}) \, d\Omega \right\} \quad (1.114)$$

s.t. $-\nabla \cdot [(\nabla \mathbf{u})^T \cdot \mathbf{T}]^T = \mathbf{f}$ in Ω , $\mathbf{n} \cdot [(\nabla \mathbf{u})^T \cdot \mathbf{T}]^T = \bar{\mathbf{t}}$ on Γ_t .

The following result is due to Gao and Strang in 1989.

Theorem 11 (Complementary-Dual Variational Principle [65])

If $(\bar{\mathbf{u}}, \bar{\mathbf{T}})$ is a critical point of the complementary variational problem (1.114), then $\bar{\mathbf{u}}$ is a critical point of the total potential energy $P(\mathbf{u})$ defined by (1.104), and

$$P(\bar{\mathbf{u}}) + P^c(\bar{\mathbf{u}}, \bar{\mathbf{T}}) = 0.$$

Moreover, if the complementary gap function

$$G_c(\mathbf{u}, \bar{\mathbf{T}}) \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}_a, \quad (1.115)$$

then $\bar{\mathbf{u}}$ is a global minimizer of $P(\mathbf{u})$ and

$$P(\bar{\mathbf{u}}) = \min_{\mathbf{u}} P(\mathbf{u}) = \max_{\mathbf{T}} \min_{\mathbf{u}} \Xi(\mathbf{u}, \mathbf{T}) = -P^c(\bar{\mathbf{u}}, \bar{\mathbf{T}}), \quad (1.116)$$

subject to $\mathbf{T}(\mathbf{x})$ being positive definite for all $\mathbf{x} \in \Omega$.

This theorem shows that the positivity of the complementary gap function $G_c(\mathbf{u}, \mathbf{T})$ provides a sufficient condition for a global minimizer, the equalities (11) and (1.116) indicate that there is no duality gap between the total potential $P(\mathbf{u})$ and its complementary energy $P^c(\mathbf{u}, \mathbf{T})$. The physical significance is also clear: a finite deformed material is stable if the second Piola-Kirchhoff stress tensor $\mathbf{T}(\mathbf{x})$ is positive definite everywhere in the domain Ω . The linear operator $B = \nabla$ in this nonconvex variational problem is a partial differential operator, therefore it is difficult to find its inverse. It took more than ten years before the canonical dual problem was finally formulated in [37, 39]. To see this, let us assume that for a given force vector field $\bar{\mathbf{t}}$ on the boundary Γ_t , the first Piola-Kirchhoff stress tensor $\boldsymbol{\tau}(\mathbf{x})$ can be defined by solving the following boundary value problem,

$$-\nabla \cdot \boldsymbol{\tau}^T(\mathbf{x}) = \mathbf{f} \quad \text{in } \Omega, \quad \mathbf{n} \cdot \boldsymbol{\tau}^T = \bar{\mathbf{t}} \quad \text{on } \Gamma_t. \quad (1.117)$$

Then the canonical dual functional $P^d(\mathbf{T})$ can be formulated as:

$$P^d(\mathbf{T}) = - \int_{\Omega} \frac{1}{2} \text{tr}(\boldsymbol{\tau} \cdot \mathbf{T}^{-1} \cdot \boldsymbol{\tau}^T + \mathbf{T}) \, d\Omega - \int_{\Omega} V^*(\mathbf{T}) \, d\Omega. \quad (1.118)$$

The criticality condition $\delta P^d(\mathbf{T}) = 0$ gives the canonical dual equation

$$\mathbf{T} \cdot (2 \delta V^*(\mathbf{T}) + I) \cdot \mathbf{T} = \boldsymbol{\tau}^T \cdot \boldsymbol{\tau}. \quad (1.119)$$

For St. Venant-Kirchhoff material, $V^*(\mathbf{T}) = \frac{1}{2} \mathbf{T} : D^{-1} : \mathbf{T}$ is a quadratic function and its Gâteaux derivative $\delta V^*(\mathbf{T}) = D^{-1} \cdot \mathbf{T}$ is linear. In this case, the canonical dual equation (1.119) is a cubic equation, which is similar to the dual algebraic equations (1.57) and (1.70).

Theorem 12 (Pure Complementary Energy Principle [37, 39])

Suppose that for a given force field $\mathbf{t}(\mathbf{x})$ on Γ_t , the first Piola-Kirchhoff stress field $\boldsymbol{\tau}(\mathbf{x})$ is defined by (1.117). Then each solution $\bar{\mathbf{T}}$ of the canonical dual equation (1.119) is a critical point of P^d , the vector defined by the line integral

$$\bar{\mathbf{u}} = \int \boldsymbol{\tau} \cdot \bar{\mathbf{T}}^{-1} \, d\mathbf{x} \quad (1.120)$$

is a critical point of $P(\mathbf{u})$, and

$$P(\bar{\mathbf{u}}) = P^d(\bar{\mathbf{T}}).$$

This theorem presents an analytic solution to the nonconvex potential variational problem (1.104). In the finite deformation theory of elasticity, this pure complementary variational principle is also known as *Gao principle* [89], which holds also for the general canonical energy function $V(\mathbf{E})$. Similar to Theorem 9, the extremality of the critical points can be identified by the complementary gap function. Applications of this pure complementary variational principle for solving nonconvex/nonsmooth boundary value problems are illustrated in [39, 40, 60].

1.7 Applications to Semi-Linear Nonconvex Systems

The canonical dual transformation and the associated triality theory can be used to solve many difficult problems in engineering and science. In this section, we present applications for solving the following nonconvex minimization problem

$$(\mathcal{P}) : \quad \min\{P(u) = W(u) + \frac{1}{2} \langle u, Au \rangle - \langle u, f \rangle \quad : \quad u \in \mathcal{U}_k\}, \quad (1.121)$$

where $W(u) : \mathcal{U}_k \rightarrow \mathbb{R}$ is a nonconvex function, and $A : \mathcal{U}_a \subset \mathcal{U} \rightarrow \mathcal{U}_a^*$ is a linear operator. If $W(u)$ is Gâteaux differentiable, the criticality condition $\delta P(u) = 0$ leads to a nonlinear Euler equation

$$Au + \delta W(u) = f. \quad (1.122)$$

The abstract form (1.122) of the primal problem (\mathcal{P}) covers many situations. In nonconvex mechanics (cf. [61, 49]), where \mathcal{U} is an infinite dimensional function space, the state variable $u(x)$ is a field function, and $A : \mathcal{U} \rightarrow \mathcal{U}^*$ is usually a partial differential operator. In this case, the governing equation (1.122) is a so-called *semi-linear equation*. For example, in the Landau-Ginzburg theory of superconductivity, $A = \Delta$ is the Laplacian over a given space domain $\Omega \subset \mathbb{R}^n$ and

$$W(u) = \int_{\Omega} \frac{1}{2} \alpha \left(\frac{1}{2} u^2 - \lambda \right)^2 d\Omega \quad (1.123)$$

is the Landau *double-well potential*, in which $\alpha, \lambda > 0$ are material constants. Then the governing equation (1.122) leads to the well-known Landau-Ginzburg equation

$$\Delta u + \alpha u \left(\frac{1}{2} u^2 - \lambda \right) = f.$$

This semi-linear differential equation plays an important role in material science and physics including: ferroelectricity, ferromagnetism, and superconductivity. In a more complicated case where $A = \Delta + \text{curl curl}$, we have

$$\Delta u + \text{curl curl } u + \alpha u \left(\frac{1}{2} u^2 - \lambda \right) = f,$$

which is the so-called Cahn-Hilliard equation in liquid crystal theory. Due to the nonconvexity of the double-well function $W(u)$, any solution of the semi-linear differential equation (1.122) is only a critical point of the total potential $P(u)$. Traditional direct analysis and related numerical methods for finding the global minimizer of the nonconvex variational problem have proven unsuccessful to date.

In dynamical systems, if $A = -\partial_{,tt} + \Delta$ is a wave operator over a given space-time domain $\Omega \subset \mathbb{R}^n \times \mathbb{R}$, then (1.122) is the well-known nonlinear Schrödinger equation

$$-u_{,tt} + \Delta u + \alpha u \left(\frac{1}{2} u^2 - \lambda \right) = f.$$

This equation appears in many branches of physics. It provides one of the simplest models of the unified field theory. It can also be found in the theory of dislocations in metals, in the theory of Josephson junctions, as well as in interpreting certain biological processes such as DNA dynamics.

In the most simple case where u depends only on time, the nonlinear Schrödinger equation reduces to the well-known Duffing equation

$$u_{,tt} = \alpha u \left(\frac{1}{2} u^2 - \lambda \right) - f.$$

Even for this one dimensional ordinary differential equation, an analytic solution is still very difficult to obtain. It is known that this equation is extremely sensitive to the initial conditions and the input (driving force) $f(t)$. Fig. 1.7 displays clearly that for the same given data, two Runge-Kutta solvers in MATLAB produce very different vibration modes and “trajectories” in the phase space $u-p$ ($p = u_t$). Mathematically speaking, due to the nonconvexity

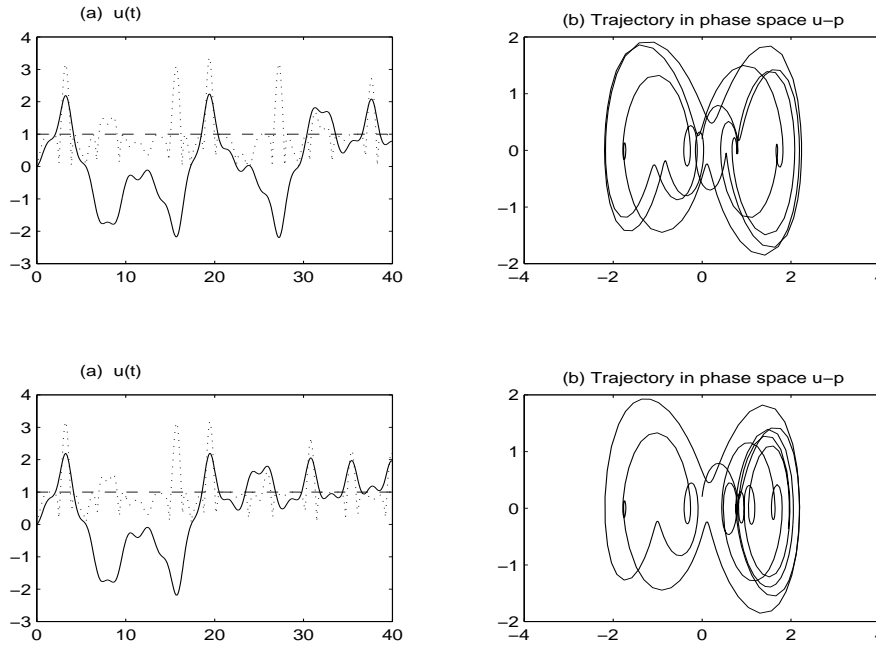


Fig. 1.7. Numerical results by *ode23* (top) and *ode15s* (bottom) solvers in MATLAB.

of the function $W(u)$, very small perturbations of the system’s initial conditions and parameters may lead the system to different local minima with significantly different performance characteristics, that is the so-called *chaotic phenomena*. Numerical results vary with the methods used. This is one of the main reasons why traditional perturbation analysis and direct approaches cannot successfully be applied to nonconvex systems [49].

Numerical discretization of the nonconvex variational problem (\mathcal{P}) in mathematical physics usually leads to a nonconvex optimization problem in finite dimensional space $\mathcal{U} = \mathbb{R}^n$, where the field variable u is simply a vector $\mathbf{x} \in \mathcal{U}$, the bilinear form $\langle \mathbf{x}, \mathbf{x}^* \rangle = \mathbf{x}^T \mathbf{x}^* = \mathbf{x} \cdot \mathbf{x}^*$ is the dot-product of two vectors, and the operator $A : \mathbb{R}^n \rightarrow \mathcal{U}^* = \mathbb{R}^n$ is a symmetrical matrix. In d.c.(difference of convex functions) programming and discrete dynamical systems, the operator $A = A^T \in \mathbb{R}^{n \times n}$ is usually indefinite. The problem

(1.121) is then one of global minimization in \mathbb{R}^n . In this section, we discuss the canonical dual transformation method for solving this type of problem.

1.7.1 Unconstrained Nonconvex Optimization Problem with Double-Well Energy

First, let us consider an unconstrained global optimization problem in finite dimensional space $\mathcal{U} = \mathbb{R}^n$, where $A = A^T \in \mathbb{R}^{n \times n}$ is a matrix, and $W(\mathbf{x})$ is a double-well function of the type $W(\mathbf{x}) = \frac{1}{2}(\frac{1}{2}|\mathbf{x}|^2 - \lambda)^2$. Then the primal problem is

$$\min \left\{ P(\mathbf{x}) = \frac{1}{2} \left(\frac{1}{2}|\mathbf{x}|^2 - \lambda \right)^2 + \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{x}^T f \quad : \quad \forall \mathbf{x} \in \mathcal{U}_k = \mathbb{R}^n \right\}. \quad (1.124)$$

The necessary condition $\delta P(\mathbf{x}) = 0$ leads to a coupled nonlinear algebraic system

$$A \mathbf{x} + \left(\frac{1}{2}|\mathbf{x}|^2 - \lambda \right) \mathbf{x} = f. \quad (1.125)$$

Clearly, a direct method for solving this nonlinear equation with n unknown is elusive. By choosing the quadratic operator $\xi = \frac{1}{2}|\mathbf{x}|^2$, the canonical function $V(\xi) = \frac{1}{2}(\xi - \lambda)^2$ is a quadratic function. By the fact that $\frac{1}{2}|\mathbf{x}|^2 = \xi \geq 0$, $\forall \mathbf{x} \in \mathbb{R}^n$, the range of the quadratic mapping $\Lambda(\mathbf{x})$ is

$$\mathcal{V}_a = \{\xi \in \mathbb{R} \mid \xi \geq 0\}.$$

Thus, on \mathcal{V}_a , the canonical duality relation $\varsigma = \delta V(\xi) = \xi - \lambda$ is one-to-one and the range of the canonical dual mapping $\delta V : \mathcal{V}_a \rightarrow \mathcal{V}^* \subset \mathbb{R}$ is

$$\mathcal{V}_a^* = \{\varsigma \in \mathbb{R} \mid \varsigma \geq -\lambda\}.$$

It turns out that (ξ, ς) is a canonical pair on $\mathcal{V}_a \times \mathcal{V}_a^*$ and the Legendre conjugate V^* is also a quadratic function:

$$V^*(\varsigma) = \text{sta}\{\xi\varsigma - V(\xi) : \xi \in \mathcal{V}_a\} = \frac{1}{2}\varsigma^2 + \lambda\varsigma.$$

For a given $\varsigma \in \mathcal{V}_a^*$, the Λ -conjugate transformation

$$\begin{aligned} U^A(\varsigma) &= \text{sta} \left\{ \frac{1}{2} \mathbf{x}^2 \varsigma - \frac{1}{2} \mathbf{x}^T A \mathbf{x} + \mathbf{x}^T f \quad : \quad \mathbf{x} \in \mathbb{R}^n \right\} \\ &= -\frac{1}{2} f^T (A + \varsigma I)^{-1} f \end{aligned}$$

is well-defined on the canonical dual feasible space \mathcal{V}_k^* , given by

$$\mathcal{V}_k^* = \{\varsigma \in \mathbb{R} \mid \det(A + \varsigma I) \neq 0, \varsigma \geq -\lambda\}. \quad (1.126)$$

Thus, the canonical dual problem can be proposed as the following [48]:

$$(\mathcal{P}^d): \max \left\{ P^d(\varsigma) = -\frac{1}{2}f^T(A + \varsigma I)^{-1}f - \frac{1}{2}\varsigma^2 - \lambda\varsigma \quad : \quad \varsigma \in \mathcal{V}_k^* \right\}. \quad (1.127)$$

This is a nonlinear programming problem with only one variable! The criticality condition of this dual problem leads to the dual algebraic equation

$$\varsigma + \lambda = \frac{1}{2}f^T(A + \varsigma I)^{-2}f. \quad (1.128)$$

For any given $A \in \mathbb{R}^{n \times n}$ and $f \in \mathbb{R}^n$, this equation can be solved by MATHEMATICA. Extremality conditions of these dual solutions can be identified by the following theorem (see [48]).

