



# Canonical Dual Transformation Method and Generalized Triality Theory in Nonsmooth Global Optimization\*

DAVID YANG GAO

Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, USA.

(Received for publication May 2000)

**Abstract.** This paper presents, within a unified framework, a potentially powerful canonical dual transformation method and associated generalized duality theory in nonsmooth global optimization. It is shown that by the use of this method, many nonsmooth/nonconvex constrained primal problems in  $\mathbb{R}^n$  can be reformulated into certain smooth/convex unconstrained dual problems in  $\mathbb{R}^m$  with  $m \leq n$  and without duality gap, and some NP-hard concave minimization problems can be transformed into unconstrained convex minimization dual problems. The extended Lagrange duality principles proposed recently in finite deformation theory are generalized suitable for solving a large class of nonconvex and nonsmooth problems. The very interesting generalized triality theory can be used to establish nice theoretical results and to develop efficient alternative algorithms for robust computations.

**Key words:** Bi-duality; Canonical dual transformation; D.C. optimization; Duality; Global optimization; Nonconvexity; Nonsmoothness, Reformulation, Triality

## 1. Introduction

The aim of this paper is to develop a powerful method and general theory for solving the following general nonconvex and nonsmooth extremum problem

$$(\mathcal{P}_{\text{ext}}) : P(x) = \Phi(x, \Lambda(x)) \rightarrow \text{extremum } \forall x \in \mathcal{X},$$

where  $\mathcal{X}$  is a locally convex topological vector space (l.c.s.),  $P : \mathcal{X} \rightarrow \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$  is a nonconvex and nonsmooth extended function with the non-empty effective domain

$$\mathcal{X}_k = \text{dom } P = \{x \in \mathcal{X} \mid |P(x)| < +\infty\}.$$

The operator  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is a continuous, generally nonlinear, mapping from  $\mathcal{X}$  to another l.c.s.  $\mathcal{Y}$ , and  $\Phi : \mathcal{X} \times \mathcal{Y} \rightarrow \bar{\mathbb{R}}$  is an extended function. Problem  $(\mathcal{P}_{\text{ext}})$  may have many locally extremum (either minimum or maximum) solutions, and it represents a general *global optimization* problem. It was shown in Gao (1999)

\* This paper is dedicated to the memory of Professor P.D. Panagiotopoulos

that this class of problems covers a great variety of situations including constrained nonconvex variational analysis, *d.c. programming*, i.e. nonconvex problems of d.c. functions (difference of convex functions), variational inequality, complementarity problems, network problems and nonconvex dynamical systems and much more.

In the history of science, mathematics and mechanics have been always complementary partners. Starting from the pioneering work of Moreau (1968) in a frictional contact mechanics problem, where the notions of the super-potential and subdifferential were originally introduced, the subject of non-smooth/non-convex global optimization has experienced significant development during the last three decades. Many problems arising in natural systems (such as engineering mechanics, chemical reactions, network flows and mathematical economics, etc.) require the considerations of nonconvexity and nondifferentiability for their mathematical modeling and the cost functions. The terminology *Non-Smooth Mechanics* was formally proposed by Moreau, Panagiotopoulos and Strang (1988). Several monographs have documented the basic theory, methods, algorithms and applications of nonsmooth/nonconvex variational analysis and global optimizations (cf., e.g., Panagiotopoulos, 1985; Horst and Pardalos, 1995; Dem'yanov et al., 1996; Mistakidis and Stavroulakis, 1998; Motreanu and Panagiotopoulos, 1999 and Gao et al., 2000). Generally speaking, traditional direct methods for solving nonsmooth and nonconvex problems are usually very difficult, or even impossible. The so-called relaxation methods can be used mainly for finding global optimal solutions (global minimizer of maximizer). However, in unilateral post-buckling analysis of nonlinear beam theory, it was shown by Gao (1998b) that the solution of actual buckling state has to be a local minimizer. In phase transitions and nonconvex dynamical programming, local extrema usually play an important role in understanding the physical mechanism of systems. In nonsmooth global optimization problems, more recent trends consist of the so-called *reformulation* and *nonlinear rescaling techniques* (cf. e.g., Fukushima and Qi, 1999; Polyak and Griva, 2000). The classical Lagrange duality theory is the main tool used in these methods.

Duality is a fundamental concept that underlies almost all natural phenomena. The duality methods in classical optimization possess beautiful theoretical properties, potentially powerful alternative performances and pleasing relationships to many other fields (see Walk, 1989; Wright, 1996). A self-contained comprehensive presentation of the mathematical theory in general nonconvex, nonsmooth systems was given recently by Gao (1999). In global optimization, duality theory falls principally into three categories:

- (1) the classical saddle Lagrange (minimax) duality in convex problems,
- (2) the nice super-Lagrangian bi-duality in geometrically linear systems and
- (3) the interesting triality and multi-duality in general nonconvex canonical systems.

In the *geometrically linear systems*, where  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is a linear operator and  $\Phi : \mathcal{X} \times \mathcal{Y} \rightarrow \overline{\mathbb{R}}$  is a *canonical function* (i.e.  $\Phi(x, y^*)$  is either convex or concave in each of its variables), the duality has been studied substantially during the last thirty years for both convex and nonconvex canonical systems. In the case that  $P : \mathcal{X} \rightarrow \overline{\mathbb{R}}$  is convex, its dual function can be well-determined by the so-called Rockafellar dual transformation:  $P^d(y^*) = \Phi^*(-\Lambda^*y^*, y^*)$ , where  $\Phi^* : \mathcal{X}^* \times \mathcal{Y}^* \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$  is the well-known Fenchel-Rockafellar conjugate function of  $\Phi$ . The Fenchel-Rockafellar duality is essentially equivalent to the classical saddle Lagrange duality, which yields a so-called mono-duality, i.e., each minimum primal problem possesses a unique maximum dual problem and  $\inf P(x) = \sup P^d(y^*)$  (see Gao, 1999). During the last decade, the so-called *primal-dual interior point method* has emerged as the most important and efficient revolutionary technique in mathematical programming (cf. e.g. Wright, 1997 for linear programming, Gay et al., 1998 and Wright, 1998 for nonconvex nonlinear programming). Actually, the primal-dual methods and ideas were studied originally by engineers at the beginning of this century (cf. e.g., Maier et al., 2000; Gao, 1999). It is well-known in engineering structural limit analysis that the direct approaches for solving minimum potential energy (primal problem) can only provide upper bounds of the so-called collapse loading factor. On the other hand, the maximum complementary energy principle (dual problem) and methods give the lower bound solutions. In safety analysis of engineering structures, the primal-dual methods provide definitely powerful and efficient tolls for solving nonsmooth, nonlinear problems (cf. e.g., Maier, 1969; Casciaro and Cascini, 1982; Gao, 1988a,b, 1999). The recent article by Maier *et al* (2000) serves as an excellent survey on the developments for applications of mathematical programming in engineering structural mechanics. Dual to the interior-point methods, the so-called *pan-penalty finite element programming* developed by Gao (1988b) is essentially a *primal-dual exterior-point method*. It was proved that in rigid-perfectly plastic limit analysis, the exterior penalty functional and the associated perturbation method possess a wonderful physical meaning, which leads to an efficient technique of dimension reduction in nonlinear mixed finite element programming by use of the saddle-Lagrange duality theory (Gao, 1988b). However, if the primal function  $P(x)$  is nonconvex, there exists a *duality gap* between the primal problem ( $\mathcal{P}_{\inf}$ ) and the Fenchel-Rockafellar dual problem ( $\mathcal{P}_{\sup}^d$ ), i.e.  $\inf P(x) \geq \sup P^d(y^*)$ . In this sense, the well-developed saddle-Lagrange duality and the Fenchel-Rockafellar duality can be used mainly for convex problems.

The duality for nonconvex minimization was first studied by Toland (1978, 1979) for d.c. functions:  $P(x) = W(\Lambda x) - F(x)$ , where  $W : \mathcal{Y} \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$  and  $F : \mathcal{X} \rightarrow \overline{\mathbb{R}}$  are two convex functions. The d.c. optimization problems arise naturally from many applications in engineering, economics and other sciences. The generalizations of Toland's duality theory were made by Auchmuty (1983–1997) to geometrically linear nonconvex variational analysis. It was shown that the

Toland's double-min duality for d.c. programming is a special case of the so-called anomalous dual problems. As a class of typical nonconvex global optimization problems, a detailed survey article on the theory, methods and algorithms of the d.c. programming was given by Tuy (1995). During last three decades several important duality concepts have been developed as studied for nonconvex optimization (cf. e.g., Ekeland, 1977; Crouzeix, 1981; Hiriart-Urruty, 1985, Penot and Volle, 1990; Singer, 1986-98; Thach et al., 1993-96; Tuy, 1991, 1995 and Rockafellar and Wets, 1997). However, it was tradition in global optimization that the primal problem is usually considered as a global minimization problem over a feasible set. This tradition obscured our sight and thus, in d.c. programming, only the double-min duality theory was studied. In nonconvex problems, the local maximizers play important roles in phase transition, unilateral bifurcations and chaotic dynamics.

Duality theory in geometrically nonlinear systems was originally studied by Gao and Strang (1989) in large deformation variational/boundary value problems governed by nonsmooth constitutive laws, where the primal function  $P(x) = W(\Lambda(x)) - F(x)$  represents the total potential of the system,  $W(y)$  and  $F(x)$  are respectively the internal and external energies. In finite deformation field theory, the nonlinear operator  $\Lambda$  is usually quadratic. In order to recover the duality gap in traditional Fenchel-Rockafellar duality theory, they introduced a so-called *complementary gap function*, which leads to a generalized complementary variational principle and a nonlinear Lagrange duality theory in *fully nonlinear variational problems*. They proved that if this complementary gap function possesses a positive sign, the generalized complementary energy  $L(x, y^*)$  is a saddle functional. Their systematical works on duality theory lead to a unified framework in applied mathematics (see Strang, 1986) and in *fully nonlinear canonical systems* (see Gao et al., 1989–1999). Recently, in the study of the post-buckling analysis of nonlinear beam theory, it was discovered by Gao (1996, 1997, 1998b) that for convex  $W(y)$  and quadratic  $\Lambda$ , if the gap function is negative,  $L(x, y^*)$  is a so-called *super-critical point functional*, and a very interesting tri-duality theory was proposed for quadratic operator  $\Lambda(x)$ . A comprehensive study on duality principles in nonsmooth and nonconvex systems is recently given by Gao (1999).

The aim of this article is to generalize the author's previous results on nonconvex variational systems into nonsmooth global optimization problems suitable for arbitrary nonlinear operator  $\Lambda$ . Actually, the key idea of the so-called canonical dual transformation method is to choose a certain nonlinear operator  $\Lambda$  such that  $\Phi : \mathcal{X} \times \mathcal{Y} \rightarrow \bar{\mathbb{R}}$  is a canonical function. Thus, the *perfect duality principles* (without duality gap) can be easily formulated by the classical Legendre transformation. The rest of this paper is divided into five main sections. The next section set up the notation used in the paper and describes the problems. A general framework in fully nonlinear, nonsmooth systems is discussed. Section 3 presents an extended Lagrangian duality theory in general global optimization. The critical points in fully nonlinear systems are classified. Section 4 is devoted mainly to the super-Lagrange duality theory. The nice bi-duality proposed in geometrically

linear systems and d.c. programming is generalized to arbitrary function  $L(x, y^*)$ . Section 5 discusses the triality in fully nonlinear systems. The very interesting triality theory is generalized for general geometrically nonlinear operator  $\Lambda$  and are illustrated by quadratic operators in global optimization. The last two sections present applications and concluding remarks.

## 2. Framework of Canonical Systems and Classification

Let  $\mathcal{X}$  and  $\mathcal{X}^*$  be two locally convex topological real linear spaces, finite- or infinite dimensional, placed in separating duality by a bilinear form  $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X}^* \rightarrow \mathbb{R}$ . For a given extended real-valued function  $P : \mathcal{X} \rightarrow \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ , the sub- and super-differentials of  $P$  at  $\bar{x} \in \mathcal{X}$  are defined by

$$\partial^- P(\bar{x}) = \{\bar{x}^* \in \mathcal{X}^* | P(x) - P(\bar{x}) \geq \langle \bar{x}^*, x - \bar{x} \rangle \forall x \in \mathcal{X}\},$$

$$\partial^+ P(\bar{x}) = \{\bar{x}^* \in \mathcal{X}^* | P(x) - P(\bar{x}) \leq \langle \bar{x}^*, x - \bar{x} \rangle \forall x \in \mathcal{X}\},$$

respectively. Clearly, we always have  $\partial^+ P = -\partial^-(-P)$ . In convex analysis, it is convention that  $\partial^-$  is simply written as  $\partial$ . In this paper,  $\partial$  stands for either  $\partial^-$  or  $\partial^+$ , i.e.

$$\partial = \{\partial^-, \partial^+\}.$$

If  $P$  is smooth, Gâteaux-differentiable at  $\bar{x} \in \mathcal{X}_a \subset \mathcal{X}$ , then

$$\partial P(\bar{x}) = \partial^- P(\bar{x}) = \partial^+ P(\bar{x}) = \{DP(\bar{x})\},$$

where  $DP : \mathcal{X}_a \rightarrow \mathcal{X}^*$  denotes the Gâteaux derivative of  $P$  at  $\bar{x}$ . The following notations and definitions, used in Gao (1999), will be of convenience in global optimization.

**DEFINITION 1.** The set of functions  $P : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  which are either convex or concave is denoted by  $\Gamma(\mathcal{X})$ . In particular, let  $\check{\Gamma}(\mathcal{X})$  denote the subset of functions  $P \in \Gamma(\mathcal{X})$  which are convex and  $\hat{\Gamma}(\mathcal{X})$  the subset of  $P \in \Gamma(\mathcal{X})$  which are concave.

The *canonical function space*  $\Gamma_G(\mathcal{X}_a)$  is a subset of functions  $P \in \Gamma(\mathcal{X}_a)$  which are Gâteaux differentiable on  $\mathcal{X}_a \subset \mathcal{X}$ .

