Canonical Duality-Triality Theory: Unified Understanding for Modeling, Problems, and NP-Hardness in Global Optimization of Multi-Scale Systems

David Gao

1 Introduction and Motivation

General problems in mathematical optimization are usually formulated in the following form:

$$(\mathcal{P}_{ap}): \quad \min f(\mathbf{x}), \quad \text{s.t. } \mathbf{h}(\mathbf{x}) = 0, \quad \mathbf{g}(\mathbf{x}) \leq 0, \tag{1}$$

where the unknown $\mathbf{x} \in \mathbb{R}^n$ is a vector, $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is the so-called "objective" function,[1] and $\mathbf{h}(\mathbf{x}) = \{h_i(\mathbf{x})\} : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g}(\mathbf{x}) = \{g_j(\mathbf{x})\} : \mathbb{R}^n \to \mathbb{R}^p$ are two vector-valued constraint functions. It must be emphasized that, different from the basic concept of *objectivity* in continuum physics and nonlinear analysis, the objective function used extensively in optimization literature is allowed to be any arbitrarily given function, even the linear function. Therefore, the $(\mathcal{P}_{ap})$ is an *abstractly (or arbitrarily) proposed problem* (APP). Although it enables one to "model" a very wide range of problems, it comes at a price: many global optimization problems are considered to be NP-hard. Without detailed information on these arbitrarily given functions, it is impossible to have a powerful theory for solving the artificial nonconvex problem (1).

Canonical duality-triality is a newly developed and continuously improved methodological theory. This theory comprises mainly: 1) a canonical transformation,

---

[1]This terminology is used mainly in the English literature. The function $f(\mathbf{x})$ is correctly called the target function in all Chinese and Japanese literature.

D. Gao (✉)
School of Science and Technology, Federation University Australia, Mt Helen, VIC 3353, Australia
e-mail: d.gao@federation.edu.au

which is a versatile methodology that can be used to model complex systems within a unified framework, 2) a complementary-dual principle, which can be used to formulate a perfect dual problem with a unified analytic solution, and 3) a triality theory, which can identify both global and local extrema and to develop effective canonical dual algorithms for solving real-world problems in both continuous and discrete systems. This theory was developed from Gao and Strang's original work on nonconvex variational/boundary-value problems in large deformation mechanics [43]. It was shown in Gao's book [18] and in recent articles [40, 53] that the (external) penalty and Lagrange multiplier methods are special applications of the canonical duality theory in convex optimization. It is now understood that this theory reveals an intrinsic multi-scale duality pattern in complex systems, many popular theories and methods in nonconvex analysis, global optimization, and computational science can be unified within the framework of the canonical duality-triality theory. Indeed, it is easy to show that the KKT theory in mathematical programming, the *semi-definite programming* (SDP) method in global optimization, and the *half-quadratic regularization* in information technology are naturally covered by the canonical duality theory [39, 56, 86].

Mathematics and mechanics have been complementary partners since the Newton times. Many fundamental ideas, concepts, and mathematical methods extensively used in calculus of variations and optimization are originally developed from mechanics. It is known that the classical Lagrangian duality theory and the associated Lagrange multiplier method were developed by Lagrange in analytical mechanics [51]. The modern concepts of super-potential and sub-differential in convex analysis were proposed by J.J. Moreau from frictional mechanics [63]. The canonical duality theory is also developed from the fundamental concepts of objectivity and work-conjugate principle in continuum physics. The Gao-Strang gap function discovered in finite deformation theory provides a global optimality condition for general nonconvex/nonsmooth variational analysis and global optimization. Application of this theory to nonlinear elasticity leads to a pure complementary energy principle which was a 50-year-old open problem [58]. Generalization to global optimization was made in 2000 [20]. Since then, this theory has been used successfully for solving a large class of challenging problems in multi-disciplinary fields of applied mathematics, computational science, engineering mechanics, operations research, and industrial and systems engineering [11–17, 22–25, 36–38, 40, 42, 44, 45, 70, 73].

However, as V.I. Arnold indicated [2]: "In the middle of the twentieth century it was attempted to divide physics and mathematics. The consequences turned out to be catastrophic." Indeed, due to the ever-increasing gap between physics and other fields, some well-defined concepts in continuum physics, such as objectivity, Lagrangian, tensor, and fully nonlinearity, etc., have been seriously misused in optimization, which leads to not only ridiculous arguments but also wrong mathematical models and many artificially proposed problems. Also, the canonical dual transformation theory and methodology have been rediscovered in different forms by researchers from different fields. The main goal of this paper is to bridge this gap by presenting the canonical duality theory in a systematical way

from a unified modeling, basic assumptions to the theory, method, and general applications. The methodology, examples, and conjectures presented in this paper are important not only for better understanding this unconventional theory but also for solving many challenging problems in complex systems. This paper will bring some fundamentally new insights into multi-scale complex systems, global optimization, and computational science.

2 Multi-Scale Modeling and Properly Posed Problems

In linguistics, a complete and grammatically correct sentence should be composed by at least three words: subject, object, and a predicate.[2] As a language of science, the mathematics should follow this rule. Based on the canonical duality principle [18], a unified mathematical problem for multi-scale complex systems was proposed in [26, 28]:

$$(\mathcal{P}): \quad \min\{\Pi(\boldsymbol{\chi}) = G(\mathbf{D}\boldsymbol{\chi}) - F(\boldsymbol{\chi}) \mid \boldsymbol{\chi} \in \mathcal{X}_c\}, \tag{2}$$

where $F : \mathcal{X}_a \subset \mathcal{X} \to \mathbb{R}$ is a subjective function such that the external duality relation $\boldsymbol{\chi}^* = \nabla F(\boldsymbol{\chi}) = \bar{\boldsymbol{\chi}}^*$ is a given input (or source), its domain $\mathcal{X}_a$ contains only geometrical constraints (such as boundary or initial conditions), which depends on each given problem; $\mathbf{D} : \mathcal{X}_a \to \mathcal{G}_a$ is a linear operator which links the configuration variable $\boldsymbol{\chi} \in \mathcal{X}_a$ with an internal variable $\mathbf{g} = \mathbf{D}\boldsymbol{\chi} \in \mathcal{G}_a$ at different physical scales; $G : \mathcal{G}_a \subset \mathcal{G} \to \mathbb{R}$ is an objective function such that the internal duality relation $\mathbf{g}^* = \nabla G(\mathbf{g})$ is governed by the constitutive law, its domain $\mathcal{G}_a$ contains only physical constraints, which depends on mathematical modeling; The feasible set is defined by:

$$\mathcal{X}_c = \{\boldsymbol{\chi} \in \mathcal{X}_a \mid \mathbf{D}\boldsymbol{\chi} \in \mathcal{G}_a\}. \tag{3}$$

The predicate in $(\mathcal{P})$ is the operator "$-$" and the difference $\Pi(\boldsymbol{\chi})$ is called the target function in general problems. The object and subject are in balance only at the optimal states.

2.1 Objectivity, Isotropy, and Symmetry in Modeling

Objectivity is a central concept in our daily life, related to reality and truth. According to Wikipedia, the objectivity in philosophy means the state or quality of

---

[2]By the facts that (object, subject) is a duality pair in a noun (or pronoun) space, which is dual to a verb space, the multi-level duality pattern $\{(object, subject); predicate\}$ is called triality, which is essential for languages and sciences.

being true even outside a subject's individual biases, interpretations, feelings, and imaginings.[3] In science, the objectivity is often attributed to the property of scientific measurement, as the accuracy of a measurement can be tested independent from the individual scientist who first reports it.[4] In continuum mechanics, the objectivity is also called the principle of frame-indifference [65, 80], which is a basic concept in mathematical modeling [8, 18, 60] but is still subjected to seriously study in continuum physics [59, 64]. Let $\mathcal{R}$ be a special orthogonal group SO(n), i.e., $\mathbf{R} \in \mathcal{R}$ if and only if $\mathbf{R}^T = \mathbf{R}^{-1}$ and det $\mathbf{R} = 1$. The following mathematical definition was given in Gao's book (Definition 6.1.2 [18]).

Definition 1 (Objectivity and Isotropy) A set $\mathcal{G}_a$ is said to be objective if $\mathbf{Rg} \in \mathcal{G}_a$ $\forall \mathbf{g} \in \mathcal{G}_a$, $\forall \mathbf{R} \in \mathcal{R}$. A real-valued function $G : \mathcal{G}_a \to \mathbb{R}$ is said to be objective if

$$G(\mathbf{Rg}) = G(\mathbf{g}) \ \ \forall \mathbf{g} \in \mathcal{G}_a, \ \forall \mathbf{R} \in \mathcal{R}. \tag{4}$$

A set $\mathcal{G}_a$ is said to be isotropic if $\mathbf{gR} \in \mathcal{G}_a$ $\ \ \forall \mathbf{g} \in \mathcal{G}_a$, $\ \forall \mathbf{R} \in \mathcal{R}$. A real-valued function $G : \mathcal{G}_a \to \mathbb{R}$ is said to be isotropic if

$$G(\mathbf{gR}) = G(\mathbf{g}) \ \ \forall \mathbf{g} \in \mathcal{G}_a, \ \forall \mathbf{R} \in \mathcal{R}. \tag{5}$$

Lemma 1 *A real-valued function $G(\mathbf{g})$ is objective if and only if there exists a real-valued function $\Phi(\mathbf{C})$ such that $G(\mathbf{g}) = \Phi(\mathbf{g}^T \mathbf{g})$.*

Geometrically speaking, an objective function is rotational symmetry, which should be an SO(n)-invariant in n-dimensional Euclidean space. Physically, an objective function doesn't depend on observers. Because of Noether's theorem,[5] rotational symmetry of a physical system is equivalent to the *angular momentum conservation law* (see Section 6.1.2 [18]). Therefore, the objectivity is essential for any real-world mathematical models. In Euclidean space $\mathcal{G}_a \subset \mathbb{R}^n$, the simplest objective function is the ℓ_2-norm $\|\mathbf{g}\|$ in $\mathbb{R}^n$ as we have $\|\mathbf{Rg}\|^2 = \mathbf{g}^T \mathbf{R}^T \mathbf{Rg} = \|\mathbf{g}\|^2 \ \ \forall \mathbf{R} \in \mathcal{R}$. In continuum physics, the objectivity implies that the equilibrium condition of angular momentum (symmetry of the Cauchy stress tensor $\boldsymbol{\sigma} = \partial G(\mathbf{g})$, Section 6.1 [18]) holds. It was emphasized by P.G. Ciarlet that *the objectivity is not an assumption, but an axiom* [8]. In Gao and Strang's work, the internal energy $W(\mathbf{g})$ must be an objective function such that its variation (Gâteaux derivative) $\boldsymbol{\sigma} = \partial W(\mathbf{g})$ is the so-called *constitutive duality law*, which depends only on the intrinsic property of the system.

---

[3]https://en.wikipedia.org/wiki/Objectivity_(philosophy).

[4]https://en.wikipedia.org/wiki/Objectivity_(science).

[5]That is, every differentiable symmetry of the action of a physical system has a corresponding conservation law.

2.2   Subjectivity, Symmetry Breaking, and Well-Posed Problem

Dual to the objective function that depends on modeling, the subjective function $F(\chi)$ depends on each problem such that its variation is governed by the *action-reaction duality law*: $\bar{\chi}^* = \partial F(\chi) \in \mathcal{X}^*$. From the point view of systems theory, the action $\bar{\chi}^* \in \mathcal{X}^*$ can be considered as the input or source of the system, and the reaction $\chi \in \mathcal{X}$ should be the output (or the configuration, the state) of the system. A system is conservative if the action is independent of the reaction. Therefore, the subjective function must be linear on its domain $\mathcal{X}_a$ and, by Riesz representation theorem, we should have $F(\chi) = \langle \chi, \bar{\chi}^* \rangle$, where the bilinear form $\langle \chi, \chi^* \rangle : \mathcal{X} \times \mathcal{X}^* \to \mathbb{R}$ puts $\mathcal{X}$ and $\mathcal{X}^*$ in duality. The target function $\Pi(\chi) = G(\mathbf{D}\chi) - F(\chi)$ can have different physical meanings in real-world applications. For example, in continuum mechanics the subjective function $F(\chi)$ is the external energy, the objective function $G(\mathbf{g})$ is the stored energy, then $\Pi(\chi)$ is the total potential energy. In this case, the minimum total potential energy principle leads to the general variational problem (2). The criticality condition $\partial \Pi(\chi) = 0$ leads to the equilibrium (Euler-Lagrange) equation:

$$A(\chi) = \mathbf{D}^* \partial G(\mathbf{D}\chi) = \bar{\chi}^* \tag{6}$$

where $\mathbf{D}^* : \mathcal{G}_a^* \to \mathcal{X}^*$ is an adjoint operator of $\mathbf{D}$ and $A : \mathcal{X}_c \to \mathcal{X}^*$ is called *equilibrium operator*. The triality structure $\mathbb{S}^e = \{\langle \mathcal{X}, \mathcal{X}^* \rangle; A\}$ forms an elementary system in Gao's book (Section 4.3, [18]). This abstract form covers the most well-known equilibrium problems in real-world applications ranging from mathematical physics in continuous analysis to mathematical programming in discrete systems [18, 34]. Particularly, if $G(\mathbf{g})$ is quadratic such that $\partial^2 G(\mathbf{g}) = \mathbf{H}$, then the operator $A : \mathcal{X}_c \to \mathcal{X}^*$ is linear and can be written in the triality form: $A = \mathbf{D}^* \mathbf{H} \mathbf{D}$, which appears extensively in mathematical physics, optimization, and linear systems (see the celebrated text by Strang [77]). Clearly, any convex quadratic function $G(\mathbf{D}\chi)$ is objective due to the Cholesky decomposition $A = \Lambda^* \Lambda \succeq 0$.

According to the action-reaction duality in physics, if there is no action or demand (i.e., $\bar{\chi}^* = 0$), the system has no reaction (i.e., $\chi = 0$). Dually, a real-world problem should have at least one nontrivial solution for any given nontrivial input.

Definition 2 (Properly and Well-Posed Problems) A problem is called *properly posed* if for any given nontrivial input it has at least one nontrivial solution. It is called *well-posed* if the solution is unique.

Clearly, this definition is more general than Hadamard's well-posed problems in dynamical systems since the continuity condition is not required. Physically speaking, any real-world problems should be well-posed since all natural phenomena exist uniquely. But practically, it is difficult to model a real-world problem precisely. Therefore, properly posed problems are allowed for the canonical duality theory. This definition is important for understanding the triality theory and NP-hard problems.

2.3   Management Optimization

In management science, the decision variable χ is simply a vector $\mathbf{x} \in \mathbb{R}^n$, which could represent the products of a manufacture company. The input $\bar{\chi}^*$ can be considered as market price (or demanding), denoted by $\mathbf{f} \in \mathbb{R}^n$. Therefore, the subjective function $\langle \mathbf{x}, \mathbf{f} \rangle = \mathbf{x}^T \mathbf{f}$ in this example is the total income of the company. The products are produced by workers $\mathbf{g} \in \mathbb{R}^m$. Due to the cooperation, we have $\mathbf{g} = \mathbf{Dx}$ and $\mathbf{D} \in \mathbb{R}^{m \times n}$ is a matrix. Workers are paid by salary $\mathbf{g}^* = \partial G(\mathbf{g})$, and therefore, the objective function $G(\mathbf{g})$ is the cost (in this example G is not necessarily to be objective since the company is a man-made system). Then, the target $\Pi(\mathbf{x}) = G(\mathbf{Dx}) - \mathbf{x}^T \mathbf{f}$ is the total loss and the minimization problem $\min \Pi(\mathbf{x})$ leads to the equilibrium equation:

$$\mathbf{D}^T \partial_\mathbf{g} G(\mathbf{Dx}) = \mathbf{f}.$$

The cost function $G(\mathbf{g})$ could be convex for a small company, but usually nonconvex for big companies to allow some people having the same salaries.

If the company has to make a profit $\frac{1}{2}\alpha \|\mathbf{x}\|^2$, where $\alpha > 0$ is a parameter, then the target function is $\Pi(\mathbf{x}) = G(\mathbf{Dx}) + \frac{1}{2}\alpha \|\mathbf{x}\|^2 - \mathbf{x}^T \mathbf{f}$ and the minimization problem $\min \Pi(\mathbf{x})$ leads to:

$$\alpha \mathbf{x} = \mathbf{f} - \mathbf{D}^T \partial_\mathbf{g} G(\mathbf{Dx}). \tag{7}$$

This is a fixed point problem. In this case, if we let $\bar{\mathbf{g}} = \bar{\mathbf{D}}\mathbf{x} = (\mathbf{Dx}, \mathbf{x})$ and $\bar{G} = G(\mathbf{g}) + \frac{1}{2}\alpha \|\mathbf{x}\|^2$, then the fixed point problem (7) can be written in the unified form of:

$$\bar{\mathbf{D}}^T \partial_{\bar{\mathbf{g}}} \bar{G}(\bar{\mathbf{D}}\mathbf{x}) = \mathbf{f}.$$

This shows that the fixed point problem is a special case of the general equilibrium equation (6), a necessary condition of the general minimization problem $(\mathcal{P}_g)$.

2.4   Nonconvex Analysis and Boundary-Value Problems

For static systems, the unknown of a mixed boundary-value problem is a vector-valued function:

$$\chi(\mathbf{x}) \in \mathcal{X}_a = \{\chi \in \mathcal{C}[\Omega, \mathbb{R}^m] | \ \chi(\mathbf{x}) = 0 \ \forall \mathbf{x} \in \Gamma_\chi\},$$

$$\Omega \subset \mathbb{R}^d, \ d \leq 3, \ m \geq 1, \ \partial\Omega = \Gamma_\chi \cup \Gamma_t,$$

and the input is $\bar{\boldsymbol{\chi}}^* = \{\mathbf{f}(\mathbf{x}) \ \forall \mathbf{x} \in \Omega, \ \mathbf{t}(\mathbf{x}) \ \forall \mathbf{x} \in \Gamma_t\}$ [43]. In this case, the external energy is $F(\boldsymbol{\chi}) = \langle \boldsymbol{\chi}, \bar{\boldsymbol{\chi}}^* \rangle = \int_\Omega \boldsymbol{\chi} \cdot \mathbf{f} \, \mathrm{d}\Omega + \int_{\Gamma_t} \boldsymbol{\chi} \cdot \mathbf{t} \, \mathrm{d}\Gamma$. In nonlinear analysis, $\mathbf{D}$ is a gradient-like partial differential operator and $\mathbf{g} = \mathbf{D}\boldsymbol{\chi} \in \mathcal{G}_a \subset \mathcal{L}^p[\Omega; \mathbb{R}^{m \times d}]$ is a *two-point tensor field* [18] over Ω. The internal energy $G(\mathbf{g})$ is defined by:

$$G(\mathbf{g}) = \int_\Omega U(\mathbf{x}, \mathbf{g}) \, \mathrm{d}\Omega, \tag{8}$$

where $U(\mathbf{x}, \mathbf{g}) : \Omega \times \mathcal{G}_a \to \mathbb{R}$ is the *stored energy density*. The system is (space) *homogeneous* if $U = U(\mathbf{g})$. Thus, $G(\mathbf{g})$ is objective if and only if $U(\mathbf{x}, \mathbf{g})$ is objective on an objective set $\mathcal{G}_a$. By the facts that $\mathbf{g} = \mathbf{D}\mathbf{u}$ is a two-point tensor, which is not considered as a strain measure, but the (right) Cauchy-Green tensor $\mathbf{C} = \mathbf{g}^T \mathbf{g}$ is an objective strain tensor, there must exist a function $\Phi(\mathbf{C})$ such that $G(\mathbf{g}) = \Phi(\mathbf{C})$. In nonlinear elasticity, the function $\Phi(\mathbf{C})$ is usually convex and the duality $\mathbf{C}^* = \partial \Phi(\mathbf{C})$ is invertible (i.e., Hill's work-conjugate principle [18]). These basic truths in continuum physics laid a foundation for the canonical duality theory.

By finite element method, the domain Ω is divided into m-elements $\{\Omega^e\}$ such that the unknown function is piecewisely discretized by $\boldsymbol{\chi}(\mathbf{x}) \simeq \mathbf{N}_e(\mathbf{x})\boldsymbol{\chi}_e \ \forall \mathbf{x} \in \Omega^e$. Thus, the nonconvex variational problem (2) can be numerically reformulated in a global optimization problem:

$$\min\{\Pi(\boldsymbol{\chi}) = G(\mathbf{D}\boldsymbol{\chi}) - \langle \boldsymbol{\chi}, \mathbf{f} \rangle \ | \ \boldsymbol{\chi} \in \mathcal{X}_c\}, \tag{9}$$

where $\boldsymbol{\chi} = \{\boldsymbol{\chi}_e\}$ is the discretized unknown $\boldsymbol{\chi}(\mathbf{x})$, $\mathbf{D}$ is a generalized matrix depending on the interpolation $\mathbf{N}_e(\mathbf{x})$, and $\mathcal{X}_c$ is a convex constraint set including the boundary conditions. The canonical dual finite element method was first proposed in 1996 [11]. Applications have been given recently in engineering and sciences [30, 45, 73].

