Canonical Duality Theory and Algorithm for Solving Bilevel Knapsack Problems With Applications

David Yang Gao

Abstract—A novel canonical duality theory (CDT) is presented for solving general bilevel mixed integer nonlinear optimization governed by linear and quadratic knapsack problems. It shows that the challenging knapsack problems can be solved analytically in term of their canonical dual solutions. The existence and uniqueness of these analytical solutions are proved. NP-hardness of the knapsack problems is discussed. A powerful CDT algorithm combined with an alternative iteration and a volume reduction method is proposed for solving the NP-hard bilevel knapsack problems. Application is illustrated by benchmark problems in optimal topology design. The performance and novelty of the proposed method are compared with the popular commercial codes.

Index Terms—Bilevel optimization, canonical duality theory (CDT), CDT algorithm, knapsack problems, NP-hardness, topology design.

I. PROBLEMS AND MOTIVATION

BILEVEL optimization is a special kind of optimization where an (upper-level) optimization problem contains another (lower-level) optimization problem as a constraint. The first bilevel optimization problem was realized in the field of game theory by a German economist Stackelberg in 1934 [51]. Since then, these problems have been found in many real-world applications, including artificial intelligence, cryptography, decision science, defensive force structure design, game theory, topology design of structures and communication networks, etc. (see [4], [6], [38], [58], and [60]). It was discovered recently by Gao [24] that the well-studied topology optimization in computational mechanics should be formulated as a bilevel knapsack problem (BKP), i.e., the upper-level optimization is a knapsack problem, while the lower-level optimization is governed by the well-known minimum total potential energy principle. The knapsack problem is one of the most fundamental problems in combinatorial optimization. It has been studied extensively for more than a century in multidisciplinary fields (see [39]). The quadratic knapsack problem (QKP) is an extension of knapsack problem that allows for quadratic terms in the target function. QKP was first introduced in 19th century (see [57]), which has a wide range of applications including telecommunication, transportation network, computer science and economics. In fact, Witzgall [57] first discussed QKP when selecting sites for satellite stations in order to maximize the global traffic with respect to a budget constraint. Similar model applies to problems like considering the location of airports, railway stations, or freight handling terminals [47]. Applications of QKP in the field of computer science is more common after the early days: compiler design problem [36], clique problem [46], very large-scale integration (VLSI) design, and manufacturing [14]. In bilevel optimization of multiscale complex systems, both linear and QKPs can appear in either upper or lower optimization. Indeed, over the last 20 years, a variety of BKPs has been proposed (see [5] and [7]).

Generally speaking, BKPs are extremely difficult from the computational point of view and cannot be expressed in terms of classical integer programs (which can only handle a single level of optimization). Due to the integer constraint and bilevel coupling, BKPs are nondifferentiable and nonconvex. It has been proven in [53] that merely evaluating a solution for optimality is also an NP-hard task. In fact, it was shown in [35] that even the linear bilevel problems are strongly NP-hard. The proof for the nonexistence of a polynomial time algorithm for linear bilevel problems can be found in [13]. Classical theories and methods cannot be used directly for solving this type of problems. Therefore, the general BKP could be one of the most challenging problems in computer science and global optimization.

During the past 40 years many approximation methods have been proposed for numerically solving bilevel optimization problems. Impressive collections of these methods and applications can be found in the books [3], [11], and [61] as well as the review article [12]. Evolutionary method [50], [54] is a popular approach, where the lower level optimization problem is solved corresponding to each and every upper level member. One of the first evolutionary algorithms for solving bilevel optimization problems was proposed in the

early 1990s. Mathieu *et al.* [44] used a nested approach with genetic algorithm at the upper level, and linear programming at the lower level. One of the early works on discrete bilevel optimization was by Vicente *et al.* [53], which focused on discrete linear bilevel programs, and analyzed the properties and existence of the optimal solution for different kinds of discretization arising from the upper and lower level variables. The authors have shown in this paper that certain compactness conditions guarantee the existence of optimal solution in continuous linear bilevel programs, discrete–continuous linear bilevel programs, and discrete–discrete linear bilevel programs. The conditions are equivalent to stating that the inducible region is nonempty. However, the existence conditions in the case of continuous-discrete linear bilevel programs are not straightforward. In the classical literature, branch-and-bound and branch-and-cut are some of the commonly used deterministic methods to handle discreteness in variables. But these well-studied methods can be computational expensive, and can be used only for solving very small-sized problems. Therefore, global optimization problems with 200 variables are referred to as "medium scale," problems with 1000 variables as "large scale," and the so-called "extra-large scale" is only around 4000 variables. However, any simple problem in topology design can easily have millions of variables [25]. Although there are some fully polynomial time approximation schemes for discrete optimization, they are not always applicable in practice due to memory requirements [56]. Also these algorithms are based on heuristic techniques, the obtained solutions come with no guarantee of global optimality and may get stuck in local minima. Therefore, it is truly important to develop a powerful deterministic method for solving large-scale general BKPs.

Canonical duality theory (CDT) is a precise methodological theory, which can be used not only for modeling complex systems within a unified framework, but also for solving a large class of challenging problems in nonconvex analysis and global optimization [17]. This theory is particularly powerful for solving integer constrained problems. In 2007, Gao [20] discovered that by simply using a standard canonical measure $\Lambda(\mathbf{z}) = \mathbf{z}^2 - \mathbf{z}$, the 0-1 constrained problems can be equivalently converted to a unified continuous concave maximization dual problem in a convex feasible space, which can be solved deterministically via well-developed convex optimization techniques. This method has been generalized for solving general discrete optimization problems [21], [34], [48], [55]. Applications of the CDT to multidisciplinary study was given recently [27] and an analytic solution to linear knapsack problem (LKP) has been obtained [24], [25].

This paper deals with a general BKP, in which, the upper-level optimization is a QKP coupled with continuous follower variables, while the lower-level optimization is a mixed integer nonlinear minimization problem involves both leader and follower variables (since an alternative iteration method is used in this paper, this order of upper and lower levels can be reversed). The main contributions of this paper include the following:

1) canonical dual solutions to QKPs;

2) improved analytic solutions to knapsack problems and criteria for NP-hardness;
3) a volume reduction method (VRM) combined with an alternative iteration for solving general BKP;
4) a powerful CDT algorithm for solving general BKP with an application to topology design.

The rest of this paper is organized as follows. The next section provides the mathematical formulation of the BKP and its special linear case. A decoupled alternative iteration (DAI) is suggested. The theoretical results of this paper are presented in Section III, including a canonical penalty-duality method, a general analytic solution form, as well as the existence and uniqueness of this solution to the QKP. In Section IV, the NP-hardness of the knapsack problems is addressed. Improved analytic solutions are provided for both QKP and LKP. Section V proposes a VRM and a CPD algorithm for solving BKP. In Section VI, applications are illustrated by both two-dimensional (2-D) and three-dimensional (3-D) benchmark problems in topology design. The performance and novelty of the CPD algorithm are verified by computational results. Finally, Section VII presents the conclusion and open problems for future research.

II. BILEVEL OPTIMIZATION MODELING AND KNAPSACK PROBLEMS

The bilevel optimization problem proposed to solve is given as

$$(\mathcal{P}) : \min_{\mathbf{x} \in \mathcal{X}_a, \mathbf{z} \in \mathcal{Z}_a} \left\{ P(\mathbf{x}, \mathbf{z}) = \frac{1}{2} \mathbf{z}^T \mathbf{Q}(\mathbf{x}) \mathbf{z} - \mathbf{z}^T \mathbf{c}(\mathbf{x}) \right\} \quad (1)$$

$$\text{s.t. } \mathbf{x} \in \arg \min_{\boldsymbol{\chi} \in \mathcal{X}_a} \left\{ \Pi(\boldsymbol{\chi}, \mathbf{z}) = G(\mathbf{D}\boldsymbol{\chi}, \mathbf{z}) - \boldsymbol{\chi}^T \mathbf{f} \right\} \quad (2)$$

where $P : \mathcal{X}_a \times \mathcal{Z}_a \subset \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ is the upper-level target function, $\Pi : \mathcal{X}_a \times \mathcal{Z}_a \to \mathbb{R}$ is the lower-level target function, $\mathbf{Q}(\mathbf{x})$ and $\mathbf{c}(\mathbf{x})$ are two pregiven matrix-valued and vector-valued functions of $\mathbf{x} \in \mathcal{X}_a$, respectively; $\mathbf{f} \in \mathbb{R}^m$ is a given vector; $\mathbf{D} : \mathcal{X}_a \to \mathcal{G}_a \subset \mathbb{R}^p$ is a linear operator, and $G : \mathcal{G}_a \times \mathcal{Z}_a \to \mathbb{R}$ is a given function, which is usually a nonconvex objective function of $\mathbf{g} = \mathbf{D}\mathbf{x}$, but linearly depends on $\mathbf{z} \in \mathcal{Z}_a$. The upper-level (or leader) variable $\mathbf{z}$ is a discrete vector, whose domain is a subset of Boolean space (i.e., a knapsack)

$$\mathcal{Z}_a = \left\{ \mathbf{z} = \{z_i\} \in \{0, 1\}^n | \mathbf{v}^T \mathbf{z} \le V_c \right\} \quad (3)$$

where $V_c > 0$ is a given size of the knapsack, $\mathbf{v} = \{v_i\} \in \mathbb{R}_+^n$ is a given positive vector. The lower-level (or the follower) variable $\mathbf{x}$ is a continuous vector, whose domain $\mathcal{X}_a$ is a convex set of $\mathbb{R}^m$, in which, only linear equality constraints are included.

Clearly, the problem $(\mathcal{P})$ is a BKP. Due to the integer constraint, the nonlinearity/nonconvexity of $\Pi(\mathbf{x}, \mathbf{z})$, and the strong coupling between the upper and lower level problems, the proposed problem $(\mathcal{P})$ could be one of the most challenging problems in global optimization and computer science. It is fundamentally difficult to solve this problem directly. Therefore, a DAI method will be adopted in this paper, i.e., the coupled bilevel optimization is split into two decoupled subproblems by alternative iteration.

1) For a given initial $\mathbf{z}_{k-1} \in \mathcal{Z}_a$, to solve the lower-level problem (2) first for

$$\mathbf{x}_k = \arg\min\{\Pi(\mathbf{x}, \mathbf{z}_{k-1})|\mathbf{x} \in \mathcal{X}_a\}. \quad (4)$$

2) For the fixed $\mathbf{x}_k \in \mathcal{X}_a$, to solve the upper-level knapsack problem (1) for

$$\mathbf{z}_k = \arg\min\{P(\mathbf{x}_k, \mathbf{z})|\mathbf{z} \in \mathcal{Z}_a\}. \quad (5)$$

The upper-level minimization (5) is the well-known QKP. The binary variable $\mathbf{z} = \{z_i\}$ represents whether item i is included in the knapsack $\mathcal{Z}_a$. The given vector $\mathbf{c}(\mathbf{x}) = \{c_i(\mathbf{x})\} \in \mathbb{R}_+^n$ must be positive for any given $\mathbf{x} \in \mathcal{X}_a$ such that c_i is the profit earned by selecting item i. The given matrix $\mathbf{Q}(\mathbf{x}) \in \mathbb{R}^{n \times n}$ must be symmetrical, its diagonal elements are usually $Q_{ii} = 0, i = 1, \ldots, n$ and $-Q_{ij}$ is the profit achieved if both item i and j are added. An important special case is the following LKP:

$$(\mathcal{P}_\ell) : \min\{P_\ell(\mathbf{z}) = -\mathbf{c}^T\mathbf{z}|\mathbf{v}^T\mathbf{z} \leq V_c, \mathbf{z} \in \{0,1\}^n\}. \quad (6)$$

A 1998 study of the Stony Brook University Algorithm Repository showed that, out of 75 algorithmic problems, the knapsack problem was the 19th most popular and the 3rd most needed after suffix trees and the bin packing problem. It is well-known that the knapsack problem is NP-hard and there is no known algorithm that can solve the problem in polynomial time. Actually, even the LKP is listed as one of Karp [37] 21 NP-complete problems. In this paper, the canonical duality method will be addressed for solving this challenging problem.