The *extended canonical function space*  $\Gamma_0(\mathcal{X})$  is a subset of functions  $P \in \Gamma(\mathcal{X})$  which are either convex, lower semicontinuous or concave, upper semicontinuous, and if  $P$  takes the values  $\pm\infty$ , then  $P$  is identically equal to  $\pm\infty$ .  $\square$

By the Legendre-Fenchel transformation, the *super-conjugate function* of an extended function  $P : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  is defined by

$$P^\sharp(x^*) = \sup_{x \in \mathcal{X}} \{\langle x, x^* \rangle - P(x)\}.$$

By the theory of convex analysis,  $P^\sharp : \mathcal{X}^* \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$  is always convex and lower semicontinuous, i.e.  $P^\sharp \in \check{\Gamma}_0(\mathcal{X}^*)$ . Dually, the *sub-conjugate function* of  $P$ , defined by

$$P^b(x^*) = \inf_{x \in \mathcal{X}} \{\langle x, x^* \rangle - P(x)\},$$

is always concave and upper semicontinuous, i.e.  $P^b \in \hat{\Gamma}_0(\mathcal{X}^*)$ , and  $P^b = -P^\sharp$ . Both the super- and sub-conjugates are called Fenchel conjugate functions and we write  $P^* = \{P^b, P^\sharp\}$ . Thus the extended Fenchel transformation can be written as

$$P^*(x^*) = \text{ext}_{x \in \mathcal{X}} \{\langle x, x^* \rangle - P(x)\}. \quad (2.1)$$

Clearly, if  $P \in \Gamma_0(\mathcal{X})$ , we have the Fenchel equivalent relations, namely,

$$x^* \in \partial P(x) \Leftrightarrow x \in \partial P^*(x^*) \Leftrightarrow P(x) + P^*(x^*) = \langle x, x^* \rangle. \quad (2.2)$$

The pair  $(x, x^*)$  is called the *Fenchel duality pair* on  $\mathcal{X} \times \mathcal{X}^*$  if and only if equation (2.2) holds on  $\mathcal{X} \times \mathcal{X}^*$ .

The conjugate pair  $(x, x^*)$  is called the *Legendre duality pair* on  $\mathcal{X}_a \times \mathcal{X}_a^* \subset \mathcal{X} \times \mathcal{X}^*$  if and only if the equivalent relations

$$x^* = DP(x) \Leftrightarrow x = DP^*(x^*) \Leftrightarrow P(x) + P^*(x^*) = \langle x, x^* \rangle \quad (2.3)$$

hold on  $\mathcal{X}_a \times \mathcal{X}_a^*$ .

Let  $(\mathcal{Y}, \mathcal{Y}^*)$  be an another pair of locally convex topological real linear spaces paired in separating duality by the second bilinear form  $\langle \cdot ; \cdot \rangle : \mathcal{Y} \times \mathcal{Y}^* \rightarrow \mathbb{R}$ . The so-called *geometrical operator*  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is a continuous, Gâteaux differentiable operator such that for any given  $x \in \mathcal{X}_a \subset \mathcal{X}$ , there exists a  $y \in \mathcal{Y}_a \subset \mathcal{Y}$  satisfying the *geometrical equation*

$$y = \Lambda(x).$$

The directional derivative of  $y$  at  $\bar{x}$  in the direction  $x \in \mathcal{X}$  is then defined by

$$\delta y(\bar{x}; x) := \lim_{\theta \rightarrow 0^+} \frac{y(\bar{x} + \theta x) - y(\bar{x})}{\theta} = \Lambda_t(\bar{x})x, \quad (2.4)$$

where  $\Lambda_t(\bar{x}) = D\Lambda(\bar{x})$  denotes the Gâteaux derivative of the operator  $\Lambda$  at  $\bar{x}$ . For a given  $y^* \in \mathcal{Y}^*$ ,  $\ell(x) = \langle \Lambda(x) ; y^* \rangle$  is a real-valued function of  $x$  on  $\mathcal{X}$ . Its Gâteaux derivative at  $\bar{x} \in \mathcal{X}_a$  in the direction  $x \in \mathcal{X}$  reads

$$\delta \ell(\bar{x}; x) = \langle \Lambda_t(\bar{x})x ; y^* \rangle = \langle x, \Lambda_t^*(\bar{x})y^* \rangle,$$

where  $\Lambda_t^*(\bar{x}) : \mathcal{Y}^* \rightarrow \mathcal{X}^*$  is the adjoint operator of  $\Lambda_t$  associated with the two bilinear forms.

Let  $\Phi : \mathcal{X} \times \mathcal{Y} \rightarrow \overline{\mathbb{R}}$  be an extended function such that

$$P(x) = \Phi(x, \Lambda(x)).$$

If  $\Phi : \mathcal{X} \times \mathcal{Y} \rightarrow \bar{\mathbb{R}}$  is an extended canonical function, i.e.  $\Phi \in \Gamma_0(\mathcal{X}) \times \Gamma_0(\mathcal{Y})$ , the duality relations between the paired spaces  $(\mathcal{X}, \mathcal{X}^*)$  and  $(\mathcal{Y}, \mathcal{Y}^*)$  can be written as

$$x^* \in \partial_x \Phi(x, y), \quad y^* \in \partial_y \Phi(x, y). \quad (2.5)$$

On the product space  $\mathcal{X}_a \times \mathcal{Y}_a \subset \mathcal{X} \times \mathcal{Y}$ , if the canonical function  $\Phi(x, y)$  is finite and Gâteaux differentiable such that the feasible space  $\mathcal{X}_k$  can be written as

$$\mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda(x) \in \mathcal{Y}_a\}, \quad (2.6)$$

then on  $\mathcal{X}_k$ , the critical condition

$$\begin{aligned} \delta P(\bar{x}; x) = \langle x, DP(\bar{x}) \rangle &= \langle x, D_x \Phi(\bar{x}, \Lambda(\bar{x})) \rangle + \\ \langle \Lambda_t(\bar{x})x; D_y \Phi(\bar{x}, \Lambda(\bar{x})) \rangle &= 0 \quad \forall x \in \mathcal{X}_x \end{aligned}$$

leads to the Euler equation:

$$D_x \Phi(\bar{x}, \Lambda(\bar{x})) + \Lambda_t^*(\bar{x}) D_y \Phi(\bar{x}, \Lambda(\bar{x})) = 0, \quad (2.7)$$

where  $D_x \Phi$  and  $D_y \Phi$  denote the partial Gâteaux derivatives of  $\Phi$  with respect to  $x$  and  $y$ , respectively. Since  $\Phi \in \Gamma_G(\mathcal{X}_a) \times \Gamma_G(\mathcal{Y}_a)$  is a canonical function, the Gâteaux derivative  $D\Phi : \mathcal{X}_a \times \mathcal{Y}_a \rightarrow \mathcal{X}_a^* \times \mathcal{Y}_a^* \subset \mathcal{X}^* \times \mathcal{Y}^*$  is a monotone mapping, i.e. there exists a pair  $(\bar{x}^*, \bar{y}^*) \in \mathcal{X}^* \times \mathcal{Y}^*$  such that

$$-\bar{x}^* = D_x \Phi(\bar{x}, \Lambda(\bar{x})), \quad \bar{y}^* = D_y \Phi(\bar{x}, \Lambda(\bar{x})). \quad (2.8)$$

Then the so-called *virtual work principle*

$$\delta \ell(\bar{x}; x) = \langle \Lambda_t(\bar{x})x; y^* \rangle = \langle x, \Lambda_t^*(\bar{x})y^* \rangle = \langle x, \bar{x}^* \rangle \quad \forall x \in \mathcal{X}_k \quad (2.9)$$

leads to the so-called *balance (or equilibrium) equation*

$$\bar{x}^* = \Lambda_t^*(\bar{x})\bar{y}^*, \quad (2.10)$$

which linearly depends on the dual variable  $\bar{y}^*$ .

In geometrically linear systems, where  $\Lambda = \Lambda_t$ , the values of the two bilinear  $\langle \cdot, \cdot \rangle$  and  $\langle \cdot; \cdot \rangle$  are equal, i.e.

$$\langle \bar{y}; \bar{y}^* \rangle = \langle \Lambda \bar{x}; \bar{y}^* \rangle = \langle \bar{x}, \Lambda^* \bar{y}^* \rangle = \langle \bar{x}, \bar{x}^* \rangle.$$

However, in geometrically nonlinear systems  $\Lambda \neq \Lambda_t$ , and the following operator decomposition is introduced by Gao and Strang (1989)

$$\Lambda(x) = \Lambda_t(x)x + \Lambda_c(x), \quad (2.11)$$

where  $\Lambda_c : \mathcal{X} \rightarrow \mathcal{Y}$  is called the *complementary operator* of the Gâteaux derivative operator  $\Lambda_t$ , which plays a key role in nonconvex duality theory. Thus, there exists a gap between the two bilinear forms, i.e.

$$\langle \Lambda(\bar{x}); \bar{y}^* \rangle = \langle \bar{x}, \Lambda_t^*(\bar{x})\bar{y}^* \rangle - G(\bar{x}, \bar{y}^*) = \langle \bar{x}, \bar{x}^* \rangle - G(\bar{x}, \bar{y}^*), \quad (2.12)$$

$$\begin{array}{ccc}
x \in \mathcal{X} & \longleftarrow \langle x, x^* \rangle & \longrightarrow \mathcal{X}^* \ni x^* \\
\Lambda_t + \Lambda_c = \Lambda \downarrow & & \uparrow \Lambda_t^* = (\Lambda - \Lambda_c)^* \\
y \in \mathcal{Y} & \longleftarrow \langle y, y^* \rangle & \longrightarrow \mathcal{Y}^* \ni y^*
\end{array}$$

Figure 1. Framework in fully nonlinear systems.

where  $G : \mathcal{X} \times \mathcal{Y}^* \rightarrow \mathbb{R}$  is so-called *complementary gap function*, defined by

$$G(x, y^*) = \langle -\Lambda_c(x); y^* \rangle : \mathcal{X} \times \mathcal{Y}^* \rightarrow \mathbb{R}. \quad (2.13)$$

This function was first introduced by Gao and Strang (1989) in finite deformation theory, which plays a key role in nonconvex variational problems.

The following classification for the global optimization problems was given by Gao (1998, 1999).

**DEFINITION 2.** Suppose that for a given problem  $(\mathcal{P}_{\text{ext}})$ , the geometrical operator  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  can be chosen in such a way that  $P(x) = \Phi(x, \Lambda(x))$ ,  $\Phi \in \Gamma_G(\mathcal{X}_a) \times \Gamma_G(\mathcal{Y}_a)$  and  $\mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda(x) \in \mathcal{Y}_a\}$ . Then

(1) the transformation  $\{P; \mathcal{X}_k\} \rightarrow \{\Phi; \mathcal{X}_a \times \mathcal{Y}_a\}$  is called the *canonical transformation*, and  $\Phi : \mathcal{X}_a \times \mathcal{Y}_a \rightarrow \mathbb{R}$  is called the *canonical function associated with*  $\Lambda$ ;

(2) the problem  $(\mathcal{P}_{\text{ext}})$  is called *geometrically nonlinear (resp. linear)* if  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is nonlinear (resp. linear); it is called *physically nonlinear (resp. linear)* if the duality mapping  $D\Phi : \mathcal{X}_a \times \mathcal{Y}_a \rightarrow \mathcal{X}_a^* \times \mathcal{Y}_a^*$  is nonlinear (resp. linear); it is called *fully nonlinear* if it is both geometrically and physically nonlinear.  $\square$

The canonical transformation plays a fundamental role in duality theory of global optimization. By this definition, the governing Equation (2.7) for fully nonlinear problems can be written in the *tri-canonical forms*, namely,

$$\begin{array}{ll}
(1) \text{ geometrical equation: } & y = \Lambda(x), \\
(2) \text{ physical relations: } & y^* \in \partial_y \Phi(x, y), \quad -x^* \in \partial_x \Phi(x, y), \\
(3) \text{ balance equation: } & x^* = \Lambda_t^*(x)y^*.
\end{array} \quad (2.14)$$

A framework for the fully nonlinear system is shown in Figure 1. Extensive illustrations of the canonical transformation and the tri-canonical forms in mathematical physics and variational analysis were given in the monograph by Gao (1999).

Very often, the extended canonical function  $\Phi$  can be written in the form

$$\Phi(x, y) = W(y) - F(x), \quad F \in \Gamma(\mathcal{X}), \quad W \in \Gamma(\mathcal{Y}).$$

The duality relations (2.5) in this special case take the forms

$$x^* \in \partial F(x), \quad y^* \in \partial W(y).$$



If  $F \in \Gamma_G(\mathcal{X}_a)$  and  $W \in \Gamma_G(\mathcal{Y}_a)$  are Gâteaux differentiable, the Euler equation (2.7) reads

$$\Lambda_t^*(\bar{x})DW(\Lambda(\bar{x})) - DF(\bar{x}) = 0.$$

If  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is linear, and  $W : \mathcal{Y} \rightarrow \mathbb{R}$  is quadratic such that  $DW = Cy$ , where  $C : \mathcal{Y} \rightarrow \mathcal{Y}^*$  is a linear operator, then the governing equations for linear system can be written as

$$\Lambda^*C\Lambda x = Ax = x^*.$$

For conservative systems, the operator  $A = \Lambda^*C\Lambda$  is usually symmetric. In static systems,  $C$  is usually positive-definite and the associated total potential  $P$  is convex. However, in dynamical systems,  $C$  is indefinite and  $P$  is called the total action, which is usually a d.c. function in convex Hamilton systems.

To demonstrate how the above scheme fits in with the finite dimensional global optimization, we list some examples in nonlinear programming.

**EXAMPLE 1. Geometrically linear nonsmooth constrained global minimization in  $\mathbb{R}^n$ .**

We first consider the following global minimization problem

$$(\mathcal{P}_{\min}) \quad \min f(x) \quad s.t. \quad x \in \mathcal{X}_k \subset \mathbb{R}^n, \quad (2.15)$$

where the primal feasible space  $\mathcal{X}_k$  is a nonempty convex subset in  $\mathbb{R}^n$ ,

$$\mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda x \in \mathcal{Y}_a\},$$

$\mathcal{X}_a \subset \mathbb{R}^n$ ,  $\mathcal{Y}_a \subset \mathbb{R}^m$  are two convex subsets,  $f \in \Gamma_0(\mathcal{X}_a)$  is a given canonical function;  $\Lambda = \{\lambda_{ij}\} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear operator (matrix) in  $\mathbb{R}^{m \times n}$ .

To reformulate this general nonsmooth global minimization problem in the geometrically linear canonical model form, we let  $\mathcal{X} = \mathcal{X}^* = \mathbb{R}^n$ ,  $\mathcal{Y} = \mathcal{Y}^* = \mathbb{R}^m$ , with the standard coordinatewise partial ordering and bilinear forms

$$\langle \Lambda x ; y^* \rangle = (\Lambda x)^T y^* = x^T (\Lambda^T y^*) = \langle x , \Lambda^* y^* \rangle = \sum_{i=1}^n \sum_{j=1}^m x_i \lambda_{ij} y_j^*.$$

Then, the adjoint of  $\Lambda$  associated with these standard bilinear forms is simply  $\Lambda^* = \Lambda^T \in \mathbb{R}^{n \times m}$ . For a given convex set  $\mathcal{X}_a$ , its indicator  $\mathcal{I}_{\mathcal{X}_a}(x) = \{0(\text{ if } x \in \mathcal{X}_a), +\infty(\text{ if } x \notin \mathcal{X}_a)\}$  is a convex, lower semicontinuous function. The two canonical functions  $F \in \Gamma_0(\mathcal{X})$  and  $W \in \check{\Gamma}(\mathcal{Y})$  can be defined by

$$F(x) = -f(x) - \mathcal{I}_{\mathcal{X}_a}(x), \quad W(y) = \mathcal{I}_{\mathcal{Y}_a}(y).$$

Clearly, the effective domains  $\text{dom } F = \mathcal{X}_a \subset \mathbb{R}^n$  and  $\text{dom } W = \mathcal{Y}_a \subset \mathbb{R}^m$  are nonempty and convex. Thus, the constrained problem (2.15) can be written in the extended (unconstrained) form

$$P(x) = \mathcal{L}_{\mathcal{Y}_a}(\Lambda x) + f(x) + \mathcal{L}_{\mathcal{X}_a}(x) \rightarrow \min \quad \forall x \in \mathbb{R}^n. \quad (2.16)$$

Its effective domain  $\text{dom } P = \mathcal{X}_k$ , and the condition  $x \in \mathcal{X}_k$  is called the *implicit constraint*.