2.5 Lagrangian Mechanics and Initial-Value Problems

In Lagrange mechanics [51, 52], the unknown $\boldsymbol{\chi}(t) \in \mathcal{X}_a \subset \mathcal{C}^1[I; \mathbb{R}^n]$ is a vector field over a time domain $I \subset \mathbb{R}$. Its components $\{\chi_i(t)\}$ $(i = 1, \ldots, n)$ are known as the *Lagrangian coordinates*. Its dual variable $\bar{\boldsymbol{\chi}}^*$ is the action vector function in $\mathbb{R}^n$, denoted by $\mathbf{f}(t)$. The external energy $F(\boldsymbol{\chi}) = \langle \boldsymbol{\chi}, \bar{\boldsymbol{\chi}}^* \rangle = \int_I \boldsymbol{\chi}(t) \cdot \mathbf{f}(t) \, \mathrm{d}t$. While the internal energy $G(\mathbf{D}\boldsymbol{\chi})$ is the so-called action:

$$G(\mathbf{D}\boldsymbol{\chi}) = \int_I L(t, \boldsymbol{\chi}, \dot{\boldsymbol{\chi}}) \, \mathrm{d}t, \quad L = T(\dot{\boldsymbol{\chi}}) - U(t, \boldsymbol{\chi}), \tag{10}$$

where $\mathbf{D}\boldsymbol{\chi} = \{1, \partial_t\}\boldsymbol{\chi} = \{\boldsymbol{\chi}, \dot{\boldsymbol{\chi}}\}$ is a vector-valued mapping, T is the kinetic energy density, U is the potential density, and $L = T - U$ is the *Lagrangian density*. Together, $\Pi(\boldsymbol{\chi}) = G(\mathbf{D}\boldsymbol{\chi}) - F(\boldsymbol{\chi})$ is called the *total action*. This standard form

holds from the classical Newton mechanics to quantum field theory.[6] Its stationary condition leads to the well-known *Euler-Lagrange equation*:

$$A(\boldsymbol{\chi}) = \mathbf{D}^* \partial G(\mathbf{D}\boldsymbol{\chi}) = \{-\partial_t, 1\} \cdot \partial L(\boldsymbol{\chi}, \dot{\boldsymbol{\chi}}) = -\partial_t \partial_{\dot{\boldsymbol{\chi}}} T(\dot{\boldsymbol{\chi}}) - \partial_{\boldsymbol{\chi}} U(t, \boldsymbol{\chi}) = \mathbf{f}. \quad (11)$$

The system is called (time) *homogeneous* if $L = L(\boldsymbol{\chi}, \dot{\boldsymbol{\chi}})$. In general, the kinetic energy T must be an objective function of the velocity[7] $\mathbf{v}_k = \dot{\mathbf{x}}_k(\boldsymbol{\chi})$ of each particle $\mathbf{x}_k = \mathbf{x}_k(\boldsymbol{\chi}) \in \mathbb{R}^3 \ \forall k \in \{1, \dots, K\}$, while the potential density U depends on each problem. For Newtonian mechanics, we have $\boldsymbol{\chi}(t) = \mathbf{x}(t)$ and $T(\mathbf{v}) = \frac{1}{2}m\|\mathbf{v}\|^2$ is quadratic. If $U = 0$, the equilibrium equation $A(\boldsymbol{\chi}) = -m\ddot{\mathbf{x}}(t) = \mathbf{f}$ includes the Newton second law: $\mathbf{F} = m\ddot{\mathbf{x}}$ and the third law: $-\mathbf{F} = \mathbf{f}$. The first law $\mathbf{v} = \dot{\mathbf{x}} = \mathbf{v}_0$ holds only if $\mathbf{f} = 0$. In this case, the system has either a trivial solution $\mathbf{x} = 0$ or infinitely many solutions $\mathbf{x}(t) = \mathbf{v}_0 t + \mathbf{x}_0$, depending on the initial conditions in $\mathcal{X}_a$. This simple fact in elementary physics plays a key role in understanding the canonical duality theory and NP-hard problems in global optimization.

By using the methods of finite difference and least squares [39, 54], the general nonlinear dynamical system (11) can also be formulated as the same global optimization problem (9), where $\boldsymbol{\chi} = \{\chi_i(t_k)\}$ is the Lagrangian coordinates $\chi_i (i = 1, \dots, n)$ at each discretized time $t_k (k = 1, \dots, m)$, $\mathbf{D}$ is a finite difference matrix, and $\mathcal{X}_c$ is a convex constraint set including the initial condition [54]. By the canonical duality theory, an intrinsic relation between chaos in nonlinear dynamics and NP-hardness in global optimization was revealed recently in [54].

2.6 Mono-Duality and Duality Gap

Lagrangian duality was developed from Lagrange mechanics since 1788 [51], where the kinetic energy $T(\mathbf{v}) = \sum_k \frac{1}{2} m_k \|\mathbf{v}_k\|^2$ is a quadratic (objective) function. For convex static systems (or dynamical systems but $U(\boldsymbol{\chi}) = 0$), the stored energy $G : \mathcal{G}_a \to \mathbb{R}$ is convex and its Legendre conjugate $G^*(\boldsymbol{\sigma}) = \{\langle \mathbf{g}; \boldsymbol{\sigma} \rangle - G(\mathbf{g}) | \ \boldsymbol{\sigma} = \partial G(\mathbf{g})\}$ is uniquely defined on $\mathcal{G}_a^*$. Thus, by $G(\mathbf{D}\boldsymbol{\chi}) = \langle \mathbf{D}\boldsymbol{\chi}; \boldsymbol{\sigma} \rangle - G^*(\boldsymbol{\sigma})$ the total action or potential $\Pi(\boldsymbol{\chi})$ can be written in the Lagrangian form[8] $L : \mathcal{X}_a \times \mathcal{G}_a^* \to \mathbb{R}$:

$$L(\boldsymbol{\chi}, \boldsymbol{\sigma}) = \langle \mathbf{D}\boldsymbol{\chi}; \boldsymbol{\sigma} \rangle - G^*(\boldsymbol{\sigma}) - \langle \boldsymbol{\chi}, \mathbf{f} \rangle = \langle \boldsymbol{\chi}, \mathbf{D}^* \boldsymbol{\sigma} - \mathbf{f} \rangle - G^*(\boldsymbol{\sigma}), \quad (12)$$

where $\boldsymbol{\chi} \in \mathcal{X}_a$ can be viewed as a Lagrange multiplier for the equilibrium equation $\mathbf{D}^* \boldsymbol{\sigma} = \mathbf{f} \in \mathcal{X}_a^*$. In linear elasticity, $L(\boldsymbol{\chi}, \boldsymbol{\sigma})$ is the well-known *Hellinger-Reissner*

---

[6]See Wikipedia: https://en.wikipedia.org/wiki/Lagrangian_mechanics.

[7]The objectivity of $T(\mathbf{v})$ is also called the isotropy in Lagrange mechanics since $\mathbf{v}$ is a vector (see [52]).

[8]In the Physics literature, the same notation L is used for both action $L(\boldsymbol{\chi}, \dot{\boldsymbol{\chi}})$ and the Lagrangian $L(\boldsymbol{\chi}, \mathbf{p})$ since both represent the same physical quantity.

complementary energy [18]. Let $\mathcal{S}_c = \{\boldsymbol{\sigma} \in \mathcal{G}_a^* | \mathbf{D}^*\boldsymbol{\sigma} = \mathbf{f}\}$ be the so-called *statically admissible space*. Then, the Lagrangian dual of the general problem $(\mathcal{P})$ is given by:

$$(\mathcal{P}^*): \quad \max\{\Pi^*(\boldsymbol{\sigma}) = -G^*(\boldsymbol{\sigma})| \ \boldsymbol{\sigma} \in \mathcal{S}_c\}, \tag{13}$$

and the saddle Lagrangian leads to a well-known min-max duality in convex (static) systems:

$$\min_{\boldsymbol{\chi} \in \mathcal{X}_c} \Pi(\boldsymbol{\chi}) = \min_{\boldsymbol{\chi} \in \mathcal{X}_a} \max_{\boldsymbol{\sigma} \in \mathcal{G}_a^*} L(\boldsymbol{\chi}, \boldsymbol{\sigma}) = \max_{\boldsymbol{\sigma} \in \mathcal{G}_a^*} \min_{\boldsymbol{\chi} \in \mathcal{X}_a} L(\boldsymbol{\chi}, \boldsymbol{\sigma}) = \max_{\boldsymbol{\sigma} \in \mathcal{S}_c} \Pi^*(\boldsymbol{\sigma}). \tag{14}$$

This one-to-one duality is the so-called *mono-duality* in Chapter 1 [18], or the *complementary-dual variational principle* in continuum physics. In finite elasticity, the Lagrangian dual is also known as the *Levison-Zubov principle*. However, this principle holds only for convex problems. In real-world problems, the stored energy $G(\mathbf{g})$ is usually nonconvex in order to model complex phenomena. Its complementary energy can't be determined uniquely by the Legendre transformation. Although its Fenchel conjugate $G^{\sharp} : \mathcal{G}_a^* \to \mathbb{R} \cup \{+\infty\}$ can be uniquely defined, the Fenchel-Moreau dual problem:

$$(\mathcal{P}^{\sharp}): \quad \max\{\Pi^{\sharp}(\boldsymbol{\sigma}) = -G^{\sharp}(\boldsymbol{\sigma})| \ \boldsymbol{\sigma} \in \mathcal{S}_c\} \tag{15}$$

is not considered as a complementary-dual problem due to Fenchel-Young inequality:

$$g_a = \min\{\Pi(\boldsymbol{\chi})| \ \boldsymbol{\chi} \in \mathcal{X}_c\} \geq \max\{\Pi^{\sharp}(\boldsymbol{\sigma})| \ \boldsymbol{\sigma} \in \mathcal{S}_c\} = g_p, \tag{16}$$

and $g_{ap} = g_a - g_p \neq 0$ is the well-known *duality gap*. This duality gap is intrinsic to all Lagrange-Fenchel-Moreau types of duality problems since the linear operator $\mathbf{D}$ can't change the nonconvexity of $G(\mathbf{D}\boldsymbol{\chi})$. It turns out that the existence of a pure stress $\boldsymbol{\sigma}$ based complementary-dual principle was a well-known debate in nonlinear elasticity for more than fifty years [58].

Remark 1 (Equilibrium Constraints and Lagrange Multiplier Law) Strictly speaking, the Lagrange multiplier method can be used mainly for equilibrium constraint in $\mathcal{S}_c$ and the Lagrange multiplier must be the solution to the primal problem (see Section 1.5.2 [18]). The equilibrium equation $\mathbf{D}^*\boldsymbol{\sigma} = \mathbf{f}$ must be an invariant under certain coordinates transformation, say the law of angular momentum conservation, which is guaranteed by the objectivity of the stored energy $G(\mathbf{D}\boldsymbol{\chi})$ in continuum mechanics (see Definition 6.1.2, [18]), or by the isotropy of the kinetic energy $T(\dot{\boldsymbol{\chi}})$ in Lagrangian mechanics [52]. Specifically, the equilibrium equation for Newtonian mechanics is an invariant under the *Galilean transformation*; while for Einstein's special relativity theory, the equilibrium equation $\mathbf{D}^*\boldsymbol{\sigma} = \mathbf{f}$ is an invariant under the *Lorentz transformation*. For linear equilibrium equation, the quadratic $G(\mathbf{g})$ is naturally an objective function for convex systems. Unfortunately, since the concept of the objectivity is misused in optimization and the notation of the Euclidian

coordinate $\mathbf{x} = \{x_i\}$ is used as the unknown, the Lagrange multiplier method and the associated augmented methods have been mistakenly used for solving general nonconvex optimization problems, which produces many artificial duality gaps [53]. ♣

2.7 Bi-Duality and Conceptual Mistakes

For convex Hamiltonian systems, the action $G(\mathbf{D}\boldsymbol{\chi})$ in (10) is a d.c. (difference of convex) functional and the Lagrangian has its standard form in Lagrangian mechanics (see Section 2.5.2 [18] with $\mathbf{q}(t) = \boldsymbol{\chi}$ and $\mathbf{p} = \boldsymbol{\sigma}$):

$$L(\mathbf{q}, \mathbf{p}) = \langle \dot{\mathbf{q}}; \mathbf{p} \rangle - \int_I [T^*(\mathbf{p}) + U(\mathbf{q})] \, dt - \langle \mathbf{q}, \mathbf{f} \rangle, \tag{17}$$

where $\mathbf{q} \in \mathcal{X}_a \subset \mathcal{C}^1[I, \mathbb{R}^n]$ is the Lagrange coordinate and $\mathbf{p} \in \mathcal{S}_a \subset \mathcal{C}[I, \mathbb{R}^n]$ is the momentum. In this case, the Lagrangian is a bi-concave functional on $\mathcal{X}_a \times \mathcal{S}_a$, but the Hamiltonian:

$$H(\mathbf{q}, \mathbf{p}) = \langle \mathbf{Dq}; \mathbf{p} \rangle - L(\mathbf{q}, \mathbf{p}) = \int_I [T^*(\mathbf{p}) + U(\mathbf{q})] \, dt \tag{18}$$

is convex.[9] The total action and its canonical dual are [18]

$$\Pi(\mathbf{q}) = \max\{L(\mathbf{q}, \mathbf{p})| \ \mathbf{p} \in \mathcal{V}_a^*\} = \int_I [T(\dot{\mathbf{q}}) - U(\mathbf{q})] \, dt - \langle \mathbf{q}, \mathbf{f} \rangle \ \ \forall \mathbf{q} \in \mathcal{X}_c \tag{19}$$

$$\Pi^d(\mathbf{p}) = \max\{L(\mathbf{q}, \mathbf{p})| \ \mathbf{q} \in \mathcal{X}_a\} = \int_I [U^*(\dot{\mathbf{p}}) - T^*(\mathbf{p})] \, dt \ \ \forall \mathbf{p} \in \mathcal{S}_c \tag{20}$$

Clearly, both Π and Π^d are d.c. functionals. In this case, the so-called *bi-duality* was first presented in author's book Chapter 2 [18]:

Theorem 1 (Bi-Duality Theory) *For a given convex Hamiltonian system, if $(\bar{\mathbf{q}}, \bar{\mathbf{p}})$ is a critical point of $L(\mathbf{q}, \mathbf{p})$ over the time interval $I \subset \mathbb{R}$, then we have*

$$\Pi(\bar{\mathbf{q}}) = \min \Pi(\mathbf{q}) \quad \Leftrightarrow \quad \min \Pi^d(\mathbf{p}) = \Pi^d(\bar{\mathbf{p}}) \tag{21}$$

$$\Pi(\bar{\mathbf{q}}) = \max \Pi(\mathbf{q}) \quad \Leftrightarrow \quad \max \Pi^d(\mathbf{p}) = \Pi^d(\bar{\mathbf{p}}). \tag{22}$$

The mathematical proof of this theory was given in Section 2.6 [18] for convex Hamiltonian systems and in Corollary 5.3.6 [18] for d.c. programming problems.

---

[9]This is the reason that instead of the Lagrangian, the Hamiltonian is extensively used in dynamics.

This bi-duality revealed not only an interesting dynamical extremum principle in periodic motion but also an important truth in convex Hamiltonian systems.

Remark 2 (Least Action Principle and Conceptual Mistakes) The least action principle plays a central role in physics from the classical Newtonian mechanics, general relativity (Einstein-Hilbert action), to the modern string theory. Credit for the formulation of this principle is commonly given to Pierre Louis Maupertuis, who felt that "Nature is thrifty in all its actions." It was historically called "least" because its solution requires finding the path that has the least value [9], say Fermat's principle in optics. However, in Hamiltonian systems it should be accurately called the principle of stationary action since its solution does not minimize the total action. Actually, the validity of the least action principle has remained obscure in physics for several centuries. As a footnote in their celebrated book (Section 1.2, [52]), Landau and Lifshitz pointed out that the least action principle holds only for a sufficient small time interval, not for the whole trajectory of the system. Unfortunately, this is not true in general since the total action could be a concave functional within a sufficient small time interval.

Theorem 1 shows that a convex Hamiltonian system is controlled by the bi-duality, which revealed the following truths (see page 77 [18]):

The least action principle is not valid for any periodic motion.
It holds for the whole trajectory of the system if the potential $U(\mathbf{q}) = 0$.

The bi-duality theory has been challenged by M.D. Voisei, C. Zălinescu, and his former student R. Strugariu in a paper published in a dynamical systems journal [78]. Instead of finding any possible mistakes in author's work, they created an artificial "Lagrangian":

$$L(x, y) := -\frac{1}{2}\alpha\|x\|^2 - \frac{1}{2}\beta\|y\|^2 + \langle a, x\rangle\langle b, y\rangle, \quad \text{(Equation (1) in [78])}$$

and the associated "total action" $f(x)$ as well as its "dual action" $g(y)$:

$$f(x) = \max\{L(x, y)| \ y \in Y\} = -\frac{1}{2}\alpha\|x\|^2 + \frac{1}{2}\beta^{-1}\langle a, x\rangle^2\|b\|^2 \ \ \forall x \in X$$

$$g(y) = \max\{L(x, y)| \ x \in X\} = -\frac{1}{2}\beta\|y\|^2 + \frac{1}{2}\alpha^{-1}\langle b, y\rangle^2\|a\|^2 \ \ \forall y \in Y$$

By using these elementary functions in linear algebra, they produced a series of strange counterexamples to against the bi-duality theory in convex Hamiltonian systems presented by the author in Chapter 2 [18]. They claimed: "Because our counter-examples are very simple, using quadratic functions defined on whole Hilbert (even finite-dimensional) spaces, it is difficult to reinforce the hypotheses of the above mentioned results in order to keep the same conclusions and not obtain trivialities."

Clearly, the quadratic function $L(x, y)$ created by Zălinescu *et al.* is totally irrelevant to the Lagrangian $L(\mathbf{q}, \mathbf{p})$ in Lagrangian mechanics and in Gao's book

[18]. Without the differential operator $\mathbf{D} = \partial_t$, the quadric d.c. function $f(x)$ (or $g(y)$) is defined on one-scale space X (or Y) and is unbounded. Therefore, its critical point does not produce any motion. This basic mistake shows that these people don't have necessary knowledge not only in Lagrangian mechanics (the time derivative $\mathbf{D} = \partial_t$ is necessary for any dynamical systems) but also in d.c. programming (unconstrained quadratic d.c. programming does not make any sense). It also shows that these people even don't know what the Lagrangian coordinate is, otherwise, they would never use a time-independent vector $x \in \mathbb{R}^n$ as an unknown in dynamical systems.

Moreover, since there is neither input in $L(x, y)$ nor initial/boundary conditions in X, all the counterexamples produced by Zălinescu *et al.* are simply not problems but only artificial "models." Since they don't follow the basic rules in mathematical modeling, such as the objectivity, symmetry, conservation, and constitutive laws, etc., these artificial "models" are very strange and even ugly (see Examples 3.3, 4.2, 4.4 [78]). This type of mistakes shows that these people don't know the difference between the modeling and problems.   ♣

3   Unified Problem and Canonical Duality-Triality Theory

In this section, we simply restrict our discussion in finite-dimensional space $\mathcal{X}$. Its element $\boldsymbol{\chi} \in \mathcal{X}$ could be a vector, a matrix, or a tensor.[10] In this case, the linear operator $\mathbf{D}$ is a generalized matrix[11] $\mathbf{D} : \mathcal{X} \rightarrow \mathcal{G}$ and $\mathcal{G}$ is a generalized matrix space equipped with a natural norm $\|\mathbf{g}\|$. Let $\mathcal{X}_a \subset \mathcal{X}$ be a convex subset (with only linear constraints) and $\mathcal{X}_a^*$ be its dual set such that for any given input $\bar{\boldsymbol{\chi}}^* = \mathbf{f} \in \mathcal{X}_a^*$ the subjective function $\langle \boldsymbol{\chi}, \mathbf{f} \rangle \geq 0 \;\; \forall \boldsymbol{\chi} \in \mathcal{X}_a$. Although the objectivity is necessary for real-world modeling, the numerical discretization of $\mathbf{D}\boldsymbol{\chi}$ could lead to a complicated function $G(\mathbf{D}\boldsymbol{\chi})$ which may not be objective in $\mathbf{g} = \mathbf{D}\boldsymbol{\chi}$. Also in operations research, many challenging problems are artificially proposed. Thus, the objectivity required in Gao and Strang's original work on nonlinear elasticity has been relaxed by the canonical duality since 2000 [20].

---

[10]Tensor is a geometrical object in mathematics and physics, which is defined as a multi-dimensional array satisfying a transformation law, see https://en.wikipedia.org/wiki/Tensor. A tensor must be independent of a particular choice of coordinate system (frame-invariance). But, this terminology has been also misused recent years in the optimization literature such that any multi-dimensional array of data is called tensor [4]. This mistake has been recognized recently in the preface of [67].

[11]A generalized matrix $\mathbf{D} = \{D_{\alpha \cdots \gamma}^{i \cdots j}\}$ is a multi-dimensional array but not necessary to satisfy a transformation law, so it is not a tensor. In order to avoid confusion, it can be called a generalized matrix, or simply *gentrix*.

3.1 Canonical Transformation and Gap Function

Definition 3 (Canonical Function and Transformation)
A real-valued function $\Phi : \mathcal{E}_a \to \mathbb{R}$ is called a canonical function if its domain $\mathcal{E}_a$ is convex and the duality relation $\boldsymbol{\xi}^* = \partial \Phi(\boldsymbol{\xi}) : \mathcal{E}_a \to \mathcal{E}_a^*$ is bijective.

For a given real-valued function $G : \mathcal{G}_a \to \mathbb{R}$, if there exist a mapping $\boldsymbol{\xi} : \mathcal{G}_a \to \mathcal{E}_a$ and a canonical function $\Phi : \mathcal{E}_a \to \mathbb{R}$ such that

$$G(\mathbf{g}) = \Phi(\boldsymbol{\xi}(\mathbf{g})) \quad \forall \mathbf{g} \in \mathcal{G}_a, \tag{23}$$

the transformation (23) is called the canonical transformation.

The canonical function is not necessary to be convex. Actually, in many real-world applications, $\Phi(\boldsymbol{\xi})$ is usually concave. For example, in differential geometry and finite deformation theory (see [18]) the objective function $G(\mathbf{g})$ is the deformation Jacobian:

$$G(\mathbf{g}) = \sqrt{\det(\mathbf{g}^T \mathbf{g})}, \quad \mathbf{g} = \nabla \boldsymbol{\chi}. \tag{24}$$

If we chose $\boldsymbol{\xi} = \det(\mathbf{g}^T \mathbf{g})$ as the geometrical measure, which is the third invariant of the *Riemannian metric tensor* $\mathbf{g}^T \mathbf{g}$ and usually denoted as I_3, the canonical function $\Phi(I_3)$ is then a concave function of the scale measure $I_3(\mathbf{g})$.

The canonical duality is a fundamental principle in sciences and oriental philosophy, which underlies all natural phenomena. Therefore, instead of the objectivity in continuum physics, a generalized objective function $G(\mathbf{g})$ is used in the canonical duality theory under the following assumption.

Definition 4 (Generalized Objective Function) For a given real-valued function $G : \mathcal{G}_a \to \mathbb{R}$, if there exist a measure $\boldsymbol{\xi} : \mathcal{G}_a \to \mathcal{E}_a$ and a canonical function $\Phi : \mathcal{E}_a \to \mathbb{R}$ such that the following conditions hold:

(D1) Positivity: $G(\mathbf{g}) \geq 0 \ \ \forall \mathbf{g} \in \mathcal{G}_a$;
(D2) Canonicality: $G(\mathbf{g}) = \Phi(\boldsymbol{\xi}(\mathbf{g})) \ \ \forall \mathbf{g} \in \mathcal{G}_a$,

then $G : \mathcal{G}_a \to \mathbb{R}$ is called to be a generalized objective function.