For a given design variable $\mathbf{z}_{k-1}$, the lower-level minimization (4) is the general nonlinear/nonconvex optimization problem proposed by the author for multiscale systems [22], [26].

Remark 1 (Objectivity, Modeling, and Canonical Duality): Objectivity is a central concept in our daily life, related to reality and truth. 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. In continuum physics, a real-valued function $G : \mathcal{G}_a \to \mathbb{R}$ is called objective if and only if it is an invariant under orthonormal transformation (see [17, Definition 6.1.2]), i.e.,

$$G(\mathbf{R}\,\mathbf{g}) = G(\mathbf{g}) \quad \forall \mathbf{g} \in \mathcal{G}_a \quad \forall \mathbf{R} \in SO(p) \quad (7)$$

where $SO(p) = \{\mathbf{R} \in \mathbb{R}^{p \times p}|\mathbf{R}^T = \mathbf{R}^T, \det \mathbf{R} = 1\}$ is a special orthogonal group in $\mathbb{R}^p$. Physically speaking, an objective function is governed by the intrinsic physical law of the system, which does not depend on observers. Therefore, the objectivity is essential for any real-world mathematical models. In continuum physics, it is also called the principle of material frame indifference [45]. Geometrically, an objective function does not depend on rigid rotation of the system considered, but only on certain measure of its variable. In the Euclidean space $\mathcal{G}_a \subset \mathbb{R}^p$, the simplest objective function is the ℓ_2-norm $\|\mathbf{g}\|$ since $\|\mathbf{R}\,\mathbf{g}\|^2 = \mathbf{g}^T\mathbf{R}^T\mathbf{R}\,\mathbf{g} = \|\mathbf{g}\|^2 \ \forall\mathbf{R} \in SO(p)$. Therefore, an objective function must be nonlinear.

Correspondingly, the linear term $F(\mathbf{x}) = \mathbf{x}^T\mathbf{f}$ in the lower-level problem (2) is called a *subjective function* [22], [25],

where the given input $\mathbf{f}$ and the constraints in $\mathcal{X}_a$ depend only on each given problem. Thus, if the operator "$-$" is considered as the predicate, the difference $\Pi(\mathbf{x}) = G(D\mathbf{x}) - F(\mathbf{x})$ between object and subject forms a complete and grammatically correct mathematical formulation. In continuum physics, if $G(\mathbf{g})$ is the free (or internal) energy and $F(\mathbf{x})$ is the input (or external) energy, then $\Pi(\mathbf{x})$ is the total potential energy and the minimum total potential energy principle leads to a general variational/optimization problem in mathematical physics [33]. By the fact that $\mathcal{G}_a \subset \mathbb{R}^p$ and $\mathcal{X}_a \subset \mathbb{R}^m$ can be in different dimensional spaces with different measures, the lower-level minimization $\min \Pi(\mathbf{x})$ presented in this paper covers general constrained nonconvex/nonsmooth optimization problems in multiscale complex systems [17], [26], [27].

According to [10], $G(\mathbf{g})$ is an objective function if and only if there exists an objective measure $\epsilon = \mathbf{g}^T\mathbf{g}$ and a real-valued function $\Phi(\epsilon)$ such that $G(\mathbf{g}) = \Phi(\mathbf{g}^T\mathbf{g})$. In continuum physics and differential geometry, if $\mathbf{g} = D\mathbf{x}$ is the deformation gradient, then the objective measure $\epsilon = \mathbf{g}^T\mathbf{g}$ is the well-known right Cauchy–Green tensor. By the fact that the free energy $\Phi(\epsilon)$ is usually convex (say the St Venant–Kirchhoff material [17]), the duality relation $\epsilon^* = \nabla\Phi(\epsilon)$ is invertible. This one-to-one constitutive relation is called the canonical duality and $\Phi(\epsilon)$ is called the canonical function. These basic truths in continuum physics laid a foundation for the CDT, in which, the key idea of the *canonical transformation* is to choose a nonlinear operator (not necessary to be objective) $\epsilon = \Lambda(\mathbf{x})$ and a canonical function $\Phi(\epsilon)$ such that the given nonconvex function $G(D\mathbf{x})$ can be written in the canonical form $G(D\mathbf{x}) = \Phi(\Lambda(\mathbf{x}))$. By the fact that the objectivity plays a fundamental role in mathematical modeling, the CDT can be powerfully used for solving many challenging problems in multidisciplinary fields [27].

However, this important concept of the objectivity has been extensively misused in optimization literature such that the general problem in nonlinear optimization (or programming) has been formulated as

$$\min f(x) \text{ s.t. } g(x) = 0, h(x) \leq 0 \quad (8)$$

where $f(x)$ is called the "objective function,"[1] which is allowed to be any arbitrarily given function, even a linear function. Clearly, this mathematical problem is too abstract. 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 artificially given constrained problem. Also, due to this conceptual mistake, the CDT has been mistakenly challenged in several publications.[2] The conceptual mistake in these false challenges revealed a big gap between physics and optimization. Interested readers are recommended to read the recent papers [23], [29] for details.

---

[1]This terminology is used mainly in English literature. The function $f(x)$ is correctly called the target function in all Chinese and Japanese literature, or the goal function in some Russian and German literature by serious researchers.

[2]See M. D. Voisei and C. Zalinescu, "Some remarks concerning Gao-Strangs complementary gap function," *Appl. Anal.*, vol. 90, pp. 1111–1121, 2011.

The CDT for solving the lower-level nonconvex continuous optimization problem (4) has been studied extensively during the past 15 years (see [18], [21], and [27]). As long as the nonconvex function $G(\mathbf{Dx}, \mathbf{z})$ in $\Pi(\mathbf{x}, \mathbf{z})$ can be written in the canonical form $\Phi(\Lambda(\mathbf{x}), \mathbf{z})$, this problem can be solved easily by the CDT to obtain both global and local optimal solutions (see many real-world applications in [27]). This paper will show how to use the CDT for solving the challenging upper-level knapsack problem (5).

III. CANONICAL DUAL SOLUTION TO UPPER-LEVEL KNAPSACK PROBLEM

For a fixed $\mathbf{x} \in \mathcal{X}_a$, the upper-level optimization (5) can be written in the typical QKP

$$(\mathcal{P}_q) : \min_{\mathbf{z} \in \{0,1\}^n} \left\{ P_q(\mathbf{z}) = \frac{1}{2} \mathbf{z}^T \mathbf{Q} \mathbf{z} - \mathbf{c}^T \mathbf{z} | \mathbf{v}^T \mathbf{z} \leq V_c \right\}. \quad (9)$$

According to the CDT for mathematical modeling [23], [24], the inequality $\mathbf{z} \cdot \mathbf{v} \leq V_c$ is a geometrical constraint, while the integer constraint $\mathbf{z} \in \{0, 1\}^n$ in $(\mathcal{P}_q)$ is a constitutive condition [25], [26], which can be equivalently replaced by the canonical constraint $\mathbf{z}^2 - \mathbf{z} = \mathbf{0} \in \mathbb{R}^n \ \forall \mathbf{z} \in \mathbb{R}^n$, where $\mathbf{z}^2 = \mathbf{z} \circ \mathbf{z} = \{z_i^2\} \in \mathbb{R}^n$. Therefore, by introducing a penalty parameter $\beta > 0$ and let $\mathcal{Z}_b = \{\mathbf{z} \in \mathbb{R}^n | \mathbf{v}^T \mathbf{z} \leq V_c\}$, the problem $(\mathcal{P}_q)$ can be relaxed by the canonical penalty function

$$(\mathcal{P}_\beta) : \min_{\mathbf{z} \in \mathcal{Z}_b} \left\{ P_\beta(\mathbf{z}) = \frac{1}{2} \mathbf{z}^T \mathbf{Q} \mathbf{z} + \frac{1}{2} \beta \|\mathbf{z}^2 - \mathbf{z}\|^2 - \mathbf{c}^T \mathbf{z} \right\}. \quad (10)$$

Clearly, we have

$$\left\{ P_q(\mathbf{z}) | \mathbf{z} \in \mathcal{Z}_a \right\} = \lim_{\beta \to \infty} \left\{ P_\beta(\mathbf{z}) | \mathbf{z} \in \mathcal{Z}_b \right\}. \quad (11)$$

Although the integer constraint is relaxed by the penalty function

$$\Psi_\beta(\mathbf{z}) = \frac{1}{2} \beta \|\mathbf{z}^2 - \mathbf{z}\|^2 \quad (12)$$

the problem (10) is a nonconvex minimization in $\mathbb{R}^n$, which is still considered to be NP-hard by traditional methods. This is the reason why the traditional external penalty method can be used mainly for linear constrained problems [40]. To solve this problem by the CDT, we need first to introduce the following canonical transformation:

$$\Psi_\beta(\mathbf{z}) = \Phi_\beta(\Lambda(\mathbf{z}))$$

$$\Phi_\beta(\boldsymbol{\varepsilon}) = \frac{1}{2} \beta \|\boldsymbol{\varepsilon}\|^2, \ \boldsymbol{\varepsilon} = \Lambda(\mathbf{z}) = \mathbf{z}^2 - \mathbf{z} \in \mathbb{R}^n.$$

Clearly, $\Phi_\beta : \mathbb{R}^n \to \mathbb{R}$ is a convex quadratic function. Its canonical dual can be simply given by the Legendre transformation

$$\Phi_\beta^*(\boldsymbol{\sigma}) = \max_{\boldsymbol{\varepsilon} \in \mathcal{E}} \left\{ \boldsymbol{\varepsilon}^T \boldsymbol{\sigma} - \Phi_\beta(\boldsymbol{\varepsilon}) \right\} = \frac{1}{2} \beta^{-1} \|\boldsymbol{\sigma}\|^2. \quad (13)$$

Thus, replacing $\Psi_\beta(\mathbf{z})$ in $P_\beta(\mathbf{z})$ by the Fenchel–Young equality $\Phi_\beta(\Lambda(\mathbf{z})) = \Lambda(\mathbf{z})^T \boldsymbol{\sigma} - \Phi_\beta^*(\boldsymbol{\sigma})$, and introducing a Lagrange multiplier $\tau \geq 0$, the canonical penalty function $P_\beta(\mathbf{z})$ can be reformed as the Gao–Strang total complementary function

$$\Xi_\beta(\mathbf{z}, \boldsymbol{\sigma}, \tau) = \frac{1}{2} \mathbf{z}^T \mathbf{G}(\boldsymbol{\sigma}) \mathbf{z} - \frac{1}{2} \beta^{-1} \|\boldsymbol{\sigma}\|^2 - \mathbf{z}^T \boldsymbol{\psi}(\boldsymbol{\sigma}, \tau) - \tau V_c \quad (14)$$

where

$$\mathbf{G}(\boldsymbol{\sigma}) = \mathbf{Q} + 2\text{Diag}(\boldsymbol{\sigma}), \ \boldsymbol{\psi}(\boldsymbol{\sigma}, \tau) = \mathbf{c} - \tau \mathbf{v} + \boldsymbol{\sigma}. \quad (15)$$