Theorem 13 (Gao [48]) *If the matrix A has r distinct non-zero eigenvalues such that*

$$a_1 < a_2 < \dots < a_r,$$

then the canonical dual algebraic equation (1.128) has at most $2r + 1$ roots

$$\varsigma_1 > \varsigma_2 \geq \varsigma_3 \geq \dots \geq \varsigma_{2r+1}.$$

For each ς_i , the vector

$$\mathbf{x}_i = (A + \varsigma_i I)^{-1}f, \quad \forall i = 1, 2, \dots, 2r + 1, \quad (1.129)$$

is a solution to the semi-linear algebraic equation (1.125) and

$$P(\mathbf{x}_i) = P^d(\varsigma_i), \quad \forall i = 1, \dots, 2r + 1. \quad (1.130)$$

Particularly, the canonical dual problem has at most one global maximizer $\varsigma_1 > -a_1$ in the open interval $(-a_1, +\infty)$, and \mathbf{x}_1 is a global minimizer of $P(\mathbf{x})$ over \mathcal{U}_k , that is

$$P(\mathbf{x}_1) = \min_{\mathbf{x} \in \mathcal{U}_k} P(\mathbf{x}) = \max_{\varsigma > -a_1} P^d(\varsigma) = P^d(\varsigma_1). \quad (1.131)$$

Moreover, in each open interval $(-a_{i+1}, -a_i)$, the canonical dual equation (1.128) has at most two real roots $-a_{i+1} < \varsigma_{2i+1} < \varsigma_{2i} < -a_i$, $\forall i = 1, \dots, 2r + 1$, ς_{2i} is a local minimizer of P^d , and ς_{2i+1} is a local maximizer of $P^d(\varsigma)$.

As an example in two-dimensional space, which is illustrated in Fig. 1.8, we simply choose $A = \{a_{ij}\}$ with $a_{11} = 0.6$, $a_{22} = -0.5$, $a_{12} = a_{21} = 0$, and $f = \{0.2, -0.1\}$. For a given parameter $\lambda = 1.5$, and $\alpha = 1.0$, the graph of $P(\mathbf{x})$ is a nonconvex surface (see Fig. 1.8a) with four potential wells and one local maximizer. The graph of the canonical dual function $P^d(\varsigma)$ is shown in Fig. 1.9. The dual canonical dual algebraic equation (1.128) has a total of five real roots:

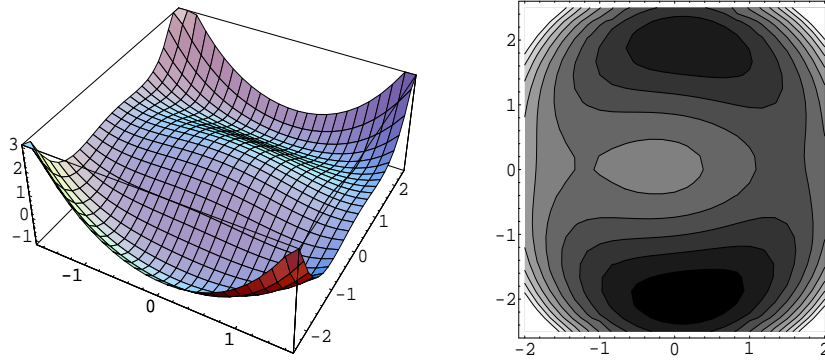


Fig. 1.8. Graph of the primal function $P(x_1, x_2)$ and its contours

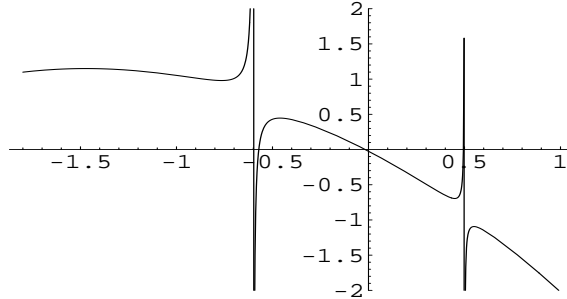


Fig. 1.9. Graph of the dual function $P^d(\varsigma)$

$$\bar{\varsigma}_5 = -1.47 < \bar{\varsigma}_4 = -0.77 < \bar{\varsigma}_3 = -0.46 < \bar{\varsigma}_2 = 0.45 < \bar{\varsigma}_1 = 0.55,$$

and we have

$$P^d(\bar{\varsigma}_5) = 1.15 > P^d(\bar{\varsigma}_4) = 0.98 > P^d(\bar{\varsigma}_3) = 0.44 > P^d(\bar{\varsigma}_2) = -0.70 > P^d(\bar{\varsigma}_1).$$

By the triality theory, we know that $\bar{\mathbf{x}}_1 = (A + \bar{\varsigma}_1 I)^{-1} f = \{0.17, -2.02\}$ is a global minimizer of $P(\bar{\mathbf{x}})$; and accordingly, $P(\bar{\mathbf{x}}_1) = P^d(\bar{\varsigma}_1) = -1.1$; and that $\bar{\mathbf{x}}_5 = \{-0.23, 0.05\}$ and $\bar{\mathbf{x}}_3 = \{1.44, 0.10\}$ are local maximizers, whereas $\bar{\mathbf{x}}_4 = \{-1.21, 0.08\}$ and $\bar{\mathbf{x}}_2 = \{0.19, 1.96\}$ are local minimizers.

The graph of $P^d(\varsigma)$ for a four-dimensional problem is shown in Fig. 1.10. It can be easily seen that $P^d(\varsigma)$ is strictly concave for $\varsigma > -a_1$. Within each interval $-a_{i-1} < \varsigma < -a_i, \forall i = 1, 2, \dots, r$, the dual function $P^d(\varsigma)$ has at most one local minimum and one local maximum. These local extrema can be identified by the triality theory [48].

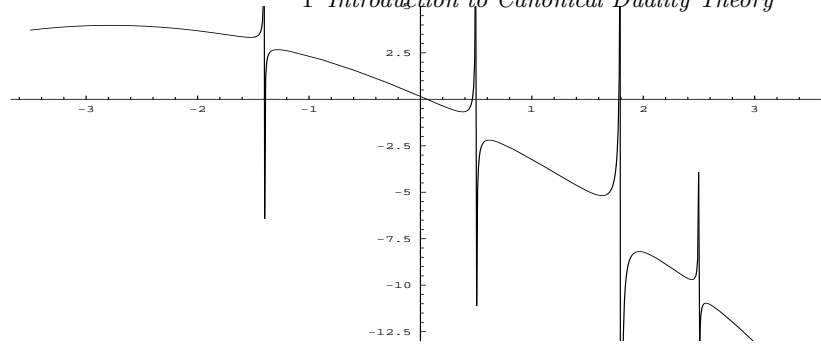


Fig. 1.10. Graph of the dual function $P^d(\varsigma)$ for a four-dimensional problem

The nonconvex function $W(\mathbf{x})$ in (1.121) could be in many other forms, for example,

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

where $B \in \mathbb{R}^{m \times n}$ is a given matrix and $\lambda > 0$ is a constant. In this case, the primal problem (\mathcal{P}) is a quadratic-exponential minimization problem

$$\min \left\{ P(\mathbf{x}) = \exp\left(\frac{1}{2}|B\mathbf{x}|^2 - \lambda\right) + \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{x}^T f : \mathbf{x} \in \mathbb{R}^n \right\}.$$

By letting $\xi = \Lambda(\mathbf{x}) = \frac{1}{2}|B\mathbf{x}|^2 - \lambda$, the canonical function $V(\xi) = \exp(\xi)$ is convex and its Legendre conjugate is $V^*(\varsigma) = \varsigma(\ln \varsigma - 1)$. The canonical dual problem was formulated in [63]:

$$(\mathcal{P}^d) : \quad \max \left\{ P^d(\varsigma) = -\frac{1}{2}f^T[G(\varsigma)]^{-1}f - (\varsigma \log \varsigma - \varsigma) - \lambda\varsigma : \varsigma \in \mathcal{V}_+^* \right\},$$

where $G(\varsigma) = A + \varsigma B^T B$ and the dual feasible space is defined by

$$\mathcal{V}_+^* = \{\varsigma \in \mathbb{R} \mid \varsigma > 0, G(\varsigma) \text{ is positive definite}\}.$$

Detailed study of this case was given in [63].

1.7.2 Constrained Quadratic Minimization over a Sphere

If the function $W(\mathbf{x})$ in problem (1.121) is an indicator of a constraint set $\mathcal{U}_k \subset \mathbb{R}^n$, that is

$$W(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \in \mathcal{U}_k, \\ +\infty & \text{otherwise,} \end{cases}$$

then the general problem (1.121) becomes a constrained nonconvex quadratic optimization problem, denoted as

$$(\mathcal{P}_q) : \min\{P(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, A\mathbf{x} \rangle - \langle \mathbf{x}, f \rangle : \mathbf{x} \in \mathcal{U}_k\}. \quad (1.132)$$

General constrained global optimization problems are discussed in the next section. Here, we consider the following quadratic minimization problem with a nonlinear constraint

$$(\mathcal{P}_q) : \min P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - f^T \mathbf{x} \quad (1.133)$$

$$s.t. \quad |\mathbf{x}| \leq r,$$

where $A = A^T \in \mathbb{R}^{n \times n}$ is a symmetric matrix, $f \in \mathbb{R}^n$ is a given vector, and $r > 0$ is a constant. The feasible space $\mathcal{U}_k = \{\mathbf{x} \in \mathbb{R}^n \mid |\mathbf{x}| \leq r\}$ is a hyper-sphere in \mathbb{R}^n . This problem often arises as a subproblem in general optimization algorithms (cf. Powell, 2002). Often, in the model trust region methods, the objective function in nonlinear programming is approximated locally by a quadratic function. In such cases, the approximation is restricted to a small region around the current iterate. These methods therefore require the solution of quadratic programming problems over spheres.

To solve this constrained nonconvex minimization by using a traditional Lagrange multiplier method, we have

$$L(\mathbf{x}, \lambda) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - f^T \mathbf{x} + \lambda(|\mathbf{x}| - r). \quad (1.134)$$

For a given $\lambda \geq 0$, the traditional dual function can be defined via the Fenchel-Moreau-Rockafellar duality theory:

$$P^*(\lambda) = \min\{L(\mathbf{x}, \lambda) : \mathbf{x} \in \mathbb{R}^n\}, \quad (1.135)$$

which is a concave function of λ . However, due to the nonconvexity of $P(\mathbf{x})$, we have only the weak duality relationship

$$\min_{|\mathbf{x}| \leq r} P(\mathbf{x}) \geq \max_{\lambda \geq 0} P^*(\lambda).$$

The duality gap θ given by the slack in the above inequality is typically nonzero indicating that the dual solution does not solve the primal problem. On the other hand, the KKT condition leads to a coupled nonlinear algebraic system

$$A\mathbf{x} + \lambda|\mathbf{x}|^{-1}\mathbf{x} = f,$$

$$\lambda \geq 0, \quad |\mathbf{x}| \leq r, \quad \lambda(|\mathbf{x}| - r) = 0.$$

As indicated by Floudas and Visweswaran (1995), due to the presence of the nonlinear sphere constraint, the solution of (\mathcal{P}_q) is likely to be irrational, which implies that it is not possible to exactly compute the solution. Therefore, many polynomial time algorithms have been suggested to compute the approximate solution to this problem (see, Sorensen, 1982; Karmarkar, 1990;

and Ye, 1992). However, by the canonical dual transformation this problem has been solved completely in [51].