The canonical transformation plays an important role in mathematical modeling and nonlinear analysis. Let $\Lambda = \boldsymbol{\xi} \circ \mathbf{D} : \mathcal{X}_a \to \mathcal{E}_a$ be the so-called *geometrically admissible operator* and $\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle : \mathcal{E} \times \mathcal{E}^* \to \mathbb{R}$ be the bilinear form which puts $\mathcal{E}$ and $\mathcal{E}^*$ in duality. By (D2), we have $\mathcal{X}_c = \{ \boldsymbol{\chi} \in \mathcal{X}_a | \ \Lambda(\boldsymbol{\chi}) \in \mathcal{E}_a \}$. The problem $(\mathcal{P})$ can be equivalently written in the following canonical form:

$$(\mathcal{P}_g) : \quad \min \{ \Pi(\boldsymbol{\chi}) = \Phi(\Lambda(\boldsymbol{\chi})) - \langle \boldsymbol{\chi}, \mathbf{f} \rangle | \ \boldsymbol{\chi} \in \mathcal{X}_c \}. \tag{25}$$

By the facts that the canonical duality is a universal principle in nature, the canonical measure $\Lambda(\boldsymbol{\chi})$ is not necessarily to be objective, and the spaces $\mathcal{X}, \mathcal{E}$ could be at

different physical scale with totally different dimensions, the canonical problem
$(\mathcal{P}_g)$ can be used to study general optimization problems in multi-scale complex
systems. The criticality condition of $(\mathcal{P}_g)$ is governed by the *fundamental principle
of virtual work*:

$$\langle \Lambda_t(\boldsymbol{\chi})\delta\boldsymbol{\chi}; \boldsymbol{\varsigma} \rangle = \langle \delta\boldsymbol{\chi}, \Lambda_t^*(\boldsymbol{\chi})\boldsymbol{\varsigma} \rangle = \langle \delta\boldsymbol{\chi}, \mathbf{f} \rangle \;\; \forall \delta\boldsymbol{\chi} \in \mathcal{X}_c, \tag{26}$$

where $\Lambda_t(\boldsymbol{\chi}) = \partial\Lambda(\boldsymbol{\chi})$ represents a generalized Gâteaux (or directional) derivative
of $\Lambda(\boldsymbol{\chi})$, its adjoint Λ_t^* is called the balance operator, $\boldsymbol{\varsigma} = \boldsymbol{\xi}^*(\boldsymbol{\xi}) = \partial\Phi(\boldsymbol{\xi})$ and
$\boldsymbol{\xi}^* : \mathcal{E}_a \rightarrow \mathcal{E}_a^*$ is a canonical dual (or constitutive) operator. The strong form of this
virtual work principle is called *the canonical equilibrium equation*:

$$\mathbf{A}(\boldsymbol{\chi}) = \Lambda_t^*(\boldsymbol{\chi})\boldsymbol{\xi}^*(\Lambda(\boldsymbol{\chi})) = \mathbf{f}. \tag{27}$$

A system governed by this equation is called a *canonical system* and is denoted as
(see Chapter 4, [18]):

$$\mathbb{S}_a = \{ \langle \mathcal{X}_a, \mathcal{X}_a^* \rangle, \langle \mathcal{E}_a; \mathcal{E}_a^* \rangle; (\Lambda, \boldsymbol{\xi}^*) \}.$$

Definition 5 (Classification of Nonlinearities) The system $\mathbb{S}_a$ is called *geometrically nonlinear* (resp., linear) if the geometrical operator $\Lambda : \mathcal{X}_a \rightarrow \mathcal{E}_a$ is nonlinear
(resp., linear); the system is called *physically (or constitutively) nonlinear* (resp.,
linear) if the canonical dual operator $\boldsymbol{\xi}^* : \mathcal{E}_a \rightarrow \mathcal{E}_a^*$ is nonlinear (resp., linear);
the system is called *fully nonlinear* (resp., linear) if it is both geometrically and
physically nonlinear (resp., linear).

Both geometrical and physics nonlinearities are basic concepts in nonlinear field
theory. The mathematical definition was first given by the author in 2000 under
the canonical transformation [20]. A diagrammatic representation of this canonical
system is shown in Figure 1.

This diagram shows a symmetry broken in the canonical equilibrium equation,
i.e., instead of Λ^*, the balance operator is Λ_t^*. It was discovered by Gao and Strang
[43] that by introducing a complementary operator $\Lambda_c(\boldsymbol{\chi}) = \Lambda(\boldsymbol{\chi}) - \Lambda_t(\boldsymbol{\chi})\boldsymbol{\chi}$, this
locally broken symmetry is recovered by a so-called complementary gap function:

Fig. 1 Diagrammatic representation for a canonical system

$$G_{ap}(\boldsymbol{\chi}, \boldsymbol{\varsigma}) = \langle -\Lambda_c(\boldsymbol{\chi}); \boldsymbol{\varsigma} \rangle, \tag{28}$$

which plays a key role in global optimization and the triality theory. Clearly, if $\Lambda = \mathbf{D}$ is linear, then $G_{ap} = 0$. Thus, the following statement is important to understand complexity:

Only the geometrical nonlinearity leads to nonconvexity in optimization, bifurcation in analysis, chaos in dynamics, and NP-hard problems in complex systems.

3.2 Complementary-Dual Principle and Analytical Solution

For a given canonical function $\Phi : \mathcal{E}_a \to \mathbb{R}$, its conjugate $\Phi^* : \mathcal{E}_a^* \to \mathbb{R}$ can be uniquely defined by the Legendre transformation:

$$\Phi^*(\boldsymbol{\varsigma}) = \text{sta}\{\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle - \Phi(\boldsymbol{\xi})| \ \boldsymbol{\xi} \in \mathcal{E}_a\}, \tag{29}$$

where $\text{sta}\{f(\boldsymbol{\chi})| \ \boldsymbol{\chi} \in \mathcal{X}\}$ stands for finding the stationary value of $f(\boldsymbol{\chi})$ on $\mathcal{X}$, and the following canonical duality relations hold on $\mathcal{E}_a \times \mathcal{E}_a^*$:

$$\boldsymbol{\varsigma} = \partial\Phi(\boldsymbol{\varepsilon}) \ \Leftrightarrow \ \boldsymbol{\varepsilon} = \partial\Phi^*(\boldsymbol{\varsigma}) \ \Leftrightarrow \ \Phi(\boldsymbol{\varepsilon}) + \Phi^*(\boldsymbol{\varsigma}) = \langle \boldsymbol{\varepsilon}; \boldsymbol{\varsigma} \rangle. \tag{30}$$

If the canonical function is convex and lower semicontinuous, the Gâteaux derivative ∂ should be replaced by the sub-differential and Φ^* is replaced by the Fenchel conjugate $\Phi^\sharp(\boldsymbol{\varsigma}) = \sup\{\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle - \Phi(\boldsymbol{\xi})| \ \boldsymbol{\xi} \in \mathcal{E}_a\}$. In this case, (30) is replaced by the generalized canonical duality:

$$\boldsymbol{\varsigma} \in \partial\Phi(\boldsymbol{\varepsilon}) \ \Leftrightarrow \ \boldsymbol{\varepsilon} \in \partial\Phi^\sharp(\boldsymbol{\varsigma}) \ \Leftrightarrow \ \Phi(\boldsymbol{\varepsilon}) + \Phi^\sharp(\boldsymbol{\varsigma}) = \langle \boldsymbol{\varepsilon}; \boldsymbol{\varsigma} \rangle \ \forall(\boldsymbol{\xi}, \boldsymbol{\varsigma}) \in \mathcal{E}_a \times \mathcal{E}_a^*. \tag{31}$$

If the convex set $\mathcal{E}_a$ contains inequality constraints, then (31) includes all the *internal KKT conditions* [14, 53]. In this sense, a KKT point of the canonical form $\Pi(\boldsymbol{\chi})$ is a generalized critical point of $\Pi(\boldsymbol{\chi})$.

By the complementarity $\Phi(\Lambda(\boldsymbol{\chi})) = \langle \Lambda(\boldsymbol{\chi}); \boldsymbol{\varsigma} \rangle - \Phi^*(\boldsymbol{\varsigma})$, the canonical form of $\Pi(\boldsymbol{\chi})$ can be equivalently written in Gao and Strang's *total complementary function* $\Xi : \mathcal{X}_a \times \mathcal{E}_a^* \to \mathbb{R}$ [43]:

$$\Xi(\boldsymbol{\chi}, \boldsymbol{\varsigma}) = \langle \Lambda(\boldsymbol{\chi}); \boldsymbol{\varsigma} \rangle - \Phi^*(\boldsymbol{\varsigma}) - \langle \boldsymbol{\chi}, \mathbf{f} \rangle. \tag{32}$$

Then, the canonical dual function $\Pi^g : \mathcal{S}_c \to \mathbb{R}$ can be obtained by the *canonical dual transformation*:

$$\Pi^g(\boldsymbol{\chi}) = \text{sta}\{\Xi(\boldsymbol{\chi}, \boldsymbol{\varsigma})| \ \boldsymbol{\chi} \in \mathcal{X}_a\} = G_{ap}^\Lambda(\boldsymbol{\varsigma}) - \Phi^*(\boldsymbol{\varsigma}), \tag{33}$$

where $G_{ap}^\Lambda(\varsigma) = \mathrm{sta}\{\langle \Lambda(\chi); \varsigma \rangle - \langle \chi, \mathbf{f} \rangle | \quad \chi \in \mathcal{X}_a\}$, which is defined on the canonical dual feasible space $\mathcal{S}_c = \{\varsigma \in \mathcal{E}_a^* | \quad \Lambda_t^*(\chi)\varsigma = \mathbf{f} \quad \forall \chi \in \mathcal{X}_a\}$. Clearly, $\mathcal{S}_c \neq \emptyset$ if $(\mathcal{P})$ is properly posed.

Theorem 2 (Complementary-Dual Principle [18]) *The pair $(\bar{\chi}, \bar{\varsigma})$ is a critical point of $\Xi(\chi, \varsigma)$ if and only if $\bar{\chi}$ is a critical point of $\Pi(\chi)$ and $\bar{\varsigma}$ is a critical point of $\Pi^g(\varsigma)$. Moreover:*

$$\Pi(\bar{\chi}) = \Xi(\bar{\chi}, \bar{\varsigma}) = \Pi^g(\bar{\varsigma}). \tag{34}$$

Proof The criticality condition $\partial \Xi(\bar{\chi}, \bar{\varsigma}) = 0$ leads to the following canonical equations:

$$\Lambda(\bar{\chi}) = \partial \Phi^*(\bar{\varsigma}), \quad \Lambda_t^*(\bar{\chi})\bar{\varsigma} = \mathbf{f}. \tag{35}$$

The theorem is proved by the canonical duality (30) and the definition of Π^g. $\square$

Theorem 2 shows a one-to-one correspondence of the critical points between the primal function and its canonical dual. In large deformation theory, this theorem solved the fifty-year-old open problem on complementary variational principle and is known as the Gao principle in literature [58]. In real-world applications, the geometrical operator Λ is usually *quadratic homogeneous*, i.e., $\Lambda(\alpha\chi) = \alpha^2 \Lambda(\chi) \ \forall \alpha \in \mathbb{R}$. In this case, we have [43] $\Lambda_t(\chi)\chi = 2\Lambda(\chi)$, $\Lambda_c(\chi) = -\Lambda(\chi)$, and

$$\Xi(\chi, \varsigma) = G_{ap}(\chi, \varsigma) - \Phi^*(\varsigma) - \langle \chi, \mathbf{f} \rangle = \frac{1}{2}\langle \chi, \mathbf{G}(\varsigma)\chi \rangle - \Phi^*(\varsigma) - \langle \chi, \mathbf{f} \rangle, \tag{36}$$

where $\mathbf{G}(\varsigma) = \partial_\chi^2 G_{ap}(\chi, \varsigma)$. Then, the canonical dual function $\Pi^g(\varsigma)$ can be written explicitly as:

$$\Pi^g(\varsigma) = \{\Xi(\chi, \varsigma)| \ \mathbf{G}(\varsigma)\chi = \mathbf{f} \ \forall \chi \in \mathcal{X}_a\} = -\frac{1}{2}\langle [\mathbf{G}(\varsigma)]^+\mathbf{f}, \mathbf{f} \rangle - \Phi^*(\varsigma), \tag{37}$$

where $\mathbf{G}^+$ represents a generalized inverse of $\mathbf{G}$.

Theorem 3 (Analytical Solution Form [18]) *If $\bar{\varsigma} \in \mathcal{S}_c$ is a critical point of $\Pi^g(\varsigma)$, then:*

$$\bar{\chi} = [\mathbf{G}(\bar{\varsigma})]^+\mathbf{f} \tag{38}$$

is a critical point of $\Pi(\chi)$ and $\Pi(\bar{\chi}) = \Xi(\bar{\chi}, \bar{\varsigma}) = \Pi^g(\bar{\varsigma})$. Dually, if $\bar{\chi} \in \mathcal{X}_c$ is a critical point of $\Pi(\chi)$, it must be in the form of (38) for a critical point $\bar{\varsigma} \in \mathcal{S}_c$ of $\Pi^g(\varsigma)$.

This unified analytical solution form holds not only for general global optimization problems in finite-dimensional systems [25] but also for a large class

of nonlinear boundary-/initial-value problems in nonconvex analysis and dynamic systems [21, 23, 54].

3.3 Triality Theory and NP-Hard Criterion

In order to study extremality property for the general problem $(\mathcal{P}_g)$, we need additional assumptions for the generalized objective function $G(\mathbf{D}\boldsymbol{\chi})$.

Assumption 1 (Canonically Convex Function) *Let $G(\mathbf{D}\boldsymbol{\chi})$ be a generalized objective function, i.e., there exist a measure $\Lambda : \mathcal{X}_a \to \mathcal{E}_a$ and a canonical function $\Phi : \mathcal{E}_a \to \mathbb{R}$ such that $G(\mathbf{D}\boldsymbol{\chi}) = \Phi(\Lambda(\boldsymbol{\chi}))$. We assume:*

(A1) Nonlinearity: $\Lambda(\boldsymbol{\chi}) : \mathcal{X}_a \to \mathcal{E}_a$ is a quadratic measure,
(A2) Regularity: $\Phi(\Lambda(\boldsymbol{\chi}))$ is twice continuously differentiable for all $\boldsymbol{\chi} \in \mathcal{X}_a$,
(A3) Convexity: $\Phi : \mathcal{E}_a \to \mathbb{R}$ is strictly convex.

One should emphasize that although $\Phi(\boldsymbol{\xi})$ is required to be strictly convex on $\mathcal{E}_a$ in this section, the composition $\Phi(\Lambda(\boldsymbol{\chi}))$ is usually nonconvex on $\mathcal{X}_a$ due to the geometrical nonlinearity. The canonical duality theory presented in this section can be generalized to the problems governed by high-order nonlinear $\Lambda\boldsymbol{\chi})$ and canonically concave function $\Phi(\boldsymbol{\xi})$ (see [18, 22, 46, 56]).

Definition 6 (Degenerate and Nondegenerate Critical Points, Morse Function) Let $\bar{\boldsymbol{\chi}} \in \mathcal{X}_c$ be a critical point of a real-valued function $\Pi : \mathcal{X}_c \to \mathbb{R}$. $\bar{\boldsymbol{\chi}}$ is called degenerate (resp. nondegenerate) if the Hessian matrix of $\Pi(\boldsymbol{\chi})$ is singular (resp., nonsingular) at $\bar{\boldsymbol{\chi}}$. The function $\Pi : \mathcal{X}_c \to \mathbb{R}$ is called a Morse function if it has no degenerate critical points.

Theorem 4 (Triality Theory [20]) *Suppose that $\Phi : \mathcal{E}_a \to \mathbb{R}$ is convex, $(\bar{\boldsymbol{\chi}}, \bar{\boldsymbol{\varsigma}})$ is a nondegenerate critical point of $\Xi(\boldsymbol{\chi}, \boldsymbol{\varsigma})$, and $\mathcal{X}_o \times \mathcal{S}_o$ is a neighborhood[12] of $(\bar{\boldsymbol{\chi}}, \bar{\boldsymbol{\varsigma}})$.*
   If $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_c^+ = \{\boldsymbol{\varsigma} \in \mathcal{S}_c | \ \mathbf{G}(\boldsymbol{\varsigma}) \succeq 0\}$, then

$$\Pi(\bar{\boldsymbol{\chi}}) = \min_{\boldsymbol{\chi} \in \mathcal{X}_c} \Pi(\boldsymbol{\chi}) = \max_{\boldsymbol{\varsigma} \in \mathcal{S}_c^+} \Pi^g(\boldsymbol{\varsigma}) = \Pi^g(\bar{\boldsymbol{\varsigma}}). \tag{39}$$

   If $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_c^- = \{\boldsymbol{\varsigma} \in \mathcal{S}_c | \ \mathbf{G}(\boldsymbol{\varsigma}) \prec 0\}$, then we have either

$$\Pi(\bar{\boldsymbol{\chi}}) = \max_{\boldsymbol{\chi} \in \mathcal{X}_o} \Pi(\boldsymbol{\chi}) = \max_{\boldsymbol{\varsigma} \in \mathcal{S}_o} \Pi^g(\boldsymbol{\varsigma}) = \Pi^g(\bar{\boldsymbol{\varsigma}}), \tag{40}$$

or (if $\dim \Pi = \dim \Pi^g$)

$$\Pi(\bar{\boldsymbol{\chi}}) = \min_{\boldsymbol{\chi} \in \mathcal{X}_o} \Pi(\boldsymbol{\chi}) = \min_{\boldsymbol{\varsigma} \in \mathcal{S}_o} \Pi^g(\boldsymbol{\varsigma}) = \Pi^g(\bar{\boldsymbol{\varsigma}}). \tag{41}$$

---

[12]The neighborhood $\mathcal{X}_o$ of $\bar{\boldsymbol{\chi}}$ means that on which, $\bar{\boldsymbol{\chi}}$ is the only critical point (see page 140 [18]).

The statement (39) is the so-called *canonical min-max duality*, which can be proved easily by Gao and Strang's work in 1989 [43]. Clearly, $\varsigma \in \mathcal{S}_c^+$ if and only if $G_{ap}(\chi, \varsigma) \geq 0 \quad \forall \chi \in \mathcal{X}$. This duality theory shows that the Gao-Strang gap function provides a global optimum criterion. The statements (40) and (41) are called the *canonical double-max* and *double-min dualities*, respectively, which can be used to find local extremum solutions. The triality theory shows that the nonconvex minimization problem $(\mathcal{P})$ is canonically dual to the following maximum stationary problem:

$$(\mathcal{P}^d): \quad \max \operatorname{sta}\{\Pi^g(\varsigma)| \ \varsigma \in \mathcal{S}_c^+\}. \tag{42}$$

Theorem 5 (Existence and Uniqueness Criteria [25]) *For a properly posed $(\mathcal{P})$, if the canonical function $\Phi : \mathcal{E}_a \to \mathbb{R}$ is convex, $\operatorname{int}\mathcal{S}_c^+ \neq \emptyset$, and*

$$\lim_{\alpha \to 0^+} \Pi^g(\varsigma_o + \alpha\varsigma) = -\infty \ \ \forall \varsigma_o \in \partial\mathcal{S}_c^+, \ \ \forall \varsigma \in \mathcal{S}_c^+, \tag{43}$$

then $(\mathcal{P}^d)$ has at least one solution $\bar{\varsigma} \in \mathcal{S}_c^+$ and $\bar{\chi} = [\mathbf{G}(\bar{\varsigma})]^+\mathbf{f}$ is a solution to $(\mathcal{P})$.' The solution is unique if $\mathbf{H} = \mathbf{G}(\bar{\varsigma}) \succ 0$.

Proof Under the required conditions, $-\Pi^g : \mathcal{S}_c^+ \to \mathbb{R}$ is convex and coercive and $\operatorname{int}\mathcal{S}_c^+ \neq \emptyset$. Therefore, $(\mathcal{P}^g)$ has at least one solution. If $\mathbf{H} \succ 0$, then $\Pi^g : \mathcal{S}_c^+ \to \mathbb{R}$ is strictly concave and $(\mathcal{P}^g)$ has a unique solution. $\qquad\square$

This theorem shows that if $\operatorname{int}\mathcal{S}_c^+ \neq \emptyset$ the nonconvex problem $(\mathcal{P}_g)$ is canonically dual to $(\mathcal{P}^g)$ which can be solved easily. Otherwise, the problem $(\mathcal{P}_g)$ is canonically dual to the following minimal stationary problem, i.e., to find a global minimum stationary value of Π^g on $\mathcal{S}_c$:

$$(\mathcal{P}^s): \quad \min \operatorname{sta}\{\Pi^g(\varsigma)| \ \varsigma \in \mathcal{S}_c\}, \tag{44}$$

which could be really NP-hard since $\Pi^g(\varsigma)$ is nonconvex on the nonconvex set $\mathcal{S}_c$. Therefore, a conjecture was proposed in [24].

Conjecture 1 (Criterion of NP-Hardness) *A properly posed problem $(\mathcal{P}_g)$ is NP-hard only if $\operatorname{int} \mathcal{S}_c^+ = \emptyset$.*

Remark 3 (History of Triality and Challenges) The triality theory was discovered by the author during his research on post-buckling of a large deformed elastic beam in 1996 [12], where the primal variable $\mathbf{u}(\mathbf{x})$ is a displacement vector in $\mathbb{R}^2$ and $\varsigma(\mathbf{x})$ is a canonical dual stress also in $\mathbb{R}^2$. Therefore, the triality theory was correctly proposed in nonconvex analysis, which provides for the first time a complete set of solutions to the post-buckling problem. Physically, the global minimizer $\bar{\mathbf{u}}(\mathbf{x})$ represents a stable buckled beam configuration (happened naturally), the local minimizer is an unstable buckled state (happened occasionally), while the local maximizer is the unbuckled beam state. Mathematical proof of the triality theory was given in [18] for one-D nonconvex variational problems (Theorem 2.6.2) and

for finite-dimensional optimization problems (Theorem 5.3.6 and Corollary 5.3.1). In 2002, the author discovered some counterexamples to the canonical double-min duality when $\dim \Pi \neq \dim \Pi^g$. Therefore, the triality theory was presented in an "either-or" form since the double-max duality is always true but the double-min duality was remarked by certain additional condition (see Remark 1 in [22] and Remark for Theorem 3 in [23]). Recently, the author and his co-workers proved that the canonical double-min duality holds weakly when $\dim \Pi \neq \dim \Pi^g$ [6, 61]. It was also discovered by using the canonical dual finite element method that the local minimum solutions in nonconvex mechanics are very sensitive not only to the input and boundary conditions of a given system but also to such artificial conditions as the numerical discretization and computational precision, etc. The triality theory provides a precise mathematical tool for studying and understanding complicated natural phenomena.

The triality theory has been repeatedly challenged by M.D. Voisei and C. Zălinescu in a set of at least 11 papers (see [29]). These papers fall into three groups. In the first group (say [78, 83]), they oppositely choose piecewise linear functions for G and quadratic functions for F as counterexamples to against the canonical duality theory with six conclusions on Gao and Strang's original work including [83]: "About the (complementary) gap function one can conclude that it is useless at least in the current context. The hope for reading an optimization theory with diverse applications is ruined . . ." Clearly, they made conceptual mistakes. In the second group, Voisei and Zălinescu chose an artificial problem with certain symmetry such that $\mathcal{S}_a^+ = \emptyset$. Such a problem can be solved easily by linear perturbation (see [62]). The counterexamples in the third group are simply those such that $\dim \Pi \neq \dim \Pi^g$. This type of counterexamples were first discovered by Gao in 2002 so it was emphasized in [22, 23] that the canonical double-min duality holds under certain additional constraints (see Remark 1 in [22] and Remark for Theorem 3 in [23]). But neither [23] nor [22] was cited by Zălinescu and his co-authors in their papers. Honest people can easily understand the motivation of these challenges.