By introducing a canonical dual feasible space

$$\mathcal{S}_a^+ = \left\{ \boldsymbol{\zeta} = (\boldsymbol{\sigma}, \tau) \in \mathbb{R}^{n+1} | \mathbf{G}(\boldsymbol{\sigma}) \succ 0, \tau > 0 \right\} \quad (16)$$

the function $\Xi_\beta(\mathbf{z}, \boldsymbol{\zeta})$ is convex in $\mathbf{z} \in \mathbb{R}^n$ for any given $\boldsymbol{\zeta} \in \mathcal{S}_a^+$. The canonical penalty-duality function $P_\beta^d : \mathcal{S}_a^+ \to \mathbb{R}$ can be defined by

$$P_\beta^d(\boldsymbol{\zeta}) = \min_{\mathbf{z} \in \mathbb{R}^n} \left\{ \Xi_\beta(\mathbf{z}, \boldsymbol{\zeta}) | \boldsymbol{\zeta} \in \mathcal{S}_a^+ \right\}$$

$$= -\frac{1}{2} \boldsymbol{\psi}(\boldsymbol{\zeta})^T \mathbf{G}(\boldsymbol{\sigma})^{-1} \boldsymbol{\psi}(\boldsymbol{\zeta}) - \frac{1}{2} \beta^{-1} \|\boldsymbol{\sigma}\|^2 - \tau V_c.$$

Thus, the canonical penalty-duality problem can be proposed in the following:

$$(\mathcal{P}_\beta^d) : \max \left\{ P_\beta^d(\boldsymbol{\zeta}) | \boldsymbol{\zeta} \in \mathcal{S}_a^+ \right\}. \quad (17)$$

Theorem 1 (Complementary-Dual Principle): For any given $\beta > 0$, if $(\mathbf{z}_\beta, \boldsymbol{\zeta}_\beta) \in \mathbb{R}^n \times \mathcal{S}_a^+$ is a KKT (Karush–Kuhn–Tucker) point of $\Xi_\beta(\mathbf{z}, \boldsymbol{\zeta})$, then $\mathbf{z}_\beta$ is a global minimum solution to the canonical penalty problem $(\mathcal{P}_\beta)$, $\boldsymbol{\zeta}_\beta$ is a solution to $(\mathcal{P}_\beta^d)$, and

$$P_\beta(\mathbf{z}_\beta) = \min_{\mathbf{z} \in \mathcal{Z}_b} P_\beta(\mathbf{z}) = \Xi_\beta(\bar{\mathbf{z}}, \bar{\boldsymbol{\xi}}) = \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} P_\beta^d(\boldsymbol{\zeta}) = P_\beta^d(\boldsymbol{\zeta}_\beta). \quad (18)$$

Proof: By the fact that $\Xi_\beta : \mathbb{R}^n \times \mathcal{S}_a^+ \to \mathbb{R}$ is a saddle function, if $(\mathbf{z}_\beta, \boldsymbol{\zeta}_\beta) \in \mathbb{R}^n \times \mathcal{S}_a^+$ is a KKT point of $\Xi_\beta(\mathbf{z}, \boldsymbol{\zeta})$, it must be a saddle point of $\Xi_\beta(\mathbf{z}, \boldsymbol{\zeta})$. Then by the definition of $P_\beta^d(\boldsymbol{\zeta})$, we have

$$\Xi_\beta(\mathbf{z}_\beta, \boldsymbol{\zeta}_\beta) = \min_{\mathbf{z} \in \mathbb{R}^n} \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \Xi(\mathbf{z}, \boldsymbol{\zeta}) = \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \min_{\mathbf{z} \in \mathbb{R}^n} \Xi(\mathbf{z}, \boldsymbol{\zeta})$$

$$= \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} P_\beta^d(\boldsymbol{\zeta}) = P_\beta^d(\boldsymbol{\zeta}_\beta). \quad (19)$$

By the KKT conditions

$$\mathbf{v}^T \mathbf{z}_\beta \leq V_c, \tau_\beta \geq 0, \tau_\beta \left(\mathbf{v}^T \mathbf{z}_\beta - V_c \right) = 0$$

and the condition $\boldsymbol{\zeta}_\beta = (\boldsymbol{\sigma}_\beta, \tau_\beta) \in \mathcal{S}_a^+$, we have $\tau_\beta > 0$ and $\mathbf{v}^T \mathbf{z}_\beta - V_c = 0$. By the convexity of the penalty function $\Psi_\beta(\boldsymbol{\epsilon})$, we have $\Psi_\beta^{**}(\boldsymbol{\epsilon}) = \Psi_\beta(\boldsymbol{\epsilon})$. Thus, for any given $\beta > 0$

$$P_\beta(\mathbf{z}_\beta) = \min_{\mathbf{z} \in \mathcal{Z}_b} P_\beta(\mathbf{z}) = \min_{\mathbf{z} \in \mathcal{Z}_b} \max_{\boldsymbol{\zeta} \in \mathcal{S}_a^+} \Xi_\beta(\mathbf{z}, \boldsymbol{\zeta}) = \Xi_\beta(\mathbf{z}_\beta, \boldsymbol{\zeta}_\beta). \quad (20)$$

Thus, $\mathbf{z}_\beta$ is a global optimal solution to $(\mathcal{P}_\beta)$. ∎

Theorem 1 shows that the nonconvex minimization problem $(\mathcal{P}_\beta)$ is canonically dual to a concave maximization problem $(\mathcal{P}_\beta^d)$ in a convex space $\mathcal{S}_a^+$, which can be solved easily by well-developed convex minimization methods.

Remark 2 (Canonical Penalty-Duality and β-Perturbation): By the facts that

$$\Phi(\boldsymbol{\varepsilon}) = \lim_{\beta \to \infty} \Phi_\beta(\boldsymbol{\varepsilon}) = \begin{cases} 0 & \text{if } \boldsymbol{\varepsilon} = \mathbf{0} \in \mathbb{R}^n \\ +\infty & \text{otherwise} \end{cases} \quad (21)$$

and

$$\Phi^*(\boldsymbol{\sigma}) = \sup_{\boldsymbol{\varepsilon} \in \mathbb{R}^n} \{ \boldsymbol{\varepsilon}^T \boldsymbol{\sigma} - \Phi(\boldsymbol{\varepsilon}) \} = 0 \ \ \forall \boldsymbol{\sigma} \in \mathbb{R}^n \quad (22)$$

we have

$$P_q^d(\boldsymbol{\zeta}) = \lim_{\beta \to +\infty} P_\beta^d(\boldsymbol{\zeta})$$

$$= -\frac{1}{2} \boldsymbol{\psi}(\boldsymbol{\zeta})^T \mathbf{G}(\boldsymbol{\sigma})^{-1} \boldsymbol{\psi}(\boldsymbol{\zeta}) - \tau V_c - \Phi^*(\boldsymbol{\sigma})$$

which is exactly the canonical dual function to the primal function $P_q(\mathbf{z})$. Therefore, $(\mathcal{P}_\beta^d)$ is actually the so-called β-perturbation of the canonical dual problem $(\mathcal{P}_q^d)$

$$(\mathcal{P}_q^d) : \max\left\{P_q^d(\boldsymbol{\zeta})|\boldsymbol{\zeta} \in \mathcal{S}_a^+\right\}. \qquad (23)$$

The penalty-duality method was first proposed by Gao [15], [16] for solving convex variational problems. The β-perturbation method for nonconvex integer constrained problems was first proposed in [24] and [28]. It was proved by [28, Th. 7] that there exists a $\beta_c > 0$ such that for any given $\beta \geq \beta_c$, both β-perturbed canonical dual problem $(\mathcal{P}_\beta^d)$ and problem $(\mathcal{P}_q^d)$ have the same solution set. This shows the relation between the canonical penalty-duality method proposed in this paper and the β-perturbation method proposed in [28]. ∎

Theorem 2 (Analytic Solution to Knapsack Problem): For any given $\beta > 0$, if $\boldsymbol{\zeta}_\beta = (\boldsymbol{\sigma}_\beta, \tau_\beta) \in \mathcal{S}_a^+$ is a solution to $(\mathcal{P}_\beta^d)$, then the vector defined by

$$\mathbf{z}_\beta = \mathbf{G}(\boldsymbol{\sigma}_\beta)^{-1}\boldsymbol{\psi}(\boldsymbol{\zeta}_\beta) \qquad (24)$$

is a global minimum solution to the canonical penalty problem $(\mathcal{P}_\beta)$.

Moreover, there exists $\beta_c \gg 0$ such that $\beta \geq \beta_c$ and $\mathbf{z}_\beta \in \mathcal{Z}_a$, then $\mathbf{z}_\beta$ is a global optimal solution to the knapsack problem $(\mathcal{P}_q)$ and

$$P_q(\mathbf{z}_\beta) = \min_{\mathbf{z}\in\mathcal{Z}_b} P_\beta(\mathbf{z}) = \Xi_\beta(\bar{\mathbf{z}}, \bar{\boldsymbol{\zeta}}) = \max_{\boldsymbol{\zeta}\in\mathcal{S}_a^+} P_\beta^d(\boldsymbol{\zeta}) = P_q^d(\boldsymbol{\zeta}_\beta). \qquad (25)$$

Proof: It is easy to verify that if $(\mathbf{z}_\beta, \boldsymbol{\zeta}_\beta)$ is a KKT point of $\Xi_\beta(\mathbf{z}, \boldsymbol{\zeta})$, the criticality condition $\nabla_{\mathbf{z}}\Xi_\beta(\mathbf{z}_\beta, \boldsymbol{\zeta}_\beta) = 0$ leads to $\mathbf{z}_\beta = \mathbf{G}(\boldsymbol{\sigma}_\beta)^{-1}\boldsymbol{\psi}(\boldsymbol{\zeta}_\beta)$. By Theorem 1 we know that if $\boldsymbol{\zeta}_\beta = (\boldsymbol{\sigma}_\beta, \tau_\beta) \in \mathcal{S}_a^+$ is a solution to $(\mathcal{P}_\beta^d)$, the vector defined by (24) must be a solution to $(\mathcal{P}_\beta)$. Since the penalty function $\Phi_\beta(\mathbf{z}) \geq 0 \; \forall \mathbf{z} \in \mathbb{R}^n$, there must exists a sufficiently big $\beta_c \gg 0$ such that $\Phi_\beta(\mathbf{z}_\beta) = 0 \; \forall \beta \geq \beta_c$, i.e., $\mathbf{z}_\beta \in \mathcal{Z}_a$. Thus, $P_q(\boldsymbol{\zeta}_\beta) = P_\beta(\mathbf{z}_\beta) = \min\{P_q(\mathbf{z})|\mathbf{z} \in \mathcal{Z}_a\}$. ∎

Since $\mathcal{S}_a^+$ is a convex open set, its boundary is

$$\partial\mathcal{S}_a^+ = \left\{\boldsymbol{\zeta} = (\boldsymbol{\sigma}, \tau) \in \mathbb{R}^{n+1}| \det \mathbf{G}(\boldsymbol{\sigma}) = 0, \tau = 0\right\}. \qquad (26)$$

The following theorem is important for understanding the NP-hardness of the knapsack problem.