The extended primal problem (2.16) covers many important special cases in constrained optimization problems.

Case I. If  $f \in \tilde{\Gamma}_G(\mathcal{X}_a)$ ,  $\mathcal{X}_a = \mathbb{R}^n$  and  $\mathcal{Y}_a = \{y \in \mathbb{R}^m \mid g_k(y) = 0, k = 1, \dots, p\}$ , where  $g : \mathcal{Y} \rightarrow \mathbb{R}^p$  is a p-vector of convex, Gâteaux differentiable functions with  $k$ th component  $g_k(x)$ , then the primal problem ( $\mathcal{P}_{\min}$ ) is a convex minimization problem with equality constraints  $g(\Lambda x) = 0$ . In this case, the canonical function  $F(x) = -f(x) - \mathcal{L}_{\mathcal{X}_a}(x) = -f(x)$  is concave and Gâteaux differentiable in  $\mathbb{R}^n$ ,  $\partial^+ F(x) = \{-Df(x)\}$ . While the subdifferential of  $W$  is a convex subset of  $\mathcal{Y}^*$ , i.e.

$$\partial^- W(y) = \begin{cases} \sum_{k=1}^p g_k^* \frac{\partial g_k(y)}{\partial y} & \text{if } y \in \mathcal{Y}_a, \\ \emptyset & \text{otherwise} \end{cases}$$

where the Lagrange multiplier  $g^* \in \mathbb{R}^p$  is the dual variable of  $g \in \mathbb{R}^p$ . In the case that  $f : \mathcal{X}_a \rightarrow \mathbb{R}$  is smooth function, then, the optimality condition  $DP(\bar{x}) = 0$  leads to the Euler-Lagrange equation

$$\Lambda^* \sum_{k=1}^p g_k^* \frac{\partial g_k(\Lambda x)}{\partial y} = Df(x), \quad g_k(\Lambda x) = 0.$$

Case II. If  $f : \mathcal{X} \rightarrow \mathbb{R}$  is nonsmooth, say, for example  $\mathcal{X} = \mathbb{R}$  and

$$f(x) = \begin{cases} \frac{1}{2}ax^2 & \text{if } x \leq x_a, \\ \frac{1}{2}ax_a^2 + \frac{1}{2}b(x - x_a)^2 + x_b^*(x - x_a) & \text{if } x > x_a, \end{cases} \quad (2.17)$$

where  $a$ ,  $b$ ,  $x_a$  and  $x_b^*$  are positive constants. In this case, the cost function  $f$  is nonsmooth (see Figure 2(a)), and its Gâteaux derivative is then a discontinuous function (see Figure 3a), i.e.

$$x^* = Df(x) = \begin{cases} ax & \text{if } x \leq x_a, \\ b(x - x_a) + x_b^* & \text{if } x > x_a, \end{cases} \quad (2.18)$$

The traditional direct approaches for solving this nonsmooth constrained optimization problem is difficult.

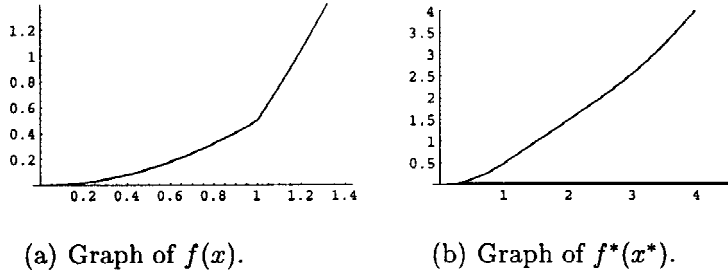


Figure 2. Nonsmooth function and its smooth Legendre conjugate.

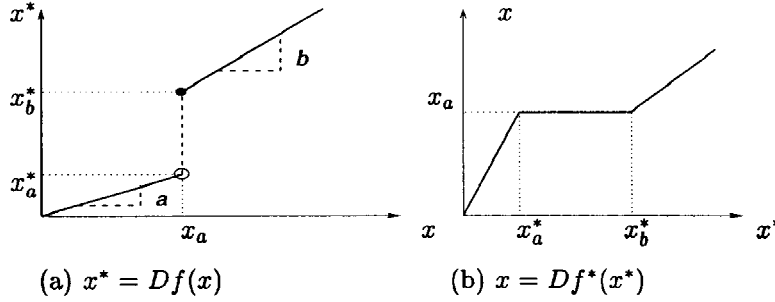


Figure 3. Discontinuous constitutive law and continuous inverse form.

By the fact that the Legendre conjugate of the nonsmooth  $f$  is a smooth function, i.e. (see Figure 2(b))

$$f^*(x^*) = \begin{cases} \frac{1}{2a}x^{*2} & \text{if } x^* \leq x_a^*, \\ \frac{1}{2a}x_a^{*2} + x_a(x^* - x_a^*) & \text{if } x_a^* < x^* \leq x_b^*, \\ \frac{1}{2a}x_a^{*2} + x_a(x^* - x_a^*) + \frac{1}{2b}(x^* - x_b^*)^2 & \text{if } x^* > x_b^*, \end{cases} \quad (2.19)$$

its Gâteaux derivative is a continuous function (see Fig. 3b)

$$x = Df^*(x^*) = \begin{cases} \frac{1}{a}x^* & \text{if } x^* \leq x_a^*, \\ x_a & \text{if } x_a^* < x^* \leq x_b^*, \\ x_a + \frac{1}{b}(x^* - x_b^*) & \text{if } x^* > x_b^*. \end{cases} \quad (2.20)$$

Thus, the dual problem will be much easier than the nonsmooth primal problem.

Case III. Concave minimization and complementarity problems.

If  $f \in \hat{\Gamma}(\mathcal{X}_a)$  is concave and for a given  $b \in \mathcal{Y} = \mathbb{R}^m$ , the feasible space  $\mathcal{Y}_a = \{y \in \mathbb{R}^m \mid y \geq b\}$  is a nonempty, closed convex cone, then the primal problem (2.15) is the so-called *concave minimization problem*. Concave minimization problems constitute one of the most fundamental and intensely-studied classes of problems in global optimization. Generally speaking, concave minimization problems are NP-hard and will possess many solutions that are local, but not global,

minima. For this reason, concave optimization problems are also called *multiextremal* global optimization problems (see Benson, 1995). The application of standard algorithms designed for solving constrained convex programming problems will generally fail to solve multiextremal global optimization problems.

Since  $F(x) = -f(x)$  is convex, the extended problem

$$P(x) = W(\Lambda x) - F(x) = \mathcal{J}_{y_a}(\Lambda x) + f(x) \rightarrow \min \quad \forall x \in \mathcal{X} \quad (2.21)$$

is a d.c. optimization problem. The classical Lagrangian associated with this nonconvex optimization with inequality constraint reads

$$L(x, \mu) = f(x) + \langle \Lambda x - b ; \mu \rangle, \quad (2.22)$$

where  $\mu \in \mathcal{Y}^*$  is a Lagrange multiplier. Since  $L(x, \mu)$  is concave in  $x$ , and we have  $\inf_x L(x, \mu) = -\infty$ , the classical saddle Lagrange duality does not work for this nonconvex problem. The extremality condition  $DL(x, \mu) = 0$  leads to the Euler-Lagrange equation

$$\Lambda^* \mu + Df(x) = 0, \quad \Lambda x = b \quad (2.23)$$

subjected to the KKT condition

$$\Lambda x - b \geq 0, \quad \langle \Lambda x - b ; \mu \rangle = 0, \quad \mu \leq 0. \quad (2.24)$$

For nonsmooth  $f$ , traditional direct methods for solving this nonlinear complementarity problem is very difficult. In this paper, we will show that by use of the super-Lagrange duality theory, this constrained nonconvex minimization can be converted into an unconstrained convex minimization dual problem.

#### EXAMPLE 2. Geometrically nonlinear problems.

Let us now consider the nonconvex optimization problem in  $\mathcal{X} = \mathbb{R}^n$

$$P(x) = \frac{1}{2}a\left(\frac{1}{2}\|Ax\|^2 - \mu\right)^2 - x^T c \rightarrow \text{sta} \quad \forall x \in \mathbb{R}^n, \quad (2.25)$$

where  $a > 0$ ,  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a matrix in  $\mathbb{R}^{m \times n}$  and  $c \in \mathbb{R}^n$  is a given vector. Clearly, for any given parameter  $\mu > 0$ ,  $P(x)$  is nonconvex on  $\mathbb{R}^n$ . The nonconvex problem (2.25) appears very often in many applications of physics, engineering and sciences. For example, in the case that  $n = m = 1$ ,  $A = 1$ ,

$$P(x) = \frac{1}{2}a\left(\frac{1}{2}x^2 - \mu\right)^2 - cx$$

is a *double-well function* (see Figure 4a), which was first studied by van der Waals in fluids mechanics in 1895. If  $n = m = 2$ ,  $A = \mathbf{I} \in \mathbb{R}^{2 \times 2}$  is an identity, then

$$P(x) = \frac{1}{2}a\left(\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - \mu\right)^2 - c_1x_1 - c_2x_2.$$

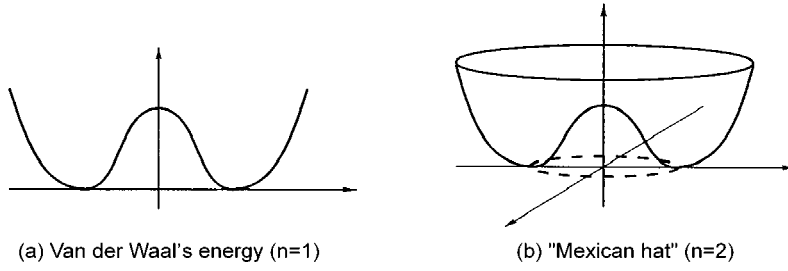


Figure 4. Illustration of the nonconvex function in problem (2.25).

For  $c = 0$ , this is the so-called ‘Mexican hat’ function (see Figure 4b) in cosmology and theoretical physics. In phase transitions of shape memory alloys, each local minimizer of the total potential  $P$  corresponding to a certain phase state of material. However, each local maximizer characterizes the critical conditions that leads to the phase transitions. In unilateral post-bifurcation analysis, the solution of the post-buckling state is usually a local minimizer (see Gao, 1998b).

Following the traditional way, we first let  $y = \Lambda = A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear operator, such that  $P(x) = W(Ax) - F(x)$  with

$$W(y) = \frac{1}{2}a\left(\frac{1}{2}y^T y - \mu\right)^2, \quad F(x) = x^T c.$$

By the Fenchel-Rockafellar dual theory, the classical dual problem associated with the linear operator  $\Lambda = A$  is

$$P^d(y^*) = -W^\sharp(y^*) \rightarrow \max \quad s.t. \quad A^* y^* = c. \quad (2.26)$$

Since the nonconvex  $W(y)$  is not a canonical function, the constitutive equation  $y^* = DW(y)$  is not one-to-one. Thus, the Legendre conjugate of  $W(y)$  does not have a simple algebraic expression. Although the Fenchel conjugate  $W^\sharp(y^*)$  is convex in  $\mathbb{R}^m$ , there exists a duality gap between the primal problem (2.25) and the Fenchel-Rockafellar dual problem (2.26), i.e.,  $\inf P(x) \geq \sup P^d(y^*)$  due to the nonconvexity of  $P$ . This duality gap shows that the Fenchel-Rockafellar duality theory can be used mainly for convex geometrically linear problems.

To put the nonconvex problem (2.25) in our canonical framework, we need to let  $\Lambda : \mathbb{R}^n \rightarrow \mathbb{R}$  be a quadratic operator

$$y = \Lambda(x) = \frac{1}{2}\|Ax\|^2 - \mu = \frac{1}{2}x^T Cx - \mu,$$

where  $C = A^T A = C^T \in \mathbb{R}^{nn}$ . In finite deformation theory, this nice symmetrical matrix  $C$  is the well-known *right Cauchy-Green strain tensor*. However, in differential geometry,  $C$  is called the *Riemannian metric tensor*. Then, in terms of  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}$ ,  $\Phi(x, y) = W(y) - F(x) = \frac{1}{2}ay^2 - x^T c$  is a canonical function on  $\mathbb{R}^n \times \mathbb{R}$ . The Legendre conjugate of the quadratic function  $W(y) = \frac{1}{2}ay^2$  is

simply defined by  $W^*(y^*) = \frac{1}{2}a^{-1}y^{*2}$ . The canonical constitutive equations

$$x^* = DF(x) = c, \quad y^* = DW(y) = ay$$

are linear. The tri-canonical equations then can be listed as

$$y = \frac{1}{2}\|Ax\|^2 - \mu, \quad y^* = ay, \quad (Cx)^T y^* = c.$$

Since the geometrical operator  $\Lambda$  is nonlinear, and the canonical constitutive equation is linear, the primal problem (2.25) is a geometrically nonlinear optimization problem in  $\mathbb{R}^n$ .

### 3. Canonical Dual Transformation and Extended Lagrangians

The goal of this section is to discuss the extended Lagrangians associated with the fully nonlinear, nonconvex primal problem

$$(\mathcal{P}_{\text{ext}}) : \quad P(x) = \Phi(x, \Lambda(x)) \rightarrow \text{ext} \quad \forall x \in \mathcal{X}$$

where  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is a Gâteaux differentiable operator such that  $\Phi \in \Gamma_0(\mathcal{X}) \times \Gamma_0(\mathcal{Y})$  is an extended canonical function, which is finite and Gâteaux differentiable on  $\mathcal{X}_a \times \mathcal{Y}_a$ , i.e.  $\Phi \in \Gamma_G(\mathcal{X}_a) \times \Gamma_G(\mathcal{Y}_a)$ . Thus, the implicit constraint of  $(\mathcal{P}_{\text{ext}})$  is  $x \in \mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda(x) \in \mathcal{Y}_a\}$ . A systematic presentation on the extended Lagrange duality for geometrically linear systems was given by Gao (1999). Our aim here is to study general fully nonlinear, global optimization problems.