First, we need to reformulate the constraint $|\mathbf{x}| \leq r$ in the canonical form

$$\xi = \Lambda(\mathbf{x}) = \frac{1}{2}|\mathbf{x}|^2.$$

Let $\lambda = \frac{1}{2}r^2$, then the canonical function $V(\Lambda(\mathbf{x}))$ can be defined as

$$V(\xi) = \begin{cases} 0 & \text{if } \xi \leq \lambda, \\ +\infty & \text{otherwise,} \end{cases}$$

whose effective domain is $\mathcal{V}_a = \{\xi \in \mathbb{R} \mid \xi \leq \lambda\}$. Letting $U(\mathbf{x}) = \mathbf{x}^T f - \frac{1}{2}\mathbf{x}^T A \mathbf{x}$, the primal problem (\mathcal{P}_q) can be reformulated in the following canonical form:

$$\min\{H(\mathbf{x}) = V(\Lambda(\mathbf{x})) - U(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}. \quad (1.136)$$

By the Fenchel transformation, the conjugate of $V(\xi)$ is

$$V^\#(\varsigma) = \max_{\xi \in \mathcal{V}_a} \{\xi\varsigma - V(\xi)\} = \begin{cases} \lambda\varsigma & \text{if } \varsigma \geq 0, \\ +\infty & \text{otherwise,} \end{cases} \quad (1.137)$$

whose effective domain is $\mathcal{V}_a^* = \{\varsigma \in \mathbb{R} \mid \varsigma \geq 0\}$. The dual feasible space \mathcal{V}_k^* in this problem is

$$\mathcal{V}_k^* = \{\varsigma \in \mathbb{R} \mid \varsigma \geq 0, \det(A + \varsigma I) \neq 0\}.$$

Thus, for a given $\varsigma \in \mathcal{V}_a^*$, the Λ -conjugate of U can be formulated as

$$\begin{aligned} U^\Lambda(\varsigma) &= \text{sta} \left\{ \frac{1}{2}|\mathbf{x}|^2\varsigma + \frac{1}{2}\mathbf{x}^T A \mathbf{x} - \mathbf{x}^T f : \mathbf{x} \in \mathbb{R}^n \right\} \\ &= -\frac{1}{2}f^T(A + \varsigma I)^{-1}f, \end{aligned}$$

and the problem (\mathcal{P}^d) , which is perfectly dual to (\mathcal{P}_q) , is given by

$$(\mathcal{P}_q^d) : \max \left\{ P^d(\varsigma) = -\frac{1}{2}f^T(A + \varsigma I)^{-1}f - \lambda\varsigma : \varsigma \in \mathcal{V}_k^* \right\}. \quad (1.138)$$

The criticality condition $\delta P^d(\bar{\varsigma}) = 0$ leads to a nonlinear algebraic equation

$$\frac{1}{2}f^T(A + \bar{\varsigma}I)^{-2}f = \lambda. \quad (1.139)$$

Similar to (1.128), this equation can also be solved easily by using MATHEMATICA. Each root $\bar{\varsigma}_i$ is a critical point of $P^d(\varsigma)$. The following theorem presents a complete set of solutions for this dual problem.

Theorem 14 (Complete Solution to (\mathcal{P}_q) [51]) *Suppose that the symmetric matrix A has $p \leq n$ distinct eigenvalues, and $i_d \leq p$ of them are negative such that*

$$a_1 < a_2 < \cdots < a_{i_d} < 0 \leq a_{i_d+1} < \cdots < a_p.$$

Then for a given vector $f \in \mathbb{R}^n$, the canonical dual problem (\mathcal{P}_q^d) has at most $2i_d + 1$ critical points $\bar{\varsigma}_i, i = 1, \dots, 2i_d + 1$, satisfying the following distribution law:

$$\bar{\varsigma}_1 > -a_1 > \bar{\varsigma}_2 \geq \bar{\varsigma}_3 > -a_2 > \cdots > -a_{i_d} > \bar{\varsigma}_{2i_d} \geq \bar{\varsigma}_{2i_d+1} > 0. \quad (1.140)$$

For each $\bar{\varsigma}_i \geq 0, i = 1, \dots, 2i_d + 1$, the vector defined by

$$\bar{\mathbf{x}}_i = (A + \bar{\varsigma}_i I)^{-1} f \quad (1.141)$$

is a KKT point of the problem (\mathcal{P}_q) and

$$P(\bar{\mathbf{x}}_i) = P^d(\bar{\varsigma}_i), \quad i = 1, 2, \dots, 2i_d + 1. \quad (1.142)$$

Moreover, if $i_d > 0$, then the problem (\mathcal{P}_q) has at most $2i_d + 1$ critical points on the boundary of the sphere, that is

$$\frac{1}{2} |\bar{\mathbf{x}}_i|^2 = \lambda, \quad i = 1, \dots, 2i_d + 1. \quad (1.143)$$

Because $A = A^T$, there exists an orthogonal matrix $R^T = R^{-1}$ such that $A = R^T D R$, where $D = (a_i \delta_{ij})$ is a diagonal matrix. For the given vector $f \in \mathbb{R}^n$, let $\mathbf{g} = Rf = (g_i)$, and define

$$\psi(\varsigma) = \frac{1}{2} f^T (A + \varsigma I)^{-2} f = \frac{1}{2} \sum_{i=1}^p g_i^2 (a_i + \varsigma)^{-2}. \quad (1.144)$$

Clearly, this real-valued function $\psi(\varsigma)$ is strictly convex within each interval $-a_{i+1} < \varsigma < -a_i$, as well as over the intervals $-\infty < \varsigma < -a_p$ and $-a_1 < \varsigma < \infty$ (see Fig. 1.11). Thus, for a given parameter $\lambda > 0$, the algebraic equation

$$\psi(\varsigma) = \frac{1}{2} \sum_{i=1}^p g_i^2 (a_i + \varsigma)^{-2} = \lambda \quad (1.145)$$

has at most $2p$ solutions $\{\bar{\varsigma}_i\}$ satisfying $-a_{j+1} < \bar{\varsigma}_{2j+1} \leq \bar{\varsigma}_{2j} < -a_j$ for $j = 1, \dots, p-1$, and $\bar{\varsigma}_1 > -a_1, \bar{\varsigma}_{2p} < -a_p$. Because A has only i_d negative eigenvalues, the equality $\psi(\varsigma) = \lambda$ has at most $2i_d + 1$ strictly positive roots $\{\bar{\varsigma}_i\} > 0, i = 1, \dots, 2i_d + 1$. By the complementarity condition $\bar{\varsigma}_i (\frac{1}{2} |\bar{\mathbf{x}}_i|^2 - \lambda) = 0$, we know that the primal problem (\mathcal{P}_q) has at most $2i_d + 1$ KKT points $\bar{\mathbf{x}}_i$ on the sphere $\frac{1}{2} |\bar{\mathbf{x}}_i|^2 = \lambda$. If $a_{i_d+1} > 0$, the equality $\psi(\varsigma) = \lambda$ may have at most $2i_d$ strictly positive roots.

By using the triality theory, the extremality conditions of the critical points of the problem (\mathcal{P}_q) can be identified by the following result.

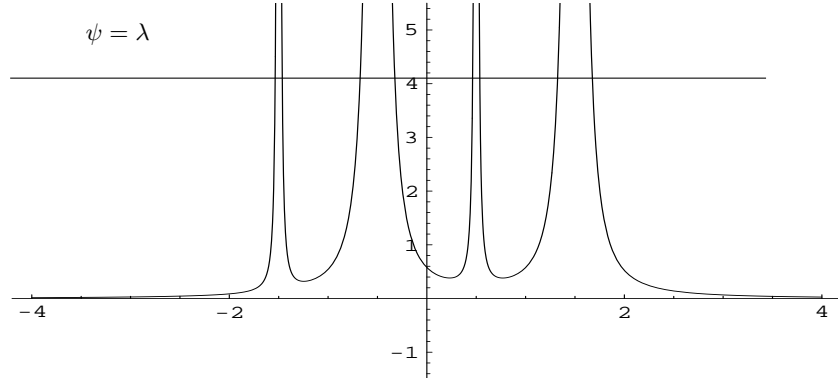


Fig. 1.11. Graph of $\psi(\zeta)$.

Theorem 15 (Global and Local Extrema [51]) *Suppose that a_1 is the smallest eigenvalue of A . Then the dual problem (\mathcal{P}_q^d) given in (1.138) has a unique solution $\bar{\zeta}_1$ over the domain $\zeta > -a_1 \geq 0$, and $\bar{\mathbf{x}}_1$ is a global minimizer of the problem (\mathcal{P}_q) , that is*

$$P(\bar{\mathbf{x}}_1) = \min_{\mathbf{x} \in \mathcal{U}_k} P(\mathbf{x}) = \max_{\zeta > -a_1} P^d(\zeta) = P^d(\bar{\zeta}_1). \quad (1.146)$$

If in each interval $(-a_{i+1}, -a_i)$, $i = 1, \dots, i_d$, the dual algebraic equation (1.139) has two roots $-a_{i+1} < \bar{\zeta}_{2i+1} < \bar{\zeta}_{2i} < -a_i$, then $\bar{\zeta}_{2i}$ is a local minimizer of $P^d(\zeta)$, and $\bar{\zeta}_{2i+1}$ is a local maximizer of $P^d(\zeta)$ over the interval $(-a_{i+1}, -a_i)$.

Proof. Since for any given $\zeta > -a_1$, the matrix $A + \zeta I$ is positive definite, that is the total complementary function $\Xi(\mathbf{x}, \zeta)$ is a saddle function, the saddle minmax theorem leads to (1.146).

The remaining statements in Theorem 15 can be proved by the graph of $P^d(\zeta)$ (see Fig. 1.12). \square

It is interesting to note that on the effective domain \mathcal{V}_a^* , the Fenchel-Young equality $V(\xi) = \langle \xi; \zeta \rangle - V^*(\zeta) = (\xi - \lambda)\zeta$ holds true. Thus, on $\mathcal{U}_a \times \mathcal{V}_a^*$, the total complementary function

$$\Xi(\mathbf{x}, \zeta) = \langle A(\mathbf{x}); \zeta \rangle - V^*(\zeta) - U(\mathbf{x}) = \zeta \left(\frac{1}{2} |\mathbf{x}|^2 - \lambda \right) + \frac{1}{2} \mathbf{x}^T A \mathbf{x} - \mathbf{x}^T f \quad (1.147)$$

can be viewed as the traditional Lagrangian of the quadratic minimization problem with the reformulated (canonical) quadratic constraint $\frac{1}{2} |\mathbf{x}|^2 \leq \lambda$, which is also called *extended Lagrangian* (see [40]). This example exhibits a connection between the nonlinear Lagrange multiplier method and the canonical dual transformation. Based on this observation, the traditional Lagrange

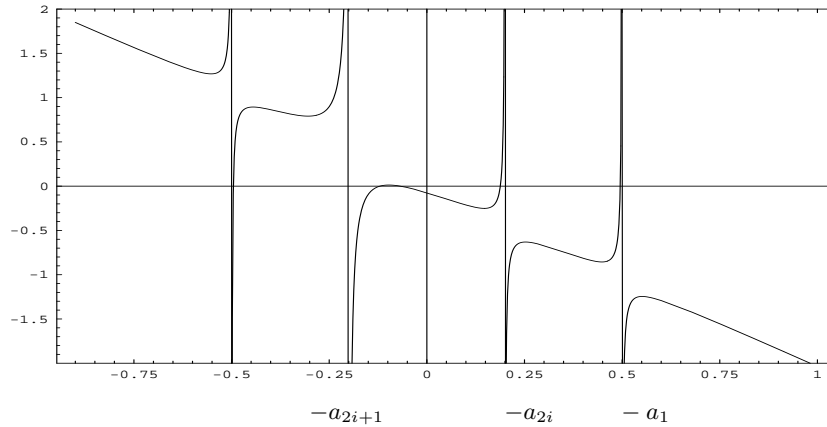


Fig. 1.12. Graph of $P^d(\zeta)$

multiplier method can be generalized to solve constrained global optimization problems.

1.8 General Constrained Global Optimization Problems

In this section, we present an important application of the canonical duality theory to the following general constrained nonlinear programming problem

$$\min \{-U(\mathbf{x}) : \mathbf{x} \in \mathcal{U}_k\}, \quad (1.148)$$

where $U(\mathbf{x})$ is a Gâteaux differentiable function, either linear or canonical function, defined on an open convex set $\mathcal{U}_a \subset \mathbb{R}^n$, and the feasible space \mathcal{U}_k is a convex subset of \mathcal{U}_a defined by

$$\mathcal{U}_k = \{\mathbf{x} \in \mathcal{U}_a \subset \mathbb{R}^n \mid g_i(\mathbf{x}) \leq 0, i = 1, \dots, p\},$$

in which $g_i(\mathbf{x}) : \mathcal{U}_a \rightarrow \mathbb{R}$ are convex functions. We will show the connection between the canonical dual transformation and nonlinear Lagrange multiplier methods and how to use the triality theory to identify global and local optima.

1.8.1 Canonical Form and Total Complementary Function

First, we need to put this problem in the framework of the canonical systems. Let the geometrical operator $\boldsymbol{\xi} = \Lambda(\mathbf{x}) = \{g_i(\mathbf{x})\} : \mathcal{U}_a \rightarrow \mathcal{V}_a \subset \mathbb{R}^p$ be a vector-valued function. The generalized canonical function

$$V(\boldsymbol{\xi}) = \begin{cases} 0 & \text{if } \boldsymbol{\xi} \leq 0 \\ \infty & \text{otherwise} \end{cases}$$

is an indicator of the convex cone $\mathcal{V}_a = \{\boldsymbol{\xi} \in \mathbb{R}^p \mid \boldsymbol{\xi} \leq 0\}$. Thus, the canonical form of the constrained problem (1.148) is

$$\min\{H(\mathbf{x}) = V(\Lambda(\mathbf{x})) - U(\mathbf{x}) : \mathbf{x} \in \mathcal{U}_a\}.$$

By the Fenchel transformation, the conjugate of $V(\boldsymbol{\xi})$ is an indicator of the dual cone $\mathcal{V}_a^* = \{\boldsymbol{\varsigma} \in \mathbb{R}^p \mid \boldsymbol{\varsigma} \geq 0\}$, that is

$$V^\sharp(\boldsymbol{\varsigma}) = \max\{\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle - V(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \mathbb{R}^p\} = \begin{cases} 0 & \text{if } \boldsymbol{\varsigma} \geq 0 \\ \infty & \text{otherwise.} \end{cases}$$

By the theory of convex analysis we have

$$\boldsymbol{\varsigma} \in \partial^- V(\boldsymbol{\xi}) \Leftrightarrow \boldsymbol{\xi} \in \partial^- V^\sharp(\boldsymbol{\varsigma}) \Leftrightarrow \langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle = V(\boldsymbol{\xi}) + V^\sharp(\boldsymbol{\varsigma}), \quad (1.149)$$

that is $(\boldsymbol{\xi}, \boldsymbol{\varsigma})$ is a generalized canonical pair on $\mathcal{U}_a \times \mathcal{V}_a^*$ [42]. Thus, the extended Lagrangian $\Xi(\mathbf{x}, \boldsymbol{\varsigma}) = \langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - V^\sharp(\boldsymbol{\varsigma}) - U(\mathbf{x})$ in this problem has a very simple form:

$$\Xi(\mathbf{x}, \boldsymbol{\varsigma}) = -U(\mathbf{x}) + \sum_{i=1}^p \varsigma_i g_i(\mathbf{x}). \quad (1.150)$$

We can see here that the canonical dual variable $\boldsymbol{\varsigma} \geq 0 \in \mathbb{R}^p$ is nothing but a Lagrange multiplier for the constraints $\Lambda(\mathbf{x}) = \{g_i(\mathbf{x})\} \leq 0$. Let

$$\mathcal{I}(\bar{\mathbf{x}}) := \{i \in \{1, \dots, p\} \mid g_i(\bar{\mathbf{x}}) = 0\}$$

be the index set of the active constraints at $\bar{\mathbf{x}}$. By the theory of global optimization (cf. [78]) we know that if $\bar{\mathbf{x}}$ is a local minimizer such that $\nabla g_i(\bar{\mathbf{x}})$, $i \in \mathcal{I}(\bar{\mathbf{x}})$, are linearly independent, then the KKT conditions hold:

$$g_i(\bar{\mathbf{x}}) \leq 0, \quad \bar{\varsigma}_i \geq 0, \quad \bar{\varsigma}_i g_i(\bar{\mathbf{x}}) = 0, \quad i = 1, \dots, p, \quad (1.151)$$

$$\nabla U(\bar{\mathbf{x}}) = \sum_{i=1}^p \bar{\varsigma}_i \nabla g_i(\bar{\mathbf{x}}). \quad (1.152)$$

Any point $(\bar{\mathbf{x}}, \bar{\boldsymbol{\varsigma}})$ that satisfies (1.151-1.152) is called a KKT stationary point of the problem (1.148). However, the KKT conditions (1.151-1.152) are only necessary for the minimization problem (1.148). They are sufficient for a constrained global minimum at $\bar{\mathbf{x}}$ provided that, for example, the functions $P(\mathbf{x}) = -U(\mathbf{x})$ and $g_i(\mathbf{x})$, $i = 1, \dots, p$, are convex. In constrained global optimization problems, the primal problems may possess many local minimizers due to the nonconvexity of the objective function and constraints. Therefore, sufficient optimality conditions play a key role in developing global algorithms. Here we show that the triality theory can provide such sufficient conditions.