Theorem 3 [Existence and Uniqueness for $(\mathcal{P}_q)$]: For any given $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}_+^n$, and $\mathbf{Q} \in \mathbb{R}^{n\times n}$, if $\mathcal{S}_a^+$ has a nonempty interior $\boldsymbol{\zeta}_o \in \mathcal{S}_a^+$ such that $P_q^d(\boldsymbol{\zeta}_o) > -\infty$ and

$$\lim_{\boldsymbol{\zeta}\to\partial\mathcal{S}_a^+} P_q^d(\boldsymbol{\zeta}) = -\infty \; \forall \boldsymbol{\zeta} \in \mathcal{S}_a^+ \qquad (27)$$

then the problem $(\mathcal{P}_q^d)$ has a unique solution $\boldsymbol{\zeta}_c \in \mathcal{S}_a^+$ and $\mathbf{z}_c = \mathbf{G}(\boldsymbol{\sigma}_c)^{-1}\boldsymbol{\psi}(\boldsymbol{\zeta}_c)$ is a unique global optimal solution to $(\mathcal{P}_q)$.

Proof: This theorem is a direct application of the CDT (see [21, Th. 4].)[3] By the facts that $P_q^d(\boldsymbol{\zeta})$ is concave and its feasible set $\mathcal{S}_a^+$ is convex and nonempty, if the condition (27) holds, then the canonical penalty-dual function $P_q^d:\mathcal{S}_a^+ \to \mathbb{R}$

---

[3]The condition $P^d(\boldsymbol{\sigma}_o) < -\infty$ in [21] is a typo. It should be $P^d(\boldsymbol{\sigma}_o) > -\infty$.

is strictly concave, bounded up and $-P_q^d(\boldsymbol{\zeta})$ is coercive on $\mathcal{S}_a^+$. Therefore, $(\mathcal{P}_q^d)$ has a unique solution $\boldsymbol{\zeta}_c$. By Theorem 2, we know that the associated $\mathbf{z}_c = \mathbf{G}(\boldsymbol{\sigma}_c)^{-1}\boldsymbol{\psi}(\boldsymbol{\zeta}_c)$ is a unique global optimal solution to $(\mathcal{P}_q)$. ∎

As an important application, let us recall the LKP

$$(\mathcal{P}_\ell) : \min_{\mathbf{z}\in\mathcal{Z}_a}\left\{P_\ell(\mathbf{z}) = -\mathbf{c}^T\mathbf{z}\right\}. \qquad (28)$$

The associated canonical penalty problem is

$$(\mathcal{P}_{\ell\beta}) : \min_{\mathbf{z}\in\mathcal{Z}_b}\left\{P_{\ell\beta}(\mathbf{z}) = \frac{1}{2}\beta\|\mathbf{z}^2 - \mathbf{z}\|^2 - \mathbf{c}^T\mathbf{z}\right\}. \qquad (29)$$

Since the matrix $\mathbf{Q}$ vanished in this case, the canonical penalty-dual function has the following simple form:

$$P_{\ell\beta}^d(\boldsymbol{\zeta}) = -\frac{1}{2}\sum_{i=1}^n\left(\frac{1}{2}\sigma_i^{-1}(\sigma_i + c_i - \tau v_i)^2 + \beta^{-1}\sigma_i^2\right) - \tau V_c. \qquad (30)$$

Let $\boldsymbol{\sigma}^{-1} = \text{Diag}(\boldsymbol{\sigma})^{-1} = \{\sigma_i^{-1}\}$ and

$$\mathcal{S}_a^+ = \left\{\boldsymbol{\zeta} = (\boldsymbol{\sigma}, \tau) \in \mathbb{R}^{n+1}|\boldsymbol{\sigma} > \mathbf{0}, \tau > 0\right\}.$$

The canonical penalty-dual problem for the LKP is

$$\left(\mathcal{P}_{\ell\beta}^d\right) : \max\left\{P_{\ell\beta}^d(\boldsymbol{\zeta})|\boldsymbol{\zeta} \in \mathcal{S}_a^+\right\}. \qquad (31)$$

Corollary 1 (Analytic Solution to Knapsack Problem [25]): For any given $\mathbf{c}, \mathbf{v} \in \mathbb{R}_+^n$ and $V_c, \beta > 0$, if $\boldsymbol{\zeta}_\beta = (\boldsymbol{\sigma}_\beta, \tau_\beta) \in \mathcal{S}_a^+$ is a solution to $(\mathcal{P}_{\ell\beta}^d)$ and $\tau_\beta\mathbf{v} - \mathbf{c} \neq \mathbf{0}$, then the penalty problem $(\mathcal{P}_{\ell\beta})$ has a unique global minimum solution which is given analytically by

$$\mathbf{z}_\beta = \frac{1}{2}\boldsymbol{\sigma}_\beta^{-1} \circ (\boldsymbol{\sigma}_\beta + \mathbf{c} - \tau_\beta\mathbf{v}). \qquad (32)$$

Moreover, there exists a $\beta_c \gg 0$ such that $\beta \geq \beta_c$ and $\mathbf{z}_\beta \in \mathcal{Z}_a$, then $\mathbf{z}_\beta$ is a global optimal solution to the LKP $(\mathcal{P}_\ell)$.

This special result for LKP was obtained recently in [25]. For $\beta = 0$, the canonical dual problem $(\mathcal{P}_\ell^d)$ of $(\mathcal{P}_\ell)$ is

$$\max_{\boldsymbol{\zeta}\in\mathcal{S}_a^+}\left\{P_\ell^d(\boldsymbol{\zeta}) = -\frac{1}{2}\sum_{i=1}^n\left(\frac{1}{2}\sigma_i^{-1}(\sigma_i + c_i - \tau v_i)^2\right) - \tau V_c\right\}. \qquad (33)$$

It is easy to prove that if $\tau_\beta\mathbf{v} - \mathbf{c} \neq \mathbf{0}$, the condition (27) is reduced by

$$\lim_{\boldsymbol{\sigma}\to\mathbf{0}^+} P_\ell^d(\boldsymbol{\sigma}, \tau_\beta) = -\infty. \qquad (34)$$

Thus, by Theorem 3 we know that the vector (32) is a unique solution to the problem $(\mathcal{P}_{\ell\beta})$ and it is a solution to the LKP $(\mathcal{P}_\ell)$ if $\beta \geq \beta_c \gg 0$ such that $\mathbf{z}_\beta \in \mathcal{Z}_a$.

Actually, for any given $\beta > 0$, the criticality condition $\nabla P_{\ell\beta}^d = 0$ leads to the following algebraic equations:

$$\beta^{-1}\sigma_i^3 + \sigma_i^2 = (\tau v_i - c_i)^2, i = 1, \ldots, n \qquad (35)$$

$$\sum_{i=1}^n \frac{1}{2}\frac{v_i}{\sigma_i}(\sigma_i - v_i\tau + c_i) - V_c = 0. \qquad (36)$$

It was proved in [17] and [20] that for any given $\beta > 0$, $\tau \geq 0$ and $\mathbf{c} \in \mathbb{R}^n$ such that $\theta_i = \tau v_i - c_i \neq 0 \; \forall i = 1, \ldots, n$, the

canonical dual algebraic equation (35) has a unique positive real solution

$$\sigma_i = \frac{1}{6}\beta\big[-1 + \phi_i(\tau) + \phi_i^c(\tau)\big] > 0, i = 1, \ldots, n \quad (37)$$

where

$$\phi_i(\varsigma) = \eta^{-1/3}\left[2\theta_i^2 - \eta + 2\theta_i\sqrt{\theta_i^2 - \eta}\right]^{1/3}, \eta = \frac{4\beta^2}{27}$$

and ϕ_i^c is the complex conjugate of ϕ_i, i.e., $\phi_i\phi_i^c = 1$.

On the other hand, for a given $\boldsymbol{\sigma} \in \mathbb{R}_+^n$, the Lagrange multiplier τ can be uniquely obtained by

$$\tau = \frac{\sum_{i=1}^n v_i(1 + c_i/\sigma_i) - 2V_c}{\sum_{i=1}^n v_i^2/\sigma_i}. \quad (38)$$

It is easy to prove that for any given $\beta > 0$ and $\tau_{k-1} > 0$, the solution $\boldsymbol{\sigma}_k$ produced by (37) satisfies $\boldsymbol{\sigma}_k < \mathbf{c}$. Thus, for a given initial $\tau_0 > 0$, an alternative iteration can be used for solving (37) and (38) and the sequence $(\boldsymbol{\sigma}_k, \tau_k) \in \mathcal{S}_a^+$ approaches to the global optimal solution of $(\mathcal{P}_\ell^d)$ in polynomial time.

Theorem 1 shows that although the canonical dual problem is a concave maximization in continuous space, it produces the analytical solution (24) to the well-known integer Knapsack problem $(\mathcal{P}_q)$. This truth was first discovered by Gao in 2007 for general quadratic integer programming problems (see [20, Th. 3]).

IV. IMPROVED SOLUTIONS AND NP-HARDNESS

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 $\mathbf{Q}_\alpha = \mathbf{Q} + 2\mathrm{Diag}(\boldsymbol{\alpha}) \succeq 0$. 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 [21]

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

Since $\mathrm{rank}\,\mathbf{Q}_\alpha = r \le n$, there must exist (see [52]) a $\mathbf{L} \in \mathbb{R}^{r \times n}$ and $\mathbf{H} \in \mathbb{R}^{r \times r}$ with $\mathrm{rank}\,\mathbf{L} = \mathrm{rank}\,\mathbf{H} = r$ and $\mathbf{H} \succ 0$ such that $\mathbf{Q}_\alpha = 4\mathbf{L}^T\mathbf{HL}$. Similar to the α-perturbed canonical dual problem $(\mathcal{P}_{ip}^g)$ given in [21], the canonical dual function in $(\mathcal{P}_q^d)$ can be reformulated as [26]

$$P_\alpha^g(\boldsymbol{\sigma}, \tau) = -\frac{1}{2}\mathrm{Abs}[\boldsymbol{\phi}(\boldsymbol{\sigma}, \tau)] - \frac{1}{2}\boldsymbol{\sigma}^T\mathbf{H}^{-1}\boldsymbol{\sigma} - \tau V_b + d$$

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

$$\boldsymbol{\phi}(\boldsymbol{\zeta}) = \mathbf{c} - \tau\mathbf{v} - 2\mathbf{L}^T\boldsymbol{\sigma} - \frac{1}{2}\mathbf{Qe}, \mathbf{e} = \{1\}^n. \quad (39)$$

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

$$\mathcal{S}_c^+ = \left\{\boldsymbol{\zeta} = (\boldsymbol{\sigma}, \tau) \in \mathbb{R}^{r+1}|\tau > 0\right\}.$$

Then the improved canonical dual problem to $(\mathcal{P}_\alpha)$ can be proposed as

$$(\mathcal{P}_\alpha^g) : \quad \max\left\{P_\alpha^g(\boldsymbol{\sigma}, \tau)|\boldsymbol{\zeta} \in \mathcal{S}_c^+\right\}. \quad (40)$$

In many real-world applications, we have $r \ll n$, thus the problem $(\mathcal{P}_\alpha^g)$ is much easier than $(\mathcal{P}_q^d)$. Similar to [21, Th. 7], we have the following improved result.