The canonical duality-triality theory has been successfully used for solving a wide class problems in both global optimization and nonconvex analysis [39], including certain challenging problems in nonconvex analysis [19], nonlinear PDEs [33], large deformation mechanics [27], and NP-hard integer programming problems [24, 31]. ♣

4   Applications in Complex Systems

Applications to nonconvex constrained global optimization have been discussed in [40, 53]. This section presents applications to two general global optimization problems.

4.1 Unconstrained Nonconvex Optimization Problem

$$(\mathcal{P}_g): \quad \min\left\{\Pi(\chi) = \sum_{s=1}^{m}\Phi_s(\Lambda_s(\chi)) - \langle\chi, \mathbf{f}\rangle \mid \chi \in \mathcal{X}_c\right\}, \qquad (45)$$

where the canonical measures $\boldsymbol{\xi}_s = \Lambda_s(\chi)$ could be either a scalar or a generalized matrix, $\Phi_k(\boldsymbol{\xi}_k)$ are any given canonical functions, such as polynomial, exponential, logarithm, and their compositions, etc. For example, if $\chi \in \mathcal{X}_c \subset \mathbb{R}^n$ and

$$G(\mathbf{D}\chi) = \sum_{i\in\mathbb{I}}\frac{1}{2}\alpha_i\chi^T\mathbf{Q}_i\chi + \sum_{j\in\mathbb{J}}\frac{1}{2}\alpha_j\left(\frac{1}{2}\chi^T\mathbf{Q}_j\chi + \beta_j\right)^2$$

$$+ \sum_{k\in\mathbb{K}}\alpha_k\exp\left(\frac{1}{2}\chi^T\mathbf{Q}_k\chi\right) + \sum_{\ell\in\mathbb{L}}\frac{1}{2}\alpha_\ell\chi^T\mathbf{Q}_\ell\chi\log(\frac{1}{2}\chi^T\mathbf{Q}_\ell\chi), \quad (46)$$

where $\{\mathbf{Q}_s\}$ are positive-definite matrices to allow the Cholesky decomposition $\mathbf{Q}_s = \mathbf{D}_s^T\mathbf{D}_s$ for all $s \in \{\mathbb{I}, \mathbb{J}, \mathbb{K}, \mathbb{L}\}$ and $\{\alpha_s, \beta_s\}$ are physical constants, which could be either positive or negative under Assumption 1. This general function includes naturally the so-called d.c. functions (i.e., difference of convex functions). Let $p = \dim\mathbb{I}, \quad q = \dim\mathbb{J} + \dim\mathbb{K} + \dim\mathbb{L}$. By using the canonical measure:

$$\boldsymbol{\xi} = \{\xi_s\} = \left\{\frac{1}{2}\alpha_i\chi^T\mathbf{Q}_i\chi, \frac{1}{2}\chi^T\mathbf{Q}_r\chi\right\} \in \mathcal{E}_a = \mathbb{R}^p \times \mathbb{R}_+^q,$$

where $\mathbb{R}_+^q = \{\mathbf{x} \in \mathbb{R}^q \mid x_i \geq 0 \ \forall i = 1, \dots, q\}, G(\mathbf{g})$ can be written in the canonical form:

$$\Phi(\boldsymbol{\xi}) = \sum_{i\in\mathbb{I}}\xi_i + \sum_{j\in\mathbb{J}}\frac{1}{2}\alpha_j(\xi_j + \beta_j)^2 + \sum_{k\in\mathbb{K}}\alpha_k\exp\xi_k + \sum_{\ell\in\mathbb{L}}\alpha_\ell\xi_\ell\log\xi_\ell.$$

Thus, $\partial\Phi(\boldsymbol{\xi}) = \{1, \varsigma_r\}$ in which $\boldsymbol{\varsigma} = \{\alpha_j(\xi_j + \beta_j), \alpha_k\exp\xi_k, \ \alpha_\ell(\log\xi_\ell - 1)\} \in \mathcal{E}_a^*$, and

$$\mathcal{E}_a^* = \{\boldsymbol{\varsigma} \in \mathbb{R}^q \mid \varsigma_j \geq -\alpha_j\beta_j \ \forall j \in \mathbb{J}, \ \varsigma_k \geq \alpha_k \ \forall k \in \mathbb{K}, \ \varsigma_\ell \in \mathbb{R} \ \forall \ell \in \mathbb{L}\}.$$

The conjugate of Φ can be easily obtained as:

$$\Phi^*(\boldsymbol{\varsigma}) = \sum_{j\in\mathbb{J}}\left(\frac{1}{2\alpha_j}\varsigma_j^2 + \beta_j\varsigma_j\right) + \sum_{k\in\mathbb{K}}\varsigma_k(\ln(\alpha_k^{-1}\varsigma_k) - 1) + \sum_{\ell\in\mathbb{L}}\alpha_\ell\exp(\alpha_\ell^{-1}\varsigma_\ell - 1).$$

$$(47)$$

Since $\Lambda(\chi)$ is quadratic homogenous, the gap function and G_{ap}^Λ in this case are

$$G_{ap}(\chi, \varsigma) = \frac{1}{2}\chi^T \mathbf{G}(\varsigma)\chi, \quad G_{ap}^\Lambda(\varsigma) = \frac{1}{2}\mathbf{f}^T[\mathbf{G}(\varsigma)]^+\mathbf{f},$$

$$\mathbf{G}(\varsigma) = \sum_{i \in \mathbb{I}} \alpha_i \mathbf{Q}_i + \sum_{s \in \{\mathbb{J}, \mathbb{K}, \mathbb{L}\}} \varsigma_s \mathbf{Q}_s.$$

Since $\Pi^g(\varsigma) = -G_{ap}^\Lambda(\varsigma) - \Phi^*(\varsigma)$ is concave and $\mathcal{S}_c^+$ is a closed convex set, if for the given physical constants and the input $\mathbf{f}$ such that $\mathcal{S}_c^+ \neq \emptyset$, the canonical dual problem $(\mathcal{P}^g)$ has at least one solution $\bar{\varsigma} \in \mathcal{S}_c^+ \subset \mathbb{R}^q$ and $\bar{\chi} = [\mathbf{G}(\bar{\varsigma})]^+\mathbf{f} \in \mathcal{X}_c \subset \mathbb{R}^n$ is a global minimum solution to $(\mathcal{P})$. If $n \gg q$, the problem $(\mathcal{P}^g)$ can be much easier than $(\mathcal{P})$.

4.2 D.C. Programming

It is known that in Euclidean space every continuous global optimization problem on a compact set can be reformulated as a d.c. optimization problem, i.e., a nonconvex problem which can be described in terms of *d.c. functions* (difference of convex functions) and *d.c. sets* (difference of convex sets) [82]. By the fact that any constraint set can be equivalently relaxed by a nonsmooth indicator function, general nonconvex optimization problems can be written in the following standard d.c. programming form:

$$\min\{f(\mathbf{x}) = g(\mathbf{x}) - h(\mathbf{x}) \mid \forall \mathbf{x} \in \mathcal{X}\}, \tag{48}$$

where $\mathcal{X} = \mathbb{R}^n$, $g(\mathbf{x}), h(\mathbf{x})$ are convex proper lower semicontinuous functions on $\mathbb{R}^n$. A more general model is that $g(\mathbf{x})$ can be an arbitrary function [82]. Clearly, this d.c. programming problem is too abstract. Although it can be used to "model" a very wide range of mathematical problems [47], it is impossible to have an elegant theory and powerful algorithms for solving this problem without detailed structures on these arbitrarily given functions. As a result of extensive studying during the last thirty years (cf. [48, 79]), even some very simple d.c. programming problems are considered as NP-hard [82].

Based on the canonical duality theory, a generalized d.c. programming problem $(\mathcal{P}_{dc})$ can be presented in a canonical d.c. minimization problem form:

$$(\mathcal{P}_{dc}): \quad \min\left\{\Pi(\chi) = \Phi(\Lambda(\chi)) - \frac{1}{2}\langle\chi, \mathbf{H}\chi\rangle - \langle\chi, \bar{\chi}^*\rangle \mid \chi \in \mathcal{X}_c\right\}, \tag{49}$$

where $\mathbf{H}$ is a given positive-definite generalized matrix.

Since the canonical measure $\boldsymbol{\xi} = \Lambda(\boldsymbol{\chi}) \in \mathcal{E}_a$ is nonlinear and $\Phi(\boldsymbol{\xi})$ is convex on $\mathcal{E}_a$, the composition $\Phi(\Lambda(\boldsymbol{\chi}))$ has a higher-order nonlinearity than a quadratic function. Therefore, the coercivity for the target function $\Pi(\boldsymbol{\chi})$:

$$\lim_{\|\boldsymbol{\chi}\| \to \infty} \Pi(\boldsymbol{\chi}) = \infty \tag{50}$$

should be naturally satisfied for many real-world problems, which is a sufficient condition for existence of a global minimal solution to $(\mathcal{P}_{dc})$ (otherwise, the set $\mathcal{X}_c$ should be bounded). Clearly, this generalized d.c. minimization problem can be used to model a reasonably large class of real-world problems in multi-disciplinary fields [34, 49].

4.3 Fixed Point Problems

Fixed point problem is a well-established subject in the area of nonlinear analysis, which is usually formulated in the following form:

$$(\mathcal{P}_{fp}): \quad \mathbf{x} = F(\mathbf{x}), \tag{51}$$

where $F : \mathcal{X}_a \to \mathcal{X}_a$ is nonlinear mapping and $\mathcal{X}_a$ is a subset of a normed space $\mathcal{X}$. Problem $(\mathcal{P}_{fp})$ appears extensively in engineering and sciences, such as equilibrium problems, mathematical economics, game theory, and numerical methods for nonlinear dynamical systems. It is realized [72] that this well-studied field is actually a special subject of global optimization.

Lemma 2 *If F is a potential operator, i.e., there exists a real-valued function $P : \mathcal{X}_a \to \mathbb{R}$ such that $F(\mathbf{x}) = \nabla P(\mathbf{x})$, then $(\mathcal{P}_{fp})$ is equivalent to the following stationary point problem:*

$$\bar{\mathbf{x}} = \arg \operatorname{sta} \left\{ \Pi(\mathbf{x}) = P(\mathbf{x}) - \frac{1}{2} \|\mathbf{x}\|^2 \mid \forall \mathbf{x} \in \mathcal{X}_a \right\}. \tag{52}$$

Otherwise, $(\mathcal{P}_{fp})$ is equivalent to the following global minimization problem:

$$\bar{\mathbf{x}} = \arg \min \left\{ \Pi(\mathbf{x}) = \frac{1}{2} \|F(\mathbf{x}) - \mathbf{x}\|^2 \mid \forall \mathbf{x} \in \mathcal{X}_a \right\}. \tag{53}$$

Proof First, we assume that $F(\mathbf{x})$ is potential operator, then $\mathbf{x}$ is a stationary point of $\Pi(\mathbf{x})$ if and only if $\nabla \Pi(\mathbf{x}) = \nabla P(\mathbf{x}) - \mathbf{x} = 0$, thus $\mathbf{x}$ is also a solution to $(\mathcal{P}_{fp})$ since $F(\mathbf{x}) = \nabla P(\mathbf{x})$.

Now, we assume that $F(\mathbf{x})$ is not a potential operator. By the fact that $\Pi(\mathbf{x}) = \frac{1}{2} \|F(\mathbf{x}) - \mathbf{x}\|^2 \geq 0 \ \forall \mathbf{x} \in \mathcal{X}$, the vector $\bar{\mathbf{x}}$ is a global minimizer of $\Pi(\mathbf{x})$ if and only if $F(\bar{\mathbf{x}}) - \bar{\mathbf{x}} = 0$. Thus, $\bar{\mathbf{x}}$ must be a solution to $(\mathcal{P}_{fp})$. $\qquad\square$

By the facts that the global minimizer of an unconstrained optimization problem must be a stationary point, and

$$\frac{1}{2}\|F(\mathbf{x})-\mathbf{x}\|^2 = P(\mathbf{x})-\frac{1}{2}\|\mathbf{x}\|^2, \quad P(\mathbf{x}) = \frac{1}{2}\langle F(\mathbf{x}), F(\mathbf{x})\rangle-\langle \mathbf{x}, F(\mathbf{x})\rangle+\|\mathbf{x}\|^2, \quad (54)$$

the global minimization problem (53) is a special case of the stationary point problem (52). Mathematically speaking, if a fixed point problem has a trivial solution, then $F(\mathbf{x})$ must be a homogeneous operator, i.e., $F(0) = 0$. For general problems, $F(\mathbf{x})$ should have a nonhomogeneous term $\mathbf{f} \in \mathbb{R}^n$. Thus, we can let $P(\mathbf{x}) = G(\mathbf{D}\mathbf{x}) - \langle \mathbf{x}, \mathbf{f}\rangle$ such that $\mathbf{D} : \mathcal{X} \to \mathcal{G} \subset \mathbb{R}^m$ is a linear operator and $G : \mathcal{G} \to \mathbb{R}$ is a generalized objective function. Thus, the fixed point problem $(\mathcal{P}_{fp})$ can be reformulated in the following stationary point problem:

$$(\mathcal{P}_{fp}): \quad \bar{\mathbf{x}} = \arg \operatorname{sta}\left\{\Pi(\mathbf{x}) = G(\mathbf{D}\mathbf{x}) - \frac{1}{2}\|\mathbf{x}\|^2 - \langle \mathbf{x}, \mathbf{f}\rangle \mid \forall \mathbf{x} \in \mathcal{X}_c\right\}. \quad (55)$$

Clearly, the fixed point problem is actually equivalent to a d.c. programming problem. Canonical duality theory for solving this fixed point problem is given recently in [72].

4.4   Mixed Integer Nonlinear Programming (MINLP)

The decision variable for (MINLP) is $\chi = \{\mathbf{y}, \mathbf{z}\} \in \mathcal{Y}_a \times \mathcal{Z}_a$, where $\mathcal{Y}_a$ is a continuous variable set and $\mathcal{Z}_a$ is a set of integers. It was shown in [69] that for any given integer set $\mathcal{Z}_a$, there exists a linear transformation $\mathbf{D}_z : \mathcal{Z}_a \to \mathbb{Z} = \{\pm 1\}^n$. Thus, based on the unified problem $(\mathcal{P}_g)$, a general MINLP problem can be proposed as:

$$(\mathcal{P}_{mi}): \quad \min\{\Pi(\mathbf{y}, \mathbf{z}) = G(\mathbf{D}_y\mathbf{y}, \mathbf{D}_z\mathbf{z}) - \langle \mathbf{y}, \mathbf{s}\rangle - \langle \mathbf{z}, \mathbf{t}\rangle \mid (\mathbf{y}, \mathbf{z}) \in \mathcal{Y}_c \times \mathcal{Z}_c\}, \quad (56)$$

where $\mathbf{f} = (\mathbf{s}, \mathbf{t})$ is a given input, $\mathbf{D}\chi = (\mathbf{D}_y\mathbf{y}, \ \mathbf{D}_z\mathbf{z}) \in \mathcal{G}_y \times \mathbb{Z}$ is a multi-scale operator, and

$$\mathcal{Y}_c = \{\mathbf{y} \in \mathcal{Y}_a \mid \mathbf{D}_y\mathbf{y} \in \mathcal{G}_y\}, \quad \mathcal{Z}_c = \{\mathbf{z} \in \mathcal{Z}_a \mid \mathbf{D}_z\mathbf{z} \in \mathbb{Z}\}.$$

In $\mathcal{Y}_a$, certain linear constraints are given. Since the set $\mathbb{Z}_a$ is bounded, by Assumption 1 either $G : \mathcal{G}_y \to \mathbb{R}$ is coercive or $\mathcal{G}_y$ is bounded. This general problem $(\mathcal{P}_{mi})$ covers many real-world applications, including the so-called fixed cost problem [41]. Let

$$\mathbf{g} = \mathbf{\Lambda}_z(\mathbf{z}) = (\mathbf{D}_z\mathbf{z}) \circ (\mathbf{D}_z\mathbf{z}) \in \mathcal{E}_z = \mathbb{R}_+^n, \quad (57)$$

where $\mathbf{x} \circ \mathbf{y} = \{x_i y_i\}^n$ is the Hadamard product in $\mathbb{R}^n$, the integer constraint in $\mathbb{Z}$ can be relaxed by the canonical function $\Psi(\mathbf{g}) = \{0$ if $\mathbf{g} \leq \mathbf{e}, \infty$ otherwise$\}$, where $\mathbf{e} = \{1\}^n$. Therefore, the canonical form of $(\mathcal{P}_{mi})$ is

$$\min\{\Pi(\mathbf{y}, \mathbf{z}) = \Phi(\mathbf{\Lambda}(\mathbf{y}, \mathbf{z})) + \Psi(\mathbf{\Lambda}_z(\mathbf{z})) - \langle \mathbf{y}, \mathbf{s} \rangle - \langle \mathbf{z}, \mathbf{t} \rangle \mid \mathbf{y} \in \mathcal{Y}_c\}. \qquad (58)$$

Since the canonical function $\Psi(\mathbf{g})$ is convex, semicontinuous, its Fenchel conjugate is

$$\Psi^\sharp(\boldsymbol{\sigma}) = \sup\{\langle \mathbf{g}; \boldsymbol{\sigma} \rangle - \Psi(\mathbf{g}) | \mathbf{g} \in \mathbb{R}^n\} = \{\langle \mathbf{e}; \boldsymbol{\sigma} \rangle \text{ if } \boldsymbol{\sigma} \geq 0, \infty \text{ otherwise}\}.$$

The generalized canonical duality relations (31) are $\boldsymbol{\sigma} \geq 0 \Leftrightarrow \mathbf{g} \leq \mathbf{e} \Leftrightarrow \langle \mathbf{g} - \mathbf{e}; \boldsymbol{\sigma} \rangle = 0$. The complementarity shows that the canonical integer constraint $\mathbf{g} = \mathbf{e}$ can be relaxed by the $\boldsymbol{\sigma} > \mathbf{0}$ in continuous space. Thus, if $\boldsymbol{\xi} = \mathbf{\Lambda}(\boldsymbol{\chi})$ is a quadratic homogenous operator and the canonical function $\Phi(\boldsymbol{\xi})$ is convex on $\mathcal{E}_a$, the canonical dual to $(\mathcal{P}_{mi})$ is

$$(\mathcal{P}_{mi}^g): \ \max\left\{\Pi^g(\boldsymbol{\varsigma}, \boldsymbol{\sigma}) = -\frac{1}{2}\langle [\mathbf{G}(\boldsymbol{\varsigma}, \boldsymbol{\sigma})]^+\mathbf{f}, \mathbf{f} \rangle - \Phi^\sharp(\boldsymbol{\varsigma}) - \langle \mathbf{e}; \boldsymbol{\sigma} \rangle | \ (\boldsymbol{\varsigma}, \boldsymbol{\sigma}) \in \mathcal{S}_c^+\right\}, \qquad (59)$$

where $\mathbf{G}(\boldsymbol{\varsigma}, \boldsymbol{\sigma})$ depends on the quadratic operators $\mathbf{\Lambda}(\boldsymbol{\chi})$ and $\mathbf{\Lambda}_z(\mathbf{z})$, $\mathcal{S}_c^+$ is a convex open set:

$$\mathcal{S}_c^+ = \{(\boldsymbol{\varsigma}, \boldsymbol{\sigma}) \in \mathcal{E}_a^* \times \mathbb{R}_+^n | \ \mathbf{G}(\boldsymbol{\varsigma}, \boldsymbol{\sigma}) \succeq 0, \ \boldsymbol{\sigma} > 0\}. \qquad (60)$$

The canonical duality-triality theory has be used successfully for solving mixed integer programming problems [35, 41]. Particularly, for the quadratic integer programming problem:

$$(\mathcal{P}_{qi}): \ \min\left\{\Pi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{x}^T\mathbf{f} | \ \mathbf{x} \in \{-1, 1\}^n\right\}, \qquad (61)$$

we have $\mathcal{S}_c^+ = \{\boldsymbol{\sigma} \in \mathbb{R}_+^n | \ \mathbf{G}(\boldsymbol{\sigma}) = \mathbf{Q} + 2\text{Diag}(\boldsymbol{\sigma}) \succeq 0, \ \boldsymbol{\sigma} > 0\}$ and

$$(\mathcal{P}_{qi}^g): \ \max\left\{\Pi^g(\boldsymbol{\sigma}) = -\frac{1}{2}\mathbf{f}^T[\mathbf{G}(\boldsymbol{\sigma})]^+\mathbf{f} - \mathbf{e}^T\boldsymbol{\sigma} | \ \boldsymbol{\sigma} \in \mathcal{S}_c^+\right\} \qquad (62)$$

which can be solved easily if $\text{int}\mathcal{S}_c^+ \neq \emptyset$. Otherwise, $(\mathcal{P}_{qi})$ could be NP-hard since $\mathcal{S}_c^+$ is an open set, which is a conjecture proposed in [24]. In this case, $(\mathcal{P}_{qi})$ is canonically dual to an unconstrained nonsmooth/nonconvex minimization problem [25].

4.5 General Knapsack Problem and Analytical Solution

Knapsack problems appear in real-world decision-making processes in a wide variety of fields, such as finding the least wasteful way to cut raw materials, resource allocation where there are financial constraints, selection of investments and portfolios, selection of assets for asset-backed securitization, and generating keys for the Merkle-Hellman and other knapsack cryptosystems. Mathematically, a general quadratic knapsack problem can be formulated as an integer programming problem:

$$(\mathcal{P}_{qk}): \quad \min\left\{\Pi_{qk}(\mathbf{z}) = \frac{1}{2}\mathbf{z}^T\mathbf{Q}\mathbf{z} - \mathbf{c}^T\mathbf{z} |\ \mathbf{z} \in \{0, 1\}^n,\ \mathbf{v}^T\mathbf{z} \le V_c\right\}, \quad (63)$$

where $\mathbf{Q} \in \mathbb{R}^{n \times n}$ is a given symmetrical, usually indefinite, matrix, $\mathbf{c},\ \mathbf{v} \in \mathbb{R}^n$ are two given vectors, and $V_c > 0$ is a design parameter.