The complementary function $V^\sharp(\boldsymbol{\varsigma}) = 0$, $\forall \boldsymbol{\varsigma} \in \mathcal{V}_a^*$, therefore in this constrained optimization problem we have

$$G_\zeta(\mathbf{x}) = \Xi(\mathbf{x}, \boldsymbol{\varsigma}) = -U(\mathbf{x}) + \boldsymbol{\varsigma}^T \Lambda(\mathbf{x}). \quad (1.153)$$

For a fixed $\boldsymbol{\varsigma} \in \mathcal{V}_a^*$, if the parametric function $G_\zeta : \mathcal{U}_a \rightarrow \mathbb{R}$ is twice Gâteaux differentiable, the space \mathcal{G} can be written as

$$\mathcal{G} = \left\{ (\mathbf{x}, \boldsymbol{\varsigma}) \in \mathcal{U}_a \times \mathcal{V}_a^* \mid \det \left(\frac{\partial^2 G_\zeta(\mathbf{x})}{\partial x_i \partial x_j} \right) \neq 0 \right\}.$$

Clearly for any given $(\mathbf{x}, \boldsymbol{\varsigma}) \in \mathcal{G}$, the dual feasible space \mathcal{V}_k^*

$$\mathcal{V}_k^* = \left\{ \boldsymbol{\varsigma} \in \mathcal{V}_a^* \mid \Lambda_t^*(\mathbf{x}) \boldsymbol{\varsigma} = \sum_{i=1}^p \varsigma_i \nabla g_i(\mathbf{x}) = \nabla U(\mathbf{x}), \forall \mathbf{x} \in \mathcal{U}_a \right\} \quad (1.154)$$

is non-empty and the Λ -conjugate transformation

$$U^\Lambda(\boldsymbol{\varsigma}) = \text{sta} \{ \langle \Lambda(\mathbf{x}); \boldsymbol{\varsigma} \rangle - U(\mathbf{x}) : \forall \mathbf{x} \in \mathcal{U}_a \}$$

can be well-formulated on \mathcal{V}_k^* . Thus, the canonical dual problem can be proposed as the following,

$$\max \{ P^d(\boldsymbol{\varsigma}) = -U^\Lambda(\boldsymbol{\varsigma}) : \boldsymbol{\varsigma} \in \mathcal{V}_k^* \}. \quad (1.155)$$

In the following, we illustrate the foregoing results using some examples.

1.8.2 Quadratic Minimization with Quadratic Constraints

Let $U(\mathbf{x}) = \mathbf{x}^T f - \frac{1}{2} \mathbf{x}^T A \mathbf{x}$ and $g(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T C \mathbf{x} - \lambda$ be quadratic functions, where A and C are two symmetrical matrices in $\mathbb{R}^{n \times n}$, $f \in \mathbb{R}^n$ is a given vector, and $\lambda \in \mathbb{R}$ is a given constant. Thus the primal problem is:

$$\min \left\{ P(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - f^T \mathbf{x} : \frac{1}{2} \mathbf{x}^T C \mathbf{x} \leq \lambda, \mathbf{x} \in \mathbb{R}^n \right\}. \quad (1.156)$$

Because we have only one constraint $g(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T C \mathbf{x} - \lambda$, the extended Lagrangian is simply

$$\Xi(\mathbf{x}, \varsigma) = \frac{1}{2} \mathbf{x}^T (A + \varsigma C) \mathbf{x} - f^T \mathbf{x} - \varsigma \lambda. \quad (1.157)$$

On the dual feasible space

$$\mathcal{V}_k^* = \{ \varsigma \in \mathbb{R} \mid \varsigma \geq 0, \det(A + \varsigma C) \neq 0 \},$$

and the canonical dual problem (1.155) can be formulated as (see [52]):

$$\max \left\{ P^d(\zeta) = -\frac{1}{2}f^T(A + \zeta C)^{-1}f - \lambda\zeta \quad : \quad \zeta \in \mathcal{V}_k^* \right\}. \quad (1.158)$$

Since in this problem both $A(\mathbf{x}) = (\frac{1}{2}\mathbf{x}^T C \mathbf{x} - \lambda)$ and $U(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T A \mathbf{x} + f^T \mathbf{x}$ are quadratic functions, $\delta^2 G_\zeta = (A + \zeta C)$. The following result was obtained recently.

Theorem 16 (Gao [52]) *Suppose that the matrix C is positive definite, and $\bar{\zeta} \in \mathcal{V}_a^*$ is a critical point of $P^d(\zeta)$. If $A + \bar{\zeta}C$ is positive definite, the vector*

$$\bar{\mathbf{x}} = (A + \bar{\zeta}C)^{-1}f$$

is a global minimizer of the primal problem (1.156). However, if $A + \bar{\zeta}C$ is negative definite, the vector $\bar{\mathbf{x}} = (A + \bar{\zeta}C)^{-1}f$ is a local minimizer of the primal problem (1.156).

In 2-D space, if we let $a_{11} = 3, a_{12} = a_{21} = .5, a_{22} = -2.0$, and $c_{11} = 1, c_{12} = c_{21} = 0, c_{22} = 0.5$, the matrix $A = \{a_{ij}\}$ is indefinite, while $C = \{c_{ij}\}$ is positive definite. Setting $f = \{1, 1.5\}$ and $\lambda = 2$, the graph of the canonical function $P(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A \mathbf{x} - \mathbf{x}^T f$ is a saddle surface (see Fig. 1.13), and the boundary of the feasible set $\mathcal{U}_k = \{\mathbf{x} \in \mathbb{R}^2 \mid \frac{1}{2}\mathbf{x}^T C \mathbf{x} \leq \lambda\}$ is an ellipse (see Fig. 1.13). In this case, the dual problem has four critical points (see Fig. 1.14):

$$\bar{\zeta}_1 = 5.22 > \bar{\zeta}_2 = 3.32 > \bar{\zeta}_3 = -2.58 > \bar{\zeta}_4 = -3.97.$$

Since $\bar{\zeta}_1 \in \mathcal{V}_+^*$ and $\bar{\zeta}_4 \in \mathcal{V}_-^*$, the triality theory tells us that $\mathbf{x}_1 = \{-0.22, 2.81\}$ is a global minimizer, and $\mathbf{x}_4 = \{-1.90, -0.85\}$ is a local minimizer. From the graph of $P^d(\zeta)$ we can see that $\mathbf{x}_2 = \{0.59, -2.70\}$ is a local minimizer, and $\mathbf{x}_3 = \{2.0, 0.15\}$ is a local maximizer. We have

$$P(\mathbf{x}_1) = -12.44 < P(\mathbf{x}_2) = -4.91 < P(\mathbf{x}_3) = 4.03 < P(\mathbf{x}_4) = 9.53.$$

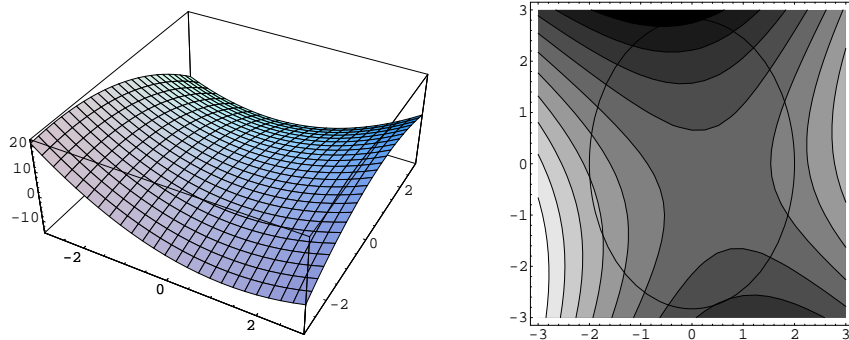
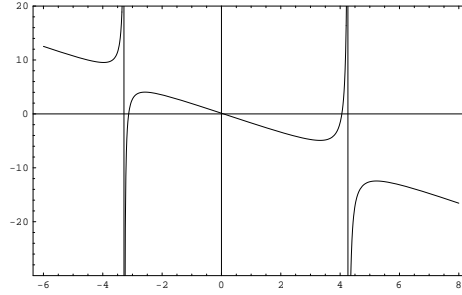


Fig. 1.13. Graph of $P(\mathbf{x})$ (left), Contours of $P(\mathbf{x})$ and boundary of \mathcal{U}_k (right)

Fig. 1.14. Graphs of $P^d(\varsigma)$

1.8.3 Quadratic Minimization with Box Constraints

The primal problem solved in this section is finding a global minimizer of a nonconvex quadratic function over a box constraint:

$$(\mathcal{P}_b) : \min \left\{ P(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - f^T \mathbf{x} \quad : \quad \ell^l \leq \mathbf{x} \leq \ell^u \right\}, \quad (1.159)$$

where $\mathbf{x} \in \mathbb{R}^n$, and ℓ^l , ℓ^u are two given vectors in \mathbb{R}^n . Problems of the form (1.159) appear frequently in partial differential equations, discretized optimal control problems, linear least squares problems, and certain successive quadratic programming methods (cf. Floudas and Visweswaran, 1995). Particularly, if $\ell^l = \mathbf{0}$ and $\ell^u = \mathbf{1}$, the problem (\mathcal{P}_b) is directly related to one of the fundamental problems of *combinatorial optimization*, namely, a continuous relaxation to the problem of minimizing a quadratic function in 0-1 variables.

In order to solve this problem, we need to reformulate the constraints in canonical form. Without loss of generality, we assume that $\ell^l = -\mathbf{1}$ and $\ell^u = \mathbf{1}$ (if necessary, a simple linear transformation can be used to convert the problem to this form).

$$\min \left\{ P(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T A \mathbf{x} - f^T \mathbf{x} \quad : \quad x_i^2 \leq 1, \quad i = 1, \dots, n \right\}. \quad (1.160)$$

The constraint in this problem is a vector-valued quadratic function

$$\boldsymbol{\xi} = A(\mathbf{x}) = \{g_i(\mathbf{x})\} = \{x_i^2 - 1\} \leq 0 \in \mathbb{R}^n. \quad (1.161)$$

Thus, the canonical dual variable $\boldsymbol{\varsigma} = \{\varsigma_i\}$ should also be a vector in \mathbb{R}^n . It has been shown recently that on the dual feasible space,

$$\mathcal{V}_k^* = \{\boldsymbol{\varsigma} \in \mathbb{R}^n \mid \boldsymbol{\varsigma} \geq 0, \det(A + 2 \text{Diag}(\boldsymbol{\varsigma})) \neq 0\},$$

where $\text{Diag}(\boldsymbol{\varsigma}) \in \mathbb{R}^{n \times n}$ represents a diagonal matrix with ς_i , $i = 1, \dots, n$ as its diagonal entries; the canonical dual problem is given by (see [55, 56])

$$\max \left\{ P^d(\boldsymbol{\varsigma}) = -\frac{1}{2}f^T(A + 2 \text{Diag}(\boldsymbol{\varsigma}))^{-1}f - \sum_{i=1}^n \varsigma_i \quad : \quad \boldsymbol{\varsigma} \in \mathcal{V}_k^* \right\}. \quad (1.162)$$

This dual problem can be solved to obtain all the critical points $\bar{\boldsymbol{\varsigma}}$. It is shown in [55, 56] that if

$$\bar{\boldsymbol{\varsigma}} \in \mathcal{V}_+^* = \{ \boldsymbol{\varsigma} \in \mathbb{R}^n \mid \boldsymbol{\varsigma} \geq 0, \ A + 2 \text{Diag}(\boldsymbol{\varsigma}) \text{ is positive definite} \},$$

then the vector $\bar{\mathbf{x}}(\bar{\boldsymbol{\varsigma}}) = (A + 2 \text{Diag}(\bar{\boldsymbol{\varsigma}}))^{-1}f$ is a global minimizer of the primal problem.