Theorem 4 [Improved Analytic Solution to $(\mathcal{P}_q)$]: For any given $V_c > 0$, $\mathbf{v}, \mathbf{c} \in \mathbb{R}_+^n$, $\mathbf{Q} \in \mathbb{R}^{n \times n}$ and an $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $\mathbf{Q}_\alpha = \mathbf{Q} + 2\mathrm{Diag}(\boldsymbol{\alpha}) = 4\mathbf{L}^T\mathbf{HL}$, and $\mathbf{H} \succ 0$, if $\boldsymbol{\zeta}_c = \{\boldsymbol{\sigma}_c, \tau_c\}$ is a solution to $(\mathcal{P}_\alpha^g)$ and

$$\phi_i(\boldsymbol{\zeta}_c) \neq 0 \; \forall i \in \{1, \ldots, n\} \quad (41)$$

then the QKP $(\mathcal{P}_q)$ has a unique global optimal solution

$$\mathbf{z}_c = \frac{1}{2}\left\{\frac{\phi_i(\boldsymbol{\zeta}_c)}{|\phi_i(\boldsymbol{\zeta}_c)|} + 1\right\}^n \quad (42)$$

and

$$P_q(\mathbf{z}_c) = \min_{\mathbf{z} \in \mathcal{Z}_a} P_q(\mathbf{z}) = \max_{\boldsymbol{\zeta} \in \mathcal{S}_c^+} P_\alpha^g(\boldsymbol{\zeta}) = P_\alpha^g(\boldsymbol{\zeta}_c). \quad (43)$$

Otherwise, if $\phi_i(\boldsymbol{\zeta}_c) = 0$ for at least one $i \in \{1, \ldots, n\}$, then $(\mathcal{P}_q)$ has at least two solutions.

It is easy to prove that $P_\alpha^g(\boldsymbol{\zeta})$ is strictly concave on the convex feasible set $\mathcal{S}_c^+$. If $\boldsymbol{\zeta}_c \in \mathcal{S}_c^+$ is a solution to $(\mathcal{P}_\alpha^g)$, this solution must be unique. Since $\dim \mathcal{S}_c^+ = r + 1 \le n + 1$, the problem $(\mathcal{P}_\alpha^g)$ is easier than $(\mathcal{P}_\beta^g)$. In the case that $(\mathcal{P}_\alpha^g)$ does not have a solution to satisfy (41), we can oppositely chose $\boldsymbol{\alpha} \in \mathbb{R}^n$ such that $\mathbf{H} \prec 0$. But in this case, $P_\alpha^g(\boldsymbol{\zeta})$ is a difference of convex (d.c.) function and the corresponding problem should be [21], [34]

$$(\mathcal{P}_\alpha^g) : \quad \min \mathrm{sta}\left\{P_\alpha^g(\boldsymbol{\zeta})|\boldsymbol{\zeta} \in \mathcal{S}_c^+\right\} \quad (44)$$

where $\min \mathrm{sta}\{f(x)\}$ means to find the minimum stationary point of $f(x)$. This is a nonsmooth d.c. programming problem. A so-called VTDIRECT parallel algorithm has been used successfully by Gao *et al.* [34] for finding global optimum solutions to quadratic integer programming problems, but not in polynomial time.

For the LKP, the improved canonical dual has a very simple form

$$(\mathcal{P}_\ell^g) : \quad \max_{\tau \ge 0}\left\{P_\ell^g(\tau) = -\frac{1}{2}\sum_{i=1}^n(|c_i - \tau v_i| - \tau v_i) - \tau V_c\right\}. \quad (45)$$

Corollary 2 [Improved Analytic Solution to $(\mathcal{P}_\ell)$]: For any given $V_c > 0$, $\mathbf{v}, \mathbf{c} \in \mathbb{R}_+^n$, if $\tau_c > 0$ is a solution to $(\mathcal{P}_\ell^g)$ and $\theta_i = \tau_c v_i - c_i \neq 0 \; \forall i \in \{1, \ldots, n\}$, then the LKP $(\mathcal{P}_\ell)$ has a unique global optimal solution

$$\mathbf{z}_c = \frac{1}{2}\left\{\frac{c_i - \tau_c v_i}{|c_i - \tau_c v_i|} + 1\right\}^n \quad (46)$$

and $P_\ell(\mathbf{z}_c) = P_\ell^g(\tau_c)$.

Actually, from the proof of Corollary 1 we know that if there exists a $\tau_\beta > 0$ such that $\theta_i = \tau_\beta v_i - c_i \neq 0 \; \forall i \in \{1, \ldots, n\}$, the canonical dual algebraic equation (35) has a unique positive solution $\boldsymbol{\sigma}_\beta$ [17], [19], [20]. For a sufficiently big $\beta \gg 0$ we have $\boldsymbol{\sigma}_\beta = \boldsymbol{\sigma}_c$ and the solution τ_β by (38) is exactly the solution τ_c of the problem $(\mathcal{P}_{\ell\alpha}^d)$ (see [25]). Therefore, $(\mathcal{P}_\ell)$ has a unique solution defined by either (24) or (46).

Remark 3 (Criteria for Nonuniqueness and NP-Hardness): Theoretical results presented so far show that if the condition (27) holds, the canonical dual problem $(\mathcal{P}_q^d)$ has a unique solution $\boldsymbol{\zeta}_c = (\boldsymbol{\sigma}_c, \tau_c) \in \mathcal{S}_a^+$, which can be obtained deterministically in polynomial time since $(\mathcal{P}_q^d)$ is equivalent to a convex minimization problem. In this case, both QKP and LKP are not NP-hard and their solutions $\mathbf{z}_c$ can be analytically given by Theorem 2 and Corollary 1, respectively. Also, we must have $\mathbf{z}_c^T \mathbf{v} = V_c$ since $\tau_c > 0$. For the LKP $(\mathcal{P}_\ell)$, the condition (27) is simply $\tau_c \mathbf{v} \neq \mathbf{c}$.

On the other hand, if the canonical dual problem $(\mathcal{P}_q^d)$ has no solution in $\mathcal{S}_a^+$, the primal problem $(\mathcal{P}_q)$ could be NP-hard, which is a conjecture first proposed by Gao [20] in 2007, i.e.,

> *Conjecture of NP-Hardness:* A global optimization problem is NP-hard only if its canonical dual has no solution in $\mathcal{S}_a^+$.

It is also an open problem left in [21] and [28]. The reason for NP-hard problems and possible solutions were discussed recently in [22] and [25]. For the LKP $(\mathcal{P}_\ell)$, as long as $\tau_c v_i = c_i$ for any one $i \in \{1, \ldots, n\}$, the canonical dual algebraic equation (35) has at least two repeated solutions $\boldsymbol{\sigma}_c$ located on the boundary of $\mathcal{S}_a^+$. In this case, the primal problem $(\mathcal{P}_\ell)$ has multiple solutions, i.e., it is not well-posed [22], [25]. A linear perturbation method for solving this case was proposed recently in [9], [25], and [55].

V. Volume Reduction Method and Canonical Duality Algorithm

Theoretically speaking, for any given $V_c < V_0 = \sum_{i=1}^n v_i$, the canonical penalty-duality method can produce global optimal solution to the bilevel minimization problem $(\mathcal{P})$. However, if $\mu_c = V_c/V_0 \ll 1$, any iteration method could lead to unreasonable numerical solutions. In order to resolve this problem, a VRM method can be proposed.

> Introduce a volume reduction control parameter $\mu \in (\mu_c, 1)$ to produce a volume reduction sequence $\{V_\gamma = \mu V_{\gamma-1}\}$ ($\gamma = 1, \ldots, \gamma_c$) such that $V_{\gamma_c} = V_c$, $\mathcal{Z}_\gamma = \{\mathbf{z} \in \{0, 1\}^n | \mathbf{v}^T \mathbf{z} \leq V_\gamma\}$.
> For a given $V_\gamma \in [V_c, V_0]$ and initial values $(\mathbf{x}^{\gamma-1}, \mathbf{z}^{\gamma-1}) \in \mathcal{X}_a \times \mathcal{Z}_a$, to find $(\mathbf{x}^\gamma, \mathbf{z}^\gamma)$ such that
>
> $(\mathcal{P})^\gamma : (\mathbf{x}^\gamma, \mathbf{z}^\gamma) = \arg\min\{P(\mathbf{x}, \mathbf{z}) | \mathbf{x} \in \mathcal{X}_a, \mathbf{z} \in \mathcal{Z}_\gamma\}$
> s.t. $\mathbf{x} = \arg\min\{\Pi(\boldsymbol{\chi}, \mathbf{z}) | \boldsymbol{\chi} \in \mathcal{X}_a\}.$

Generally speaking, for any given sequence $\{V_\gamma\}$ we should have

$$(\mathcal{P}) = \lim_{\gamma \to \gamma_c} (\mathcal{P})^\gamma. \tag{47}$$

Numerically, different volume sequence $\{V_\gamma\}$ may produce totally different numerical results as long as the DAI is used. This is an intrinsic difficulty for all coupled bilevel optimal design problems. Based on the DAI and the VRM, the canonical duality algorithm (CDT) for solving the general bilevel optimization problem $(\mathcal{P})$ can be proposed.

Algorithm 1 Canonical Dual Algorithm for BKP (CDT)

1: **Input parameters:** μ, error allowances $\omega > 0$, $\mathbf{z}^0 = \{1\} \in \mathbb{R}^n$, and $V_0 = \sum_{i=1}^n v_i$. Let $\gamma = 1$.
2: **Solve** the lower-level optimization problem (4):

$$\mathbf{x}^\gamma = \arg\min\{\Pi(\mathbf{x}, \mathbf{z}^{\gamma-1}) \mid \mathbf{x} \in \mathcal{X}_a\}. \tag{48}$$

3: **Compute** $\mathbf{Q} = \mathbf{Q}(\mathbf{x}^\gamma)$, $\mathbf{c} = \mathbf{c}(\mathbf{x}^\gamma)$, and $V_\gamma = \max\{V_c, \mu V_{\gamma-1}\}$.
4: **Solve** the canonical dual problem $(\mathcal{P}_q^d)^\gamma$ to obtain

$$\boldsymbol{\zeta}^\gamma = (\boldsymbol{\sigma}^\gamma, \tau^\gamma) = \arg\min\{P_q^d(\boldsymbol{\zeta}) | \boldsymbol{\zeta} \in \mathcal{S}_a^+\}. \tag{49}$$

5: **Compute** the upper-level solution $\mathbf{z}^\gamma$ by

$$\mathbf{z}^\gamma = \mathbf{G}(\boldsymbol{\sigma}^\gamma)^{-1} \boldsymbol{\psi}(\boldsymbol{\zeta}^\gamma).$$

6: **If** $|P_q(\mathbf{z}^\gamma) - P_q(\mathbf{z}^{\gamma-1})| \leq \omega$ and $V_\gamma \leq V_c$, then stop; otherwise, continue.
7: **Let** $\gamma = \gamma + 1$. Go to Step 2.

The canonical dual problem $(\mathcal{P}_q^d)^\gamma$ in this algorithm could be either $(\mathcal{P}_\alpha^g)^\gamma$ or $(\mathcal{P}_\beta^d)^\gamma$ for QKP $(\mathcal{P}_q)$. For LKP $(\mathcal{P}_\ell)^\gamma$, its canonical dual could be either $(\mathcal{P}_\ell^g)^\gamma$ or $(\mathcal{P}_{\ell\beta}^d)^\gamma$.