For any fixed  $x \in \mathcal{X}$ , the partial conjugate function of  $\Phi$  with respect to  $y$  is defined by

$$\Phi_y^*(x, y^*) = \text{ext}\{\langle y ; y^* \rangle - \Phi(x, y) \mid \forall y \in \mathcal{Y}\}.$$

Clearly, if  $\Phi(x, \cdot) \in \Gamma_G(\mathcal{Y}_a)$ , and  $\mathcal{Y}_a^* \subset \mathcal{Y}^*$  is the range of the mapping  $D_y\Phi : \mathcal{Y}_a \rightarrow \mathcal{Y}$ , then the Legendre duality relation

$$y^* = D_y\Phi(x, y) \Leftrightarrow y = D_{y^*}\Phi_y^*(x, y^*) \Leftrightarrow \Phi(x, y) + \Phi_y^*(x, y) = \langle y ; y^* \rangle$$

holds on  $\mathcal{Y}_a \times \mathcal{Y}_a^*$ . For the canonical function  $\Phi \in \Gamma_0(\mathcal{X}) \times \Gamma_0(\mathcal{Y})$ ,  $\Phi_y^{**}(x, y) = \Phi(x, y)$  holds on  $\mathcal{X} \times \mathcal{Y}$ . Thus, on the so-called *canonical phase space*  $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}^*$ , the function  $H : \mathcal{Z} \rightarrow \bar{\mathbb{R}}$  defined by

$$H(x, y^*) = \Phi_y^*(x, y^*) \in \Gamma(\mathcal{X}) \times \Gamma(\mathcal{Y}^*) \quad (3.27)$$

is called the *canonical Hamiltonian* associated with  $\Phi$ .

Symmetrically, for a fixed  $y \in \mathcal{Y}$ , the partial conjugate of  $\Phi$  with respect to  $x$  is

$$\Phi_x^*(x^*, y) = \text{ext}_{x \in \mathcal{X}}\{\langle x ; x^* \rangle - \Phi(x, y)\}.$$

If  $\Phi(\cdot, y) \in \Gamma_G(\mathcal{X}_a)$ , and  $\mathcal{X}_a^* \subset \mathcal{X}^*$  is the range of the mapping  $D_x \Phi : \mathcal{X}_a \rightarrow \mathcal{X}$ , then the Legendre duality relation

$$\Phi(x, y) + \Phi_x^*(x, y) = \langle x, x^* \rangle$$

holds on  $\mathcal{X}_a \times \mathcal{X}_a^*$ .

If the geometrical operator  $\Lambda : \mathcal{X}_a \rightarrow \mathcal{Y}_a$  is linear and its adjoint operator  $\Lambda^* : \mathcal{Y}_a^* \rightarrow \mathcal{X}_a^*$  is onto, then the complementary Hamiltonian  $H^c : \mathcal{X} \times \mathcal{Y}^* \rightarrow \bar{\mathbb{R}}$  can be defined by

$$H^c(x, y^*) = -\Phi_x^*(\Lambda^* y^*, \Lambda x). \quad (3.28)$$

DEFINITION 3. For a given problem  $(\mathcal{P}_{\text{ext}})$ , if there exists a Gâteaux differentiable operator  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  and an extended canonical function  $\Phi \in \Gamma_0(\mathcal{X}) \times \Gamma_0(\mathcal{Y})$  such that  $P(x) = \Phi(x, \Lambda(x))$ , then the function  $L : \mathcal{Z} = \mathcal{X} \times \mathcal{Y}^* \rightarrow \bar{\mathbb{R}}$  defined by

$$L(x, y^*) = \langle \Lambda(x); y^* \rangle - H(x, y^*) \quad (3.29)$$

is called the *extended Lagrangian form* of  $(\mathcal{P}_{\text{ext}})$  associated with  $\Lambda$ . It is called the *canonical Lagrangian* if  $L \in \Gamma(\mathcal{X}) \times \Gamma(\mathcal{Y}^*)$ .  $\square$

Clearly, for any given  $x \in \mathcal{X}$ , the extended Lagrangian  $L(x, \cdot) \in \Gamma(\mathcal{Y}^*)$  is a canonical function of  $y^*$  and

$$P(x) = \text{ext}_{y^* \in \mathcal{Y}^*} L(x, y^*) \quad \forall x \in \mathcal{X}.$$

Thus, for linear  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$ ,  $L$  defined by (3.29) is always a canonical Lagrangian form. However, in geometrically nonlinear systems the convexity of  $L(\cdot, y^*) : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  will depend on the operator  $\Lambda$  and the canonical dual variable  $y^*$ .

A point  $(\bar{x}, \bar{y}^*) \in \mathcal{X} \times \mathcal{Y}^*$  is said to be a critical point of  $L$  if  $L$  is Gâteaux-differentiable at  $(\bar{x}, \bar{y}^*)$  and

$$D_x L(\bar{x}, \bar{y}^*) = 0, \quad D_{y^*} L(\bar{x}, \bar{y}^*) = 0.$$

It is easy to find out that the criticality condition  $DL(\bar{x}, \bar{y}^*) = 0$  is equivalent to the following *canonical Lagrange equations*

$$DL(\bar{x}, \bar{y}^*) = 0 \Rightarrow \begin{cases} \Lambda_t^*(\bar{x})\bar{y}^* = D_x \Phi_y^*(\bar{x}, \bar{y}^*), \\ \Lambda(\bar{x}) = D_{y^*} \Phi_y^*(\bar{x}, \bar{y}^*). \end{cases} \quad (3.30)$$

In global optimization, the following definitions are needed for the purpose of studying the generalized Lagrange duality (see, Gao, 1998).

DEFINITION 4. Let  $L : \mathcal{Z} \rightarrow \bar{\mathbb{R}}$  be an arbitrary given function, and  $\mathcal{Z}_a = \mathcal{X}_a \times \mathcal{Y}_a^*$  an open set in  $\mathcal{Z}$ .

A point  $(\bar{x}, \bar{y}^*)$  is said to be a *right-saddle point* of  $L$  on  $\mathcal{Z}_a$  if

$$L(\bar{x}, y^*) \leq L(\bar{x}, \bar{y}^*) \leq L(x, \bar{y}^*) \quad \forall (x, y^*) \in \mathcal{Z}_a. \quad (3.31)$$

A point  $(\bar{x}, \bar{y}^*)$  is said to be a *left-saddle point* of  $L$  on  $\mathcal{Z}_a$  if

$$L(\bar{x}, y^*) \geq L(\bar{x}, \bar{y}^*) \geq L(x, \bar{y}^*) \quad \forall (x, y^*) \in \mathcal{Z}_a. \quad (3.32)$$

A point  $(\bar{x}, \bar{y}^*)$  is said to be a *sub-critical* (or  $\partial^-$ -critical) point of  $L$  on  $\mathcal{Z}_a$  if

$$L(\bar{x}, y^*) \geq L(\bar{x}, \bar{y}^*) \leq L(x, \bar{y}^*) \quad \forall (x, y^*) \in \mathcal{Z}_a. \quad (3.33)$$

A point  $(\bar{x}, \bar{y}^*)$  is said to be a *super-critical* (or  $\partial^+$ -critical) point of  $L$  on  $\mathcal{Z}_a$  if

$$L(\bar{x}, y^*) \leq L(\bar{x}, \bar{y}^*) \geq L(x, \bar{y}^*) \quad \forall (x, y^*) \in \mathcal{Z}_a. \quad (3.34)$$

In convex analysis, the right-saddle point is simply called the saddle point. By the definitions of the extended differentials, the following results show the reason why the names of the super- and sub-Lagrangians were introduced.

1. A point  $(\bar{x}, \bar{y}^*)$  is a right-saddle point of  $L$  on  $\mathcal{Z}$  if and only if

$$0 \in \partial_x^- L(\bar{x}, \bar{y}^*), \quad 0 \in \partial_{y^*}^+ L(\bar{x}, \bar{y}^*). \quad (3.35)$$

2. A point  $(\bar{x}, \bar{y}^*)$  is a sub-critical point of  $L$  on  $\mathcal{Z}$  if and only if

$$0 \in \partial_x^- L(\bar{x}, \bar{y}^*), \quad 0 \in \partial_{y^*}^- L(\bar{x}, \bar{y}^*). \quad (3.36)$$

3. A point  $(\bar{x}, \bar{y}^*)$  is a super-critical point of  $L$  on  $\mathcal{Z}$  if and only if

$$0 \in \partial_x^+ L(\bar{x}, \bar{y}^*), \quad 0 \in \partial_{y^*}^+ L(\bar{x}, \bar{y}^*). \quad (3.37)$$

In geometrically linear systems ( $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is linear), the inequalities (3.34) are equivalent to following symmetrical canonical Hamilton forms:

$$\Lambda \bar{x} \in \partial_{y^*}^- H(\bar{x}, \bar{y}^*), \quad \Lambda^* \bar{y}^* \in \partial_x^- H(\bar{x}, \bar{y}^*).$$

This is the definition of the so-called *anomalous critical points*, introduced by Auchmuty in geometrically linear problems, which is a special case of the super-critical points.

Let  $L : \mathcal{Z} \rightarrow \bar{\mathbb{R}}$  be a given arbitrary extended function, which is Gâteaux differentiable on  $\mathcal{Z}_a = \mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{Z}$ . Two functions associated with  $L(x, y^*)$  can be defined by

$$P(x) = \text{sta}_{y^* \in \mathcal{Y}_a^*} L(x, y^*) \quad \forall x \in \mathcal{X}_a, \quad (3.38)$$

$$P^d(y^*) = \text{sta}_{x \in \mathcal{X}_a} L(x, y^*) \quad \forall y^* \in \mathcal{Y}_a^*. \quad (3.39)$$

The following lemma plays a key role in duality theory for global optimization.



LEMMA 1. Let  $L(x, y^*)$  be an arbitrary function, partially Gâteaux differentiable on an open subset  $\mathcal{Z}_a = \mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{X} \times \mathcal{Y}^*$ . If  $(\bar{x}, \bar{y}^*) \in \mathcal{X}_a \times \mathcal{Y}_a^*$  is one of the (either right- or left-) saddle points, the super- or sub-critical points of  $L$ , then  $(\bar{x}, \bar{y}^*)$  is a critical point of  $L$  on  $\mathcal{Z}_a$ .

Moreover, if  $P$  is Gâteaux differentiable at  $\bar{x}$ , and  $P^d$  is Gâteaux differentiable at  $\bar{y}^*$ , then  $DP(\bar{x}) = 0$ ,  $DP^d(\bar{y}^*) = 0$ , and

$$P(\bar{x}) = L(\bar{x}, \bar{y}^*) = P^d(\bar{y}^*). \quad (3.40)$$

The proof of this lemma can be found in Gao (1998b, 1999) in parametrical variational analysis.

Any critical point of a Gâteaux differentiable saddle-Lagrangian (resp. super-Lagrangian) is a saddle-critical (resp. super-critical) point. However, if  $(\bar{x}, \bar{y}^*)$  is a saddle-critical (or super-critical) point of  $L$ , it does not follow that the extended Lagrangian  $L$  is a saddle-Lagrangian (or super-Lagrangian) since  $L$  is not necessary to be a canonical function.

Clearly,  $(\bar{x}, \bar{y}^*)$  is a left-saddle (resp. sub-critical) critical point of  $L$  if and only if it is a right-saddle (resp. super-critical) point of  $-L$ . In the following, we only discuss the right and super-Lagrangians. Let  $\mathcal{Z}_r = \mathcal{X}_r \times \mathcal{Y}_r^* \subset \mathcal{X} \times \mathcal{Y}^*$  be an open subset. In global optimization, the following statements are of important theoretical value.

(S1) Under certain necessary and sufficient conditions, if

$$\inf_{x \in \mathcal{X}_r} \sup_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = \sup_{y^* \in \mathcal{Y}_r^*} \inf_{x \in \mathcal{X}_r} L(x, y^*) \quad (3.41)$$

holds, then a statement of this type is called a *saddle-minimax theorem* and the pair  $(\bar{x}, \bar{y}^*)$  is called a *saddle-minimax point* of  $L$  on  $\mathcal{Z}_r$ .

(S2) Under certain necessary and sufficient conditions if

$$\inf_{x \in \mathcal{X}_r} \sup_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = \inf_{y^* \in \mathcal{Y}_r^*} \sup_{x \in \mathcal{X}_r} L(x, y^*). \quad (3.42)$$

A statement of this type is called a *super-minimax theorem* and the pair  $(\bar{x}, \bar{y}^*)$  is called a *super-minimax point* of  $L$  on  $\mathcal{Z}_r$ .

(S3) Under certain conditions, a pair  $(\bar{x}, \bar{y}^*) \in \mathcal{Z}_r$  exists such that

$$L(x, \bar{y}^*) \leq L(\bar{x}, \bar{y}^*) \geq L(\bar{x}, y^*) \quad (3.43)$$

holds for all  $(x, y^*) \in \mathcal{Z}_r$ . A statement of this type is called a *super-critical point theorem*.

By the fact that the suprema of  $L(x, y^*)$  can be taken in either order on  $\mathcal{X}_r \times \mathcal{Y}_r^*$ , the equality

$$\sup_{x \in \mathcal{X}_r} \sup_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = \sup_{y^* \in \mathcal{Y}_r^*} \sup_{x \in \mathcal{X}_r} L(x, y^*) \quad (3.44)$$

always holds. This fact is trivial in convex systems but important in global optimization. A pair  $(\bar{x}, \bar{y}^*)$  which maximizes  $L$  on  $\mathcal{Z}_r$  is called a local *super-maximum point* of  $L$  on  $\mathcal{Z}_r$ .

In classical saddle Lagrange duality theory, the primal and dual functions associated with  $L$  are defined by the saddle Lagrange dual transformation:

$$P(x) = \sup_{y^* \in \mathcal{Y}^*} L(x, y^*) \quad \forall x \in \mathcal{X}, \quad (3.45)$$

$$P^d(y^*) = \inf_{x \in \mathcal{X}} L(x, y^*) \quad \forall y^* \in \mathcal{Y}^*. \quad (3.46)$$

The *weak minimax duality*

$$\inf_{x \in \mathcal{X}} P(x) \geq \sup_{y^* \in \mathcal{Y}^*} P^d(y^*)$$

is always held for any function  $L(x, y^*)$ . For saddle Lagrangian, the following theorem is well-known (cf. e.g., Walk, 1989; Gao, 1999).