1.8.4 Integer Programming

The most simple integer constrained quadratic programming problem could be

$$(\mathcal{P}_{ip}) : \min \left\{ P(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, A\mathbf{x} \rangle - \langle \mathbf{x}, f \rangle \mid \mathbf{x} \in \{-1, 1\}^n \right\}, \quad (1.163)$$

where the notation $\{-1, 1\}^n$ denotes integer vectors of \mathbb{R}^n with components either -1 or 1 . From the point view of the canonical duality, the integer programming problem (\mathcal{P}_{ip}) is actually a special case of the box constrained minimization problem (\mathcal{P}_b) if the inequality constraint $\boldsymbol{\xi} \in \mathbb{R}_+^n$ in (1.161) is replaced by $\boldsymbol{\xi}(\mathbf{x}) = \{x_i^2 - 1\}^n = 0$. From the complementarity condition

$$\boldsymbol{\xi} = \{x_i^2 - 1\}^n \leq 0, \quad \boldsymbol{\varsigma} \geq 0, \quad (x_i^2 - 1)\varsigma_i = 0, \quad i = 1, \dots, n \quad (1.164)$$

we know that if the canonical dual variable $\boldsymbol{\varsigma} > 0 \in \mathbb{R}^n$, then the primal variable $\mathbf{x} \in \{-1, 1\}^n$ has to be integer. Thus, on the dual feasible space

$$\mathcal{S}_{ip}^+ = \{ \boldsymbol{\varsigma} \in \mathbb{R}^n \mid \boldsymbol{\varsigma} > \mathbf{0}, \ f \in \mathcal{C}_{ol}(\mathbf{G}_b(\boldsymbol{\varsigma})), \ \mathbf{G}_b(\boldsymbol{\varsigma}) \succeq 0 \}, \quad (1.165)$$

the canonical dual problem of (\mathcal{P}_{ip}) was formulated in [56] as

$$(\mathcal{P}_{ip}^d) : \max_{\boldsymbol{\varsigma} \in \mathcal{S}_{ip}^+} \left\{ P^d(\boldsymbol{\varsigma}) = -\frac{1}{2}\langle \mathbf{G}_b^\dagger(\boldsymbol{\varsigma})f, f \rangle - \langle \mathbf{e}, \boldsymbol{\varsigma} \rangle \right\}. \quad (1.166)$$

Theorem 17 *If $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_{ip}^+$ is a critical point of the canonical dual function $P^d(\boldsymbol{\varsigma})$, then the vector $\bar{\mathbf{x}} = \mathbf{G}_b^\dagger(\bar{\boldsymbol{\varsigma}})f$ is a global minimizer of the primal problem (\mathcal{P}_{ip}) , i.e.,*

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x} \in \{-1, 1\}^n} P(\mathbf{x}) = \max_{\boldsymbol{\varsigma} \in \mathcal{S}_{ip}^+} P^d(\boldsymbol{\varsigma}) = P^d(\bar{\boldsymbol{\varsigma}}). \quad (1.167)$$

This theorem shows that by the canonical duality theory, the discrete integer programming problems can be transformed to a concave maximization problem in continuous convex space. Detailed study on canonical duality theory for solving constrained Boolean and multi-integer programming problems can be found in [19, 56, 59, 140].

1.8.5 Concave Minimization

The primal problem in this case is given by

$$(\mathcal{P}_c) : \min\{P(\mathbf{x}) = -U(\mathbf{x}) : B\mathbf{x} \leq \mathbf{b}, \mathbf{x} \in \mathbb{R}^n\}, \quad (1.168)$$

where $U(\mathbf{x})$ is a convex, or even nonsmooth function, and where $B \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are given. It is well-known that this problem is NP-hard. Concave minimization problems constitute one of the most fundamental and intensely studied classes of problems in global minimization. A comprehensive review/survey of the mathematical properties, common applications, and solution methods is given by Benson [7]. By the use of the canonical dual transformation, a perfect dual problem has been formulated in [52]. In order to provide insights into the connection between the canonical dual transformation and the traditional Lagrange multiplier method, we demonstrate here how this perfect dual formulation can also be reproduced by the classical Lagrangian duality approach when executed in a particular fashion inspired by the canonical duality.

First, let us introduce a parameter μ such that

$$\min\{P(\mathbf{x}) : B\mathbf{x} \leq \mathbf{b}\} \leq \mu \leq \max\{P(\mathbf{x}) : B\mathbf{x} \leq \mathbf{b}\}.$$

Then the parameterized canonical form of this problem can be formulated as (see [52])

$$(\mathcal{P}_\mu) : \min\{P(\mathbf{x}) = -U(\mathbf{x}) : \{U + \mu, B\mathbf{x} - \mathbf{b}\} \leq 0 \in \mathbb{R}^{1+m}, \mathbf{x} \in \mathbb{R}^n\}. \quad (1.169)$$

In this case, the constraint $g_1(\mathbf{x}) = U(\mathbf{x}) + \mu$ is convex and $\{g_i(\mathbf{x}), i = 2, \dots, m+1\} = B\mathbf{x} - \mathbf{b}$ are linear. By introducing Lagrange multipliers $(\varsigma, \mathbf{y}) \in \mathbb{R}^{1+m}$, and letting

$$\mathcal{V}_a^* = \{(\varsigma, \mathbf{y}) \in \mathbb{R}^{1+m} \mid \varsigma \geq 0, \mathbf{y} \geq \mathbf{0} \in \mathbb{R}^m\},$$

the Lagrangian dual to the parameterized canonical problem (1.169) is given by

$$\Xi(\mathbf{x}, \varsigma, \mathbf{y}) = (\varsigma - 1)U(\mathbf{x}) + \mu\varsigma + \mathbf{y}^T(B\mathbf{x} - \mathbf{b}).$$

Thus, by the classical Lagrangian duality, the dual problem to (\mathcal{P}_μ) is

$$(LD) : \max_{(\varsigma, \mathbf{y}) \in \mathcal{V}_a^*} \{\mu\varsigma - \mathbf{y}^T \mathbf{b} + \min_{\mathbf{x}} \{(\varsigma - 1)U(\mathbf{x}) + \mathbf{y}^T B\mathbf{x}\}\}. \quad (1.170)$$

Because $U(\mathbf{x})$ is convex, the inner minimization problem in this dual form has a unique solution $\bar{\mathbf{x}}$ if $\varsigma > 1$.

Remark 1. Assume that

- (1) $U(\mathbf{x})$ is a convex function such that $\mathbf{x}^* = \delta U(\mathbf{x})$ is invertible for each $\mathbf{x} \in \mathbb{R}^n$, and the Legendre conjugate function $U^*(\mathbf{x}^*) = \text{sta}\{\mathbf{x}^T \mathbf{x}^* - U(\mathbf{x}) : \delta U(\mathbf{x}) = \mathbf{x}^*\}$ is uniquely defined in \mathbb{R}^n
- (2) An optimum solution $\bar{\mathbf{x}}$ to the problem (\mathcal{P}_μ) is a KKT solution with Lagrange multipliers $\bar{\varsigma} > 1, \bar{\mathbf{y}} \geq \mathbf{0} \in \mathbb{R}^m$.

Let

$$\mathcal{V}_+^* = \{(\varsigma, \mathbf{y}) \in \mathbb{R}^{1+m} \mid \varsigma > 1, \mathbf{y} \geq \mathbf{0} \in \mathbb{R}^m\}.$$

Under Remark 1, thus, we can write (LD) in (1.170) as

$$(LD) : \max_{(\varsigma, \mathbf{y}) \in \mathcal{V}_+^*} \left\{ \mu\varsigma - \mathbf{y}^T \mathbf{b} + (\varsigma - 1) \min_{\mathbf{x}} \left\{ \frac{\mathbf{y}^T B \mathbf{x}}{\varsigma - 1} + U(\mathbf{x}) \right\} \right\}. \quad (1.171)$$

Observe that the effect of having introduced $U(\mathbf{x}) + \mu \leq 0$ is to convexify the inner minimization problem in (1.171), which, by the assumption of Remark 1, reduces (LD) to the following equivalent dual problem.

$$(\mathcal{P}_\mu^d) : \max_{(\varsigma, \mathbf{y}) \in \mathcal{V}_+^*} \left\{ P^d(\varsigma, \mathbf{y}) = \mu\varsigma - \mathbf{y}^T \mathbf{b} + (1 - \varsigma) U^* \left(\frac{B^T \mathbf{y}}{1 - \varsigma} \right) \right\}. \quad (1.172)$$

This is the dual problem proposed by the canonical dual transformation in [52]. By the fact that the Legendre conjugate $U^*(\mathbf{x}^*)$ of the convex function $U(\mathbf{x})$ is also convex, this canonical dual is a concave maximization problem over the dual feasible space \mathcal{V}_+^* , which can be solved uniquely for a given parameter $\mu \in \mathbb{R}$ if \mathcal{V}_+^* is nonempty.

Under Remark 1, note that $\bar{\mathbf{x}}$ solves the primal problem (\mathcal{P}_μ) because $P(\bar{\mathbf{x}}) = \mu$, and satisfies the KKT conditions

$$(\bar{\varsigma} - 1)\delta U(\bar{\mathbf{x}}) + B^T \bar{\mathbf{y}} = 0, \quad (1.173)$$

$$B\bar{\mathbf{x}} \leq \mathbf{b}, \quad U(\bar{\mathbf{x}}) + \mu = 0, \quad \bar{\mathbf{y}}^T (B\bar{\mathbf{x}} - \mathbf{b}) = 0, \quad \bar{\mathbf{y}} \geq \mathbf{0}, \quad \bar{\varsigma} > 1. \quad (1.174)$$

Writing the (LD) in (1.170) as

$$\max_{(\varsigma, \mathbf{y}) \in \mathcal{V}_+^*} P_\theta^d(\varsigma, \mathbf{y}),$$

where

$$P_\theta^d(\varsigma, \mathbf{y}) = \mu\varsigma - \mathbf{y}^T \mathbf{b} + \min_{\mathbf{x}} \{ (\varsigma - 1)U(\mathbf{x}) + \mathbf{y}^T B \mathbf{x} \},$$

we get

$$P_\theta^d(\bar{\varsigma}, \bar{\mathbf{y}}) = \bar{\varsigma}\mu - \mathbf{b}^T \bar{\mathbf{y}} + (\bar{\varsigma} - 1)U(\hat{\mathbf{x}}) + \bar{\mathbf{y}}^T B \hat{\mathbf{x}}, \quad (1.175)$$

where $\hat{\mathbf{x}}$ satisfies $\delta U(\hat{\mathbf{x}}) = B^T \bar{\mathbf{y}} / (1 - \bar{\varsigma})$. By (1.173) and the assumed invertibility of the canonical dual relation $\mathbf{x}^* = \delta U(\mathbf{x})$, we get $\hat{\mathbf{x}} = \bar{\mathbf{x}}$. Substituting this into (1.175) and using (1.174) yields $P_\theta^d(\bar{\varsigma}, \bar{\mathbf{y}}) = P(\bar{\mathbf{x}})$, that is there is zero duality gap. Furthermore, letting

$$\mathcal{U}_\mu = \{ \mathbf{x} \in \mathbb{R}^n \mid B \mathbf{x} \leq \mathbf{b}, \quad -U(\mathbf{x}) = \mu \},$$

we have the following result.

Theorem 18 (KKT Condition and Global Optimality) Under Remark 1, for a given parameter μ , if $(\bar{\varsigma}, \bar{\mathbf{y}}) \in \mathcal{V}_a^*$ is a KKT point of (\mathcal{P}_μ^d) such that

$$\bar{\mathbf{x}}^* = \frac{B^T \bar{\mathbf{y}}}{1 - \bar{\varsigma}},$$

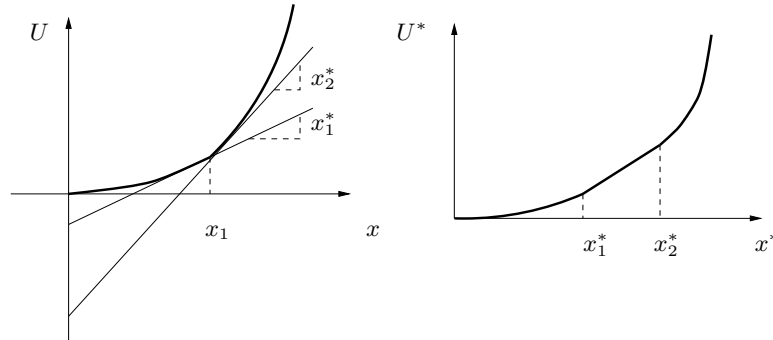
then the vector $\bar{\mathbf{x}} = \delta U^*(\bar{\mathbf{x}}^*)$ is a KKT point of (\mathcal{P}_μ) , and $P(\bar{\mathbf{x}}) = P^d(\bar{\varsigma}, \bar{\mathbf{y}})$.

Moreover, if $\bar{\varsigma} > 1$, then $(\bar{\varsigma}, \bar{\mathbf{y}})$ is a global maximizer of $P^d(\varsigma, \mathbf{y})$ on \mathcal{V}_+^* , $\bar{\mathbf{x}}$ is a global minimizer of $P(\mathbf{x})$ on the feasible space \mathcal{U}_μ , and

$$\min_{\mathbf{x} \in \mathcal{U}_\mu} P(\mathbf{x}) = \max_{(\varsigma, \mathbf{y}) \in \mathcal{V}_+^*} P^d(\varsigma, \mathbf{y}).$$

This example shows again that when a nonconvex constrained optimization problem can be written in a canonical form, the classical Lagrange multiplier method can be used to formulate a perfect dual problem. A detailed study on the canonical duality theory for solving general constrained nonconvex minimization problems and its connections with Lagrangian duality appears in [64].

One advantage of the canonical duality approach is that if the convex $U(\mathbf{x})$ is nonsmooth on \mathcal{U}_a , its Fenchel-Legendre conjugate U^* is a smooth function on \mathcal{U}_a^* (see Fig. 1.15). Such an idea has also been used in the study of geometrical dual analysis for solving nonsmooth “shape-preserving” design problems (see [9, 85, 143]).



(a) Graph of $U(x)$.

(b) Graph of the Legendre conjugate $U^*(x^*)$.