The knapsack problem has been studied for more than a century, with early works dating as far back as 1897. The main difficulty in this problem is the integer constraint $\mathbf{z} \in \{0, 1\}^n$, so that even the most simple linear knapsack problem:

$$(\mathcal{P}_{lk}): \quad \max\left\{\Pi_{lk}(\mathbf{z}) = -\mathbf{c}^T\mathbf{z} |\ \mathbf{z} \in \{0, 1\}^n,\ \mathbf{v}^T\mathbf{z} \le V_c\right\}, \quad (64)$$

is listed as one of Karp's 21 NP-complete problems [50].

By the fact that $\boldsymbol{\alpha} \circ \mathbf{z}^2 = \boldsymbol{\alpha} \circ \mathbf{z}\ \forall \mathbf{z} \in \{0, 1\}^n,\ \forall \boldsymbol{\alpha} \in \mathbb{R}^n$, for any given symmetrical $\mathbf{Q} \in \mathbb{R}^{n \times n}$ we can choose an $\boldsymbol{\alpha}$ such that $\mathbf{Q}_\alpha = \mathbf{Q} + 2\text{Diag}(\boldsymbol{\alpha}) \succeq 0$. Thus, by $\mathbf{c}_\alpha = \mathbf{c} + \boldsymbol{\alpha}$, the problem $(\mathcal{P}_q)$ can be equivalently written in the so-called α-perturbation form [25]:

$$(\mathcal{P}_\alpha): \quad \min\left\{\Pi_\alpha(\mathbf{z}) = \frac{1}{2}\mathbf{z}^T\mathbf{Q}_\alpha\mathbf{z} - \mathbf{c}_\alpha^T\mathbf{z}\ |\ \mathbf{v}^T\mathbf{z} \le V_c,\ \mathbf{z} \in \{0, 1\}^n\right\}. \quad (65)$$

Let rank $\mathbf{Q}_\alpha = r \le n$, there must exist (see [77]) an $\mathbf{L} \in \mathbb{R}^{r \times n}$ and $\mathbf{H} \in \mathbb{R}^{r \times r}$ with rank $\mathbf{L} = \text{rank } \mathbf{H} = r$ and $\mathbf{H} \succ 0$ such that $\mathbf{Q}_\alpha = 4\mathbf{L}^T\mathbf{H}\mathbf{L}$. Similar to the α-perturbed canonical dual problem $(\mathcal{P}_{ip}^g)$ given in [25], the canonical dual problem $(\mathcal{P}_q^g)$ can be reformulated as:

$$(\mathcal{P}_\alpha^g): \quad \max_{\boldsymbol{\zeta} \in \mathcal{S}_c^+}\left\{\Pi_\alpha^g(\boldsymbol{\sigma}, \tau) = -\frac{1}{2}\text{Abs}[\boldsymbol{\phi}(\boldsymbol{\sigma}, \tau)] - \frac{1}{2}\boldsymbol{\sigma}^T\mathbf{H}^{-1}\boldsymbol{\sigma} - \tau V_b + d\right\}, \quad (66)$$

where $V_b = V_c - \frac{1}{2}\sum_{i=1}^n v_i$, $d = \frac{1}{8}\sum_{i=1}^n(2\alpha_i + \sum_{j=1}^n Q_{ij}) - \frac{1}{2}\sum_{i=1}^n(c_i + \alpha_i)$,

$$\mathcal{S}_c^+ = \{\boldsymbol{\zeta} = (\boldsymbol{\sigma}, \tau) \in \mathbb{R}^{m+1}|\quad \tau \ge 0\}. \quad (67)$$

$$\boldsymbol{\phi}(\boldsymbol{\sigma}, \tau) = \mathbf{c} - \tau \mathbf{v} - 2\mathbf{L}^T \boldsymbol{\sigma} - \frac{1}{2}\mathbf{Qe}, \tag{68}$$

The notation $\text{Abs}[\boldsymbol{\phi}(\boldsymbol{\sigma}, \tau)]$ denotes $\text{Abs}[\boldsymbol{\phi}(\boldsymbol{\sigma}, \tau)] = \sum_{i=1}^{n} |\phi_i(\boldsymbol{\sigma}, \tau)|$.

Theorem 6 (Analytical Solution to Quadratic Knapsack Problem) *For any given* $V_c > 0$, $\mathbf{v}, \mathbf{c} \in \mathbb{R}^n_+$, $\boldsymbol{\alpha} \in \mathbb{R}^n_+$ *such that* $\mathbf{Q}_\alpha = \mathbf{Q} + 2Diag\,(\boldsymbol{\alpha}) = 4\mathbf{L}^T\mathbf{HL}$ *and* $\mathbf{H} \succ 0$, *if* $\bar{\boldsymbol{\zeta}} = \{\bar{\boldsymbol{\sigma}}, \bar{\tau}\}$ *is a solution to* $(\mathcal{P}^g_\alpha)$, *then*

$$\bar{\mathbf{z}} = \frac{1}{2}\left\{\frac{\phi_i(\bar{\boldsymbol{\sigma}}, \bar{\tau})}{|\phi_i(\bar{\boldsymbol{\sigma}}, \bar{\tau})|} + 1\right\}^n \tag{69}$$

is a global optimal solution to $(\mathcal{P}_\alpha)$ *and*

$$\Pi_\alpha(\bar{\mathbf{z}}) = \min_{\mathbf{z} \in \mathcal{Z}_a} \Pi_\alpha(\mathbf{z}) = \max_{\boldsymbol{\zeta} \in \mathcal{S}^+_c} \Pi^g_\alpha(\boldsymbol{\zeta}) = \Pi^g_\alpha(\bar{\boldsymbol{\zeta}}). \tag{70}$$

Theorem 7 (Existence and Uniqueness Theorem to Quadratic Knapsack Problem) *For any given* $V_c > 0$, $\mathbf{v}, \mathbf{c} \in \mathbb{R}^n_+$, $\boldsymbol{\alpha} \in \mathbb{R}^n_+$ *such that* $\mathbf{Q}_\alpha = \mathbf{Q} + 2Diag\,(\boldsymbol{\alpha}) = 4\mathbf{L}^T\mathbf{HL}$, $\mathbf{H} \succ 0$, *and* $\bar{\boldsymbol{\zeta}} = \{\bar{\boldsymbol{\sigma}}, \bar{\tau}\}$ *is a solution to* $(\mathcal{P}^g_\alpha)$, *if*

$$\phi_i(\bar{\boldsymbol{\sigma}}, \bar{\tau}) \neq 0 \quad \forall i = 1, \ldots, n \tag{71}$$

then the canonical dual feasible set $\mathcal{S}^+_c \neq \emptyset$ *and the knapsack problem* $(\mathcal{P}_\alpha)$ *has a unique solution. Otherwise, if* $\phi_i(\bar{\boldsymbol{\sigma}}, \bar{\tau}) = 0$ *for at least one* $i \in \{1, \ldots, n\}$, *then* $\mathcal{S}^+_c = \emptyset$ *and* $(\mathcal{P}_\alpha)$ *has at least two solutions.*

The canonical dual for the linear knapsack problem has a very simple form:

$$(\mathcal{P}^g_{lk}): \quad \max_{\tau \geq 0}\left\{\Pi^g_{lk}(\tau) = -\frac{1}{2}\sum_{i=1}^{n}(|c_i - \tau v_i| - \tau v_i) - \tau V_c\right\}. \tag{72}$$

Corollary 1 (Analytical Solution to Linear Knapsack Problem) *For any given* $V_c > 0$, $\mathbf{v}, \mathbf{c} \in \mathbb{R}^n_+$, *if* $\bar{\tau} > 0$ *is a solution to* $(\mathcal{P}^g_{lk})$, *then*

$$\bar{\mathbf{z}} = \frac{1}{2}\left\{\frac{c_i - \bar{\tau} v_i}{|c_i - \bar{\tau} v_i|} + 1\right\}^n \tag{73}$$

is a global optimal solution to $(\mathcal{P}_l)$ *and*

$$\Pi_{lk}(\bar{\mathbf{z}}) = \Pi^g_{lk}(\bar{\tau}) \tag{74}$$

Corollary 2 (Existence and Uniqueness Theorem to Linear Knapsack Problem) *For any given* $\mathbf{v}, \mathbf{c} \in \mathbb{R}^n_+$, *if there exists a constant* $\tau_c > 0$ *such that*

$$\psi_i(\tau_c) = \tau_c v_i - c_i \neq 0 \quad \forall i = 1, \ldots, n \tag{75}$$

then the knapsack problem ($\mathcal{P}_{lk}$) has a unique solution. Otherwise, if $\psi_i(\tau_c) = 0$ for at least one $i \in \{1, \ldots, n\}$, then ($\mathcal{P}_{lk}$) has at least two solutions.

Detailed proof of these results is given by Gao in [31].

The so-called multi-dimensional knapsack problem (MKP) is a generalization of the linear knapsack problem, that is:

$$(\mathcal{P}_{mk}): \quad \max \mathbf{c}^T \mathbf{z}, \quad s.t. \ \mathbf{Wz} \leq \boldsymbol{\omega}, \quad \mathbf{z} \in \{0, 1\}^n, \tag{76}$$

where $\mathbf{c} \in \mathbb{R}_+^n$ and $\boldsymbol{\omega} \in \mathbb{R}_+^m$ ($m < n$) are two given nonnegative vectors,

$$\mathbf{W} \in \mathbb{R}_+^{m \times n} = \{\mathbf{W} = \{w_{ij}\} \in \mathbb{R}^{m \times n} | \ w_{ij} \geq 0 \ \forall i = 1, \ldots, m, \ j = 1, \ldots, n\}$$

is a given nonnegative matrix such that $w_{ij} \leq \omega_j$, $\sum_{j=1}^n w_{ij} \geq \omega_i$. Clearly, this problem has multi-knapsacks $\{\omega_i\}^m$. Therefore, instead of the multi-dimensional, the correct name for ($\mathcal{P}_{mk}$) should be the *multi-kanpsacks problem*. This problem has applications in many fields including capital budgeting problems and resource allocation [66]. The canonical dual problem for ($\mathcal{P}_{mk}$) is

$$(\mathcal{P}_{mk}^g): \quad \max_{\boldsymbol{\tau} \in \mathbb{R}_+^m} \left\{ \Pi_{mk}^g(\boldsymbol{\tau}) = -\frac{1}{2} \sum_{i=1}^n (|c_i - \sum_{j=1}^m w_{ji}\tau_j| - \sum_{j=1}^m w_{ji}\tau_j) - \boldsymbol{\omega}^T \boldsymbol{\tau} \right\}. \tag{77}$$

Thus, if $\bar{\boldsymbol{\tau}} = \{\bar{\tau}_i\}$ is a global maximizer of ($\mathcal{P}_{mk}^g$), the analytic solution to ($\mathcal{P}_{mk}$) is

$$\mathbf{z} = \frac{1}{2} \left(\frac{c_i - \sum_{j=1}^m w_{ji}\bar{\tau}_j}{|c_i - \sum_{j=1}^m w_{ji}\bar{\tau}_j|} + 1 \right). \tag{78}$$

4.6 Bilevel Optimization and Optimal Control

Bilevel optimization appears extensively in optimal design and control of complex systems. A general formulation of the bilevel optimization problem can be written as follows:

$$(\mathcal{P}_{bo}): \quad \min \{T(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \in \mathcal{X}_a, \ \mathbf{y} \in \mathcal{Y}_a\}, \tag{79}$$

$$\text{s.t. } \mathbf{y} \in \arg \min \{\Pi(\mathbf{v}, \mathbf{x}) \mid \mathbf{v} \in \mathcal{Y}_a\}, \tag{80}$$

where T represents the top-level target (or leader) function, Π is the lower-level target (or follower) function. Similarly, $\mathbf{x} \in \mathcal{X}_a$ represents upper-level decision

vector and $\mathbf{y} \in \mathcal{Y}_a$ represents the lower-level variable. Clearly, this is a coupled nonlinear optimization problem, which is fundamentally difficult even for convex systems.

To solve this coupling problem numerically, an alternative iteration method can be used [30]:

(1) For a given $\mathbf{x}_{k-1}$, solve the lower-level problem first to obtain

$$\mathbf{y}_k \in \arg\min\{\Pi(\mathbf{y}, \mathbf{x}_{k-1})| \;\; \mathbf{y} \in \mathcal{Y}_a\}. \tag{81}$$

(2) Then for the fixed $\mathbf{y}_k$, solve the upper-level problem for

$$\mathbf{x}_k = \arg\min\{T(\mathbf{x}, \mathbf{y}_k) \;\; | \;\; \mathbf{x} \in \mathcal{X}_a\}. \tag{82}$$

These two single-level optimization problems can be solved by the canonical duality theory, and the sequence $\{\mathbf{x}_k, \mathbf{y}_k\}$ can converge to an optimal solution of $(\mathcal{P}_{bo})$ under certain conditions.

As an example, let us consider the following optimal control problem:

$$(\mathcal{P}_{oc}): \quad \min \{\Phi(\mathbf{v}, \boldsymbol{\chi}) \;\; | \;\; \mathbf{v} \in \mathcal{U}, \;\; \boldsymbol{\chi} \in \mathcal{X}\}, \tag{83}$$

$$\text{s.t. } \dot{\boldsymbol{\chi}} = \mathbf{a}(\boldsymbol{\chi}, \mathbf{v}, t) \;\; \forall t \in I = [t_o, t_b], \tag{84}$$

where $\boldsymbol{\chi}(t)$ is the state, $\mathbf{v}(t)$ is the control; $\mathcal{X}$ is a feasible set including the boundary conditions: $\boldsymbol{\chi}(t_o) = \boldsymbol{\chi}_o$ and $\psi(\boldsymbol{\chi}(t_b), t_b) = 0$. The upper-level target Φ is usually a quadratic continuous-time cost functional:

$$\Phi(\mathbf{v}, \boldsymbol{\chi}) = \frac{1}{2}\int_I \left[\boldsymbol{\chi}^T(t)\mathbf{Q}(t)\boldsymbol{\chi}(t) + \mathbf{v}^T(t)\mathbf{R}(t)\mathbf{v}(t) - 2\boldsymbol{\chi}^T(t)\mathbf{P}(t)\mathbf{v}(t)\right] dt + \Phi_b(\boldsymbol{\chi}(t_b)) \tag{85}$$

where $\mathbf{Q}(t) \in \mathbb{R}^{d \times d}$, $\mathbf{R}(t) \in \mathbb{R}^{p \times p}$ are positive semi-definite and positive definite, respectively, on the time domain $I = [t_o, t_b]$, $\Phi_b(\boldsymbol{\chi}(t_b)) = \frac{1}{2}\boldsymbol{\chi}^T(t_b)\mathbf{Q}_b\boldsymbol{\chi}(t_b)$. New to this cost function is the coupling term $\boldsymbol{\chi}^T(t)\mathbf{P}(t)\mathbf{v}(t)$, where $\mathbf{P}(t) \in \mathbb{R}^{d \times p}$ is a given matrix function of t, which plays an important role in alternative iteration methods for solving the general nonlinear optimal control problem $(\mathcal{P}_{oc})$.

For conservative systems, the nonlinear operator $\mathbf{a}(\boldsymbol{\chi}, \mathbf{v}, t)$ is a potential operator, i.e., there exists an action (or Lagrangian) $\Pi(\boldsymbol{\chi}, \dot{\boldsymbol{\chi}}, \mathbf{v})$ such that for any given control $\mathbf{v}(t) \in \mathcal{U}$ the differential equation (84) can be written in the following least action form:

$$\boldsymbol{\chi} \in \arg\min\{\Pi(\boldsymbol{\chi}, \dot{\boldsymbol{\chi}}, \mathbf{v}) \;|\; \boldsymbol{\chi} \in \mathcal{X}\} \tag{86}$$

Although such a Lagrangian does not exist for dissipative systems, the least squares method can always be used so that (84) can also be written in this minimization form.

In order to reformulate the challenging control problem ($\mathcal{P}_{oc}$) in function space, the finite element method can be used such that the time domain I is divided into n elements $\{I_e = [t_k, t_{k+1}]$ and in each I_e, the unknown fields can be numerically discretized as:

$$\boldsymbol{v}(t) = \mathbf{N}_u^e(t)\mathbf{u}_e, \quad \boldsymbol{\chi}(t) = \mathbf{N}_x^e(t)\mathbf{x}_e \quad \forall t \in I_e, \quad e = 1, \ldots, n, \tag{87}$$

where $\mathbf{N}_u^e(t)$ is an interpolation matrix for $\boldsymbol{v}(t)$, $\mathbf{u}_e = (\boldsymbol{v}(t_k), \boldsymbol{v}(t_{k+1}))$ is a nodal control vector; similarly, $\mathbf{N}_x^e(t)$ is an interpolation matrix for $\boldsymbol{\chi}(t)$ and $\mathbf{x}_e = (\boldsymbol{\chi}(t_k), \boldsymbol{\chi}(t_{k+1}))$ is a nodal state vector. Let $\mathcal{U}_a \subset \mathbb{R}^{p \times n}$ be an admissible nodal control space, $\mathcal{X}_a \subset \mathbb{R}^{d \times n}$ be an admissible state space, $\mathbf{u} = \{\boldsymbol{v}_k\} \in \mathcal{U}_a$, and $\mathbf{x} = \{\boldsymbol{\chi}_k\} \in \mathcal{X}_a$, then both the cost functional Φ and the action Π can be numerically written as:

$$\Phi(\boldsymbol{v}, \boldsymbol{\chi}) \approx \Phi_h(\mathbf{u}, \mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}_h\mathbf{x} + \frac{1}{2}\mathbf{u}^T\mathbf{R}_h\mathbf{u} - \mathbf{x}^T\mathbf{P}_h\mathbf{u}, \tag{88}$$

$$\Pi(\boldsymbol{\chi}, \dot{\boldsymbol{\chi}}, \boldsymbol{v}) \approx \Pi_h(\mathbf{x}, \mathbf{u}) = G(\mathbf{Dx}, \mathbf{u}) - F(\mathbf{x}, \mathbf{u}), \tag{89}$$

where $G(\mathbf{Dx}, \mathbf{u})$ and $F(\mathbf{x}, \mathbf{u})$ depend on the action $\Pi(\boldsymbol{\chi}, \dot{\boldsymbol{\chi}}, \mathbf{u})$,

$$\mathbf{Q}_h = \sum_{e=1}^{n} \int_{I_e} \mathbf{N}_x^T(t)\mathbf{Q}(t)\mathbf{N}_x(t) \, dt + \frac{1}{2}\mathbf{N}_x^T(t_b)\mathbf{Q}_b\mathbf{N}_x(t_b),$$

$$\mathbf{R}_h = \sum_{e=1}^{n} \int_{I_e} \mathbf{N}_u^T(t)\mathbf{R}(t)\mathbf{N}_u(t) \, dt,$$

$$\mathbf{P}_h = \sum_{e=1}^{n} \int_{I_e} \mathbf{N}_x^T(t)\mathbf{P}(t)\mathbf{N}_u(t) \, dt.$$

Therefore, the optimal control problem ($\mathcal{P}_{oc}$) can be written in a bilevel optimization problem:

$$(\mathcal{P}_{oc}^h): \quad \min \{\Phi_h(\mathbf{u}, \mathbf{x}) \mid \mathbf{u} \in \mathcal{U}_a, \quad \mathbf{x} \in \mathcal{X}_a\}, \tag{90}$$

$$\text{s.t. } \mathbf{x} \in \arg\min\{\Pi_h(\mathbf{y}, \mathbf{u}) \mid \mathbf{y} \in \mathcal{X}_a\} \tag{91}$$

The canonical duality theory has been successfully applied for solving nonlinear dynamical systems [54, 71] and the relation between chaos and NP-hardness was first discovered by Latorre and Gao [54]. Combined with an alternative iteration method, the canonical duality theory can be used to efficiently solve the general bilevel optimization problems.

4.7 Multi-Level Multi-Targets MINLP and Topology Optimization

Multi-target optimization is concerned with mathematical optimization problems involving more than one target function to be optimized simultaneously. Since the target is a vector-valued function, it is also known as vector optimization, multi-criteria optimization, or Pareto optimization. By the fact that the objectivity has been misused in optimization literature, this important research area has been misguidedly called multi-objective optimization. Multi-target optimization problems appear extensively in multi-scale complex systems where optimal decisions need to be taken in the presence of trade-offs between two or more conflicting targets. Therefore, the multi-level optimization and MINLP problems are naturally involved with the multi-target optimization. In real-world applications, the multi-level multi-target mixed integer nonlinear programming (MMM) could have many different formulations. Based on the canonical duality theory, a simple form of MMM problems can be proposed as the following:

$$(\mathcal{P}_{3m}): \quad \min \{T(\mathbf{z}, \bar{\mathbf{x}}, \bar{\mathbf{y}}) \mid \bar{\mathbf{x}} \in \mathcal{X}_a, \;\; \bar{\mathbf{y}} \in \mathcal{Y}_a, \;\; \mathbf{z} \in \mathcal{Z}_a\}, \tag{92}$$

$$\text{s.t. } \bar{\mathbf{x}} \in \arg\min\{\Pi_x(\mathbf{x}, \mathbf{y}, \mathbf{z}) \mid \mathbf{x} \in \mathcal{X}_a\}, \tag{93}$$

$$\bar{\mathbf{y}} \in \arg\min\{\Pi_y(\mathbf{x}, \mathbf{y}, \mathbf{z}) \mid \mathbf{y} \in \mathcal{Y}_a\}. \tag{94}$$

Without loss of generality, we assume that the leader variable $\mathbf{z} \in \mathcal{Z}_a$ is a discrete vector, the follower variables $\mathbf{x} \in \mathcal{X}_a$ and $\mathbf{y} \in \mathcal{Y}_a$ are continuous vectors; the top-level (leader) target $T : \mathcal{Z}_a \times \mathcal{X}_a \times \mathcal{Y}_a \to \mathbb{R}^m$ is a vector-valued function, which is not necessary to be objective, while the lower-level (follower) targets Π_x and Π_y are real-valued functions such that the follower problems can be written respectively in the canonical form $(\mathcal{P})$, where the objectivity and subjectivity are required. If we let

$$\mathbf{u} = \{\mathbf{x}, \mathbf{y}, \cdots\} \in \mathcal{U}_a = \mathcal{X}_a \times \mathcal{Y}_a \times \cdots,$$

$$\Pi(\mathbf{u}, \mathbf{z}) = \{\Pi_x(\mathbf{u}, \mathbf{z}), \Pi_y(\mathbf{u}, \mathbf{z}), \cdots\} : \mathcal{U}_a \times \mathcal{Z}_a \to \mathbb{R}^d, \;\; d \geq 2,$$

then the MMM problem can be written in a general form:

$$(\mathcal{P}_{3m}): \quad \min \{T(\mathbf{z}, \mathbf{u}) \mid \mathbf{u} \in \mathcal{U}_a, \;\; \mathbf{z} \in \mathcal{Z}_a\}, \tag{95}$$

$$\text{s.t. } \mathbf{u} \in \arg\min\{\Pi(\mathbf{v}, \mathbf{z}) \mid \mathbf{v} \in \mathcal{U}_a\}. \tag{96}$$

Clearly, the $(\mathcal{P}_{3m})$ should be one of the most challenging problems proposed so far in global optimization even if both T and Π are linear vector-valued functions.