**THEOREM 1.** *Let  $L : \mathcal{X} \times \mathcal{Y}^* \rightarrow \bar{\mathbb{R}}$  be a saddle-Lagrangian such that the functions  $P : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  and  $P^d : \mathcal{Y}^* \rightarrow \bar{\mathbb{R}}$  are well-defined by (3.45) and (3.46), respectively, and that the effective domains  $\mathcal{X}_k = \text{dom } P \subset \mathcal{X}$ ,  $\mathcal{Y}_s^* = \text{dom } P^d \subset \mathcal{Y}^*$  are not empty. Then the strong saddle-minimax duality theorem in the form*

$$\inf_{x \in \mathcal{X}_k} P(x) = \inf_{x \in \mathcal{X}_k} \sup_{y^* \in \mathcal{Y}_a^*} L(x, y^*) = \sup_{y^* \in \mathcal{Y}_s^*} \inf_{x \in \mathcal{X}_a} L(x, y^*) = \sup_{y^* \in \mathcal{Y}_s^*} P^d(y^*) \quad (3.47)$$

holds.

In engineering mechanics, the primal feasible set  $\mathcal{X}_k$  is called the *kinetically admissible space*, the dual feasible set  $\mathcal{Y}_s^*$  is referred as the *statically admissible space*. In the case that  $\Phi(x, y) = W(y) - F(x)$  with  $W \in \check{\Gamma}(\mathcal{Y})$  and  $F \in \hat{\Gamma}(\mathcal{X})$ , the extended Lagrangian takes the form

$$L(x, y^*) = \langle \Lambda(x) ; y^* \rangle - W^*(y^*) - F(x). \quad (3.48)$$

By the Fenchel transformation, for any given  $x \in \mathcal{X}$  we have

$$P(x) = \sup_{y^* \in \mathcal{Y}^*} L(x, y^*) = W^{\sharp\sharp}(\Lambda(x)) - F(x) = W(\Lambda(x)) - F(x)$$

for all  $W \in \check{\Gamma}_0(\mathcal{Y}_a)$ . The effective domain of  $P$  is  $\mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda(x) \in \mathcal{Y}_a\}$ . On the other hand, if  $\Lambda$  is a linear operator, then for any given  $y^* \in \mathcal{Y}^*$ , the Fenchel-Rockafellar dual function takes the form

$$P^d(y^*) = \inf_{x \in \mathcal{X}} L(x, y^*) = F^{\flat}(\Lambda^* y^*) - W^{\sharp\sharp}(y^*). \quad (3.49)$$

The effective domain of  $P^d$  is  $\mathcal{Y}_s^* = \text{dom } P = \{y^* \in \mathcal{Y}_a^* \mid \Lambda^* y^* \in \mathcal{X}_a^*\}$ .

However, in geometrically nonlinear systems, the dual function  $P^d$  and the dual feasible set  $\mathcal{Y}_s^*$  will depend on the nonlinear operator  $\Lambda$ .

#### 4. Bi-Duality Theory in Global Optimization

In this section, we study the bi-duality theory for general nonconvex systems. We assume that  $L : \mathcal{X}_a \times \mathcal{Y}_a^* \rightarrow \mathbb{R}$  is a given arbitrary function. We let  $\mathcal{X}_k \subseteq \mathcal{X}_a$  and  $\mathcal{Y}_s^* \subseteq \mathcal{Y}_a^*$  be two subsets such that

$$\begin{aligned} \sup_{y^* \in \mathcal{Y}_a^*} L(x, y^*) &< +\infty \quad \forall x \in \mathcal{X}_k, \\ \sup_{x \in \mathcal{X}_a} L(x, y^*) &< +\infty \quad \forall y^* \in \mathcal{Y}_s^*. \end{aligned}$$

The super-critical point duality theorem proposed by Gao (1999) is also true for global optimization problems.

**THEOREM 2.** *Let the Lagrangian  $L : \mathcal{X} \times \mathcal{Y}^* \rightarrow \bar{\mathbb{R}}$  be a given arbitrary function. If there exists either a super-maximum point  $(\bar{x}, \bar{y}^*) \in \mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{X} \times \mathcal{Y}^*$  such that*

$$\max_{x \in \mathcal{X}_a} \max_{y^* \in \mathcal{Y}_a^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \max_{y^* \in \mathcal{Y}_a^*} \max_{x \in \mathcal{X}_a} L(x, y^*), \quad (4.50)$$

or a super-minimax point  $(\bar{x}, \bar{y}^*) \in \mathcal{X}_a \times \mathcal{Y}_a^*$  such that

$$\min_{x \in \mathcal{X}_a} \max_{y^* \in \mathcal{Y}_a^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \min_{y^* \in \mathcal{Y}_a^*} \max_{x \in \mathcal{X}_a} L(x, y^*), \quad (4.51)$$

then  $(\bar{x}, \bar{y}^*)$  is a super-critical point of  $L$  on  $\mathcal{X}_a \times \mathcal{Y}_a^*$ .

Dually, if  $L$  is partially Gâteaux differentiable on an open set  $\mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{X} \times \mathcal{Y}^*$ , and  $(\bar{x}, \bar{y}^*)$  is a super-critical point of  $L$  on the open subset  $\mathcal{X}_k \times \mathcal{Y}_s^* \subset \mathcal{X}_a \times \mathcal{Y}_a^*$ , then either the super-maximum theorem in the form

$$\max_{x \in \mathcal{X}_k} \max_{y^* \in \mathcal{Y}_s^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \max_{y^* \in \mathcal{Y}_s^*} \max_{x \in \mathcal{X}_a} L(x, y^*), \quad (4.52)$$

holds, or the super-minimax theorem in the form

$$\min_{x \in \mathcal{X}_k} \max_{y^* \in \mathcal{Y}_s^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \min_{y^* \in \mathcal{Y}_s^*} \max_{x \in \mathcal{X}_a} L(x, y^*) \quad (4.53)$$

holds.

The proof of this theorem can be found in Gao (1999). This theorem plays an important role in d.c. programming and dynamical systems. In particular, if  $L \in \hat{\Gamma}(\mathcal{X}) \times \hat{\Gamma}(\mathcal{Y}^*)$ , then we have the following *super-Lagrangian duality theorem*.

**THEOREM 3.** *Let  $L \in \hat{\Gamma}(\mathcal{X}) \times \hat{\Gamma}(\mathcal{Y}^*)$  be partially Gâteaux differentiable on an open set  $\mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{X} \times \mathcal{Y}^*$ , and  $(\bar{x}, \bar{y}^*)$  is a critical point of  $L$  on the open subset  $\mathcal{X}_k \times \mathcal{Y}_s^* \subset \mathcal{X}_a \times \mathcal{Y}_a^*$ , then either the super-maximum theorem in the form*

$$\sup_{x \in \mathcal{X}_k} \sup_{y^* \in \mathcal{Y}_a^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \sup_{y^* \in \mathcal{Y}_s^*} \sup_{x \in \mathcal{X}_a} L(x, y^*) \quad (4.54)$$

*holds, or the super-minimax theorem in the form*

$$\inf_{x \in \mathcal{X}_k} \sup_{y^* \in \mathcal{Y}_a^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \inf_{y^* \in \mathcal{Y}_s^*} \sup_{x \in \mathcal{X}_a} L(x, y^*) \quad (4.55)$$

*holds.*

*Proof.* Since  $L \in \hat{\Gamma}(\mathcal{X}) \times \hat{\Gamma}(\mathcal{Y}^*)$  is a super-critical function on  $\mathcal{X} \times \mathcal{Y}^*$ , then its critical points must be the super-critical point on  $\mathcal{X} \times \mathcal{Y}^*$ . The theorem can be easily proved by use of Theorem 2.  $\square$

For a given arbitrary function  $L : \mathcal{Z}_a \rightarrow \mathbb{R}$ , we let

$$P(x) = \sup_{y^* \in \mathcal{Y}_a^*} L(x, y^*) \quad \forall x \in \mathcal{X}_a, \quad (4.56)$$

$$P^d(y^*) = \sup_{x \in \mathcal{X}_a} L(x, y^*) \quad \forall y^* \in \mathcal{Y}_a^*. \quad (4.57)$$

Both  $P : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  and  $P^d : \mathcal{Y}^* \rightarrow \bar{\mathbb{R}}$  are generally nonconvex. Thus, the primal and dual problems associated with  $L$  can be proposed as

$$(\mathcal{P}_{\text{ext}}) : P(x) \rightarrow \text{ext} \quad \forall x \in \mathcal{X}, \quad (4.58)$$

$$(\mathcal{P}_{\text{ext}}^d) : P^d(y^*) \rightarrow \text{ext} \quad \forall y^* \in \mathcal{Y}^*. \quad (4.59)$$

The problems  $(\mathcal{P}_{\text{ext}})$  and  $(\mathcal{P}_{\text{ext}}^d)$  are realisable if their effective domains  $\mathcal{X}_k$  and  $\mathcal{Y}_s^*$  are not empty. In classical convex optimization, the maximization problem of  $P$  is usually replaced by the minimization problem of  $-P$ . However, this is not true in global optimization, and in general,  $(\mathcal{P}_{\text{inf}})$  and  $(\mathcal{P}_{\text{sup}})$  are two different problems.

**THEOREM 4 (Bi-duality theorem).** *Let  $L : \mathcal{X}_a \times \mathcal{Y}_a^* \rightarrow \mathbb{R}$  be a given arbitrary function such that  $P$  and  $P^d$  are well-defined by (4.58) and (4.59) on the nonempty open effective domains  $\mathcal{X}_k$  and  $\mathcal{Y}_s^*$ , respectively. If  $(\bar{x}, \bar{y}^*) \in \mathcal{X}_k \times \mathcal{Y}_s^*$  is a super-critical point of  $L$  on the open domain  $\mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{X} \times \mathcal{Y}^*$ , then*

$$P(\bar{x}) = \inf_{x \in \mathcal{X}_k} P(x) \quad \text{if and only if} \quad \inf_{y^* \in \mathcal{Y}_s^*} P^d(y^*) = P^d(\bar{y}^*); \quad (4.60)$$

$$P(\bar{x}) = \sup_{x \in \mathcal{X}_k} P(x) \quad \text{if and only if} \quad \sup_{y^* \in \mathcal{Y}_s^*} P^d(y^*) = P^d(\bar{y}^*). \quad (4.61)$$

*Proof.* This theorem follows from the combination of the Lemma 1 and Theorem 2.  $\square$

In the case that  $L(x, y^*) = \langle \Lambda(x) ; y^* \rangle - W^*(y^*) - F(x)$  is an extended Lagrange form associated with a d.c. function  $P(x) = W(\Lambda(x)) - F(x)$ , then the dual function reads

$$P^d(y^*) = F_\Lambda^d(y^*) - W^*(y^*), \quad (4.62)$$

where  $F_\Lambda^d : \mathcal{Y}^* \rightarrow \bar{\mathbb{R}}$  is the so-called  $\Lambda$ -dual function of  $F$  defined by the following  $\Lambda$ -dual transformation:

$$F_\Lambda^d(y^*) = \text{sta}\{\langle \Lambda(x) ; y^* \rangle - F(x) \mid \forall x \in \mathcal{X}\}. \quad (4.63)$$

In geometrically linear systems, the statement (4.60) reduces to Auchmuty's anomalous duality theorem. In particular, if the primal function can be written as  $P(x) = W(\Lambda x) - F(x)$  with  $W \in \check{\Gamma}_G(\mathcal{Y}_a)$  and  $F \in \check{\Gamma}_G(\mathcal{X}_a)$ , then the effective domain  $\text{dom } P = \mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda x \in \mathcal{Y}_a\}$ . The dual function

$$P^d(y^*) = F^\sharp(\Lambda^* y^*) - W^\sharp(y^*)$$

is also a d.c. function with effective domain  $\text{dom } P^d = \mathcal{Y}_s^* = \{y^* \in \mathcal{Y}_a^* \mid \Lambda^* y^* \in \mathcal{X}_a\}$ . In this special case, the statement (4.60) is a more precise version of the Toland's double-min duality theorem. In convex Hamilton systems, the total action  $P$  of the system is a d.c. functional (the difference of the total kinetic energy and the total potential energy). Since  $P$  is not convex, the problem may have many local extrema. In periodic dynamics, both local minima and local maxima are the equilibrium states of the systems, and have to be considered simultaneously. As a traditional minimization problem, the well-known least action principle is in fact a misnomer. The bi-duality theory, however, gives a complete picture for this type of problems.

## 5. Triality Theory in Fully Nonlinear Problems

The triality theory was originally proposed by the author (Gao, 1996, 1997, 1999) from post-buckling problems in finite deformation theory, where the geometrical operator  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  is a quadratic mapping (the right Cauchy-Green tensor). In this section, we will generalize this interesting result into global optimization problems. We assume that for any given nonconvex extended function  $P : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ , there exists a general nonlinear operator  $\Lambda : \mathcal{X} \rightarrow \mathcal{Y}$  and a canonical function  $W \in \Gamma(\mathcal{Y})$  such that the canonical transformation can be written as

$$P(x) = W(\Lambda(x)) - \langle x, c \rangle, \quad c \in \mathcal{X}^*. \quad (5.64)$$

Since  $F(x) = \langle x, c \rangle$  is a linear function, the Hamiltonian  $H(x, y^*) = W^*(y^*) + \langle x, c \rangle$  is a canonical function on  $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}^*$  and the extended Lagrangian reads

$$L(x, y^*) = \langle \Lambda(x) ; y^* \rangle - W^*(y^*) - \langle x, c \rangle. \quad (5.65)$$

For a fixed  $y^* \in \mathcal{Y}^*$ , the convexity of  $L(\cdot, y^*) : \mathcal{X} \rightarrow \bar{\mathbb{R}}$  depends on  $\Lambda(x)$  and  $y^* \in \mathcal{Y}^*$ .

Let  $\mathcal{Z}_a = \mathcal{X}_a \times \mathcal{Y}_a^* \subset \mathcal{Z}$  be the effective domain of  $L$ , and let  $\mathcal{L}_c \subset \mathcal{Z}_a$  be a critical point set of  $L$ , i.e.

$$\mathcal{L}_c = \{(\bar{x}, \bar{y}) \in \mathcal{X}_a \times \mathcal{Y}_a^* \mid \delta L(\bar{x}, \bar{y}^*; x, y^*) = 0 \quad \forall (x, y^*) \in \mathcal{X}_a \times \mathcal{Y}_a^*\}.$$

For any given critical point  $(\bar{x}, \bar{y}^*) \in \mathcal{L}_c$ , we let  $\mathcal{X}_r \times \mathcal{Y}_r^*$  be its neighborhood such that on  $\mathcal{X}_r \times \mathcal{Y}_r^*$ , the pair  $(\bar{x}, \bar{y}^*)$  is the only critical point of  $L$ . The following result is of fundamental importance in global optimization.