Fig. 1.15. Nonsmooth function and its smooth Legendre conjugate

1.9 Sequential Canonical Dual Transformation and Solutions to Polynomial Minimization Problems

The canonical dual transformation method can be generalized in different ways to solve the global optimization problem:

$$\min\{P(\mathbf{x}) = W(\mathbf{x}) - U(\mathbf{x}) : \mathbf{x} \in \mathcal{U}_a\} \quad (1.176)$$

with different types of nonconvex functions $W(\mathbf{x}) = V(\Lambda(\mathbf{x}))$ and geometrical operators Λ . If the geometrical operator $\Lambda : \mathcal{U} \rightarrow \mathcal{V}$ is a general nonlinear, nonconvex mapping, we can continue to use the canonical dual transformation such that the general nonconvex function $W(\mathbf{x})$ can be written in the canonical form (see [40]):

$$W(\mathbf{x}) = V(\Lambda(\mathbf{x})) = V_n(\xi_n(\xi_{n-1}(\dots(\xi_1(u))\dots))), \quad (1.177)$$

where $\xi_k(\xi_{k-1})$ is either a convex or a concave function of ξ_{k-1} , and we write

$$V_k(\xi_k) = \xi_{k+1}(\xi_k), \quad k = 1, \dots, n-1.$$

Thus, the geometrical operator $\Lambda : \mathcal{U} \rightarrow \mathcal{V}$ in this problem is a sequential composition of nonlinear mappings $\Lambda^{(k)} : \mathcal{V}_{k-1} \rightarrow \mathcal{V}_k, \quad k = 1, \dots, n, \mathcal{V}_0 = \mathcal{U},$ and $\mathcal{V}_n = \mathcal{V}$, that is

$$\xi_n(\mathbf{x}) = \Lambda(\mathbf{x}) = \left[\Lambda^{(n)} \circ \Lambda^{(n-1)} \circ \dots \circ \Lambda^{(1)} \right] (\mathbf{x}).$$

Because each $V_k(\xi_k)$ is a canonical function of ξ_k , the canonical duality relation $\varsigma_k = \delta V_k(\xi_k) : \mathcal{V}_k \rightarrow \mathcal{V}_k^*$ is one-to-one. It turns out that the Legendre conjugate

$$V_k^*(\varsigma_k) = \langle \xi_k; \varsigma_k \rangle - V_k(\xi_k)$$

can be uniquely defined. Letting $\boldsymbol{\varsigma} = \{\varsigma_i\} \in \mathbb{R}^n$, the sequential canonical Lagrangian associated with the general nonconvex problem (1.176) can be written as (see [40])

$$\Xi(\mathbf{x}, \boldsymbol{\varsigma}) = \langle \Lambda^{(1)}(\mathbf{x}); \varsigma_n! \rangle - V_w^*(\boldsymbol{\varsigma}) - U(\mathbf{x}), \quad (1.178)$$

where $\varsigma_p! := \varsigma_p \varsigma_{p-1} \cdots \varsigma_2 \varsigma_1$ and

$$V_w^*(\boldsymbol{\varsigma}) = V_n^*(\varsigma_n) + \varsigma_n V_{n-1}^*(\varsigma_{n-1}) + \cdots + \frac{\varsigma_n!}{\varsigma_1} V_1^*(\varsigma_1). \quad (1.179)$$

Thus, the canonical dual problem can be formulated as:

$$\max\{P^d(\boldsymbol{\varsigma}) = U^{\Lambda^{(1)}}(\boldsymbol{\varsigma}) - V_w^*(\boldsymbol{\varsigma}) : \boldsymbol{\varsigma} \in \mathcal{V}_k^*\}. \quad (1.180)$$

For certain given canonical functions V , and U , and the geometrical operator $A^{(1)}$, the A -conjugate transformation

$$U^{A^{(1)}}(\varsigma) = \text{sta}\{\langle A^{(1)}(\mathbf{x}); \varsigma_n! \rangle - U(\mathbf{x}) : \delta A^{(1)}(\mathbf{x})\varsigma_n! = \delta U(\mathbf{x})\}$$

can be well-defined on certain dual feasible spaces \mathcal{V}_k^* , and the canonical dual variables ς_k linearly depend on ς_1 . This canonical dual problem can be solved very easily. Two sequential canonical dual transformation methods have been proposed in Chapter 4 of [40]. Applications to general nonconvex differential equations and chaotic dynamical systems have been given in [35, 41].

As an application, let us consider the following polynomial minimization problem

$$\min\{P(\mathbf{x}) = W(\mathbf{x}) - \mathbf{x}^T f : \mathbf{x} \in \mathbb{R}^n\}, \quad (1.181)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ is a real vector, $f \in \mathbb{R}^n$ is a given vector, and $W(\mathbf{x})$ is a so-called *canonical polynomial* of degree $d = 2^{p+1}$ (see [40]), defined by

$$W(\mathbf{x}) = \frac{1}{2}\alpha_p \left(\frac{1}{2}\alpha_{p-1} \left(\dots \left(\frac{1}{2}\alpha_1 \left(\frac{1}{2}|\mathbf{x}|^2 - \lambda_1 \right)^2 \dots \right)^2 - \lambda_{p-1} \right)^2 - \lambda_p \right)^2, \quad (1.182)$$

where α_i, λ_i are given parameters. It is known that the general polynomial minimization problem is NP-hard even when $d = 4$ (see [97]). Many numerical methods and algorithms have been suggested recently for finding tight lower bounds of general polynomial optimization problems (see [84, 104]).

For the current canonical polynomial minimization problem, the dual problem has been formulated in [54]; that is

$$(\mathcal{P}^d) : \max_{\varsigma} \left\{ P^d(\varsigma) = -\frac{|f|^2}{2\varsigma_p!} - \sum_{k=1}^p \frac{\varsigma_p!}{\varsigma_k!} V_k^*(\varsigma_k) \right\}, \quad (1.183)$$

where

$$\varsigma_1 = \varsigma, \quad \varsigma_k = \alpha_k \left(\frac{1}{2\alpha_{k-1}} \varsigma_{k-1}^2 - \lambda_k \right), \quad k = 2, \dots, p. \quad (1.184)$$

In this case, $V_k^*(\varsigma_k)$ is a quadratic function of ς_k defined by

$$V_k^*(\varsigma_k) = \frac{1}{2\alpha_k} \varsigma_k^2 + \lambda_k \varsigma_k.$$

The dual problem is a nonlinear program having only one variable $\varsigma \in \mathbb{R}$, which is much easier to solve than the primal problem. Clearly, for any $\varsigma \neq 0$ and $\varsigma_k^2 \neq 2\alpha_k \lambda_{k+1}$, the dual function P^d is well-defined and the criticality condition $\delta P^d(\varsigma) = 0$ leads to a dual algebraic equation

$$2(\varsigma_p!)^2(\alpha_1^{-1}\varsigma + \lambda_1) = |f|^2. \quad (1.185)$$

Theorem 19 (Complete Solution Set to Canonical Polynomial [54])

For any parameters α_k , and $\lambda_k, k = 1, \dots, p$, and input f , the dual algebraic equation (1.185) has at most $s = 2^{p+1} - 1$ real solutions: $\bar{\zeta}^{(i)}, i = 1, \dots, s$. For each dual solution $\bar{\zeta} \in \mathbb{R}$, the vector $\bar{\mathbf{x}}$ defined by

$$\bar{\mathbf{x}}(\bar{\zeta}) = (\bar{\zeta}_p!)^{-1} f \quad (1.186)$$

is a critical point of the primal problem (\mathcal{P}) and

$$P(\bar{\mathbf{x}}) = P^d(\bar{\zeta}).$$

Conversely, every critical point $\bar{\mathbf{x}}$ of the polynomial $P(\mathbf{x})$ can be written in the form (1.186) for some dual solution $\bar{\zeta} \in \mathbb{R}$.

In the case that $p = 1$, the nonconvex function $W(\mathbf{x}) = \frac{1}{2}\alpha_1(\frac{1}{2}|\mathbf{x}|^2 - \lambda_1)^2$ is a double-well function. The global and local extrema can be identified by the triality theory given in Theorem 6. For the general case of $p > 1$, the sufficient condition for global minimizer was obtained recently in [54].

Theorem 20 (Sufficient Condition for Global Minimzer) Suppose that for any arbitrarily given positive parameters $\alpha_k, \lambda_k \geq 0, \forall k \in \{1, \dots, p\}$, $\bar{\zeta}$ is a solution of the dual algebraic equation (1.185). If

$$\bar{\zeta} > \zeta_+ = \sqrt{2\alpha_1 \left(\lambda_2 + \sqrt{\frac{2}{\alpha_2} \left(\lambda_3 + \dots + \sqrt{\frac{2}{\alpha_{p-2}} \left(\lambda_{p-1} + \sqrt{\frac{2}{\alpha_{p-1}} \lambda_p} \right)} \right)} \right)} \right)},$$

then $\bar{\zeta}$ is a global maximizer of P^d on the open domain $(\zeta_+, +\infty)$, the vector $\bar{\mathbf{x}} = (\bar{\zeta}_p!)^{-1} f$ is a global minimizer of the polynomial minimization problem (1.181), and

$$P(\bar{\mathbf{x}}) = \min_{\mathbf{x} \in \mathbb{R}^n} P(\mathbf{x}) = \max_{\zeta > \zeta_+} P^d(\zeta) = P^d(\bar{\zeta}). \quad (1.187)$$

In the case of $p = 2$, the nonconvex function $W(\mathbf{x})$ is a canonical polynomial of degree eight. The dual function $P^d(\zeta)$ has the form of

$$\Pi^d(\zeta) = -\frac{|f|^2}{2\zeta\zeta_2} - \left(\frac{1}{\alpha_2}\zeta_2^2 + \lambda_2\zeta_2 + \zeta_2 \left(\frac{1}{2\alpha_1}\zeta^2 + \lambda_1\zeta \right) \right), \quad (1.188)$$

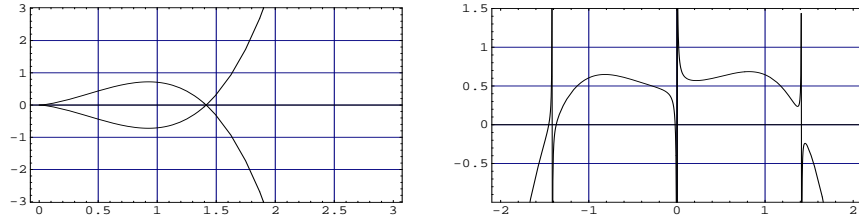
where $\zeta_2 = \alpha_2\zeta^2/(2\alpha_1) - \lambda_2\alpha_2$. In this case, the dual algebraic equation (1.185)

$$2\zeta^2 \left(\frac{\alpha_2}{2\alpha_1}\zeta^2 - \lambda_2\alpha_2 \right)^2 \left(\frac{1}{\alpha_1}\zeta + \lambda_1 \right) = |f|^2 \quad (1.189)$$

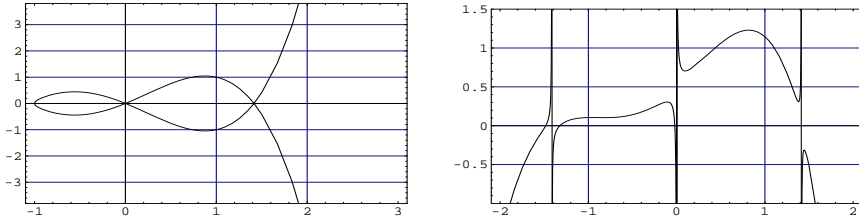
has at most seven real roots $\bar{\zeta}_i, i = 1, \dots, 7$. Let

$$\phi_2(\zeta) = \pm \zeta \left(\frac{\alpha_2}{2\alpha_1}\zeta^2 - \lambda_2\alpha_2 \right) \sqrt{2 \left(\frac{1}{\alpha_1}\zeta + \lambda_1 \right)},$$

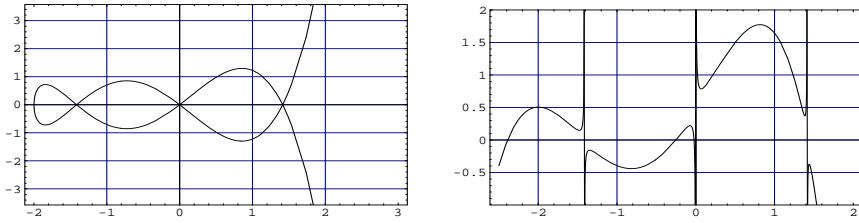
and $f = \{0.1, -0.1\}$, $\alpha_1 = 1$, $\alpha_2 = 1$, and $\lambda_2 = 1$. Then, for different values of λ_1 , the graphs of $\phi_2(\zeta)$ and $P^d(\zeta)$ are shown in Fig. 1.16. The graphs of $P(\mathbf{x})$ are shown in Fig. 1.17 (for $\lambda_1 = 0$ and $\lambda_1 = 1$) and Fig. 1.18 (for $\lambda_1 = 2$). Since $\zeta_+ = \sqrt{2\alpha_1\lambda_2} = \sqrt{2}$, we can see that the dual function $P^d(\zeta)$ is strictly concave for $\zeta > \zeta_+ = \sqrt{2}$. The dual algebraic equation (1.189) has a total of seven real solutions when $\lambda_1 = 2$, and the largest $\zeta_1 = 2.10 > \zeta_+ = 2$ gives the global minimizer $\mathbf{x}_1 = f/\zeta_1 = \{2.29, -0.92\}$, and $P(\mathbf{x}_1) = -1.32 = P^d(\zeta_1)$. The smallest $\zeta_7 = -4.0$ gives a local maximizer $\mathbf{x}_7 = \{-0.04, 0.02\}$ and $P(\mathbf{x}_7) = 4.51 = P^d(\zeta_7)$ (see Fig. 1.18).



(a) $\lambda_1 = 0$: Three solutions $\zeta_3 = 0.22 < \zeta_2 = 1.37 < \zeta_1 = 1.45$



(b) $\lambda_1 = 1$: Five solutions $\{-0.96, -0.11, 0.096, 1.38, 1.45\}$



(c) $\lambda_1 = 2$: Seven solutions $\{-2.0, -1.45, -1.35, -0.072, 0.07, 1.39, 1.44\}$

Fig. 1.16. Graphs of the algebraic curve $\phi_2(\zeta)$ (left) and dual function $P^d(\zeta)$ (right)

Detailed studies on solving general polynomial minimization problems are given in [40, 54, 84, 117, 118].