Topology optimization is a mathematical tool that optimizes the best mass density distribution $\rho(\mathbf{x})$ within a design domain $\Omega \subset \mathbb{R}^d$ in order to obtain the best structural performance governed by the minimum total potential principle:

$$\min \left\{ \Pi(\mathbf{u}, \rho) = \int_\Omega U(\nabla \mathbf{u}) \rho \, d\Omega - \int_{\Gamma_t} \mathbf{u}^T \mathbf{t} \, d\Gamma \ \mid \ \mathbf{u} \in \mathcal{U} \right\} \tag{97}$$

where $\mathbf{u} : \Omega \to \mathbb{R}^d$ is a displacement vector field, the design variable $\rho(\mathbf{x}) \in \{0, 1\}$ is a discrete scalar field, which takes $\rho(\mathbf{x}) = 1$ at a solid material point $\mathbf{x} \in \Omega$ and $\rho(\mathbf{x}) = 0$ at a void point $\mathbf{x} \in \Omega$. By using finite element method, the design domain Ω is meshed with n disjointed finite elements $\{\Omega_e\}$ and let

$$\mathbf{u}(\mathbf{x}) = \mathbf{N}(\mathbf{x})\mathbf{u}_e \ \ \rho(\mathbf{x}) = z_e \in \{0, 1\} \ \ \forall \mathbf{x} \in \Omega_e,$$

the total potential energy can be numerically written as:

$$\Pi(\mathbf{u}, \rho) \approx \Pi_h(\mathbf{u}, \mathbf{z}) = \mathbf{z}^T \mathbf{c}(\mathbf{u}) - \mathbf{u}^T \mathbf{f}$$

where $\mathbf{u} = \{\mathbf{u}_e\} \in \mathcal{U}_a \subset \mathbb{R}^m$ is a nodal displacement vector, $\mathbf{z} \in \mathcal{Z}_a \subset \{0, 1\}^n$ is a discretized design vector, and

$$\mathbf{c}(\mathbf{u}) = \left\{ \int_{\Omega_e} U(\nabla \mathbf{N}(\mathbf{x})\mathbf{u}_e) \, d\Omega \right\} \in \mathbb{R}^n_+, \ \ \mathbf{f} = \left\{ \int_{\Gamma_{te}} \mathbf{N}(\mathbf{x})^T \mathbf{t}(\mathbf{x}) \, d\Gamma \right\} \in \mathbb{R}^m.$$

Let

$$\mathcal{Z}_a = \{\mathbf{z} \in \{0, 1\}^n \mid \ \mathbf{v}^T \mathbf{z} \leq \omega\},$$

where $\mathbf{v} = \{v_e\} \in \mathbb{R}^n_+$ and $v_e \geq 0$ is the volume of the e-th element, and $\omega > 0$ is the desired volume of structure. The correct mathematical problem for general topology optimization has been proposed recently by Gao [30, 31]:

Problem 1 (Topology Optimization for General Materials) *For a given external load* $\mathbf{f}$ *and the desired volume* $\omega > 0$, *to solve the bilevel MINLP problem:*

$$(\mathcal{P}_{to}): \quad \min \{T(\mathbf{z}, \mathbf{u}) \ \mid \ \mathbf{u} \in \mathcal{U}_a, \ \mathbf{z} \in \mathcal{Z}_a\}, \tag{98}$$

$$s.t. \ \mathbf{u} \in \arg\min\{\Pi_h(\mathbf{v}, \mathbf{z}) \ \mid \ \mathbf{v} \in \mathcal{U}_a\}. \tag{99}$$

The top-level target function T depends on each design problem, which could be

$$T_1(\mathbf{z}, \mathbf{u}) = \mathbf{u}^T \mathbf{f} - \mathbf{z}^T \mathbf{c}(\mathbf{u}), \tag{100}$$

$$T_2(\mathbf{z}, \mathbf{u}) = \frac{1}{2}\mathbf{z}^T \mathbf{Q}(\mathbf{u})\mathbf{z} - \mathbf{z}^T \mathbf{c}(\mathbf{u}). \tag{101}$$

where $\mathbf{Q}(\mathbf{u})$ is a symmetrical matrix whose diagonal elements $\{Q_{ii}\}_{i=1}^n = 0$ and $Q_{ij}(\mathbf{u})$ is the negative effect to the structure if the i-th and j-th elements are elected. Clearly, the top-level is a linear knapsack problem for $T = T_1$, or a quadratic

knapsack problem if $T = T_2$. If $\mathcal{Z}_a = \{\mathbf{z} \in \{0, 1\}^n | \quad \mathbf{W}\mathbf{z} \leq \boldsymbol{\omega}\}$, then the top level is a multi-knapsack problem.

In topology limit design, the top-level target could be a vector-valued function depending on certain design parameter α, say:

$$T_3(\alpha, \mathbf{z}, \mathbf{u}) = \{\alpha, T_i(\mathbf{z}, \mathbf{u})\}, \quad i = 1, 2. \tag{102}$$

If α is the volume ω, then $(\mathcal{P}_{to})$ is a topology optimization for lightweight design problem:

Problem 2 (Topology Lightweight Design) *For the given external load and $\omega^b > \omega^a > 0$, to solve*

$$(\mathcal{P}_{lw}): \min \{T_3(\omega, \mathbf{z}, \mathbf{u}) \mid \mathbf{v}^T \mathbf{z} \leq \omega, \ \omega \in [\omega^a, \omega^b], \ \mathbf{u} \in \mathcal{U}_a, \ \mathbf{z} \in \{0, 1\}^n\}, \tag{103}$$

$$\text{s.t. } \mathbf{u} \in \arg\min\{\Pi_h(\mathbf{v}, \mathbf{z}) \mid \mathbf{v} \in \mathcal{U}_a\}. \tag{104}$$

This is a bilevel multi-target knapsack problem.

If $\alpha = -\eta$ and $\eta > 0$ is the external loading factor, then by simply choosing $T_3(\alpha, \mathbf{z}, \mathbf{u}) = -\mathbf{z}^T \mathbf{c}(\mathbf{u})$ we have the following problem.

Problem 3 (Topology Limit Design) *For the given external load distribution $\mathbf{f}$ and the plastic yield condition in $\mathcal{U}_a$, to solve*

$$(\mathcal{P}_{ld}): \quad \max \{\mathbf{z}^T \mathbf{c}(\mathbf{u}) \mid \eta > 0, \ \mathbf{u} \in \mathcal{U}_a, \ \mathbf{z} \in \mathcal{Z}_a\}, \tag{105}$$

$$\text{s.t. } \mathbf{u} \in \arg\min\{\mathbf{z}^T \mathbf{c}(\mathbf{v}) - \eta \mid \mathbf{v}^T \mathbf{f} = 1, \ \mathbf{v} \in \mathcal{U}_a\}. \tag{106}$$

If $\alpha = \{\omega, -\eta\}$, the combination of $(\mathcal{P}_{lw})$ and $(\mathcal{P}_{ld})$ forms a new problem:

Problem 4 (Topology Lightweight Limit Design) *For the given $\omega^b > \omega^a > 0$, the external load distribution $\mathbf{f}$ and the plastic yield condition in $\mathcal{U}_a$, to solve*

$$(\mathcal{P}_{ll}): \quad \min \{\omega, -\mathbf{z}^T \mathbf{c}(\mathbf{u})\} \ \forall \ \omega \in [\omega^a, \omega^b], \ \eta > 0, \ \mathbf{u} \in \mathcal{U}_a, \ \mathbf{z} \in \mathcal{Z}_a, \tag{107}$$

$$\text{s.t. } \mathbf{u} \in \arg\min\{\mathbf{z}^T \mathbf{c}(\mathbf{v}) - \eta \mid \mathbf{v}^T \mathbf{f} = 1, \ \mathbf{v} \in \mathcal{U}_a\}. \tag{108}$$

Due to a conflict between $\min \omega$ and $\max\{\eta = \mathbf{z}^T \mathbf{c}(\mathbf{u})\}$, this MMM problem could exist a (possibly infinite) number of Pareto optimal solutions.

The canonical duality theory is particularly useful in topology optimization for full-stress (or plastic limit) design. In this type of problems, it is much more convenient to use the stress as the unknown in analysis. Therefore, dual to $(\mathcal{P}_{to})$ the problem for full-stress design can be proposed as:

$$(\mathcal{P}_{to}^*): \quad \max \{T^*(\mathbf{z}, \boldsymbol{\sigma}) \mid \mathbf{z} \in \mathcal{Z}_a, \ \boldsymbol{\sigma} \in \mathcal{S}_a^+\}, \tag{109}$$

$$\text{s.t. } \boldsymbol{\sigma} \in \arg\max\{\Pi_h^d(\boldsymbol{\tau}, \mathbf{z}) \mid \boldsymbol{\tau} \in \mathcal{S}_a^+\}, \tag{110}$$

The top-level dual target $T^d(\mathbf{z}, \boldsymbol{\sigma})$ can be

$$T_1^d(\mathbf{z}, \boldsymbol{\sigma}) = \mathbf{z}^T \mathbf{c}^d(\boldsymbol{\sigma}), \tag{111}$$

$$T_2^d(\mathbf{z}, \boldsymbol{\sigma}) = \mathbf{z}^T \mathbf{c}^d(\boldsymbol{\sigma}) + \frac{1}{2}\mathbf{z}^T \mathbf{Q}^*(\boldsymbol{\sigma})\mathbf{z}, \tag{112}$$

where $\mathbf{c}^d(\boldsymbol{\sigma}) \in \mathbb{R}_+^n$ is a positive vector such that each of its components $c_e^d(\boldsymbol{\sigma})$ is the pure complementary energy in the e-th element Ω_e. The feasible space $\mathcal{S}_a^+$ is a bounded convex set with an inequality constraint $\|\boldsymbol{\sigma}\|_g \leq \sigma_c$, where $\|\boldsymbol{\sigma}\|_g$ is a generalized norm which depends on the yield condition adopted, say either Trisca or von Mises criterion [11]. Corresponding to T_3, we have

$$T_3^d(\alpha^*, \mathbf{z}, \boldsymbol{\sigma}) = \{\alpha^*, T_i^d(\mathbf{z}, \boldsymbol{\sigma})\}, \quad i = 1, 2, \tag{113}$$

where α^* could be $-\omega$, η, or other design parameters.

For linear elastic structures, the total potential energy is a quadratic function of $\mathbf{u}$:

$$\Pi_h(\mathbf{u}, \mathbf{z}) = \frac{1}{2}\mathbf{u}^T \mathbf{K}(\mathbf{z})\mathbf{u} - \mathbf{u}^T \mathbf{f}, \tag{114}$$

where $\mathbf{K}(\mathbf{z}) = \{z_e \mathbf{K}_e\} \in \mathbb{R}^{n \times n}$ is the overall stiffness matrix, obtained by assembling the sub-matrix $z_e \mathbf{K}_e$ for each element Ω_e. Accordingly, $\mathbf{c}(\mathbf{u}) = \frac{1}{2}\{\mathbf{u}_e^T \mathbf{K}_e \mathbf{u}_e\}$ is the strain energy vector. In this case, the global optimal solution for the lower-level minimization problem (99) is simply governed by a linear equilibrium equation $\mathbf{K}(\mathbf{z})\mathbf{u} = \mathbf{f}$. Then for $T = T_1$, the bilevel knapsack problem $(\mathcal{P}_{to})$ can be written in the single-level reduction:

$$(\mathcal{P}_{le}): \quad \min \left\{ \mathbf{f}^T \mathbf{u} - \mathbf{z}^T \mathbf{c}(\mathbf{u}) \mid \mathbf{K}(\mathbf{z})\mathbf{u} = \mathbf{f}, \; \mathbf{v}^T \mathbf{z} \leq V_c, \; \mathbf{z} \in \{0, 1\}^n \right\}. \tag{115}$$

This knapsack-type problem makes a perfect sense in topology optimization, i.e., among all elements $\{\Omega_e\}$ with the given volume vector $\mathbf{v} = \{v_e\}$, one should keep only those who stored more strain energy density $\mathbf{c}(\mathbf{u})$. Based on the canonical dual solution to the knapsack problem, a canonical duality algorithm (CDT) is developed with successful applications.

In term of the stress, the full-stress design problem $(\mathcal{P}_{fs}^*)$ for linear elastic structures can be simply given as:

$$(\mathcal{P}_{fs}^*): \quad \max\{\mathbf{z}^T \mathbf{c}^d(\boldsymbol{\sigma}) \mid \mathbf{z} \in \mathcal{Z}_a, \; \boldsymbol{\sigma} \in \mathcal{S}_a^+\}, \tag{116}$$

where $\mathbf{c}^d(\boldsymbol{\sigma}) = \{\frac{1}{2}\boldsymbol{\sigma}_e^T \mathbf{C}_e \boldsymbol{\sigma}_e\} \in \mathbb{R}_+^n$ is the stress energy vector, $\mathbf{C}_e$ is the compliance matrix of the e-th element,

$$\mathcal{S}_a^+ = \{\boldsymbol{\sigma} \in \mathbb{R}^p \mid \mathbf{D}^* \boldsymbol{\sigma} = \mathbf{f}, \; \|\boldsymbol{\sigma}\|_g \leq \sigma_c\}, \tag{117}$$
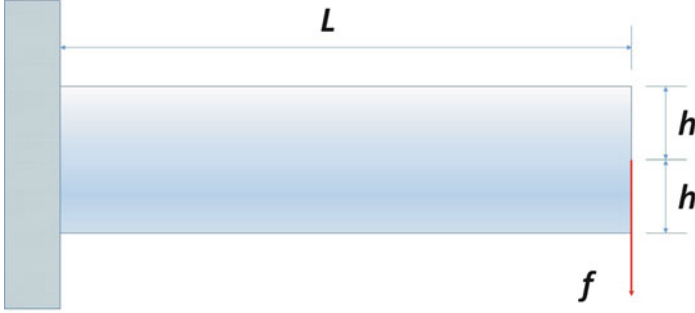
Fig. 2 The design domain for a long cantilever beam with external load

in which $\sigma_c > 0$ is a material constant, and $\mathbf{D}^* \in \mathbb{R}^{m \times p}$ is a balance operator depending on the polynomial interpolation in the mixed finite element method [10, 11].

Example Let us consider the 2-D classical long cantilever beam (see Figure 2). The correct topology optimization model for this benchmark problem should be $(\mathcal{P}_{le})$ [31]. We let $\omega = 0.4$ and the pre-given domain $\omega_0 = 1$ is discretized by $nex \times ney = 180 \times 60$ elements. Computational results obtained by the CDT and by the popular methods SIMP and BESO are summarized in Figure 3, where the $C = \mathbf{z}^T \mathbf{c}(\mathbf{u})$ is the total strain energy. The parameters used are $penal = 3$, $rmin = 1$ for BESO and $penal = 3$, $rmin = 1.5$, $ft = 1$ for SIMP. Clearly, the precise solid-void solution produced by the CDT method is much better than the approximate results produced by other methods. In order to look the strain energy distribution $\mathbf{c} = \{c_e(\mathbf{u})\}$ in the optimal structure, we let $nex \times ney = 80 \times 30$. Figure 4 shows clearly that the CDT can produce mechanically sound structure with homogeneous distribution of strain energy density. Detailed study on canonical duality theory for solving topology optimization problems is given recently in [30–32].

5   Symmetry, NP-Hardness, and Perturbation Methods

The concept of symmetry is closely related to the duality and, in certain sense, can be viewed as a *geometric duality*. Mathematically, symmetry means invariance under transformation. By the canonicality, the object $G(\mathbf{g})$ possesses naturally certain symmetry. If the subject $F(\boldsymbol{\chi}) = 0$, then $\Pi(\boldsymbol{\chi}) = G(\mathbf{D}\boldsymbol{\chi}) = \Phi(\Lambda(\boldsymbol{\chi}))$ and $(\mathcal{P}_g)$ should have either a trivial solution or multiple solutions due to the symmetry. In this case, $\Pi^d(\boldsymbol{\varsigma}) = -\Phi^*(\boldsymbol{\varsigma})$ is concave and, by the triality theory, its critical point $\bar{\boldsymbol{\varsigma}} \in \mathcal{S}_c^-$ is a global maximizer, and $\bar{\boldsymbol{\chi}} = [\mathbf{G}(\bar{\boldsymbol{\varsigma}})]^+\mathbf{f} = 0$ is the biggest local maximizer of $\Pi(\boldsymbol{\chi})$, while the global minimizers must be $\bar{\boldsymbol{\chi}}(\bar{\boldsymbol{\varsigma}})$ for those $\bar{\boldsymbol{\varsigma}} \in \partial \mathcal{S}_c^+$ such that
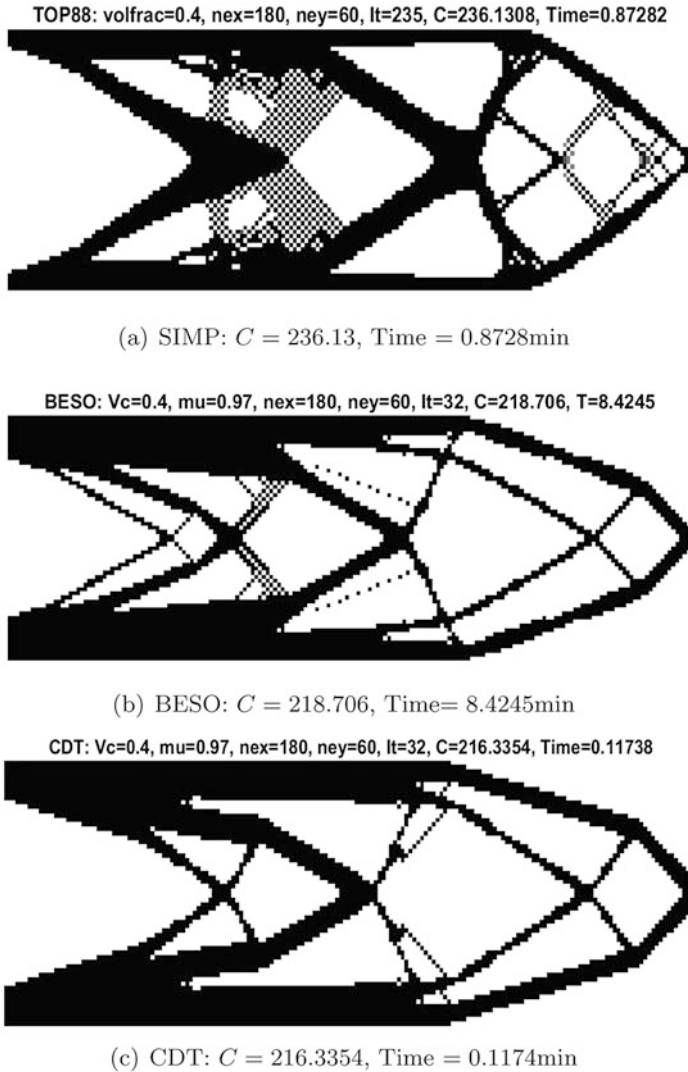
TOP88: volfrac=0.4, nex=180, ney=60, It=235, C=236.1308, Time=0.87282

(a) SIMP: $C = 236.13$, Time $= 0.8728$min

BESO: Vc=0.4, mu=0.97, nex=180, ney=60, It=32, C=218.706, T=8.4245

(b) BESO: $C = 218.706$, Time$= 8.4245$min

CDT: Vc=0.4, mu=0.97, nex=180, ney=60, It=32, C=216.3354, Time=0.11738

(c) CDT: $C = 216.3354$, Time $= 0.1174$min

Fig. 3 Computational results by SIMP (a), BESO (b), and CDT (c) with $\omega = 0.4$.

$$\Pi^g(\bar{\varsigma}) = \min\{-\Phi^*(\varsigma)|\ \det \mathbf{G}(\varsigma) = 0 \ \forall \varsigma \in \mathcal{S}_c\}. \tag{118}$$

Clearly, this nonconvex constrained concave minimization problem could be really NP-hard. Therefore, many well-known NP-hard problems in computer science and global optimization are not well-posed problems. Such as the *max-cut problem*, which is a special case of quadratic integer programming problem ($\mathcal{P}_{qi}$). Due to the symmetry $\mathbf{Q} = \mathbf{Q}^T$ and $\mathbf{f} = 0$, its canonical dual problem has multiple solutions
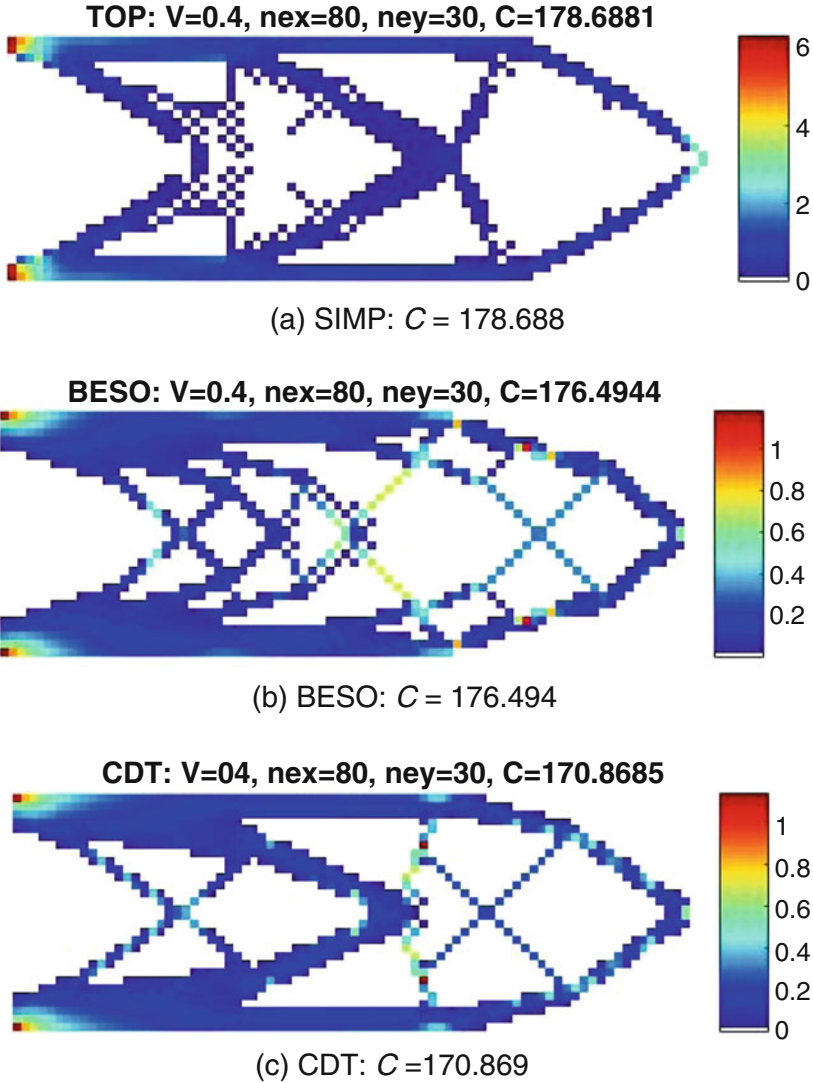
TOP: V=0.4, nex=80, ney=30, C=178.6881

(a) SIMP: $C = 178.688$

BESO: V=0.4, nex=80, ney=30, C=176.4944

(b) BESO: $C = 176.494$

CDT: V=04, nex=80, ney=30, C=170.8685

(c) CDT: $C = 170.869$

Fig. 4 Strain energy distributions by SIMP (a), BESO (b), and CDT (c) with $\omega = 0.4$.

on the boundary of $\mathcal{S}_c^+$. The problem is considered as NP-complete even if $Q_{ij} = 1$ for all edges. Strictly speaking, this is not a real-world problem but only a perfect geometrical model. Without sufficient geometrical constraints in $\mathcal{X}_a$, the graph is not physically fixed and any rigid motion is possible. However, by adding a linear perturbation $\mathbf{f} \neq 0$, this problem can be solved efficiently by the canonical duality theory [85]. Also, it was proved by the author [25, 35] that the general quadratic integer problem $(\mathcal{P}_{qi})$ has a unique solution as long as the input $\mathbf{f} \neq 0$ is big enough.
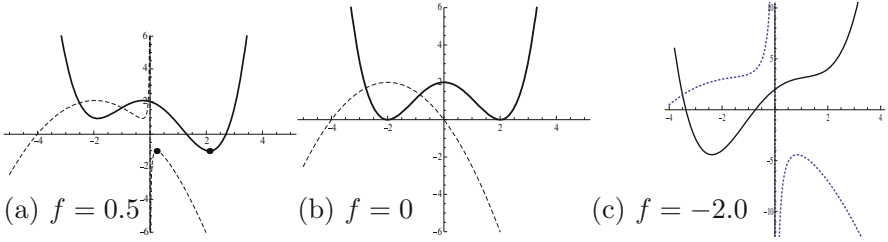
(a) $f = 0.5$ (b) $f = 0$ (c) $f = -2.0$

Fig. 5 Graphs of $\Pi(\mathbf{x})$ (solid) and $\Pi^g(\varsigma)$ (dashed) ($\alpha = 1, \lambda = 2$)

These results show that the subjective function plays an essential role for symmetry breaking which leads to a well-posed problem. To explain the theory and understand the NP-hard problems, let us consider a simple problem.