VI. Application to Optimal Topology Design

Topology design is the arrangement of the various elements (links, nodes, volumes, etc.) of a complex network or a discretized continuous system in multidisciplinary fields of communication, electronics, optics, structural, bio- and nano-mechanics (see [4], [6], [24], and [38]). It was discovered recently by Gao [24] that the optimal topology design for general elastic structures should be formulated as a BKP

$$(\mathcal{P}_{to}) : \min\{P(\mathbf{x}, \mathbf{z}) = -\mathbf{c}(\mathbf{x})^T \mathbf{z} | \mathbf{x} \in \mathcal{X}_a, \mathbf{z} \in \mathcal{Z}_a\} \tag{50}$$

$$\text{s.t.} \quad \mathbf{x} \in \arg\min\{\Pi(\boldsymbol{\chi}, \mathbf{z}) | \boldsymbol{\chi} \in \mathcal{X}_a\}. \tag{51}$$

The leader design variable $\mathbf{z} = \{0, 1\}^n$ in this problem is the so-called density distribution such that if $z_e = 1$, then the eth element of the elastic structure is solid, otherwise, this element is void. The follower variable $\mathbf{x} \in \mathcal{X}_a \subset \mathbb{R}^m$ represents the displacement vector, whose domain $\mathcal{X}_a$ is a convex set, in which the boundary condition is given (i.e., certain components of $\mathbf{x}$ are fixed). The vector $\mathbf{c}(\mathbf{x}) = \{c_e(\mathbf{x}_e)\}$ and its component represents the strain energy stored in the eth element. For linear elastic structural, the total potential energy $\Pi(\mathbf{x}, \mathbf{z})$ is a quadratic function of $\mathbf{x}$

$$\Pi(\mathbf{x}, \mathbf{z}) = \frac{1}{2}\mathbf{x}^T \mathbf{K}(\mathbf{z})\mathbf{x} - \mathbf{x}^T \mathbf{f} = \sum_{e=1}^n \frac{z_e}{2} \mathbf{x}_e^T \mathbf{K}_e \mathbf{x}_e - \mathbf{x}^T \mathbf{f} \tag{52}$$

where $\mathbf{K}_e$ and $\mathbf{x}_e$ are, respectively, the stiffness matrix and the nodal displacement vector of the eth element; $\mathbf{v} = \{v_e\} \in \mathbb{R}_+^n$ and $v_e \geq 0$ represents the volume of the eth element; and V_c is a desired volume. Since $\mathbf{K}_e$ is positive definite, we have $\mathbf{c}(\mathbf{x}) = (1/2)\{\mathbf{x}_e^T \mathbf{K}_e \mathbf{x}_e\}^n \in \mathbb{R}_+^n$ and the global stiffness matrix $\mathbf{K}(\mathbf{z}) = \{z_e \mathbf{K}_e\} \in \mathbb{R}^{m \times m}$ is also positive definite for any given $\mathbf{z} \in \mathcal{Z}_a$. The given vector $\mathbf{f} \in \mathbb{R}^m$ is the external force.

Clearly, for a given displacement vector $\mathbf{x} \in \mathcal{X}_a$, the upper-level problem is a typical LKP, which can be solved analytically by the CDT. While for a fixed design variable $\mathbf{z} \in \mathcal{Z}_a$,
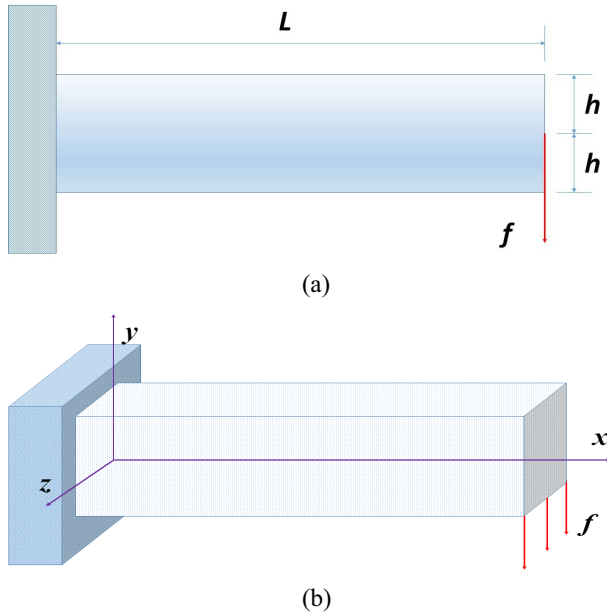
Fig. 1. Design domains for long cantilever beams with external load. (a) 2-D beam. (b) 3-D beam.

Fig. 2. Optimal topology by different mesh resolutions with $\mu_c = V_c/V_0 = 0.5$ and $\mu = 0.975$. Discrete variables $n =$ mesh, continuous variables $m = 2(nex+1)(ney+1)$. (a) Mesh $= 40 \times 10$, $C = 416.577$, Time $= 1.2165$ s. (b) Mesh $= 100 \times 30$, $C = 232.6$, Time $= 2.96$ s. (c) Mesh $= 180 \times 60$, $C = 171.77$, Time $= 7.69$ s.

the lower-level minimization is the well-known minimum total potential energy principle for linear elastic structures. Since the total potential energy $\Pi(\mathbf{x}, \mathbf{z})$ is a quadratic function of $\mathbf{x}$, the lower-level solution can be given analytically by $\mathbf{x} = \mathbf{K}(\mathbf{z})^{-1}\mathbf{f}$. Thus, this challenging topology design problem can be solved by the combination of the DAI-VRM and the canonical duality method. Since the LKP can be solved analytically by Corollary 1 or 2, the computational complexity for solving the lower-level solution $\mathbf{x} = \mathbf{K}(\mathbf{z})^{-1}\mathbf{f}$ is about $O(m^3)$, the CDT is a polynomial time algorithm.

The proposed CDT algorithm for topology design has been implemented in MATLAB. The test examples are the 2-D and 3-D benchmark cantilever beams in optimal topology design of elastic structure (see Fig. 1). Performance of the CDT method is first tasted for different mesh resolutions. Results in Fig. 2 show that for any given mesh resolutions, the CDT method produces precise integer solutions. Clearly, the finer the resolution, the smaller the stored energy $C = \mathbf{z}^T\mathbf{c}(\mathbf{x})$ with better material distributions. This means that the fine structure has more capacity for external load. For the mesh resolution 180×60 elements, we have $n = 180 \times 60 = 10\,800$ discrete variables and $m = 2 \times (180 + 1) \times (60 + 1) = 21\,960$ continuous variables, but the total computing time by an HP laptop computer (with Processor Intel Core I7-4810, CPU @ 2.80 GHz and memory 2.80 GB) is only $t = 7.69$ s.

To compare with the commercial codes TOP88 (i.e., SIMP) and BESO [59], we use the same mesh resolution of 180×60. The volume reduction rate $\mu = 0.97$ for both BESO and CDT. Computational results are reported in Fig. 3, which show clearly that the CDT method produces geometrically simple and mechanically sound structure. Since the SIMP method is a continuous relaxation approach, which cannot produce integer solutions, the structure obtained by the SIMP code is ugly with a large area of grayscaling checkerboard patterns. By the
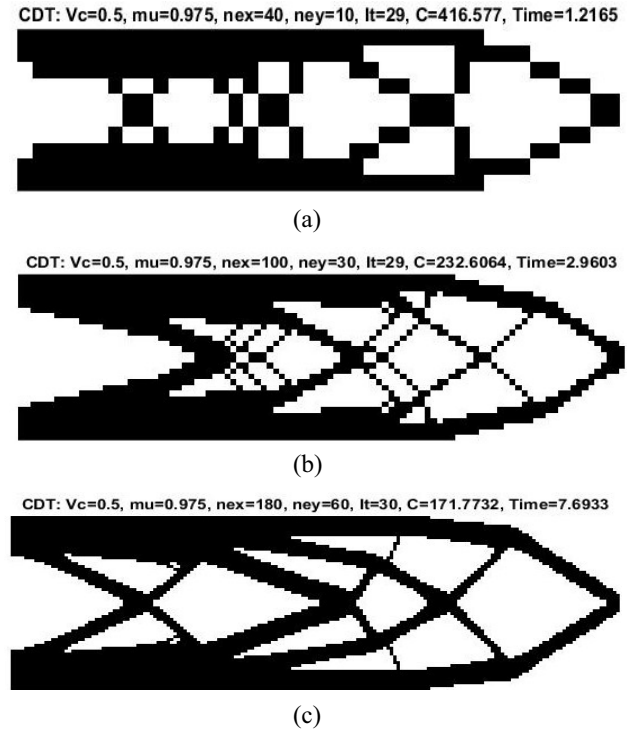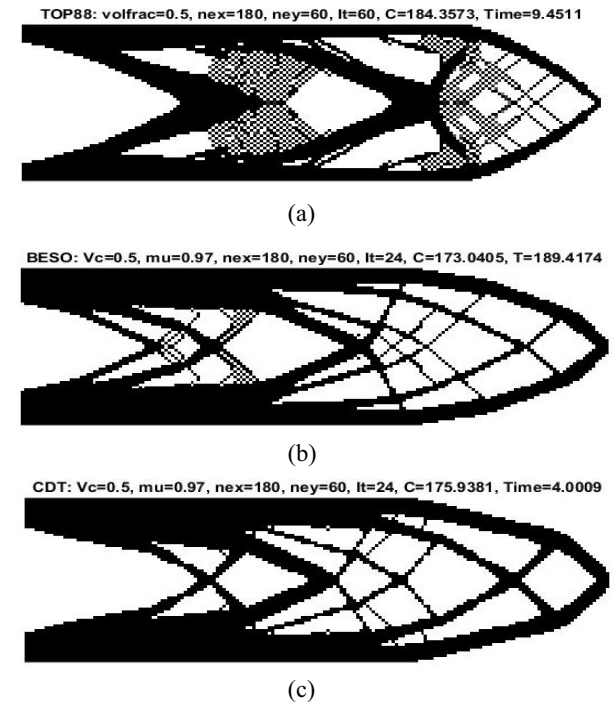

Fig. 3. 2-D beam structures by (a) SIMP, (b) BESO, and (c) CDT with $n = 180 \times 60$ and $m = 2(180 + 1) \times (60 + 1)$. The CDT is twice faster than the SIMP, and about 47 times faster than BESO. Computational result by (a) SIMP, (b) BESO, and (c) CDT.

fact that the BESO code is an evolutionary method for solving the LKP by simply using comparison algorithm, although it can produce integer solution similar to that by CDT, it is not
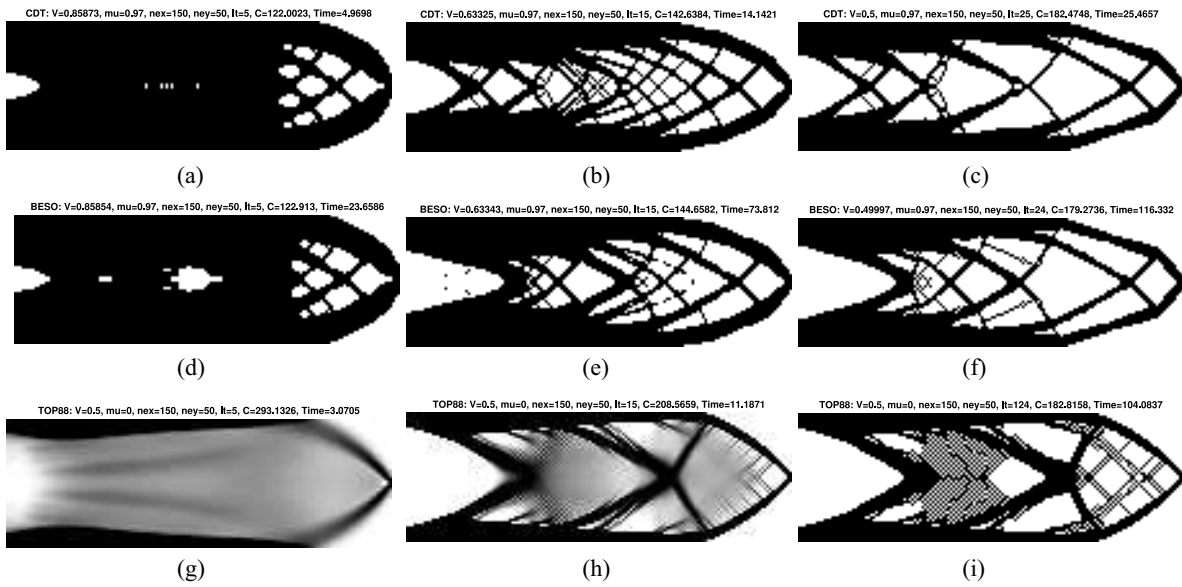
Fig. 4. Topology evolutions for 2-D beam by CDT (top), BESO (middle), and SIMP (bottom) with $V_c = 0.5$ and mesh resolution 150×50. The CDT is about four times faster than both BESO and SIMP. The SIMP can produce only grayscale topology since it is a continuous approach, i.e., $\mathbf{z} \in (0, 1)^n$. (a) CDT: $V_5 = 0.8587$, Time = 4.97. (b) BESO: $V_5 = 0.8585$, Time = 23.66. (c) SIMP: Iteration = 5, Time = 3.07. (d) $V_{15} = 0.63325$, Time = 14.1421. (e) $V_{15} = 0.63343$, Time = 73.812. (f) Iteration = 15, Time = 11.1871. (g) $V_c = 0.5$, Time = 25.4657. (h) $V_c = 0.5$, Time = 116.332. (i) Final Iteration, Time = 104.0837.

a polynomial time algorithm [25], therefore, it takes much longer time (more than 30 times) than the CDT method.