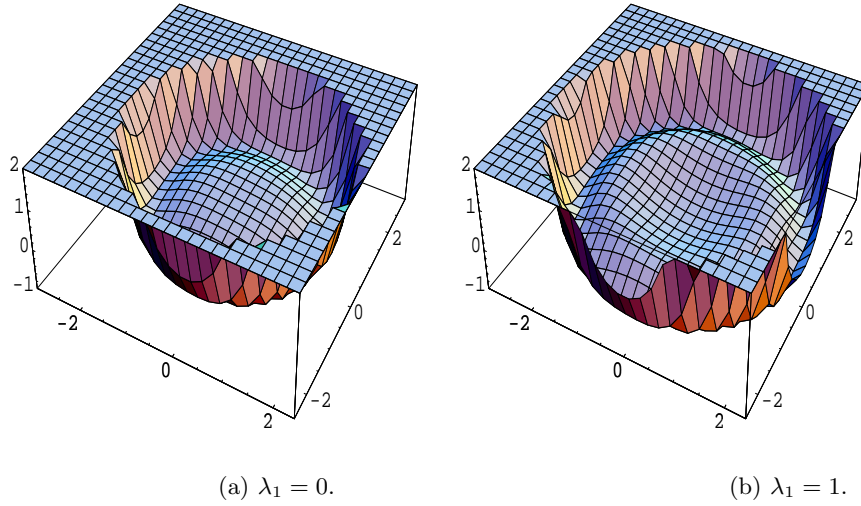


Fig. 1.17. Graphs of $P(x)$

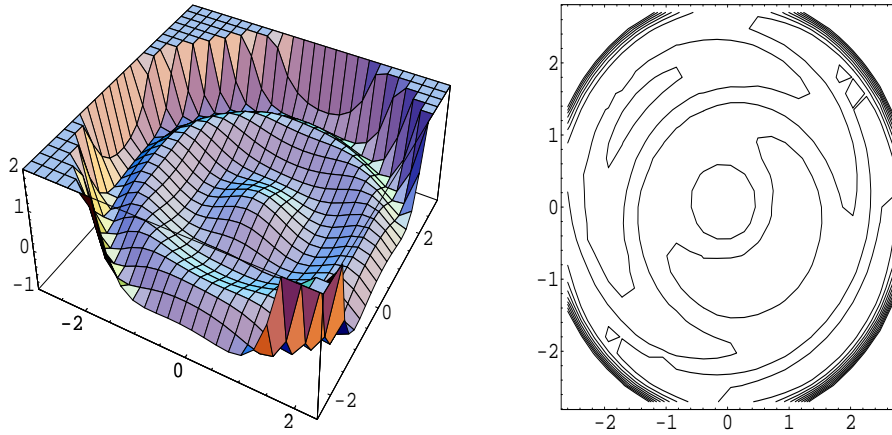


Fig. 1.18. Graph of $P(x)$ with $\lambda_1 = 2$

1.10 Concluding Remarks

We have presented a detailed review on the canonical dual transformation and its associated triality theory, with specific applications to nonconvex analysis and global optimization problems. Duality plays a key role in modern mathematics and science. The inner beauty of duality theory owes much to the fact that many different natural phenomena can be cast in the unified mathematical framework of Fig. 1. According to the traditional philosophical principle

of ying-yang duality; *The Complementarity of One Ying and One Yang is the Dao* (see [33, 83]), i.e., the constitutive relations in any physical system should be one-to-one. Niels Bohr realized its value in quantum mechanics. His complementarity theory and philosophy laid a foundation on which the field of modern physics was developed [103]. In nonconvex analysis and optimization, this one-to-one canonical duality relation serves as the foundation for the canonical dual transformation method. For any given nonconvex problem, as long as the geometrical operator Λ is chosen properly and the tri-canonical forms can be characterized correctly, the canonical dual transformation can be used to establish elegant theoretical results and to develop efficient algorithms for robust computations. The extended Lagrangian duality and triality theories show promise of having significance in many diverse fields.

As indicated in [40], duality in natural systems is a very broad and rich field. To theoretical scientists and philosophical thinkers as well as great artists, duality has always played a central role in their respective fields. It is really “a splendid feeling to realize the unity of a complex of phenomena that by physical perception appear to be completely separated” (Albert Einstein). It is pleasing to see that more and more knowledgeable researchers and scientists are working in this wonderland and exploring the intrinsic beauty of nature, often revealed via duality theory.

Acknowledgement. This work is supported by the National Science Foundation by Grant Numbers CCF-0514768, and AFOSR/NL.

References

1. Arthurs, A.M. (1980). *Complementary Variational Principles*, Clarendon Press, Oxford.
2. Atai, A.A. and Steigmann, D. (1998), Coupled deformations of elastic curves and surfaces, *Int. J. Solids and Structures*, **35**, 1915-1952.
3. Aubin, J.P. and Ekeland, I. (1976). Estimates of the duality gap in nonconvex optimization. *Math. Oper. Res.*, **1**, no. 3, 225-245.
4. Auchmuty, G. (1983). Duality for non-convex variational principles, *J. Diff. Equations*, **50**, pp 80-145.
5. Auchmuty, G. (1986). Dual variational principles for eigenvalue problems, *Proceedings of Symposia in Pure Math.*, **45**, Part 1, 55-71.
6. Auchmuty, G. (2001). Variational principles for self-adjoint elliptic eigenproblems, in *Nonconvex/Nonsmooth Mechanics: Modelling, Methods and Algorithms*, Gao, D.Y., R.W. Ogden and G. Stavroulakis (ed.) Kluwer Academic Publishers, 2000, 478pp.
7. Benson, H. (1995). Concave minimization: theory, applications and algorithms, in *Handbook of Global Optimization*, eds. R. Horst and P. Pardalos, Kluwer Academic Publishers, 43-148.
8. Casciaro, R. and Cascini, A. (1982). A mixed formulation and mixed finite elements for limit analysis, *Int. J. Solids and Struct.*, **19**, 169-184.
9. Cheng, H., Fang, S.C., and Lavery, J. (2005). Shape-preserving properties of univariate cubic L_1 splines, *J. Comput. Appl. Math.*, **174**, 361-382.
10. Chien, Wei-zang (1980). *Variational Methods and Finite Elements* (in Chinese). Science Press, 608pp.
11. Clarke, F.H. (1983). *Optimization and Nonsmooth Analysis*, John Wiley, New York.
12. Clarke, F.H. (1985). The dual action, optimal control, and generalized gradients, *Mathematical Control Theory*, Banach Center Publ., **14**, PWN, Warsaw, pp. 109-119.
13. Crouzeix, J.P. (1981). Duality framework in quasiconvex programming, in *Generalized Convexity in Optimization and Economics*, eds. S. Schaible and W.T. Ziemba, Academic Press, 207-226.
14. Dacorogna, D. (1989). *Direct Methods in the Calculus of Variations*. Springer-Verlag, New York.
15. Ekeland, I. (1977). Legendre duality in nonconvex optimization and calculus of variations, *SIAM J. Control and Optimization*, **15**, 905-934.
16. Ekeland, I (1990). *Convexity Methods in Hamiltonian Mechanics*, Springer-Verlag, New York, 247 pp.
17. Ekeland, I. (2003). Nonconvex duality, in *Proceedings of IUTAM Symposium on Duality, Complementarity and Symmetry in Nonlinear Mechanics*, D.Y. Gao (ed.), Kluwer Academic Publishers, Dordrecht/Boston/London, pp. 13-19.
18. Ekeland, I. and Temam, R. (1976). *Convex Analysis and Variational Problems*, North-Holland.
19. Fang, S.-C., Gao, D.Y., Sheu, R.-L., and Wu, S.Y. (2007). Canonical dual approach for solving 0-1 quadratic programming problems, to appear in *J. Industrial and Management Optimization*, **4**(1), 125-142.
20. Fang, S.-C., Gao, D.Y., Sheu, R.-L., and Xing, W.X. (2008). Global optimization for a class of fractional programming problems, *J. Global Optimization*, in press.

21. Floudas, C.A. and Visweswaran, V. (1995). Quadratic optimization, in *Handbook of Optimization*, R. Horst and P.M. Pardalos (eds), Kluwer Academic Publishers, Dordrecht/Boston/London, pp. 217-270.
22. Gao, D.Y. (1986). *Complementarity Principles in Nonsmooth Elastoplastic Systems and Pan-penalty Finite Element Methods*, Ph.D. Thesis, Tsinghua University, Beijing, China, 236 pp.
23. Gao, D.Y. (1988). On the complementary bounding theorems for limit analysis, *Int. J. of Solids Struct.*, **24**, 545-556.
24. Gao, D.Y. (1988). Panpenalty finite element programming for limit analysis, *Computers & Structures*, **28**, pp. 749-755.
25. Gao, D.Y. (1990). Dynamically loaded rigid-plastic analysis under large deformation, *Quart. Appl. Math.* **48**, pp. 731-739.
26. Gao, D.Y. (1990). On the extremum potential variational principles for geometrical nonlinear thin elastic shell, *Science in China (Scientia Sinica) (A)*, **33** (1), pp. 324-331.
27. Gao, D.Y. (1990). On the extremum variational principles for nonlinear elastic plates, *Quart. Appl. Math.*, **48**, pp. 361-370.
28. Gao, D.Y. (1990). Complementary principles in nonlinear elasticity, *Science in China (Scientia Sinica) (A) (Chinese Ed.)*, **33**(4), pp. 386-394.
29. Gao, D.Y. (1990). Bounding theorem on finite dynamic deformations of plasticity, *Mech. Research Commun.*, **17**, pp. 33-39.
30. Gao, D.Y. (1991). Extended bounding theorems for nonlinear limit analysis, *Int. J. Solids Structures*, **27**, pp. 523-531.
31. Gao, D.Y. (1992). Global extremum criteria for nonlinear elasticity, *J. Appl. Math. Physics (ZAMP)*, **43**, 924-937.
32. Gao, D.Y. (1996). Nonlinear elastic beam theory with applications in contact problem and variational approaches, *Mech. Research Commun.*, **23** (1), 11-17.
33. Gao, D.Y. (1996). Complementarity and duality in natural sciences, *Philosophical Study in Modern Science and Technology* (in Chinese). Tsinghua University Press, Beijing, China, 12-25.
34. Gao, D.Y. (1997). Dual extremum principles in finite deformation theory with applications to post-buckling analysis of extended nonlinear beam theory, *Appl. Mech. Rev.*, **50**, 11, November 1997, S64-S71.
35. Gao, D.Y. (1998). Duality, triality and complementary extremum principles in nonconvex parametric variational problems with applications, *IMA J. Appl. Math.*, **61**, 199-235.
36. Gao, D.Y. (1998). Bi-complementarity and duality: A framework in nonlinear equilibria with applications to the contact problems of elastoplastic beam theory, *J. Appl. Math. Anal.*, **221**, 672-697.
37. Gao, D.Y. (1999). Pure complementary energy principle and triality theory in finite elasticity, *Mech. Res. Comm.* **26** (1), 31-37.
38. Gao, D.Y. (1999). *Duality-Mathematics*, Wiley Encyclopedia of Electrical and Electronical Engineering, **6**, 68-77.
39. Gao, D.Y. (1999). General Analytic Solutions and Complementary Variational Principles for Large Deformation Nonsmooth Mechanics. *Meccanica* **34**, 169-198.
40. Gao, D.Y. (2000). *Duality Principles in Nonconvex Systems: Theory, Methods and Applications*, Kluwer Academic Publishers, Dordrecht /Boston /London, xviii + 454pp.

41. Gao, D.Y. (2000). Analytic solution and triality theory for nonconvex and nonsmooth variational problems with applications, *Nonlinear Analysis*, 42, 7, 1161-1193.
42. Gao, D.Y. (2000). Canonical dual transformation method and generalized triality theory in nonsmooth global optimization, *J. Global Optimization*, 17 (1/4), pp. 127-160.
43. Gao, D.Y.(2000). Finite deformation beam models and triality theory in dynamical post-buckling analysis, *Int. J. Non-Linear Mechanics*, 5, 103-131.
44. Gao, D.Y. (2001). Bi-Duality in Nonconvex Optimization, in *Encyclopedia of Optimization*, C. A. Floudas and P.D. Pardalos (eds). Kluwer Academic Publishers, Dordrecht/Boston/London, Vol. 1, pp. 477-482.
45. Gao, D.Y. (2001). Gao, D.Y., Tri-duality in Global Optimization, in *Encyclopedia of Optimization*, C. A. Floudas and P.D. Pardalos (eds). Kluwer Academic Publishers, Dordrecht/Boston/London, Vol. 1, pp. 485-491.
46. Gao, D.Y. (2001). Complementarity, polarity and triality in nonsmooth, nonconvex and nonconservative Hamilton systems, *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences*, 359, 2347-2367.
47. Gao, D.Y. (2002). Duality and triality in non-smooth, nonconvex and non-conservative systems: A survey, new phenomena and new results, in *Nonsmooth/Nonconvex Mechanics with Applications in Engineering*, edited by C. Baniotopoulos. Thessaloniki, Greece. pp. 1-14.
48. Gao, D.Y. (2003). Perfect duality theory and complete solutions to a class of global optimization problems, *Optimisation*, 52 (4-5), 467-493.
49. Gao, D.Y. (2003). Nonconvex semi-linear problems and canonical duality solutions, *Advances in Mechanics and Mathematics*, Kluwer Academic Publishers, Dordrecht/Boston/London, Vol. II, 261-312.
50. Gao, D.Y. (2004). Complementary variational principle, algorithm, and complete solutions to phase transitions in solids governed by Landau-Ginzburg equation, *Mathematics and Mechanics of Solids*, 9, 285-305.
51. Gao, D.Y. (2004). Canonical duality theory and solutions to constrained nonconvex quadratic programming, *J. Global Optimization*, 29, 377-399.
52. Gao, D.Y.(2005). Sufficient conditions and perfect duality in nonconvex minimization with inequality constraints, *J. Industrial and Management Optimization*, 1, 59-69.
53. Gao, D.Y. (2005). Canonical duality in nonsmooth, concave minimization with inequality constraints, *Advances in Nonsmooth Mechanics, a special volume in honor of Professor J.J. Moreau's 80th birthday*, P. Alart and O. Maisonneuve (eds). Springer, New York, 305-314.
54. Gao, D.Y. (2006). Complete solutions to a class of polynomial minimization problems, *J. Global Optimization*, 35, 131-143.
55. Gao, D.Y. (2007). Duality-Mathematics, *Wiley Encyclopidia of Electrical and Electronics Engineering*, Vol. 6, (Second Edition), John G. Webster (eds).
56. Gao, D.Y. (2007). Solutions and optimality to box constrained nonconvex minimization problems *J. Indust. and Manage. Optim.*, 3(2), 293-304.
57. Gao, D.Y. and Cheung, Y.K. (1989). On the extremum complementary energy principles for nonlinear elastic shells. *Int. J. Solids & Struct.*, 26, pp. 683-693.
58. Gao, D.Y. and Hwang, K.C. (1988). On the complementary variational principles for elasto-plasticity. *Scientia Sinica (A)*, 31, pp. 1469-1476.