Example 1 (Nonconvex Minimization in $\mathbb{R}^n$)

$$\min \left\{ \Pi(\mathbf{x}) = \frac{1}{2}\alpha(\frac{1}{2}\|\mathbf{x}\|^2 - \lambda)^2 - \mathbf{x}^T\mathbf{f} \quad \forall \mathbf{x} \in \mathbb{R}^n \right\}, \tag{119}$$

where $\alpha, \lambda > 0$ are given parameters. Let $\Lambda(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 \in \mathbb{R}$, and the canonical dual function is $\Pi^d(\varsigma) = -\frac{1}{2}\varsigma^{-1}\|\mathbf{f}\|^2 - \lambda\varsigma - \frac{1}{2}\alpha^{-1}\varsigma^2$, which is defined on $\mathcal{S}_c = \{\varsigma \in \mathbb{R}| \ \varsigma \neq -\lambda, \ \varsigma = 0 \text{ iff } \mathbf{f} = 0\}$. The criticality condition $\partial \Pi^d(\varsigma) = 0$ leads to a canonical dual equation:

$$(\alpha^{-1}\varsigma + \lambda)\varsigma^2 = \frac{1}{2}\|\mathbf{f}\|^2. \tag{120}$$

This cubic equation has at most three real solutions satisfying $\varsigma_1 \geq 0 \geq \varsigma_2 \geq \varsigma_3$, and, correspondingly, $\{\mathbf{x}_i = \mathbf{f}/\varsigma_i\}$ are three critical points of $\Pi(\mathbf{x})$. By the fact that $\varsigma_1 \in \mathcal{S}_a^+ = \{\varsigma \in \mathbb{R} \mid \varsigma \geq 0\}$, $\mathbf{x}_1$ is a global minimizer of $\Pi(\mathbf{x})$. While for $\varsigma_2, \varsigma_3 \in \mathcal{S}_a^- = \{\varsigma \in \mathbb{R} \mid \varsigma < 0\}$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are local min (for $n = 1$) and local max of $\Pi(\mathbf{x})$, respectively (see Figure 5(a)).

If we let $\mathbf{f} = 0$, the graph of $\Pi(\mathbf{x})$ is symmetric (i.e., the so-called double-well potential or the Mexican hat for $n = 2$ [23]) with infinite number of global minimizers satisfying $\|\mathbf{x}\|^2 = 2\lambda$. In this case, the canonical dual $\Pi^g(\varsigma) = -\frac{1}{2}\alpha^{-1}\varsigma^2 - \lambda\varsigma$ is strictly concave with only one critical point (local maximizer) $\varsigma_3 = -\alpha\lambda < 0$. The corresponding solution $\mathbf{x}_3 = \mathbf{f}/\varsigma_3 = 0$ is a local maximizer. By the canonical dual equation (120) we have $\varsigma_1 = \varsigma_2 = 0$ located on the boundary of $\mathcal{S}_a^+$, which corresponding to the two global minimizers $x_{1,2} = \pm\sqrt{2\lambda}$ for $n = 1$, see Figure 1(b). If we let $f = -2$, then the graph of $\Pi(\mathbf{x})$ is quasi-convex with only one critical point and (120) has only one solution $\varsigma_1 \in \mathcal{S}_c^+$ (see Figure 1(c)).

This simple example reveals an important truth, i.e., the symmetry is the key that leads to multiple solutions. Theoretically speaking, nothing is perfect in this real

world, a perfect symmetry is not allowed for any real-world problem. Thus, any real-world problem must be *well-posed* [29]. In reality, it is impossible to precisely model any real-world problem; although most of the NP-hard problems are artificial, they appear extensively not only in global optimization and computer science but also in chaotic dynamical systems, decision science, and philosophy, say, the well-known Buridan's ass paradox in its most simple version.

Example 2 (Paradox of Buridan's Ass and Perturbation)   A donkey facing two identical hay piles starves to death because reason provides no grounds for choosing to eat one rather than the other.

The mathematical problem of this paradox was formulated in [31]:

$$\max\{c_1 z_1 + c_2 z_2 |\quad c_1 = c_2 = c, \quad z_1 + z_2 \leq 1, \quad (z_1, z_2) \in \{0, 1\}^2\}. \tag{121}$$

Clearly, this is a linear knapsack problem in $\mathbb{R}^2$. Due to the symmetries: $v_1 = v_2 = 1$ and $c_1 = c_2 = c$, the solution to (77) is $\tau_c = c$. Therefore, $\psi_i(\tau_c) = 0 \quad \forall i = 1, 2$ and by Theorem 2 this problem has multiple (two) solutions, which is NP-hard to this donkey.

In order to solve such NP-hard problems, the key idea is to break the symmetry. A linear perturbation method has been proposed by the author and his co-workers. This method is based on a simple truth, i.e., it is impossible to have the two identical hay piles. Thus, by adding a linear perturbation term $\epsilon \rho_1$ to the cost function to break the symmetry, then for $c = 2, \quad \epsilon = 0.05$, the solution to (77) is $\tau_c = 2.0184$. So, the condition (75) holds for $i = 1, 2$ and by the canonical duality theory, the perturbed Buridan's ass problem has a unique solution $\mathbf{z} = (1, 0)$.

Perturbation method has been successfully applied for solving many challenging problems including hard cases of trust region method [7], NP-hard problems in integer programming [84, 85], and nonconvex constrained optimization in Euclidean geometry [62]. By the fact that the subjective function $F(\chi) = \langle \chi, \mathbf{f} \rangle$ plays a key role in real-world problems, the following conjecture was proposed recently [26, 28].

Conjecture 2 *For any given properly posed problem $(\mathcal{P}_g)$ under the Assumption 1, there exists a constant $f_c > 0$ such that $(\mathcal{P}^g)$ has a unique solution in $\mathcal{S}_c^+$ as long as $\|\mathbf{f}\| \geq f_c$.*

This conjecture shows that any properly posed problems are not NP-hard if the input $\|\mathbf{f}\|$ is big enough. Generally speaking, most NP-hard problems have multiple solutions located either on the boundary or the outside of $\mathcal{S}_c^+$. Therefore, a quadratic perturbation method can be suggested as:

$$\begin{aligned}
\Xi_{\delta_k}(\chi, \varsigma) &= \Xi(\chi, \varsigma) + \frac{1}{2}\delta_k \|\chi - \chi_k\|^2 \\
&= \frac{1}{2}\langle \chi, \mathbf{G}_{\delta_k}(\varsigma)\chi \rangle - \Phi^*(\varsigma) - \langle \chi, \mathbf{f}_{\delta_k} \rangle + \frac{1}{2}\delta_k \langle \chi_k, \chi_k \rangle,
\end{aligned}$$

where $\delta_k > 0$, $\boldsymbol{\chi}_k$ $(k = 1, 2, \dots)$ are perturbation parameters, $\mathbf{G}_{\delta_k}(\boldsymbol{\varsigma}) = \mathbf{G}(\boldsymbol{\varsigma}) + \delta_k \mathbf{I}$, and $\mathbf{f}_{\delta_k} = \mathbf{f} + \delta_k \boldsymbol{\chi}_k$. Thus, the original canonical dual feasible space $\mathcal{S}_c^+$ can be enlarged to $\mathcal{S}_{\delta_k}^+ = \{\boldsymbol{\varsigma} \in \mathcal{S}_c | \mathbf{G}_{\delta_k}(\boldsymbol{\varsigma}) \succ 0\}$ such that a perturbed canonical dual problem can be proposed as:

$$(\mathcal{P}_k^g): \quad \max\left\{\min\{\Xi_{\delta_k}(\boldsymbol{\chi}, \boldsymbol{\varsigma}) | \ \boldsymbol{\chi} \in \mathcal{X}_a\} | \ \boldsymbol{\varsigma} \in \mathcal{S}_{\delta_k}^+\right\}. \tag{122}$$

Based on this problem, a canonical primal-dual algorithm has been developed with successful applications for solving sensor network optimization problems [70] and chaotic dynamics [54].

6 Connections with Popular Methods and Techniques

By the fact that the canonical duality-triality theory is a unified mathematical methodology with solid foundation in physics, it is naturally connected to many other powerful methods and techniques in different fields. This paper discusses only two well-known methods in optimization and a so-called composite minimization problem. Connections with other theories and methodologies can be found in [34, 56].

6.1 Relation with SDP Programming

Now, let us show the relation between the canonical duality theory and the popular semi-definite programming relaxation.

Theorem 8 *Suppose that* $\Phi : \mathcal{E}_s \to \mathbb{R}$ *is convex and* $\bar{\boldsymbol{\varsigma}} \in \mathcal{E}_a^*$ *is a solution of the problem:*

$$(\mathcal{P}^{sd}): \quad \min\{g + \Phi^*(\boldsymbol{\varsigma})\} \ s.t. \begin{pmatrix} \mathbf{G}(\boldsymbol{\varsigma}) & \mathbf{f} \\ \mathbf{f}^T & 2g \end{pmatrix} \succeq 0 \ \forall \boldsymbol{\varsigma} \in \mathcal{E}_a^*, \ g \in \mathbb{R}, \tag{123}$$

then $\boldsymbol{\chi} = [\mathbf{G}(\boldsymbol{\varsigma})]^+ \mathbf{f}$ *is a global minimum solution to the nonconvex problem* $(\mathcal{P})$.

Proof The problem $(\mathcal{P}^d)$ can be equivalently written in the following problem (see [86]):

$$\min\left\{g + \Phi^*(\boldsymbol{\varsigma}) | \ g \geq G_{ap}^\Lambda(\boldsymbol{\varsigma}), \ \mathbf{G}(\boldsymbol{\varsigma}) \succeq 0 \ \forall \boldsymbol{\varsigma} \in \mathcal{E}_a^*\right\}. \tag{124}$$

Then, by using the Schur complement lemma, this problem is equivalent to $(\mathcal{P}^{sd})$. The theorem is proved by the triality theory. $\qquad\square$

It was proved [35] that for the same problem $(\mathcal{P}_{qi})$, if we use different geometrical operator:

$$\Lambda(\mathbf{x}) = \mathbf{x}\mathbf{x}^T \in \mathcal{E}_a = \{\boldsymbol{\xi} \in \mathbb{R}^{n \times n} | \ \boldsymbol{\xi} = \boldsymbol{\xi}^T, \ \ \boldsymbol{\xi} \succeq 0,$$

$$\text{rank } \boldsymbol{\xi} = 1, \ \ \xi_{ii} = 1 \ \ \forall i = 1, \dots, n\},$$

and the associated canonical function $\Phi(\boldsymbol{\xi}) = \frac{1}{2}\langle \boldsymbol{\xi}; \mathbf{Q} \rangle + \{0 \text{ if } \boldsymbol{\xi} \in \mathcal{E}_a, +\infty \text{ otherwise}\}$, where $\langle \boldsymbol{\xi}; \boldsymbol{\varsigma} \rangle = \text{tr}(\boldsymbol{\xi}^T \boldsymbol{\varsigma})$, we should obtain the same canonical dual problem $(\mathcal{P}_{qi}^d)$. Particularly, if $\mathbf{f} = 0$, then $(\mathcal{P}_{qi})$ is a typical linear semi-definite programming:

$$\min \frac{1}{2}\langle \boldsymbol{\xi}; \mathbf{Q} \rangle \ \ s.t. \ \boldsymbol{\xi} \in \mathcal{E}_a.$$

Since $\mathcal{E}_a$ is not bounded and there is no input, this problem is not properly posed, which could have either no solution or multiple solutions for a given indefinite $\mathbf{Q} = \mathbf{Q}^T$.

The SDP programming has been used for solving a canonical dual problem in post-buckling analysis of a large deformed elastic beam [1].

6.2 Relation to Reformulation-Linearization/Convexification Technique

The *Reformulation-Linearization/Convexification Technique* (RLT) proposed by H. Sherali and C.H. Tuncbilek [75] is one well-known novel approach for efficiently solving general polynomial programming problems. The key idea of this technique is also to introduce a geometrically nonlinear operator $\boldsymbol{\xi} = \Lambda(\mathbf{x})$ such that the higher-order polynomial object $G(\mathbf{x})$ can be reduced to a lower-order polynomial $\Phi(\boldsymbol{\xi})$. Particularly, for the quadratic minimization problems with linear inequality constraints in $\mathcal{X}_a$:

$$(\mathcal{P}_q): \ \ \min \left\{ \Pi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{Q}\mathbf{x} - \mathbf{x}^T \mathbf{f} | \ \mathbf{x} \in \mathcal{X}_a \right\}, \tag{125}$$

by choosing the quadratic transformation:

$$\boldsymbol{\xi} = \Lambda(\mathbf{x}) = \mathbf{x}\overrightarrow{\otimes}\mathbf{x} \in \mathcal{E}_a \subseteq \mathbb{R}^{n \times n}, \ \ i.e., \ \ \boldsymbol{\xi} = \{\xi_{ij}\} = \{x_i x_j\}, \ \ \forall 1 \le i \le j \le n, \tag{126}$$

where $\overrightarrow{\otimes}$ represents the Kronecker product (avoiding symmetric terms, i.e., $\xi_{ij} = \xi_{ji}$), the quadratic object $G(\mathbf{g})$ can be reformulated as the following *first-level RLT linear relaxation*:

$$G(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} = \frac{1}{2}\sum_{k=1}^{n}q_{kk}\xi_{kk} + \sum_{k=1}^{n-1}\sum_{l=k+1}^{n}q_{kl}\xi_{kl} = \Phi(\boldsymbol{\xi}). \qquad (127)$$

The linear $\Phi(\boldsymbol{\xi})$ can be considered as a special canonical function since $\boldsymbol{\varsigma} = \partial\Phi(\boldsymbol{\xi})$ is a constant and $\Phi^*(\boldsymbol{\varsigma}) = \langle\boldsymbol{\xi};\boldsymbol{\varsigma}\rangle - \Phi(\boldsymbol{\xi}) \equiv 0$ is uniquely defined. Thus, using $\Phi(\boldsymbol{\xi}) = \langle\boldsymbol{\xi};\boldsymbol{\varsigma}\rangle$ to replace $G(\mathbf{x})$ and considering $\boldsymbol{\xi}$ as an independent variable, the problem $(\mathcal{P}_q)$ can be relaxed by the following *RLT linear program*:

$$(\mathcal{P}_{RLT}): \quad \min\{\Phi(\boldsymbol{\xi}) - \langle\mathbf{x},\mathbf{f}\rangle|\ \mathbf{x}\in\mathcal{X}_a,\ \ \boldsymbol{\xi}\in\mathcal{E}_a\}. \qquad (128)$$

Based on this RLT linear program, a branch and bound algorithm was designed [76]. It is proved that if $(\bar{\mathbf{x}},\bar{\boldsymbol{\xi}})$ solves $(\mathcal{P}_{RLT})$, then its objective value yields a lower bound of $(\mathcal{P}_q)$ and $\bar{\mathbf{x}}$ provides an upper bound for $(\mathcal{P}_q)$. Moreover, if $\bar{\boldsymbol{\xi}} = \Lambda(\bar{\mathbf{x}}) = \bar{\mathbf{x}}\overrightarrow{\otimes}\bar{\mathbf{x}}$, then $\bar{\mathbf{x}}$ solves $(\mathcal{P}_q)$.

This technique has been significantly adapted along with supporting approximation procedures to solve a variety of more general nonconvex constrained optimization problems having polynomial or more general factorable objective and constraint functions [74].

By the fact that for any symmetric $\mathbf{Q}$, there exists $\mathbf{D}\in\mathbb{R}^{n\times m}$ such that $\mathbf{Q} = \mathbf{D}^T\mathbf{H}\mathbf{D}$ with $\mathbf{H} = \{h_{kk} = \pm 1,\quad h_{kl} = 0\quad\forall k\neq l\}\in\mathbb{R}^{m\times m}$, the canonicality condition (127) can be simplified as:

$$G(\mathbf{D}\mathbf{x}) = \frac{1}{2}(\mathbf{D}\mathbf{x})^T\mathbf{H}(\mathbf{D}\mathbf{x}) = \frac{1}{2}\sum_{k=1}^{m}h_{kk}\xi_{kk} = \Phi(\boldsymbol{\xi}), \qquad (129)$$

$$\boldsymbol{\xi} = \Lambda(\mathbf{x}) = (\mathbf{D}\mathbf{x})\overrightarrow{\otimes}(\mathbf{D}\mathbf{x})\in\mathbb{R}^{m\times m}. \qquad (130)$$

Clearly, if the scale $m\ll n$, the problem $(\mathcal{P}_{RLT})$ will be much easier than the problems using the geometrically nonlinear operator $\boldsymbol{\xi} = \mathbf{x}\overrightarrow{\otimes}\mathbf{x}$. Moreover, if we are using the Lagrange multiplier $\boldsymbol{\varsigma}\in\mathcal{E}_a^* = \{\boldsymbol{\varsigma}\in\mathbb{R}^{m\times m}|\ \langle\Lambda(\mathbf{x});\boldsymbol{\varsigma}\rangle\geq 0\ \forall\mathbf{x}\in\mathbb{R}^n\}$ to relax the ignored geometrical condition $\boldsymbol{\xi} = \Lambda(\mathbf{x})$ in $(\mathcal{P}_{RLT})$, the problem $(\mathcal{P}_q)$ can be equivalently relaxed as:

$$(\mathcal{P}_\Upsilon): \quad \min_{\mathbf{x}\in\mathcal{X}_a}\min_{\boldsymbol{\xi}\in\mathcal{E}_a}\max_{\boldsymbol{\varsigma}\in\mathcal{E}_a^*}\{\Upsilon(\mathbf{x},\boldsymbol{\xi},\boldsymbol{\varsigma}) = \Phi(\boldsymbol{\xi}) + \langle\Lambda(\mathbf{x}) - \boldsymbol{\xi};\boldsymbol{\varsigma}\rangle - \langle\mathbf{x},\mathbf{f}\rangle\}. \qquad (131)$$

Thus, if $(\bar{\mathbf{x}},\bar{\boldsymbol{\xi}},\bar{\boldsymbol{\varsigma}})$ is a solution to $(\mathcal{P}_\Upsilon)$, then $\bar{\mathbf{x}}$ should be a solution to $(\mathcal{P}_q)$. By using the sequential canonical quadratic transformation $\Lambda(\mathbf{x}) = \Lambda_p(\ldots(\Lambda_1(\mathbf{x})\ldots)$ (see Chapter 4, [18]), this technique can be used for solving general global optimization problems.

6.3 Relation to Composite Minimization

The so-called composite minimization in optimization literature is given in the
following form [57]:

$$\min_x h(c(x)), \tag{132}$$

where $c : \mathbb{R}^n \to \mathbb{R}^m$ is called the inner function, and $h : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty, +\infty\}$
is called the outer function. Although there are some mathematical assumptions,
such that $c(x)$ is smooth and $h(y)$ may be nonsmooth but is usually convex, this
is another abstractly proposed problem. Therefore, this problem appears mainly
from numerical approximation methods, for example, the least squares method for
solving the fixed point problem (53).