In order to see the structures' evolution during the iterations, we tested the 2-D beam with $V_c = 0.5$, nex = 150, and ney = 50. Fig. 4 shows the outputs of the structures by all methods at the fifth, 15th, and the final iterations. Since the initial volume is $V_0 = 1$, the volume reduction rate is $\mu = 0.97$ for both CDT and BESO, therefore, the volumes at the fifth and the 15th iteration should be $V_5 = 0.859$ and $V_{15} = 0.633$, respectively. Results in Fig. 4 (a) and (b) shows clearly that the structures produced by CDT algorithm are final knapsack solutions if the desired knapsacks V_c are V_5 and V_{15}, respectively. However, the structures produced by BESO are broken with disconnected branches Fig. 4(b), the structures by SIMP are uncertain with large areas of grayscales.

In order to look the energy density distribution $\mathbf{c} = \{c_e(\mathbf{x})\}$ in the optimal structures, we use the mesh resolution nex $\times$ ney = 80×30. Fig. 5 shows clearly that the CDT can produce mechanically sound structure with homogeneous distribution of strain energy density. Since the material distribution produced by SIMP is not mechanically reasonable, the overall strain energy stored in each element is about five times higher than that by CDT (see the color bars in Fig. 5). If we consider the dark-red level (= 1) is the elastic limit, then the structure produced by SIMP is far beyond this limit. This shows that the optimal structure by CDT has much potential to support even more external load. Although the BESO can produce better result than the SIMP, it can be used only for small-sized problems since it is not a polynomial time algorithm.

For the 3-D beam, we only compare the CDT with the SIMP since the BESO is computationally expensive. First, we use the mesh resolution $40 \times 15 \times 4$ and volume fraction $V_c = 0.2$. Fig. 6 shows that the optimal topology produced by CDT is not only geometrically elegant, but also mechanically sound
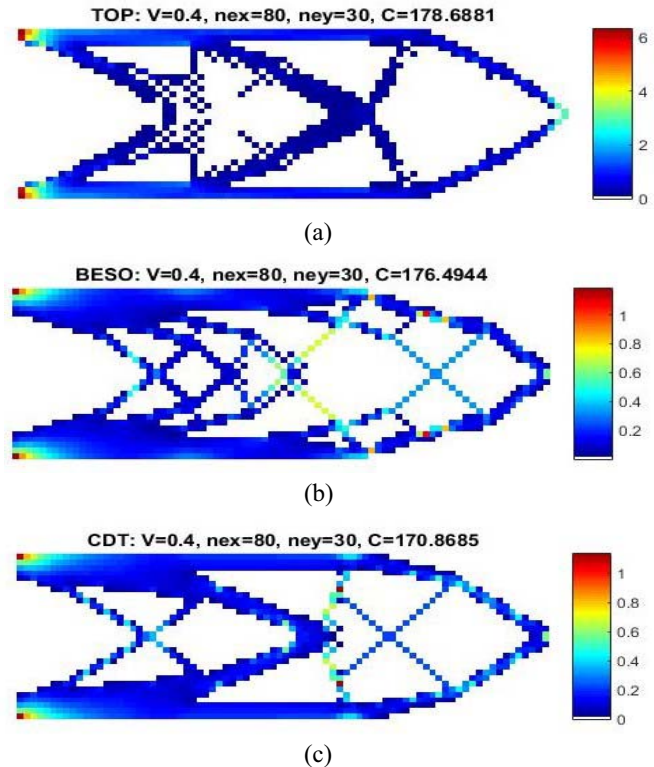

Fig. 5. Structures and energy density distributions by (a) SIMP, (b) BESO, and (c) CDT with $V_c = 0.4$.
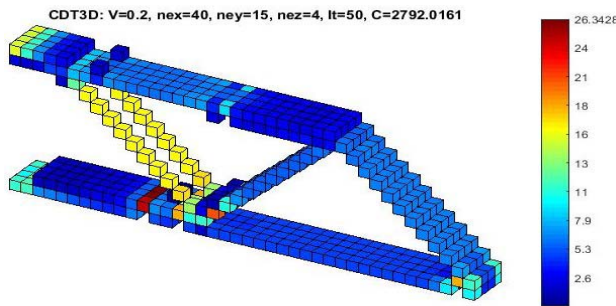
with homogeneous distribution of energy density and very few red elements. From the color bar we can see that the red-level is only at scale of 26.34. However, the topology produced by SIMP is overstaffed with many red elements at scale of 69.4. This shows that the structure produced by the CDT can support external load three times than the structure by SIMP.
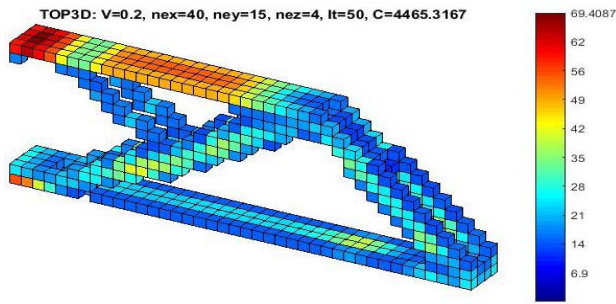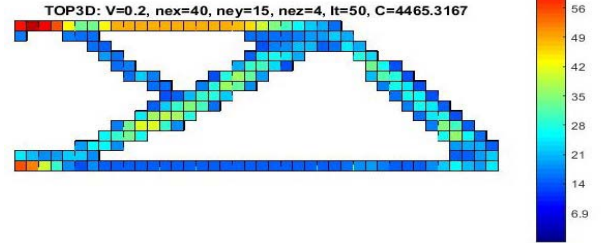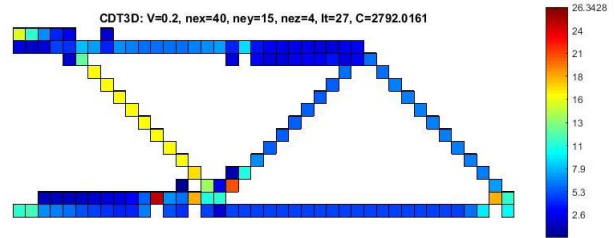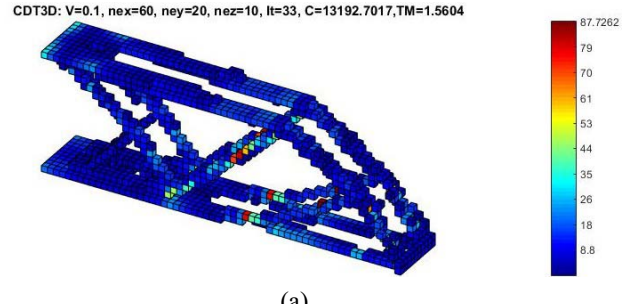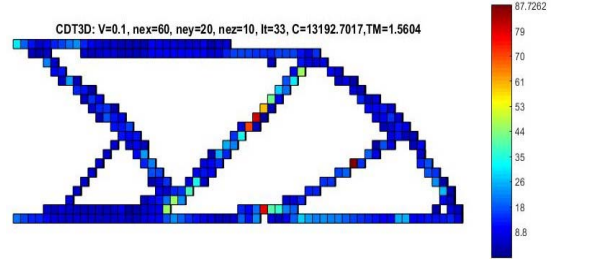
Fig. 6. Structures for 3-D beam with mesh $= 40 \times 15 \times 4$ and $V_c = 0.2$. The SIMP beam is overstaffed with many red elements at scale $= 69.4$ (highly stressed). The CDT beam is geometrically elegant with very few red elements at scale $= 26.34$, so it can support more external load. (a) SIMP solution: $C = 4465.3$. (b) SIMP solution: front view. (c) CDT solution: $C = 2792$. (d) Front view of CDT solution.

Fig. 7. 3-D beam structures by CDT and SIMP with $V_c = 0.1$ and mesh resolution $60 \times 20 \times 10$. Since the strain energy stored in the CDT beam ($C = 13192.7$) is only about a half of that stored in the SIMP beam ($C = 21960$), the CDT beam has much more potential to support more external load. (a) CDT solution: $C = 13192.7$, Time $= 1.56$ min. (b) Front view of CDT solution. (c) SIMP solution at the 50th iteration: $C = 21960$, Time $= 12.2$ min. (d) Front view of SIMP solution.
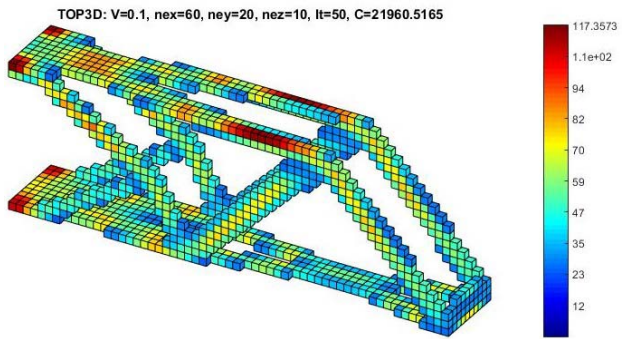
Now we increase the mesh resolution to $60 \times 20 \times 10$ and reduce the volume fraction to $V_c = 0.1$. Fig. 7 shows results obtained by the commercial code TOP3D (SIMP) and CDT3D (with $\mu = 0.93$). Since the 0-1 density distribution produced by the CDT is much more mechanically reasonable than the one by SIMP, most of elements in Fig. 7(a) are in blue color with only a few red elements at the scale $= 88$. The beam by SIMP has more red elements at the scale $= 117$. The structure shown in Fig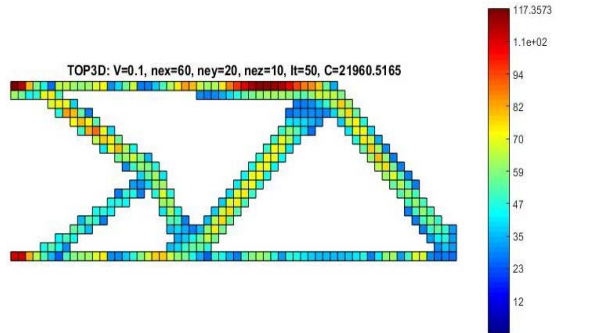. 7(c) is the out put of TOP3D at the It $= 50$th iteration, which is almost the same as the final result (It $= 199$) shown in Fig. 7(e). This result shows that the SIMP has a very slow rate of convergence. Also the CDT is more than 10 times faster than the TOP3D. The uncolored Fig. 8 shows that most elements are grayscale (i.e., $0 < z_i < 1$) since the SIMP is a continuous approximation.
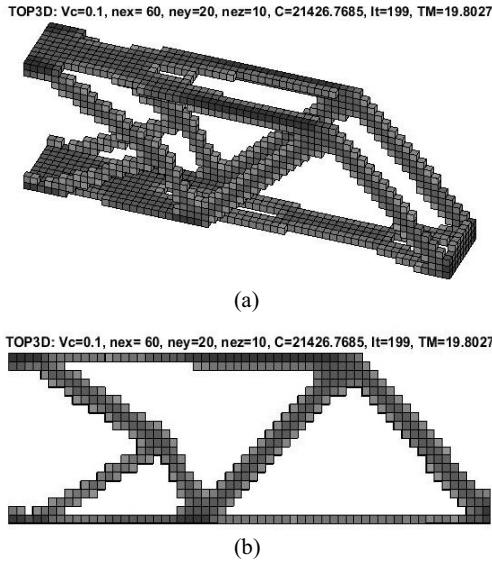
Fig. 8. 3-D beam structures by SIMP with $V_c = 0.1$ and mesh resolution $60 \times 20 \times 10$. (a) SIMP solution ($It = 199$): $C = 21426$, Time = 19.8 min. (b) Front view of SIMP solution.