59. Gao, D.Y. and Ruan, N. (2009). Solutions to quadratic minimization problems with box and integer constraints, to appear in *J. Global Optimization*.
60. Gao, D.Y. and Ogden, R.W. (2007). Closed-form solutions, extremality and nonsmoothness criteria in a large deformation elasticity problem, to appear in *Zeitschrift fr angewandte Mathematik und Physik*.
61. Gao, D. Y., R.W. Ogden and G. Stavroulakis (2001). *Nonsmooth and Nonconvex Mechanics: Modelling, Analysis and Numerical Methods*. Kluwer Academic Publishers, Boston/Dordrecht/London, 2001, xliv+471pp.
62. Gao, D.Y. and Onate, E.T. (1990). Rate variational extremum principles for finite elastoplasticity. *Appl. Math. Mech.*, 11(7), pp. 659-667.
63. Gao, D.Y. and Ruan, N. (2007). Complete solutions and optimality criteria for nonconvex quadratic-exponential minimization problem, to appear in *Math. Meth. of Oper. Res.*.
64. Gao, D.Y. and Sherali, H.D. (2007). Canonical duality theory for solving constrained global optimization problems, Manuscript, Department of Mathematics and ISE, Virginia Tech, Blacksburg.
65. Gao, D.Y. and Strang, G. (1989). Geometric nonlinearity: Potential energy, complementary energy, and the gap function, *Quart. Appl. Math.*, 47(3), 487-504.
66. Gao, D.Y. and Strang, G. (1989). Dual extremum principles in finite deformation elastoplastic analysis, *Acta Appl. Math.*, 17, pp. 257-267.
67. Gao, D.Y. and Wierzbicki, T. (1989). Bounding theorem in finite plasticity with hardening effect. *Quart. Appl. Math.*, 47, pp. 395-403.
68. Gao, D.Y. and Yang, W.-H. (1995). Multi-duality in minimal surface type problems, *Studies in Appl. Math.*, 95, 127-146.
69. Guo, Z.H. (1980). The unified theory of variational principles in nonlinear elasticity. *Archive of Mechanics*, 32, pp. 577-596.
70. Gasimov, R.N. (2002). Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming. *J. Global Optimization*, 24, 187-203.
71. Goh, C.J. and Yang, X.Q. (2002). *Duality in Optimization and Variational Inequalities*, Taylor and Francis, 329pp.
72. Greenberg, H.J. (1949). On the variational principles of plasticity. Brown University, ONR, NR-041-032, March.
73. Haar, A. and von Kármán, Th. (1909). Zur theorie der spannungszustände in plastischen und sandartigen medien. *Nachr. den Gesellsch der Wissensch. zu Göttingen*, 204-218.
74. Han, Weimin (2005). *A Posteriori Error Analysis Via Duality Theory With Applications in Modeling and Numerical Approximations*. Springer Book Series: Advances in Mechanics and Mathematics , Vol. 8, Springer, 302pp.
75. Hellinger, E. (1914). Die allgemeine Ansätze der Mechanik der Kontinua. *Encyklopädie der Mathematischen Wissenschaften IV*, 4, 602-94.
76. Hill, R. (1978), Aspects of invariance in solids mechanics, *Adv. in Appl. Mech.*, 18, 1-75.
77. Hiriart-Urruty, J.-B. (1985). Generalized differentiability, duality and optimization for problems dealing with difference of convex functions, *Appl. Mathematics and Optimization*, 6, 257-269.
78. Horst, R., Pardalos, P.M., and Thoai, N.V. (2000). *Introduction to Global Optimization*, Kluwer Academic Publishers.

79. Hu, H.-C. (1955). On some variational principles in the theory of elasticity and the theory of plasticity, *Scientia Sinica* **4**, 33-54.
80. Huang X.X. and Yang X.Q. (2003). A unified augmented Lagrangian approach to duality and exact penalization. *Mathematics of Operations Research*. **28**, 524-532.
81. Koiter, W.T. (1973). On the principle of stationary complementary energy in the nonlinear theory of elasticity, *SIAM, J. Appl. Math.*, **25**, 424-434.
82. Koiter, W.T. (1976). On the complementary energy theorem in nonlinear elasticity theory, Trends in Appl. of Pure Math. to Mech. ed. G. Fichera, Pitman.
83. Lao Zhi (400BC). *Dao De Jing* (or *Tao Te Ching*), English edition by D.C. Lau, Penguin Classics, 1963.
84. Lasserre, J. (2001). Global optimization with polynomials and the problem of moments. *SIAM J. Optimization*, **11** (3), 796-817.
85. Lavery, J. (2004). Shape-preserving approximation of multiscale univariate data by cubic L_1 spline fits, *Comput. Aided Geom. Design*, **21**, 43-64.
86. Lee, S. J. and Shield, R. T. (1980). Variational principles in finite elastics, *J. Appl. Math. Physics (ZAMP)*, **31**, 437-453.
87. Lee, S. J. and Shield, R. T. (1980). Applications of variational principles in finite elasticity, *J. Appl. Math. Physics (ZAMP)*, **31**, pp 454-472.
88. Levinson, M. (1965). The complementary energy theorem in finite elasticity, *Trans. ASME, ser. E, J. Appl. Mech.*, **87**, pp. 826-828.
89. Li, S.F. and Gupta, A. (2006). On dual configuration forces, *J. of Elasticity*, **84**:13-31.
90. Maier, G. (1969). Complementarity plastic work theorems in piecewise-linear elastoplasticity, *Int. J. Solids and Struct.*, **5**, 261-270.
91. Maier, G. (1970). A matrix structural theory of piecewise-linear plasticity with interacting yield planes, *Meccanica*, **5**, 55-66.
92. Maier, G., Carvelli, V. and Cocchetti, G. (2000). On direct methods for shake-down and limit analysis, Plenary lecture at the 4th EUROMECH Solid Mechanics Conference, Metz, France, June 26-30, *European J. of Mechanics, A/Solids*, **19**, Special Issue, S79-S100.
93. Marsden, J. and Ratiu, T. (1995). *Introduction to Mechanics and Symmetry*, Springer.
94. Moreau, J.J. (1968). La notion de sur-potentiel et les liaisons unilatérales en élastostatique, *C.R. Acad. Sc. Paris*, **267** A, 954-957.
95. Moreau, J.J., Panagiotopoulos, P.D. and Strang, G. (1988). Topics in nonsmooth mechanics. Birkhuser Verlag, Basel-Boston, MA.
96. Murty, K.G. and Kabadi, S.N. (1987). Some NP-complete problems in quadratic and nonlinear programmings, *Math. Progr.*, **39**, 117-129.
97. Nesterov, Y. (2000). Squared functional systems and optimization problems. *High Performance Optimization* (H. Frenk *et al*, eds), Kluwer Academic Publishers, pp. 405-440.
98. Noble, B. and Sewell, M.J. (1972). On dual extremum principles in applied mathematics, *J. Inst. Math. Appl.*, **9**, 123-193.
99. Oden, J. T. and Lee, J. K. (1977). Dual-mixed hybrid finite element method for second-order elliptic problems. Mathematical aspects of finite element methods (Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975). In *Lecture Notes in Math.*, Vol. 606, Springer, Berlin, 275-291.
100. Oden, J.T. and Reddy, J.N. (1983). *Variational Methods in Theoretical Mechanics*. Springer-Verlag.

101. Ogden, R.W. (1975). A note on variational theorems in non-linear elastostatics. *Math. Proc. Camb. Phil. Soc.*, 77, pp. 609-615.
102. Ogden, R.W. (1977). Inequalities associated with the inversion of elastic stress-deformation relations and their implications, *Math. Proc. Camb. Phil. Soc.*, 81, 313-324.
103. Pais, A. (1991). *Niels Bohr's Times, In Physics, Philosophy, and Polity*, Clarendon Press, Oxford. 565pp.
104. Parrilo, P. and Sturmfels, B. (2001). Minimizing polynomial functions, *Proceedings of DIMACS Workshop on Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science*, (eds. S. Basu and L. Gonzalez-Vega), American Mathematical Society, 2003, pp. 83-100.
105. Pian, T.H.H. and Tong, P. (1980). Reissner's principle in finite element formulations, in *Mechanics Today*, 5, S. Nemat-Nasser (ed.), Pergamon Press, 377-395.
106. Pian, T.H.H. and Wu, C.C. (2006). *Hybrid and Incompatible Finite Element Methods*, Chapman & Hall/CRC, 378 pp.
107. Rockafellar, R.T. (1967). Duality and stability in extremum problems involving convex functions, *Pacific J. Math.* 21, 167-187.
108. Rockafellar, R.T. (1970). *Convex Analysis*, Princeton University Press.
109. Rockafellar, R.T. (1974). *Conjugate Duality and Optimization*, SIAM, Philadelphia.
110. Rockafellar, R.T. and Wets, R.J.B. (1998). *Variational Analysis*, Springer: Berlin, New York.
111. Powell, M.J.D (2002). UOBYQA: unconstrained optimization by quadratic approximation, *Mathematical Programming, Series B*, 92 (3), pp. 555-582.
112. Rubinov, A.M. and Yang X.Q. (2003). *Lagrange-Type Functions in Constrained Non-Convex Optimization*. Kluwer Academic Publishers, Boston / Dordrecht / London, 285 pp.
113. Rubinov, A.M., Yang X.Q. and Glover, B.M. (2001). Extended Lagrange and penalty functions in optimization. *J. Optim. Theory Appl.*, 111 (2), 381-405.
114. Penot, J.-P. and Volle, M. (1990). On quasiconvex duality, *Mathematics of Operations Research*, 14, 597-625.
115. Sahni, S. (1974). Computationally related problems, *SIAM J. Comp.*, 3, 262-279.
116. Sewell, M.J. (1987). *Maximum and Minimum Principles*, Cambridge Univ. Press, 468pp.
117. Sherali, H.D. and Tuncbilek, C. (1992). A global optimization for polynomial programming problem using a reformulation-linearization technique. *J. of Global Optimization*, 2, 101-112.
118. Sherali, H.D. and Tuncbilek, C. (1997). New reformulation-linearization technique based relaxation for univariate and multivariate polynomial programming problems. *Operations Research Letters*, 21(1), 1-10.
119. Silverman, H.H. and Tate, J. (1992). *Rational Points on Elliptic Curves*, Springer-Verlag.
120. Singer, I. (1998). Duality for optimization and best approximation over finite intersections. *Numer. Funct. Anal. Optim.*, 19, no. 7-8, 903-915.
121. Strang, G. (1979). A minimax problem in plasticity theory, *Functional Analysis Methods in Numerical Analysis*, M.Z. Nashed (ed.), Springer Lecture Notes 701 319-333.
122. Strang, G. (1986). *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, 758 pp.

123. Strang, G. (1984). Duality in the classroom, *American Math. Monthly*, 91, 250-254.
124. Strang, G. (1983). Maximal flow through a domain, *Mathematical Programming*, 26, 123-143.
125. Strang, G. (1982). L^1 and L^∞ and approximation of vector fields in the plane, *Nonlinear Partial Differential Equations in Applied Science*, H. Fujita, P. Lax, and G. Strang (eds.), *Lecture Notes in Num. Appl. Anal.* 5, 273-288, Springer, New York.
126. Tabarrok, B. and Rimrott, F.P.J. (1994). *Variational methods and complementary formulations in dynamics*. Kluwer Academic Publishers: Dordrecht.
127. Temam, R. and Strang, G. (1980). Duality and relaxation in the variational problems of plasticity, *J. de Mecanique*, 19, 1-35.
128. Thach, P.T. (1993). Global optimality criterion and a duality with a zero gap in nonconvex optimization. *SIAM J. Math. Anal.* 24, no. 6, 1537-1556.
129. Thach, P. T. (1995). Diewert-Crouzeix conjugation for general quasiconvex duality and applications. *J. Optim. Theory Appl.*, 86, no. 3, 719-743.
130. Thach, P. T., Konno, H. and Yokota, D. (1996). Dual approach to minimization on the set of Pareto-optimal solutions. *J. Optim. Theory Appl.*, 88, no. 3, 689-707.
131. Toland, J.F. (1978). Duality in nonconvex optimization, *J. Mathematical Analysis and Applications*, 66, 399-415.
132. Toland, J.F. (1979). A duality principle for non-convex optimization and the calculus of variations, *Arch. Rational Mech. Anal.* 71, . 41-61.
133. Tonti, E. (1972). A mathematical model for physical theories, *Accad. Naz. dei Lincei*, Serie III, LII, I, 175-181; II, 350-356.
134. Tonti, E. (1972). On the mathematical structure of a large class of physical theories, *Accad. Naz. dei Lincei*, Serie VIII, LII, 49-56.
135. Tuy, H (1995). D.C. optimization: theory, methods and algorithms, in *Handbook of Global Optimization*, eds. R. Horst and P. Pardalos, Kluwer Academic Publishers, 149-216.
136. Vavasis, S. (1990). Quadratic programming is in NP, *Info. Proc. Lett.*, 36, 73-77.
137. Vavasis, S. (1991). *Nonlinear Optimization: Complexity Issues*, Oxford University Press, New York, NY.
138. Veubeke, B.F. (1972). A new variational principle for finite elastic displacements. *Int. J. Engineering Science*, 10, pp. 745-763.
139. Walk, M (1989). *Theory of duality in mathematical programming*, Springer-Verlag, Wien.
140. Wang, Z.B., Fang, S.-C., Gao, D.Y., and Xing, W.X., Global extremal conditions for multi-integer quadratic programming, *J. Industrial and Management Optimization*, 4(2), 213-225.
141. Wright, M. H. (1998). The interior-point revolution in constrained optimization, in *High-Performance Algorithms and Software in Nonlinear Optimization* (R. DeLeone, A. Murli, P. M. Pardalos, and G. Toraldo, eds.) 359-381, Kluwer Academic Publishers, Dordrecht., The Netherlands.
142. Ye, Y. (1992). A new complexity result on minimization of a quadratic function with a sphere constraint, in *Recent Advances in Global Optimization* (C. Floudas and P. Pardalos eds.), Princeton University Press, Princeton, NJ.

143. Zhao, Y.B., Fang, S.C., and Lavery, J. (2006). Geometric dual formulation of the first derivative based C^1 -smooth univariate cubic L_1 spline functions, to appear in *Complementarity, Duality, and Global Optimization*, a special issue of *J. Global Optimization*, D.Y. Gao and H.D. Sherali, Eds.
144. Zhou, Y.Y. and Yang, X.Q. (2004). Some results about duality and exact penalization. *J. Global Optimization*, 29, 497-509.
145. Zubov, L.M. (1970), The stationary principle of complementary work in non-linear theory of elasticity, *Prikl. Mat. Mech.*, 34, pp. 228-232.