**THEOREM 5 (Triality theorem).** *Let  $(\bar{x}, \bar{y}^*) \in \mathcal{L}_c$  be a critical point of  $L$  and  $\mathcal{X}_r \times \mathcal{Y}_r^*$  a neighborhood of  $(\bar{x}, \bar{y}^*)$ .*

*I. Suppose that  $W \in \check{\Gamma}(\mathcal{Y}_a)$  is convex. If  $\langle \Lambda(x) ; \bar{y}^* \rangle$  is convex on  $\mathcal{X}_r$ , then*

$$\min_{x \in \mathcal{X}_r} \max_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \max_{y^* \in \mathcal{Y}_r^*} \min_{x \in \mathcal{X}_r} L(x, y^*). \quad (5.66)$$

*However, if  $\langle \Lambda(x) ; \bar{y}^* \rangle$  is concave on  $\mathcal{X}_r$ , then either*

$$\min_{x \in \mathcal{X}_r} \max_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \min_{y^* \in \mathcal{Y}_r^*} \max_{x \in \mathcal{X}_r} L(x, y^*), \quad (5.67)$$

*or*

$$\max_{x \in \mathcal{X}_r} \max_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \max_{y^* \in \mathcal{Y}_r^*} \max_{x \in \mathcal{X}_r} L(x, y^*). \quad (5.68)$$

*II. Suppose that  $W \in \hat{\Gamma}(\mathcal{Y}_a)$  is concave. If  $\langle \Lambda(x) ; \bar{y}^* \rangle$  is concave on  $\mathcal{X}_r$ , then*

$$\max_{x \in \mathcal{X}_r} \min_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \min_{y^* \in \mathcal{Y}_r^*} \max_{x \in \mathcal{X}_r} L(x, y^*). \quad (5.69)$$

*However, if  $\langle \Lambda(x) ; \bar{y}^* \rangle$  is convex on  $\mathcal{X}_r$ , then either*

$$\max_{x \in \mathcal{X}_r} \min_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \max_{y^* \in \mathcal{Y}_r^*} \min_{x \in \mathcal{X}_r} L(x, y^*), \quad (5.70)$$

*or*

$$\min_{x \in \mathcal{X}_r} \min_{y^* \in \mathcal{Y}_r^*} L(x, y^*) = L(\bar{x}, \bar{y}^*) = \min_{y^* \in \mathcal{Y}_r^*} \min_{x \in \mathcal{X}_r} L(x, y^*). \quad (5.71)$$

*Proof.* For convex  $W(y)$ , its Fenchel conjugate  $W^*(y^*)$  is also convex. If  $\langle \Lambda(x) ; \bar{y}^* \rangle$  is convex on  $\mathcal{X}_r$ , then  $L \in \check{\Gamma}(\mathcal{X}_r) \times \hat{\Gamma}(\mathcal{Y}_r^*)$  is a saddle function and  $(\bar{x}, \bar{y}^*)$  is a saddle point of  $L$  on  $\mathcal{X}_r \times \mathcal{Y}_r^*$ . Thus (5.66) follows from the saddle-Lagrangian duality theorem. However, if  $\langle \Lambda(x) ; \bar{y}^* \rangle$  is concave on  $\mathcal{X}_r$ , then  $L \in \hat{\Gamma}(\mathcal{X}_r) \times \hat{\Gamma}(\mathcal{Y}_r^*)$ , and  $(\bar{x}, \bar{y}^*)$  is a super-critical point of  $L$  on  $\mathcal{X}_r \times \mathcal{Y}_r^*$ . By the super-Lagrangian duality theorem (Theorem 3), we have either (5.67) or (5.68).

Similarly for concave  $W(y)$ . □

Since  $W \in \Gamma(\mathcal{Y}_a)$  is a canonical function, we always have

$$P(x) = \text{ext}\{L(x, y^*) \mid y^* \in \mathcal{Y}^*\} \quad \forall x \in \mathcal{X}_k. \quad (5.72)$$

On the other hand, for a given Gâteaux differentiable geometrical mapping  $\Lambda : \mathcal{X}_a \rightarrow \mathcal{Y}_a$ , the criticality condition  $D_x L(\bar{x}, y^*) = 0$  leads to the equilibrium equation

$$\Lambda_r^*(\bar{x})y^* = c. \quad (5.73)$$

If there exists a subspace  $\mathcal{Y}_s^* \subset \mathcal{Y}_a^*$  such that for any  $y^* \in \mathcal{Y}_s^*$  and a given source variable  $c \in \mathcal{X}^*$ , the equation (5.73) can be solved for  $\bar{x} = \bar{x}(y^*)$ , then by Gao-Strang's decomposition  $\Lambda(x) = \Lambda_r(x)x + \Lambda_c(x)$ , the dual function  $P^d : \mathcal{Y}_s^* \rightarrow \mathbb{R}$  can be written explicitly in the form

$$P^d(y^*) = \text{sta}\{L(x, y^*) \mid x \in \mathcal{X}\} = -G^d(y^*) - W^*(y^*) \quad \forall y^* \in \mathcal{Y}_s^*, \quad (5.74)$$

where  $G^d : \mathcal{Y}^* \rightarrow \mathbb{R}$  is the so-called pure complementary gap function, defined by

$$G^d(y^*) = G(\bar{x}(y^*), y^*) = -\langle \Lambda_c(\bar{x}(y^*)); y^* \rangle. \quad (5.75)$$

For any given critical point  $(\bar{x}, \bar{y}^*) \in \mathcal{L}_c$ , we have  $G^d(\bar{y}^*) = \langle \bar{x}, c \rangle - \langle \Lambda(\bar{x}(\bar{y}^*)); \bar{y}^* \rangle$ . Thus, the Legendre duality relations among the canonical functions  $W$  and  $W^*$  lead to

$$P(\bar{x}) - P^d(\bar{y}^*) = 0 \quad \forall (\bar{x}, \bar{y}^*) \in \mathcal{L}_c. \quad (5.76)$$

This identity shows that there is no duality gap between the nonconvex function  $P$  and its canonical dual function  $P^d$ . Actually the duality gap, which exists in classical duality theories, is now recovered by the complementary gap function  $G(\bar{x}, \bar{y}^*)$ .

**THEOREM 6 (Tri-duality theorem).** *Suppose that  $W \in \check{\Gamma}(\mathcal{Y}_a)$ ,  $(\bar{x}, \bar{y}^*) \in \mathcal{L}_c$  is a critical point of  $L$  and  $\mathcal{X}_r \times \mathcal{Y}_r^*$  is a neighborhood of  $(\bar{x}, \bar{y}^*)$ . If  $\langle \Lambda(x); \bar{y}^* \rangle$  is convex on  $\mathcal{X}_r$ , then*

$$P(\bar{x}) = \min_{x \in \mathcal{X}_r} P(x) \quad \text{if and only if} \quad P^d(\bar{y}^*) = \max_{y^* \in \mathcal{Y}_r^*} P^d(y^*). \quad (5.77)$$

*However, if  $\langle \Lambda(x); \bar{y}^* \rangle$  is concave on  $\mathcal{X}_r$ , then*

$$P(\bar{x}) = \min_{x \in \mathcal{X}_r} P(x) \quad \text{if and only if} \quad P^d(\bar{y}^*) = \min_{y^* \in \mathcal{Y}_r^*} P^d(y^*); \quad (5.78)$$

$$P(\bar{x}) = \max_{x \in \mathcal{X}_r} P(x) \quad \text{if and only if} \quad P^d(\bar{y}^*) = \max_{y^* \in \mathcal{Y}_r^*} P^d(y^*). \quad (5.79)$$

*Proof.* This is a special case of the triality theorem. □

In numerical analysis of many engineering problems (such as finite deformation theory and computational differential geometry), the nonlinear mapping  $\Lambda : \mathcal{X} = \mathbb{R}^n \rightarrow \mathcal{Y} = \mathbb{R}^m$  is usually a symmetrical quadratic operator from  $\mathbb{R}^n$  to  $\mathbb{R}^m$

$$\Lambda(x) = \frac{1}{2}x^T \mathbf{\Lambda} x = \frac{1}{2} \left\{ \sum_{i,j=1}^n \Lambda_{ij}^k x_i x_j \right\} \in \mathbb{R}^m, \quad (5.80)$$

where  $\mathbf{\Lambda} \in \mathbb{R}^{mnn}$  is the third order tensor

$$\mathbf{\Lambda} = \{\Lambda_{ij}^k\} = \{\Lambda_{ji}^k\} \quad i, j = 1, \dots, n, \quad k = 1, \dots, m.$$

By the decomposition  $\Lambda(x) = \Lambda_t(x)x + \Lambda_c(x)$ , the operators  $\Lambda_t$  and its complementary  $\Lambda_c$  have the forms

$$\Lambda_t(x)x = \left\{ \sum_{i,j=1}^n \Lambda_{ij}^k x_i x_j \right\}, \quad \Lambda_c(x) = -\frac{1}{2} \left\{ \sum_{i,j=1}^n \Lambda_{ij}^k x_i x_j \right\} \in \mathbb{R}^m. \quad (5.81)$$

The complementary gap function

$$G(x, y^*) = \langle -\Lambda_c(x); y^* \rangle = \frac{1}{2} \sum_{k=1}^m \sum_{i,j=1}^n \Lambda_{ij}^k x_i x_j y_k^* = \frac{1}{2} x^T \mathbf{H}(y^*) x \quad (5.82)$$

is a quadratic function of  $x \in \mathbb{R}^n$ . Its convexity depends on the Hessian matrix

$$\mathbf{H}(y^*) = \left\{ \sum_{k=1}^m \Lambda_{ij}^k y_k^* \right\} \in \mathbb{R}^{nn}.$$

In finite element analysis of large deformation mechanics problems,  $\mathbf{H}(y^*)$  is usually a sparse matrix. Let  $\mathcal{Y}_s^* \subset \mathcal{Y}_a^*$  be a convex set such that on which, the generalized inverse  $\mathbf{H}^+(y^*)$  of  $\mathbf{H}$  exists and satisfies

$$\mathbf{H}(y^*) = \mathbf{H}(y^*)\mathbf{H}^+(y^*)\mathbf{H}(y^*), \quad \mathbf{H}^+(y^*) = \mathbf{H}^+(y^*)\mathbf{H}(y^*)\mathbf{H}^+(y^*), \quad \forall y^* \in \mathcal{Y}_s^*.$$

Thus, the solution for the equilibrium equation (5.73) is

$$\bar{x} = \mathbf{H}^+(y^*)c \quad \forall y^* \in \mathcal{Y}_s^*.$$

In this case, the canonical dual function associated with the quadratic operator can be written as

$$P^d(y^*) = -\frac{1}{2}c^T \mathbf{H}^+(y^*)c - W^*(y^*), \quad (5.83)$$

which is, in general, a nonconvex function on the dual feasible space  $\mathcal{Y}_s^* \subset \mathbb{R}^m$ . Very often, we have  $n \geq m$ . This dimension reduction is of extremely important in large scale nonconvex programming.



Since  $\Lambda : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a pure quadratic operator, we have  $\Lambda(x) = -\Lambda_c(x)$ . This leads to  $G(x, y^*) = \langle \Lambda(x) ; y^* \rangle$ . In this case, Theorems 5 and 6 reduce into the triality theory proposed in finite deformation mechanics by Gao (1997, 1999). Thus, for a given  $\bar{y}^* \in \mathcal{Y}_a^*$ , the quadratic gap function  $G(\cdot, \bar{y}^*) : \mathcal{X}_a \rightarrow \mathbb{R}$  is convex if and only if the Hessian matrix  $\mathbf{H}(\bar{y}^*)$  is positive-definite.

## 6. Applications

**EXAMPLE 3.** We first consider the geometrically linear nonconvex problems. Recall the constrained minimization of Concave Function in  $\mathbb{R}^n$  discussed in Section 2

$$(\mathcal{P}_{\min}) : \quad \min_{x \in \mathbb{R}^n} f(x) \quad s.t. \quad x \geq 0, \quad \Lambda x \geq b \in \mathbb{R}^m, \quad (6.84)$$

where  $f \in \hat{\Gamma}_0(\mathbb{R}^n)$  is a concave function,  $\Lambda = \{\lambda_{ij}\} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear operator (matrix) in  $\mathbb{R}^{m \times n}$ . To solve this NP-hard problem, we let

$$\mathcal{X}_a = \{x \in \mathcal{X} = \mathbb{R}^n \mid x \geq 0\}, \quad \mathcal{Y}_a = \{y \in \mathcal{Y} = \mathbb{R}^m \mid y \geq b\}.$$

The feasible set  $\mathcal{X}_k$

$$\mathcal{X}_k = \{x \in \mathbb{R}^n \mid x \geq 0, \quad \Lambda x \geq b\}$$

is a convex subset in  $\mathbb{R}^n$ . By letting

$$F(x) = -f(x) + \mathcal{I}_{\mathcal{X}_a}(x), \quad W(y) = \mathcal{I}_{\mathcal{Y}_a}(y).$$

the extended problem can be written as

$$P(x) = \mathcal{I}_{\mathcal{Y}_a}(\Lambda x) + f(x) - \mathcal{I}_{\mathcal{X}_a}(x) \rightarrow \min \quad \forall x \in \mathbb{R}^n.$$

Since  $f \in \hat{\Gamma}(\mathbb{R}^n)$ , the extended function  $P : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is indeed a d.c. function.

For the convex function  $W(y) = \mathcal{I}_{\mathcal{Y}_a}(y)$ , its Fenchel conjugate can be computed as

$$\begin{aligned} W^\sharp(y^*) &= \sup_{y \in \mathbb{R}^m} \{\langle y ; y^* \rangle - W(y)\} = \sup_{y \geq b} \langle y ; y^* \rangle \\ &= \langle b ; y^* \rangle + \mathcal{I}_{\mathcal{Y}_a^*}(y^*) \quad \forall y^* \in \mathbb{R}^m, \end{aligned}$$

where

$$\mathcal{Y}_a^* = \{y^* \in \mathbb{R}^m \mid y^* \leq 0\}$$

is a negative cone in  $\mathbb{R}^m$ . Then, the extended Lagrangian associated with this nonconvex optimization with inequality constraint reads

$$L(x, y^*) = \langle \Lambda x - b ; y^* \rangle - \mathcal{I}_{\mathcal{Y}_a^*}(y^*) + f(x) - \mathcal{I}_{\mathcal{X}_a}(x). \quad (6.85)$$

Clearly,  $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$  is a super-Lagrangian. Its effective domain

$$\mathcal{Z}_a = \mathcal{X}_a \times \mathcal{Y}_a^* = \{(x, y^*) \in \mathbb{R}^n \times \mathbb{R}^m \mid x \geq 0, y^* \leq 0\}$$

is a convex set in  $\mathbb{R}^n \times \mathbb{R}^m$ . On  $\mathcal{Z}_a$ , the criticality condition  $DL(\bar{x}, \bar{y}^*) = 0$  will lead to a so-called *bi-complementarity problem* (see Gao, 1998a). For any given  $y^* \in \mathcal{Y}_a^*$ , the dual function of  $P$  can be obtained by the super-Lagrangian dual transformation

$$P^d(y^*) = \sup_{x \in \mathbb{R}^n} L(x, y^*) = -f^\flat(-\Lambda^* y^*) - \langle b ; y^* \rangle \quad \forall y^* \in \mathcal{Y}_a^*, \quad (6.86)$$

where  $f^\flat \in \hat{\Gamma}(\mathbb{R}^n)$  is the Fenchel sub-conjugate of the concave function  $f \in \hat{\Gamma}(\mathbb{R}^n)$ . Thus, the dual problem associated with  $(\mathcal{P}_{\min})$  is a convex minimization problem

$$(\mathcal{P}_{\min}^d) : P^d(y^*) = -f^\flat(-\Lambda^* y^*) - \langle b ; y^* \rangle \rightarrow \min \quad \text{s.t. } y^* \leq 0 \in \mathbb{R}^m. \quad (6.87)$$

By the fact that the Fenchel conjugate of a nonsmooth function could be smooth, the solution of this convex dual problem is much easier than the primal one. Since  $L(x, y^*)$  is a super-Lagrangian on  $\mathbb{R}^n \times \mathbb{R}^m$ , the bi-duality theorem holds on  $\mathcal{X}_a \times \mathcal{Y}_a^*$ .