In numerical analysis or *matrix completion* [5], the variable x is a $d \times n$ matrix
$\mathbf{X} = \{\mathbf{x}_i\} = \{\mathbf{x}_i^\alpha\}$ ($\alpha = 1, \ldots, d, \ i = 1, \ldots, n$). In sensor network communication
systems, the component $\mathbf{x}_i \in \mathbb{R}^d$ is the position of the i-th sensor and the well-
studied sensor localization problem is to find the sensor locations $\{\mathbf{x}_i\}$ by solving
the following nonlinear system [70]:

$$\|\mathbf{x}_i - \mathbf{x}_j\| = d_{ij} \ \forall (i, j) \in \mathcal{A}_d, \ \ \|\mathbf{x}_i - \mathbf{a}_k\| = e_{ik} \ \forall (i, k) \in \mathcal{A}_e \tag{133}$$

where $\{d_{ij}\}$ and $\{e_{ik}\}$ are given distances, $\mathbf{a}_k \in \mathbb{R}^d$ ($k = 1, \cdots, m$) are specified
anchors, and $\mathcal{A}_d$ and $\mathcal{A}_e$ are two index sets. By the least squares method, this
problem can be formulated as a fourth-order polynomial minimization:

$$\min \left\{ \Pi(\mathbf{X}) = \sum_{(i,j) \in \mathcal{A}_d} \frac{1}{2} w_{ij} (\|\mathbf{x}_i - \mathbf{x}_j\|^2 - d_{ij})^2 + \sum_{(i,k) \in \mathcal{A}_e} \frac{1}{2} \omega_{ik} (\|\mathbf{x}_i - \mathbf{a}_k\|^2 - e_{ik})^2 \right\},$$

where w_{ij} and ω_{ik} are given weights. Clearly, this is a composite minimization if
we let

$$c(\mathbf{X}) = \{\mathbf{c}_{ij}(\mathbf{X}), \ \mathbf{c}_{ik}(\mathbf{X})\}, \ \ \mathbf{c}_{ij} = \mathbf{x}_i - \mathbf{x}_j, \ \mathbf{c}_{ik} = \mathbf{x}_i - \mathbf{a}_k\}, \tag{134}$$

$$h(c) = \sum_{(i,j) \in \mathcal{A}_d} \frac{1}{2} w_{ij} (\|\mathbf{c}_{ij}\|^2 - d_{ij})^2 + \sum_{(i,k) \in \mathcal{A}_e} \frac{1}{2} \omega_{ik} (\|\mathbf{c}_{ik}\| - e_{ik})^2. \tag{135}$$

In this case, the matrix-valued function $c(\mathbf{X}) = \mathbf{g}(\mathbf{X}) = \mathbf{DX} = \{\mathbf{x}_i - \mathbf{x}_j, \ \mathbf{x}_i - \mathbf{a}_k\}$
is the finite difference operator in numerical analysis and $h(c) = G(\mathbf{g})$ is a fourth-
order nonconvex polynomial of the linear operator $\mathbf{g} = \mathbf{DX}$.

We can also let

$$c(\mathbf{X}) = \{c_{ij}(\mathbf{X}), c_{ik}(\mathbf{X})\}, \ \ c_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 - d_{ij}, \ c_{ik} = \|\mathbf{x}_i - \mathbf{a}_k\|^2 - e_{ik}, \tag{136}$$

$$h(c) = \sum_{(i,j)\in\mathcal{A}_d} \frac{1}{2} w_{ij} c_{ij}^2 + \sum_{(i,k)\in\mathcal{A}_e} \frac{1}{2} \omega_{ik} c_{ik}^2. \tag{137}$$

In this case, $\Pi(\mathbf{X}) = h(c(\mathbf{X}))$ is also a composite function but now $c(\mathbf{X}) = \boldsymbol{\xi}(\mathbf{X}) = \Lambda(\mathbf{X})$ is a nonlinear operator and $\Phi(\boldsymbol{\xi}) = h(\boldsymbol{\xi})$ is a convex function. Therefore, the composition:

$$\Pi(\mathbf{X}) = h(c(\mathbf{X})) = G(\mathbf{DX}) = \Phi(\Lambda(\mathbf{X}))$$

is indeed a canonical transformation. The sensor localization problem is considered to be NP-hard by traditional theories and methods even if $d = 1$ [3]. From the point view of the canonical duality theory, this problem has usually multiple global minimizers due to the lacking of the subjective function. Therefore, by introducing a linear perturbation $F(\mathbf{X}) = \langle \mathbf{X}, \mathbf{T} \rangle = \text{tr}(\mathbf{X}^T \mathbf{T})$, the perturbed sensor localization problem $\min\{\Pi(\mathbf{X}) = \Phi(\Lambda(\mathbf{X})) - F(\mathbf{X})\}$ can be solved deterministically by the canonical duality theory in polynomial time [55, 68, 70].

Generally speaking, the composite function is a special case of the canonical transformation $G(\mathbf{g}) = \Phi \circ \Lambda(\mathbf{g})$ if $h(y) = \Phi(y)$ is convex, $x = \mathbf{g}$, and $\Lambda(x) = c(x)$ as the geometrical measure. It is an objective function if $c(x) = x^T x$. In this case, $h(c(x))$ is the so-called convex composite function. In real-world applications, $\mathbf{g}(\mathbf{x})$ could be again a composite function. For multi-scale systems, $\mathbf{g}$ can be defined by (see [45]):

$$\mathbf{g}(\mathbf{x}) = (\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_k)\mathbf{x} = \{\mathbf{g}_i(\mathbf{x})\}, \quad \mathbf{g}_i(\mathbf{x}) = \mathbf{D}_i \mathbf{x}, \tag{138}$$

each $\mathbf{g}_i$ is a geometrical measure with dimension different from other $\mathbf{g}_j$, $j \neq i$. Correspondingly:

$$G(\mathbf{Dx}) = \Phi(\Lambda(\mathbf{x})), \quad \Lambda(\mathbf{x}) = \Lambda_k \circ \Lambda_{k-1} \circ \cdots \circ \Lambda_1(\mathbf{x}) \tag{139}$$

is called the *sequential canonical transformation* (see Chapter 4, [18]). Particularly, if every $\Lambda_i(\boldsymbol{\xi}_{i-1})$ is a convex polynomial function of $\boldsymbol{\xi}_{i-1} = \Lambda_{i-1}$ ($i = 1, \ldots, k$, $\Lambda_0 = \mathbf{x}$), the composition $\Phi(\Lambda(\mathbf{x}))$ is the *canonical polynomial function*. The sequential canonical transformation for solving high-order polynomial minimization problems have been studied in [18, 46].

7 Conclusions

Based on the necessary conditions and basic laws in physics, a unified multi-scale global optimization problem is proposed in the canonical form:

$$\Pi(\boldsymbol{\chi}) = G(\mathbf{D}\boldsymbol{\chi}) - F(\boldsymbol{\chi}) = \Phi(\Lambda(\boldsymbol{\chi})) - \langle \boldsymbol{\chi}, \mathbf{f} \rangle. \tag{140}$$

The object G depends only on the model and $G(\mathbf{g}) \geq 0 \ \forall \mathbf{g} \in \mathcal{G}_a$ is necessary; G should be an objective function for physical systems, but it is not necessary for artificial systems (such as management/manufacturing processes and numerical simulations, etc.). The subject F depends on each properly posed problem and must satisfy $F(\chi) \geq 0$ together with necessary geometrical constraints for the output $\chi \in \mathcal{X}_a$ and equilibrium conditions for the input $\mathbf{f} \in \mathcal{X}_a^*$. The geometrical nonlinearity of $\Lambda(\chi)$ is necessary for nonconvexity in global optimization, bifurcation in nonlinear analysis, chaos in dynamics, and NP-hardness in computer science.

Developed from large deformation nonconvex analysis/mechanics, the canonical duality-triality is a precise mathematical theory with solid foundation in physics and natural root in philosophy, so it is naturally related to the traditional theories and powerful methods in global optimization and nonlinear analysis. By the fact that the canonical duality is a universal law of nature, this theory can be used not only to model real-world problems but also for solving a wide class of challenging problems in multi-scale complex systems. The conjectures proposed in this paper can be used for understanding and clarifying NP-hard problems.

It is author's hope that by reading this paper, the readers can have a clear understanding not only on the canonical duality-triality theory and its potential applications in multi-disciplinary fields, but also on the generalized duality-triality principle and its role in modeling/understanding real-world problems.

References

1. Ali, E.J. and Gao, D.Y. (2017). Improved canonical dual finite element method and algorithm for post-buckling analysis of nonlinear Gao beam. In *Canonical Duality Theory*, D.Y. Gao, N. Ruan and V. Latorre (Eds), pages 277–289. Springer, 2017.
2. Anorld, VI (1998) On teaching mathematics, *Russian Math. Surveys*, 53 (1), 229–236.
3. Aspnes, J., Goldberg, D. and Yang, Y.R. On the computational complexity of sensor network localization, in: Lecture Notes in Computer Science, 3121, Springer-Verlag, 2004, pp. 3244.
4. Bader B W, Kolda T G (2006). Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Trans Math Software*, 32(4): 635–653.
5. Candés, E. and Recht, B. (2008). Exact matrix completion via convex optimization, Technical Report, California Institute of Technology, 2008.
6. Chen Y. and Gao, D.Y. (2016). Global solutions to nonconvex optimization of 4th-order polynomial and log-sum-exp functions, *J. Global Optimization*, 64(3), 417–431.
7. Chen Y. and Gao, D.Y. (2017). Global solutions to spherically constrained quadratic minimization via canonical duality theory, in *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, DY Gao, V. Latorre and N. Ruan (eds), Springer, New York, pp 291–314.

8. Ciarlet, PG (2013). *Linear and Nonlinear Functional Analysis with Applications,* SIAM, Philadelphia.

9. Feynman R, Leighton R, and Sands M. *The Feynman Lectures on Physics*, Volume II, 1964.

10. Gao, DY (1988). Panpenalty finite element programming for limit analysis, *Computers & Structures,* 28, pp. 749–755.

11. Gao, DY (1996). Complementary finite element method for finite deformation nonsmooth mechanics. *J. Eng. Math.* 30, 339–353.

12. Gao, D.Y. (1997). Dual extremum principles in finite deformation theory with applications to post-buckling analysis of extended nonlinear beam theory. *Applied Mechanics Reviews*, 50 (11), S64-S71 (1997).

13. Gao, D.Y. (1998). Duality, triality and complementary extremum principles in nonconvex parametric variational problems with applications, *IMA J. Appl. Math.*, 61, 199–235.

14. Gao, D.Y. (1998). Bi-complementarity and duality: A framework in nonlinear equilibria with applications to the contact problems of elastoplastic beam theory, *J. Math. Anal. Appl.*, 221, 672–697.

15. Gao, D.Y. (1999). Pure complementary energy principle and triality theory in finite elasticity, *Mech. Res. Comm.* 26 (1), 31–37.

16. Gao, D.Y. (1999). *Duality-Mathematics,* Wiley Encyclopedia of Electronical and Electronical Engineering, **6**, 68–77.

17. Gao, D.Y. (1999). General Analytic Solutions and Complementary Variational Principles for Large Deformation Nonsmooth Mechanics. *Meccanica* 34, 169–198.

18. Gao, D.Y. (2000). *Duality Principles in Nonconvex Systems: Theory, Methods and Applications.* Springer, New York/Boston, 454pp.

19. Gao, D.Y. (2000). Analytic solution and triality theory for nonconvex and nonsmooth variational problems with applications, *Nonlinear Analysis*, 42, 7, 1161–1193.

20. Gao, D. Y. *Canonical dual transformation method and generalized triality theory in nonsmooth global optimization.* J. Glob. Optim. 17(1/4), pp. 127–160 (2000).

21. Gao, D.Y. (2001). Complementarity, polarity and triality in nonsmooth, nonconvex and nonconservative Hamiltonian systems. *Phil. Trans. R. Soc. London A* 359, 2347–2367.

22. Gao, D.Y.(2003). Perfect duality theory and complete solutions to a class of global optimization problems. Optimization 52(4–5), 467–493

23. Gao, D.Y. (2003). Nonconvex semi-linear problems and canonical duality solutions. *Advances in Mechanics and Mathematics*, II, Springer, 261–311.

24. Gao, D.Y. (2007). Solutions and optimality to box constrained nonconvex minimization problems *J. Indust. Manage. Optim.*, 3(2), 293–304.

25. Gao, D.Y. (2009). Canonical duality theory: unified understanding and generalized solutions for global optimization. *Comput. & Chem. Eng.* 33, 1964–1972.

26. Gao, D.Y. (2014). Unified modeling and theory for global optimization. Plenary Lecture at *16th Baikal Int. Seminar on Methods of Optimization and Their Applications*, June 30 - July 6, 2014, Olkhon, Russia.

27. Gao, D.Y. (2016). Analytical solutions to general anti-plane shear problems in finite elasticity, *Continuum Mech. Thermodyn. 28:175–194*

28. Gao, D.Y. (2016). On unified modeling, theory, and method for solving multi-scale global optimization problems, in *Numerical Computations: Theory And Algorithms,* (Editors) Y. D. Sergeyev, D. E. Kvasov and M. S. Mukhametzhanov, AIP Conference Proceedings 1776, 020005.

29. Gao, D.Y. (2016). On unified modeling, canonical duality-triality theory, challenges and breakthrough in optimization, https://arxiv.org/abs/1605.05534 .

30. Gao, D.Y. (2017). Canonical Duality Theory for Topology Optimization, *Canonical Duality-Triality: Unified Theory and Methodology for Multidisciplinary Study*, D.Y. Gao, N. Ruan and V. Latorre (Eds). Springer, New York, pp.263–276.

31. Gao, D.Y. (2018). On topology optimization and canonical duality method, *Computer Methods in Applied Mechanics and Engineering*, 341, 249–277.

32. Gao, D.Y. and Ali, E.J. (2018). A novel canonical duality theory for 3-D topology optimization, *Emerging Trends in Applied Mathematics and High-Performance Computing*, V.K. Singh, D.Y. Gao and A. Fisher (Eds). Springer, New York.

33. Gao, D.Y. and Hajilarov, E. Analytic solutions to 3-D finite deformation problems governed by St Venant–Kirchhoff material. In DY Gao, V. Latorre, and N. Ruan, editors, *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, pages 69–88. Springer, New York, 2017.

34. Gao, D.Y., Latorre, V. and Ruan, N. (2017). *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, Springer, New York, 377pp.

35. Gao, D.Y. and Ruan, N.: Solutions to quadratic minimization problems with box and integer constraints. J. Glob. Optim. 47, 463–484 (2010).

36. Gao, D.Y., Ogden, R.W. (2008). Multi-solutions to non-convex variational problems with implications for phase transitions and numerical computation. *Q. J. Mech. Appl. Math.* 61, 497–522.

37. Gao, D.Y., Ogden, R.W. (2008). Closed-form solutions, extremality and nonsmoothness criteria in a large deformation elasticity problem. *Zeits. Ang. Math. Physik* 59, 498–517.

38. Gao, DY, Ogden, RW, Stavroulakis, G (2001). *Nonsmooth and Nonconvex Mechanics: Modelling, Analysis and Numerical Methods.* Kluwer Academic Publishers.

39. Gao, DY, Ruan, N, and Latorre, V (2017). Canonical duality-triality theory: Bridge between nonconvex analysis/mechanics and global optimization in complex systems, in *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, DY Gao, V. Latorre and N. Ruan (eds), Springer, New York, pp 1–48.

40. Gao, D.Y., Ruan, N., Sherali, H. (2009). Solutions and optimality criteria for nonconvex constrained global optimization problems with connections between canonical and Lagrangian duality. *J. Glob. Optim.* 45, 473–497.

41. Gao, D.Y., Ruan, N., Sherali, H. (2010). Canonical dual solutions for fixed cost quadratic programs, *Optimization and Optimal Control,* A. Chinchuluun et al. (eds.), Springer Optimization and Applications 39.

42. Gao, D.Y. and Sherali, H.D. (2009). Canonical Duality Theory: Connection between nonconvex mechanics and global optimization, *Advances in Appl. Math. and Global Optimization,* D.Y. Gao and H. Sherali (eds). Springer.

43. Gao, D.Y., Strang, G.: Geometric nonlinearity: Potential energy, complementary energy, and the gap function. Quart. Appl. Math. 47(3), 487–504 (1989).

44. Gao, D.Y. and Wu, C.Z. (2012). On the triality theory for a quartic polynomial optimization problem, *J. Ind. Manag. Optim.* 8(1), 229–242.

45. Gao, D.Y., Yu, H.F. (2008). Multi-scale modelling and canonical dual finite element method in phase transitions of solids. *Int. J. Solids Struct.* 45, 3660–3673.

46. Gao, T.K. (2013). Complete solutions to a class of eighth-order polynomial optimization problems, *IMA J Appl Math*, 80(1), 158–176.

47. Hiriart-Urruty, J.B.(1985). Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. *Lecture Note Econ. Math. Syst.*, 256: 37–70.

48. Horst, R., Thoai, N.V.(1999). DC Programming: overview. *J. Opt. Theory Appl.*, 103: 1–43.

49. Jin, Z. and Gao, D.Y. (2017). On modeling and global solutions for d.c. optimization problems by canonical duality theory, *Applied Mathematics and Computation*, 296, 168–181

50. Karp, R.M. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103, New York: Plenum, 1972.

51. Lagrange, Joseph-Louis (1811). *Mecanique Analytique.* Courcier, (reissued by Cambridge Univ. Press, 2009).

52. Landau, L.D. and Lifshitz, E.M. (1976). *Mechanics*. Vol. 1 (3rd ed.). Butterworth-Heinemann.

53. Latorre, V. and Gao, D.Y. (2016). Canonical duality for solving general nonconvex constrained problems, *Optimization Letters*, 10(8):1763–1779. http://link.springer.com/article/10.1007/s11590-015-0860-0/fulltext.html

54. Latorre, V. and Gao, D.Y. (2016). Global optimal trajectory in chaos and NP-Hardness. *Int. J. Birfurcation and Chaos*, 26, 1650142 (2016) https://doi.org/10.1142/S021812741650142X

55. Latorre, V. and Gao, D.Y. (2018). Canonical duality theory for solving large-sized noisy sensor network localization problems. to appear in *IEEE Transactions on Cybernetics*

56. Latorre, V. and Gao, D.Y. (2018). On half-quadratic reformulation and canonical duality theory in image restoration. Submitted.

57. Lewis A. S. and Wright, S. J. (2016). A proximal method for composite minimization, *Mathematical Programming: Series A and B archive*, 158 (1–2): 501–546

58. Li, S.F. and Gupta, A. (2006). On dual configuration forces, *J. of Elasticity*, 84:13–31.

59. Liu, I.-S. (2005). Further remarks on Euclidean objectivity and the principle of material frame-indifference. *Continuum Mech. Thermodyn.*, 17, 125–133

60. Marsden, J.E. and Hughes, T.J.R. : *Mathematical Foundations of Elasticity,* Prentice-Hall, 1983.

61. Morales, D.M. and Gao, D.Y. (2015). Canonical duality theory and triality for solving general nonconstrained global optimization problems, *Math. Mech. Complex Systems*, Vol. 3 (2015), No. 2, 139–161.

62. Morales, D.M. and Gao, D.Y. (2017). On minimum distance between two surfaces, in *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, DY Gao, V. Latorre and N. Ruan (eds), Springer, New York, pp 359–372.

63. Moreau, J.J. (1968). La notion de sur-potentiel et les liaisons unilatérales en élastostatique, *C.R. Acad. Sc. Paris,* 267 A, 954–957.

64. Murdoch, A.I.(2005). On criticism of the nature of objectivity in classical continuum physics, *Continuum Mech. Thermodyn.,* 17(2):135–148

65. Oden, J.T. *An Introduction to Mathematical Modeling*. John Wiley & Sons, 2011.

66. Puchinger, J., Raidl, G.R., Pferschy, U. (2010). The multidimensional knapsack problem: structure and algorithms, *INFORMS J. Comput.* 22 (2): 250–265 .

67. Qi, L.Q., Chen, H. and Chen, Y. (2018) *Tensor Eigenvalues and Their Applications*, Springer, New York, 329pp

68. Ruan, N. and Gao, D.Y. (2014). Global optimal solutions to a general sensor network localization problem, *Performance Evaluations*, 75–76: 1–16.

69. Ruan, N. and Gao, D.Y. (2018). Global optimal solution to quadratic discrete programming problem with inequality constraints. In *Canonical Duality-Triality: Unified Theory and Methodology for Multidisciplinary Study*, D.Y. Gao, V. Latorre and N. Ruan (Eds). Springer, New York, pp.315–338. http://arxiv.org/abs/1205.0856

70. Ruan, N. and Gao, D.Y. (2014). Global optimal solutions to a general sensor network localization problem, *Performance Evaluations*, 75–76: 1–16.

71. Ruan, N. and Gao, D.Y.(2014). Canonical duality approach for nonlinear dynamical systems, *IMA J. Appl. Math.*, 79: 313–325.

72. Ruan, N. and Gao, D.Y. (2018). On Modelling and Complete Solutions to General Fixpoint Problems in Multi-Scale Systems with Applications, https://arxiv.org/abs/1801.08651

73. Santos, H.A.F.A. and Gao D.Y. (2011) Canonical dual finite element method for solving post-buckling problems of a large deformation elastic beam, *Int. J. Nonlinear Mechanics,* 47: 240–247.

74. Sherali, H. D. (2002). Tight Relaxations for Nonconvex Optimization Problems Using the Reformulation- Linearization/Convexification Technique (RLT), *Handbook of Global Optimization, Volume 2: Heuristic Approaches,* P. M. Pardalos and H. E. Romeijn, Eds., Kluwer Academic Publishers, 1–63.

75. Sherali, H.D. and Tuncbilek, C.H. (1992). A Global Optimization Algorithm for Polynomial Programming Problems Using a Reformulation-Linearization Technique, *J. Global Optim.,* Vol. 2, No. 1, pp. 101–112.

76. Sherali, H.D. and Tuncbilek, C.H. (1995). A reformulation-convexification approach for solving nonconvex quadratic programming problems. *J. Global Optimization,* 7:1–31.

77. Strang, G. (1986). *Introduction to Applied Mathematics*, Wellesley-Cambridge Press.

78. Strugariu, R. , Voisei, M.D. and Zalinescu, C. : Counter-examples in bi-duality, triality and tri-duality, *Discrete & Continuous Dynamical Systems - A,* 2011, 31(4): 1453–1468.

79. Pham Dinh Tao, Le Thi Hoai An(2014). Recent Advances in DC Programming and DCA. *Transactions on Computational Collective Intelligence*, 13: 1–37.
80. Truesdell, C. and Noll, W. (1965). *The Nonlinear Field Theories of Mechanics*, Springer-Verlag, 591pp.
81. Toland, J.F.(1979). A duality principle for non-convex optimisation and the calculus of variations. *Arch. Ration. Mech. Anal.*, 71: 41–61.
82. Tuy, H.(1995). D.C. optimization: Theory, methods and algorithms. In: Horst, R., Pardalos, P.M. (eds.) *Handbook of Global Optimization*, pp. 149–216. Kluwer Academic Publishers, Dordrecht.
83. Voisei, M.D. and Zălinescu , C.(2011). Some remarks concerning Gao-Strang's complementary gap function, *Applicable Analysis*, Vol. 90, No. 6, 1111–1121.
84. Wang, Z.B., Fang, S.C., Gao, D.Y. and Xing, W.X. (2008). Global extremal conditions for multi-integer quadratic programming. *J Industrial and Management Optimization,* 4(2):213.
85. Wang, Z.B., Fang, S.C., Gao, D.Y. and Xing, W.X. (2012). Canonical dual approach to solving the maximum cut problem, *J. Glob. Optim.*, 54, 341–351.
86. Zhou, X.J., Gao, D.Y. and Yang, C.H. (2016) Global solutions to a class of CEC benchmark constrained optimization problems, *Optim Lett*, 10:457–472 https://doi.org/10.1007/s11590-014-0784-0