In particular, if the inequality constraint  $\Lambda x \geq b$  in  $(\mathcal{P}_{\min})$  is replaced by the equality constraint  $\Lambda x = b$ , then  $\mathcal{Y}_a^* = \mathbb{R}^m$ . In this case, the dual problem  $(\mathcal{P}_{\min}^d)$  of the constrained, nonconvex/nonsmooth primal problem  $(\mathcal{P}_{\min})$  in  $\mathbb{R}^n$  is a unconstrained, smooth convex minimization problem in  $\mathbb{R}^m$ ! Very often, we have  $n \geq m$ . This dimension reduction technique is extremely important in large scale nonlinear programming in finite element analysis (see Gao, 1988b). More interesting examples can be found in Gao (1999).  $\square$

**EXAMPLE 4.** As a special case, let us consider the constrained extremum problem of a given concave function in one-dimension:

$$(\mathcal{P}_{\text{ext}}) : f(x) = cx - \frac{1}{2}ax^2 \quad \rightarrow \text{ext} \quad \forall x \in \bar{I} = [x_a, x_b] \quad (6.88)$$

where  $a > 0$ ,  $c \in \mathbb{R}$  are given constraints. We assume that  $-\infty < x_a < 0 < x_b < \infty$ . Since  $f(x)$  is strictly concave on the open domain  $I = (x_a, x_b)$ , the minima are attained only on the boundary of  $I$ , i.e.

$$\inf_{x \in [x_a, x_b]} f(x) = \min\{f(x_a), f(x_b)\} > -\infty.$$

On the other hand, if the critical point  $\bar{x} = c/a$  of  $f(x)$  is in  $I = (x_a, x_b)$ , then the maximization problem  $(\mathcal{P}_{\max})$  is realizable and

$$\sup_{x \in \bar{I}} f(x) = \max_{x \in \bar{I}} f(x) = f\left(\frac{c}{a}\right).$$

There are many ways to set this problem within our framework, but each of them will lead to a different dual problem. Here we let  $\mathcal{X} = \mathbb{R}$ ,  $\mathcal{X}_a = [x_a, x_b]$  and  $\Lambda = 1$ , then

$$y = \Lambda x = x \in \mathcal{Y} = \mathbb{R}.$$

Thus, the range of the mapping  $\Lambda : \mathcal{X}_a \rightarrow \mathcal{Y} = \mathbb{R}$  is  $\mathcal{Y}_a = [x_a, x_b]$ . Let  $F(x) = -f(x)$  and

$$W(y) = \begin{cases} 0 & \text{if } y \in \mathcal{Y}_a, \\ +\infty & \text{if } y \notin \mathcal{Y}_a. \end{cases}$$

It is not difficult to check that  $W : \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$  is convex. On  $\mathcal{Y}_a$ ,  $W$  is finite and differentiable. Thus, the primal feasible set can be defined by

$$\mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda x = x \in \mathcal{Y}_a\} = [x_a, x_b].$$

The constrained primal problem ( $\mathcal{P}_{\text{ext}}$ ) is then equivalent to the unconstrained nonconvex extended global optimization problem

$$(\mathcal{P}_{\text{ext}}) : P(x) = W(\Lambda x) - F(x) \rightarrow \text{ext } \forall x \in \mathbb{R}. \quad (6.89)$$

Since  $F(x) = -f(x)$  is strictly convex and differentiable on  $\mathcal{X}_a = [x_a, x_b]$ , and

$$x^* = DF(x) = ax - c \in \mathcal{X}_a^* = [x_a a - c, x_b a - c] \subset \mathcal{X}^* = \mathbb{R}$$

is invertible, the Legendre conjugate  $F^* : \mathcal{X}_a^* \rightarrow \mathbb{R}$  can easily be obtained as

$$F^*(x^*) = \max_{x \in \mathcal{X}_a} \{xx^* - F(x)\} = \frac{1}{2a}(x^* + c)^2.$$

By the Legendre-Fenchel transformation, the conjugate of the nonsmooth function  $W$  can be obtained as

$$W^*(y^*) = \sup_{y \in \mathcal{Y}} \{yy^* - W(y)\} = \max_{y \in \mathcal{Y}_a} yy^* = \begin{cases} x_b y^* & \text{if } y^* > 0, \\ 0 & \text{if } y^* = 0, \\ x_a y^* & \text{if } y^* < 0. \end{cases}$$

It is convex and differentiable on  $\mathcal{Y}_a^* = \mathcal{Y}^* = \mathbb{R}$ .

On  $\mathcal{X}_a \times \mathcal{Y}_a^* = [x_a, x_b] \times \mathbb{R}$ , the extended Lagrangian associated with the problem ( $\mathcal{P}_{\text{ext}}$ ) is well-defined by

$$\begin{aligned} L(x, y^*) &= y^* \Lambda x - W^*(y^*) - F(x) \\ &= \begin{cases} xy^* - x_b y^* - \frac{1}{2}ax^2 + cx & \text{if } y^* \geq 0, \\ xy^* - x_a y^* - \frac{1}{2}ax^2 + cx & \text{if } y^* < 0. \end{cases} \end{aligned} \quad (6.90)$$

Since both  $W^*$  and  $P$  are convex,  $L(x, y^*)$  is a super-critical point function. If  $x \in \mathcal{X}_k = [x_a, x_b]$ , then

$$P(x) = \sup_{y^* \in \mathcal{Y}_a^*} L(x, y^*).$$

On the other hand, for any  $y^*$  in the dual feasible set

$$\mathcal{Y}_s^* = \{y^* \in \mathcal{Y}_a^* = \mathbb{R} \mid \Lambda^* y^* = y^* \in \mathcal{X}_a^*\} = [x_a a - c, x_b a - c],$$

the dual function is obtained by

$$\begin{aligned} P^d(y^*) &= \sup_{x \in \mathcal{X}_a} L(x, y^*) = \sup_{x \in \mathbb{R}} \{\Lambda x y^* - F(x)\} - W^*(y^*) \\ &= F^*(\Lambda^* y^*) - W^*(y^*), \end{aligned}$$

where

$$\begin{aligned} F^*(\Lambda^* y^*) &= \sup_{x \in \mathcal{X}_a} \{\Lambda x y^* - F(x)\} = \sup_{x \in \mathbb{R}} \{x(y^* + c) - \frac{1}{2} a x^2\} \\ &= \frac{1}{2a} (y^* + c)^2 = F^*(y^*). \end{aligned}$$

Thus, the dual action  $P^d$  is well defined on  $\mathcal{Y}_s^*$  by

$$P^d(y^*) = \begin{cases} \frac{1}{2a} (y^* + c)^2 - x_b y^* & \text{if } y^* > 0, \\ \frac{1}{2a} c^2 & \text{if } y^* = 0, \\ \frac{1}{2a} (y^* + c)^2 - x_a y^* & \text{if } y^* < 0. \end{cases} \quad (6.91)$$

This is a double-well function on  $\mathbb{R}$  (see Figure 5). The dual problem

$$(\mathcal{P}_{\text{ext}}^d): P^d(y^*) \rightarrow \text{ext} \quad \forall y^* \in \mathcal{Y}_s^*$$

is a convex optimization problem on either

$$\mathcal{Y}_s^{*+} = \{y^* \in \mathcal{Y}_s^* \mid y^* > 0\} \quad \text{or} \quad \mathcal{Y}_s^{*-} = \{y^* \in \mathcal{Y}_s^* \mid y^* < 0\}.$$

In  $n$ -dimensional problems, this dual problem is much easier than the primal problem. The criticality condition of  $(\mathcal{P}_{\text{ext}}^d)$  leads to

$$\bar{y}^* = \begin{cases} x_b a - c & \text{if } \bar{y}^* > 0, \\ x_a a - c & \text{if } \bar{y}^* < 0. \end{cases}$$

It is easy to check that the following duality theorems hold:

$$\max_{x \in \mathcal{X}_k} P(x) = \max_{y^* \in \mathcal{Y}_s^*} P^d(y^*),$$

$$\min_{x \in \mathcal{X}_k^\pm} P(x) = \min_{y^* \in \mathcal{Y}_s^{*\pm}} P^d(y^*),$$

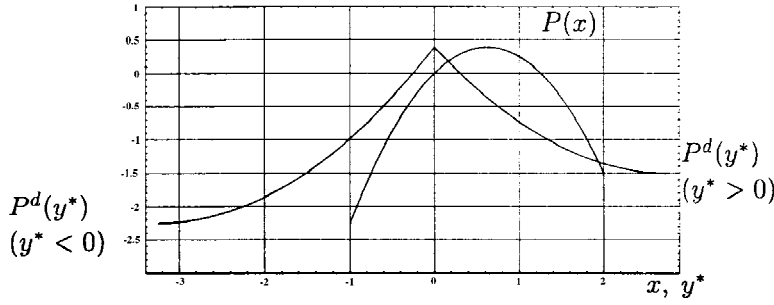


Figure 5. Bi-duality in constrained nonconvex optimization.

where  $\mathcal{X}_k^+ = \{x \in \mathbb{R} \mid 0 \leq x \leq x_b\}$ ,  $\mathcal{X}_k^- = \{x \in \mathbb{R} \mid x_a \leq x \leq 0\}$ . The graphs of  $P(x)$  and  $P^d(y^*)$  are shown in Fig. 5. If  $\bar{I} = [0, x_b]$ , then the primal minimization problem ( $\mathcal{P}_{\text{inf}}$ ) is equivalent to a nonconvex variational inequality problem (or unilateral variational problem). In multi-dimensional systems, traditional direct approaches are very difficult. However, the super-Lagrange dual problem ( $\mathcal{P}_{\text{inf}}^d$ ) is a strictly convex minimization problem on  $\mathcal{Y}_s^{*+}$ , which is substantially easier than the primal one.

EXAMPLE 5. We now illustrate the application of the interesting tri-duality theory for solving the nonconvex optimization problem (2.25)

$$P(x) = \frac{1}{2}a\left(\frac{1}{2}\|Ax\|^2 - \mu\right)^2 - x^T c \rightarrow \text{sta } \forall x \in \mathbb{R}^n. \tag{6.92}$$

The Euler equation associated with this nonconvex stationary problem is a nonlinear algebraic equation in  $\mathbb{R}^n$

$$a\left(\frac{1}{2}\|A\bar{x}\|^2 - \mu\right)C\bar{x} = c,$$

where  $C = A^T A = C^T \in \mathbb{R}^{nn}$ . We are interested in finding all the critical points of  $P$ . Let  $\mathcal{X} = \mathbb{R}^n = \mathcal{X}^*$ , and  $\Lambda : \mathbb{R}^n \rightarrow \mathcal{Y} = \mathbb{R}$  a quadratic operator

$$y = \Lambda(x) = \frac{1}{2}\|Ax\|^2 - \mu = \frac{1}{2}x^T Cx - \mu.$$

Since  $F(x) = \langle x, c \rangle = x^T c$  is a linear function on  $\mathbb{R}^n$ , the admissible space  $\mathcal{X}_a = \mathcal{X} = \mathbb{R}^n$ . By the fact that  $x^* = DF(x) = c$ , the range for the canonical mapping  $DF : \mathcal{X} \rightarrow \mathcal{X}^* = \mathbb{R}$  is a hyperplane in  $\mathbb{R}^n$ , i.e.

$$\mathcal{X}_a^* = \{x^* \in \mathbb{R}^n \mid x^* = c\}.$$

The feasible set for the primal problem is  $\mathcal{X}_k = \{x \in \mathcal{X}_a \mid \Lambda(x) \in \mathcal{Y}_a\} = \mathbb{R}^n$ .

By the fact that  $x^T Cx \geq 0 \forall x \in \mathcal{X}_a = \mathcal{X} = \mathbb{R}^n$ , the range for the geometrical mapping  $\Lambda : \mathcal{X}_a \rightarrow \mathbb{R}$  is a closed convex set in  $\mathbb{R}$

$$\mathcal{Y}_a = \{y \in \mathbb{R} \mid y \geq -\mu\} \subset \mathcal{Y} = \mathbb{R}.$$

On the admissible subset  $\mathcal{Y}_a \subset \mathcal{Y} = \mathbb{R}$ , the canonical function  $W(y) = \frac{1}{2}ay^2$  is quadratic. The range for the constitutive mapping  $DW : \mathcal{Y}_a \rightarrow \mathcal{Y}^* = \mathbb{R}$  is also a closed convex set in  $\mathbb{R}$

$$\mathcal{Y}_a^* = \{y^* \in \mathbb{R} \mid y^* \geq -a\mu\}.$$

On  $\mathcal{Y}_a^*$ , the Legendre conjugate of  $W$  is also strictly convex

$$W^*(y^*) = \frac{1}{2}a^{-1}y^{*2}, \quad (6.93)$$

and the Legendre duality relations hold on  $\mathcal{Y}_a \times \mathcal{Y}_a^*$ .

On  $\mathcal{X}_a \times \mathcal{Y}_a^* = \mathbb{R}^n \times \mathbb{R}$ , the extended Lagrangian in this case reads

$$L(x, y^*) = \frac{1}{2}y^*x^T Cx - \mu y^* - \frac{1}{2}a^{-1}y^{*2} - x^T c. \quad (6.94)$$

It is easy to check that the dual function associated with  $L$  is

$$P^d(y^*) = \frac{1}{2}(y^*)^{-1}c^T Cc - \mu y^* - \frac{1}{2a}y^{*2}.$$

The dual Euler-Lagrange equation is an algebraic equation in  $\mathbb{R}$ :

$$(\mu + a^{-1}y^*)y^{*2} = \frac{1}{2}\sigma^2, \quad \sigma^2 = c^T Cc. \quad (6.95)$$

Since  $C \in \mathbb{R}^{nn}$  is positive-definite, this equation holds only on  $\mathcal{Y}_a^*$ . For a given parameter  $\mu$  and  $c \in \mathbb{R}^n$ , this dual equation has at most three real roots  $y_k^* \in \mathcal{Y}_a^*$ ,  $k = 1, 2, 3$ , which leads to the primal solution

$$x_k = y_k^* C^+ c, \quad k = 1, 2, 3.$$

By Lemma 1 we know that each  $(x_k, y_k^*)$  is a critical point of  $L$  and

$$P(x_k) = L(x_k, y_k^*) = P^d(y_k^*), \quad k = 1, 2, 3.$$

In the case of  $n = 1$ , the graphs of  $P$  and  $P^d$  are shown in Figure 6. It was proved in Gao (1998b) that if  $\mu < \mu_c = 1.5(\sigma/a)^{2/3}$  the problem has only one global minimizer (see Figure 6(a)). However, if  $\mu > \mu_c$ , the dual Euler-Lagrange Equation (6.95) has three roots  $y_1^* > 0 > y_2^* > y_3^*$ , corresponding to three critical points of  $P^d$  (see Figure 6(b)). Then,  $y_1^*$  is a global maximizer of  $P^d$ ,  $x_1 = \sigma/y_1^*$  is a global minimizer of  $P$ ,  $P^d$  takes local minimum and local maximum values at  $y_2^*$  and  $y_3^*$ , respectively,  $x_2 = \sigma/y_2^*$  is a local maximizer of  $P$ , while  $x_3 = \sigma/y_3^*$  is a local minimizer.

The Lagrangian associated with this double-well energy is

$$L(x, y^*) = \frac{1}{2}x^2 y^* - \left(\frac{1}{2a}y^{*2} + \mu y^*\right) - y^* x.$$

It is a saddle function for  $y^* > 0$ . If  $y^* < 0$ , it is a super-critical point function (see Figure 7).

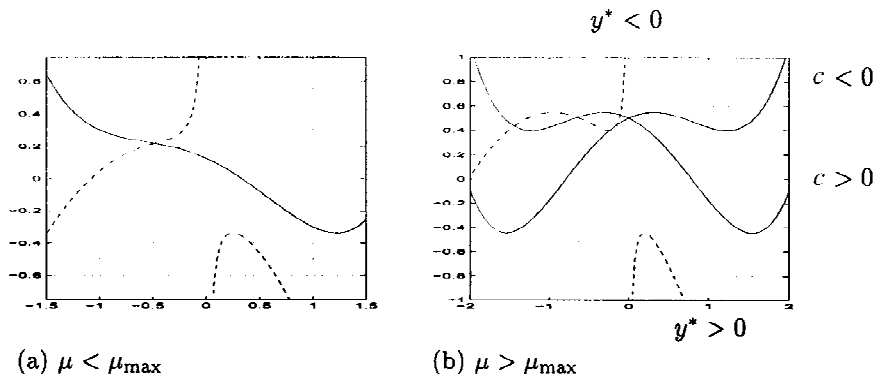


Figure 6. Double-well energy  $P(x)$  (solid lines) and its dual  $P^d(y^*)$  (dashed lines).

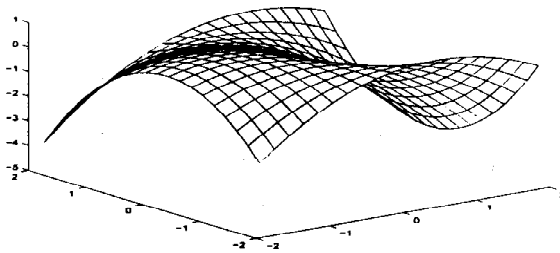


Figure 7. Lagrangian for the double-well energy in the Example 5.

## 7. Concluding Remarks

The concept of duality is one of the most successful ideas in modern optimization. The inner beauty of duality theory owes much to the fact that the nature was originally created in a splendid harmonious way. By the fact that the canonical physical variables appear always in pairs, the canonical dual transformation method can be used to solve many problems in natural systems. The associated extended Lagrange duality and triality theories have profound computational impacts. Compared with the traditional direct methods in global optimization problems, the main advantages of the canonical dual transformation method can be listed as the following.

1. Provides powerful and efficient primal-dual alternative approaches;
2. Converts nonsmooth constrained problems into smooth unconstrained dual problems;
3. Reduces the dimensions in nonlinear programming.

For any given nonlinear problem, as long as there exists a geometrical operator  $\Lambda$  such that the tri-canonical forms can be characterized correctly, the canonical dual transformation method and associated duality and triality principles can be used to establish nice theoretical results and to develop powerful alternative algorithms for robust computations. For a given nonlinear operator  $\Lambda(x)$  and associated canonical dual variable  $y^* \in \mathcal{Y}_a^*$ , the extended Lagrangian  $L(\cdot, y^*) : \mathcal{X}_a \rightarrow \mathbb{R}$  may not be a canonical function of  $x \in \mathcal{X}_a$ . In this case, the so-called *sequential*

*canonical dual transformation*, proposed by Gao (1999) in one-dimensional functional spaces, can be used to construct a *high order canonical Lagrangian*  $L_n$  for solving problems with multi-well cost functions.

## References

1. Auchmuty, G (1983), Duality for non-convex variational principles, *J. Diff. Equations*, 50: 80–145
2. Auchmuty, G (1989), Duality algorithms for nonconvex variational principles, *Numer. Funct. Anal. and Optim.*, 10: 211–264.
3. Auchmuty, G. (1997), Min-max problems for non-potential operator equations. in *Optimization Methods in Partial Differential Equations* (South Hadley, MA, 1996), 19–28, Contemp. Math., 209, Amer. Math. Soc., Providence, RI, 1997.
4. Benson, H. (1995), Concave minimization: theory, applications and algorithms, in R. Horst and P. Pardalos, (eds.) *Handbook of Global Optimization*, Kluwer Academic Publishers, pp. 43–148.
5. Casciaro, R. and Cascini, A. (1982), A mixed formulation and mixed finite elements for limit analysis, *Int. J. Solids and Struct.*, 19: 169–184.
6. Clarke, F.H. (1985), The dual action, optimal control, and generalized gradients, *Mathematical Control Theory*, Banach Center Publ., 14, PWN, Warsaw, pp. 109–119.
7. Crouzeix, J.P. (1981), Duality framework in quasiconvex programming, in S. Schaible and W.T. Ziemba, (eds.) *Generalized Convexity in Optimization and Economics*, Academic Press, pp. 207–226.
8. Dem'yanov, V.F., Stavroulakis, G.E., Polyakova, L.N. and Panagiotopoulos, P.D. (1996), *Quasidifferentiability and Nonsmooth Modelling in Mechanics, Engineering and Economics*. Kluwer Academic Publishers: Dordrecht.
9. Ekeland, I. (1977), Legendre duality in nonconvex optimization and calculus of variations, *SIAM J. Control and Optimiz.*, 15: 905–934.
10. Ekeland, I (1990), *Convexity Methods in Hamiltonian Mechanics*, Springer-Verlag, 247pp.
11. Ekeland, I and Temam, R (1976), *Convex Analysis and Variational Problems*, North-Holland.
12. Ericksen, J.L. (1975). Equilibrium of bars, *J. Elasticity* 5: 191–202.
13. Fukushima, M. and Qi, L.Q. (eds) (1999), *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*. Kluwer Academic Publishers, Dordrecht /Boston / London, 161–180.
14. Gao, D.Y. (1988a). On the complementary bounding theorems for limit analysis, *Int. J. Solids Struct.*, 24: 545–556.
15. Gao, D.Y. (1988b), Panpenalty finite element programming for plastic limit analysis, *Compute & Struct.*, 28: 749–755.
16. Gao, D.Y. (1996), Post-buckling analysis of nonlinear extended beam theory and dual variational principles, in L.A. Godoy, M. Rysz and L. Suarez, (eds.), *Applied Mechanics in Americas* Vol. 4, eds. Proc of PACAM V, San Juan, Puerto Rico, Iowa Univ. Press.
17. Gao, D.Y. (1997), Dual extremum principles in finite deformation theory with applications in post-buckling analysis of nonlinear beam model, *Appl. Mech. Reviews, ASME*, **50** (11): S64–S71.
18. Gao, D.Y. (1998a), Bi-complementarity and duality: A framework in nonlinear equilibria with application to contact problem, *J. Math. Anal. Appl.*, 221: 672–697.
19. Gao, D.Y. (1998b), Duality, triality and complementary extremum principles in nonconvex parametric variational problems with applications, *IMA J. Applied Math.*, **61**, 199–235.
20. Gao, D.Y. (1998c), Minimax and triality theory in nonsmooth variational problems, in M. Fukushima and L.Q. Qi (eds.), *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth*



- and *Smoothing Methods*, Kluwer Academic Publishers, Dordrecht / Boston / London, pp. 161–180.
21. Gao, D.Y. (1999a), *Duality Principles in Nonconvex Systems: Theory, Methods and Applications*, Kluwer Academic Publishers, Dordrecht / Boston / London, 472pp.
  22. Gao, D.Y. (1999b), Duality-Mathematics, in John G. Webster (ed), *Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons, Inc., 6, 68–77.
  23. Gao, D.Y., Ogden, R.W. and Stavroulakis, G.E. (eds.) (2000), *Nonsmooth and Nonconvex Mechanics*, Kluwer Academic Publishers, Dordrecht / Boston / London.
  24. Gao, D.Y. and Strang, G. (1989), Geometric nonlinearity: Potential energy, complementary energy, and the gap function, *Quart. Appl. Math.*, XLVII(3): 487–504.
  25. Gay, D.M., Overton, M.L. and Wright, M.H. (1998), A primal-dual interior method for nonconvex nonlinear programming. In *Advances in nonlinear programming* (Beijing, 1996), 31–56, Appl. Optim., 14, Kluwer Acad. Publ., Dordrecht, 1998.
  26. Hiriart-Urruty, J.-B. (1985), Generalized differentiability, duality and optimization for problems dealing with differences of convex functions, in F.H. Clarke V.F. Demyanov and F. Giannessi, (eds.), *Lecture Notes in Economics and Mathematical Systems*, Plenum, New York.
  27. Horst, R. and Pardalos, P.M. (1995), *Handbook in Global Optimization*, Kluwer Academic Publishers, Boston.
  28. Maier, G. (1969), Complementarity plastic work theorems in piecewise-linear elastoplasticity, *Int. J. Solids Struct.*, 5: 261–270.
  29. Maier, G., Carvelli, V. and Cocchetti, G. (2000), On direct methods for shakedown and limit analysis, Plenary lecture at the 4th EUROMECH Solid Mechanics Conf., Metz, France, June 26–30, to appear in *Eur. J. Mech., A/Solids*.
  30. Martínez-Legaz, J.-E. and Singer, I. (1995), Dualities associated to binary operations on  $\overline{R}$ . *J. Convex Anal.*, 2, (1-2): 185–209.
  31. Martínez-Legaz, J.-E.; and Singer, I. (1998). On  $\Phi$ -convexity of convex functions. *Linear Algebra Appl.*, 278 (1-3): 163–181.
  32. Mistakidis, E.S. and Stavroulakis, G.E. (1998), *Nonconvex Optimization in Mechanics, Algorithms, Heuristics and Engineering Applications by the F.E.M.*, Kluwer Academic Publishers, Dordrecht / Boston / London, 285pp.
  33. Moreau, J.J. (1968), La notion de sur-potentiel et les liaisons unilatérales en élastostatique, *C.R. Acad. Sc. Paris*, 267 A, 954–957.
  34. Moreau, J.J., Panagiotopoulos, P.D. and Strang, G. (1988), *Topics in nonsmooth mechanics*. Birkhuser Verlag, Basel-Boston, MA, 329pp.
  35. Motreanu, D. and Panagiotopoulos, P.D. (1999). *Minimax Theorems and Qualitative Properties of the Solutions of Hemivariational Inequalities*. Kluwer Academic Publishers, Dordrecht.
  36. Panagiotopoulos, P.D. (1985), *Inequality Problems in Mechanics and Applications*, Birkhäuser, Boston.
  37. Penot, J.P. and Volle, M. (1990), On quasiconvex duality, *Math. Oper. Res.*, 14: 195–227.
  38. Polyak, R.A. and Griva, I. (2000), Nonlinear rescaling in discrete minimax, in D.Y. Gao, R.W. Ogden and G. Stavroulakis (eds.), *Nonsmooth and Nonconvex Mechanics*, Kluwer Academic Publishers, Dordrecht / Boston / London.
  39. Rockafellar, R.T. (1974), *Conjugate Duality and Optimization*, SIAM, J.W. Arrowsmith Ltd., Bristol 3, England.
  40. Rockafellar, R.T. and Wets, R.J.B. (1997). *Variational analysis*, Springer: Berlin, New York.
  41. Sewell, M.J. (1987), *Maximum and Minimum Principles*, Cambridge Univ. Press, 468pp.
  42. Singer, I. (1986), A general theory dual optimization problems, *J. Math. Anal. Appl.*, 116: 77–130.
  43. Singer, I. (1992). Some further duality theorems for optimization problems with reverse convex constraint sets. *J. Math. Anal. Appl.*, 171(1): 205–219

44. Singer, I. (1996), On dualities between function spaces. *Math. Methods Oper. Res.*, 43(1): 35–44.
45. Singer, I. (1998), Duality for optimization and best approximation over finite intersections. *Numer. Funct. Anal. Optim.*, 19(7–8), 903–915.
46. Strang, G. (1986), *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, 758pp.
47. Thach, P.T. (1993), Global optimality criterion and a duality with a zero gap in nonconvex optimization. *SIAM J. Math. Anal.* 24(6): 1537–1556.
48. Thach, P.T. (1995), Diewert-Crouzeix conjugation for general quasiconvex duality and applications. *J. Optim. Theory Appl.*, 86(3): 719–743.
49. 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(3): 689–707.
50. Toland, J.F. (1978), Duality in nonconvex optimization, *J. Math. Anal. and Appl.*, 66: 399–415.
51. Toland, J.F. (1979), A duality principle for non-convex optimization and the calculus of variations, *Arch. Rat. Mech. Anal.* 71: 41–61
52. Tuy, H. (1991), Polyhedral annexation, dualization and dimension reduction technique in global optimization, *J. Global Optim.*, 1: 229–244.
53. Tuy, H. (1995), D.C. optimization: theory, methods and algorithms, in R. Horst and P. Pardalos (eds.), *Handbook of Global Optimization*, Kluwer Academic Publishers, 149–216.
54. Walk, M. (1989), *Theory of duality in mathematical programming*, Springer-Verlag, Wien / New York.
55. Wright, M.H. (1998), The interior-point revolution in constrained optimization, in R. DeLeone, A. Murli, P. M. Pardalos and G. Toraldo (eds.), *High-Performance Algorithms and Software in Nonlinear Optimization* 359–381, Kluwer Academic Publishers, Dordrecht.
56. Wright, S.J. (1997), *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, PA, 289pp.