Detailed study on the CDT for solving topology design problems and a 66-line MATLAB code of the CDT algorithm will be given in [30].

VII. CONCLUSION

Based on a combination of an alternative iteration, a volume reduction and the novel CDT, a new powerful method is proposed for solving bilevel optimization governed by knapsack problems. The main theoretical contributions are presented in Sections III and IV, i.e., by using the canonical duality method, the well-known knapsack problems can be solved analytically in terms of their canonical dual solutions. The existence and uniqueness of this general analytical solution are proved for both QKP and LKP. It shows that for any given $\mathbf{c}, \mathbf{v} \in \mathbb{R}^n$, if there exists a canonical dual solution $\tau_c > 0$ such that $\tau_c \mathbf{c} \neq \mathbf{v}$, the LKP can be solved uniquely in polynomial time. Therefore, a polynomial algorithm CDT is proposed for solving the BKPs. Its novelty is demonstrated by solving a benchmark problem in topology design. Numerical results indicated that the VRM is essential for solving bilevel optimization governed by knapsack problems. Both theoretical and numerical results verified that the CDT is a powerful method for studying general bilevel optimization problems.

Due to the intrinsic coupling effect in the bilevel optimization problem $(\mathcal{P})$, although for each given lower level variable $\mathbf{x}_k \in \mathcal{X}_a$, the upper level knapsack problem $\mathbf{z}_k = \arg\min\{P(\mathbf{x}_k, \mathbf{z}) | \mathbf{z} \in \mathcal{Z}_a\}$ can be solved analytically by the CDT, the solution sequence $\{\mathbf{x}_k, \mathbf{z}_k\}$ may not converge to the global optimal solution of $(\mathcal{P})$ since the DAI method is adopted in Algorithm 1. This is the main reason that the numerical solutions produced by the CDT depend sensitively on the volume reduction rate μ. This is an open problem and deserves theoretical study in the future.

ACKNOWLEDGMENT

The author would like to thank the Editor and the anonymous reviewers for the important comments and suggestions.

REFERENCES

[1] E. Aiyoshi and K. Shimizu, "A solution method for the static constrained Stackelberg problem via penalty method," *IEEE Trans. Autom. Control*, vol. AC-29, no. 12, pp. 1111–1114, Dec. 1984.

[2] J. Angelo, E. Krempser, and H. Barbosa, "Differential evolution for bilevel programming," in *Proc. Congr. Evol. Comput. (CEC)*, 2013, pp. 470–477.

[3] J. F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*. Dordrecht, The Netherlands: Kluwer, 1998.

[4] O. Botero and H. Chaouchi, "RFID network topology design based on genetic algorithms," in *Proc. IEEE Int. Conf. RFID Technol. Appl.*, 2011, pp. 300–305.

[5] L. Brotcorne, S. Hanafi, and R. Mansi, "One-level reformulation of the bilevel knapsack problem using dynamic programming," *Discr. Optim.*, vol. 10, no. 1, pp. 1–10, 2013.

[6] J. F. Camacho-Vallejo, J. Mar-Ortiz, F. López-Ramos, and R. P. Rodríguez, "A genetic algorithm for the bi-level topological design of local area networks," *PLoS ONE*, vol. 10, no. 6, 2015, Art. no. e0128067. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128067

[7] A. Caprara, M. Carvalho, A. Lodi, and G. J. Woeginger, "A study on the computational complexity of the bilevel knapsack problem," *SIAM J. Optim.*, vol. 24, no. 2, pp. 823–838, 2014.

[8] A. Chaabani, S. Bechikh, and L. Ben Said, "A co-evolutionary decomposition-based algorithm for bi-level combinatorial optimization," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2015, pp. 1659–1666.

[9] Y. Chen and D. Y. Gao, "Global solutions to spherical constrained quadratic minimization via canonical duality theory," in *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, D. Y. Gao, V. Latorre, and N. Ruan, Eds. New York, NY, USA: Springer, 2017, pp. 291–314.

[10] P. G. Ciarlet, *Linear and Nonlinear Functional Analysis With Applications*. Philadelphia, PA, USA: SIAM, 2013.

[11] S. Dempe, "Foundations of bilevel programming," in *Nonconvex Optimization and Its Applications Series*. Dordrecht, The Netherlands: Kluwer, 2002.

[12] S. Dempe, "Annotated bibliography on bilevel programming and mathematical problems with equilibrium constraints," *Optimization*, vol. 52, pp. 333–359, 2003.

[13] X. T. Deng, "Complexity issues in bilevel linear programming," *Multilevel Optim. Algorithms Appl.*, vol. 20, pp. 149–164, 1998.

[14] C. E. Ferreira, A. Martin, C. C. De Souza, R. Weismantel, and L. A. Wolsey, "Formulations and valid inequalities for the node capacitated graph partitioning problem," *Math. Program.*, vol. 74, no. 3, pp. 247–266, 1996.

[15] D. Y. Gao, *On Complementary-Dual Principles in Elastoplastic Systems and Pan-Penalty Finite Element Method*, Ph.D. dissertation, Tsinghua Univ., Beijing, China, 1986.

[16] D. Y. Gao, "Panpenalty finite element programming for limit analysis," *Comput. Struct.*, vol. 28, pp. 749–755, 1988.

[17] D. Y. Gao, *Duality Principles in Nonconvex Systems: Theory, Methods and Applications*. New York, NY, USA: Springer, 2000, p. 454.

[18] D. Y. Gao, "Canonical dual transformation method and generalized triality theory in nonsmooth global optimization," *J. Glob. Optim.*, vol. 17, nos. 1–4, pp. 127–160, 2000.

[19] D. Y. Gao, "Analytic solution and triality theory for nonconvex and nonsmooth variational problems with applications," *Nonlin. Anal.*, vol. 42, no. 7, pp. 1161–1193, 2000.

[20] D. Y. Gao, "Solutions and optimality criteria to box constrained nonconvex minimization problems," *J. Ind. Manag. Optim.*, vol. 3, no. 2, pp. 293–304, 2007.

[21] D. Y. Gao, "Canonical duality theory: Unified understanding and generalized solutions for global optimization," *Comput. Chem. Eng.*, vol. 33, no. 12, pp. 1964–1972, 2009.

[22] D. Y. Gao, "On unified modeling, theory, and method for solving multiscale global optimization problems," *AIP Conf. Proc.*, vol. 1776, 2016, Art. no. 020005.

[23] D. Y. Gao. (2016). *On Unified Modeling, Canonical Duality-Triality Theory, Challenges and Breakthrough in Optimization*. [Online]. Available: http://arxiv.org/abs/1605.05534

[24] D. Y. Gao, "Canonical duality theory for topology optimization," in *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, D. Y. Gao, N. Ruan, and V. Latorre, Eds. New York, NY, USA: Springer, 2017, pp. 263–276.

[25] D. Y. Gao, "On topology optimization and canonical dual method," *Comput. Methods Appl. Mech. Eng.*, vol. 341, pp. 249–277, Nov. 2018.

[26] D. Y. Gao, "Canoical duality-triality: Unified understanding modeling, problems, and NP-hardness in multi-scale optimization," in *Emergent Trends in Applied Mathematics and High-Preference Computing*, V. K. Singh, D. Y. Gao, and A. Fisher, Eds. New York, NY, USA: Springer, 2018.

[27] D. Y. Gao, V. Latorre, and N. Ruan, *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*. New York, NY, USA: Springer, 2017, p. 377.

[28] D. Y. Gao and N. Ruan, "Solutions to quadratic minimization problems with box and integer constraints," *J. Glob. Optim.*, vol. 47, no. 3, pp. 463–484, 2010.

[29] D. Y. Gao, N. Ruan, and V. Latorre, "Canonical duality-triality theory: Bridge between nonconvex analysis/mechanics and global optimization in complex systems," in *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*. New York, NY, USA: Springer, 2017, pp. 1–48.

[30] D. Y. Gao, N. Ruan, and V. Latorre, "On pure complementary energy principle for topology optimization and a 66-line MATLAB code," *Comput. Methods Appl. Mechanics Eng.*, submitted for publication.

[31] D. Y. Gao, N. Ruan, and H. D. Sherali, "Solutions and optimality criteria for nonconvex constrained global optimization problems with connections between canonical and Lagrangian duality," *J. Glob. Optim.*, vol. 45, no. 3, pp. 473–497, 2009.

[32] D. Y. Gao and H. Sherali, "Canonical duality: Connection between nonconvex mechanics and global optimization," in *Advances in Applied Mathematics and Global Optimization*. New York, NY, USA: Springer, 2009, pp. 249–316.

[33] D. Y. Gao and G. Strang, "Geometric nonlinearity: Potential energy, complementary energy, and the gap function," *Quart. Appl. Math.*, vol. 47, no. 3, pp. 487–504, 1989.

[34] D. Y. Gao, L. T. Watson, D. R. Easterling, W. I. Thacker, and S. C. Billups, "Solving the canonical dual of box- and integer-constrained nonconvex quadratic programs via a deterministic direct search algorithm," *Optim. Methods Softw.*, vol. 28, no. 2, pp. 313–326, 2013.

[35] P. Hansen, B. Jaumard, and G. Savard, "New branch-and-bound rules for linear bilevel programming," *SIAM J. Sci. Statist. Comput.*, vol. 13, no. 5, pp. 1194–1217, 1992.

[36] C. Helmberg, F. Rendl, and R. Weismantel, "Quadratic knapsack relaxations using cutting planes and semidefinite programming," in *Integer Programming and Combinatorial Optimization* (Lecture Notes in Computer Science). Heidelberg, Germany: Springer, 1996, pp. 175–189.

[37] R. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*, R. E. Miller and J. W. Thatcher, Eds. New York, NY, USA: Plenum, 1972, pp. 85–103.

[38] J. R. Kim, J. B. Jo, and H. K. Yang, "A solution for bi-level network design problem through Nash genetic algorithm," in *Proc. Int. Conf. Hybrid Inf. Technol.*, Nov. 2006, pp. 269–280.

[39] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*. Berlin, Germany: Springer-Verlag, 2004.

[40] V. Latorre and D. Y. Gao, "Canonical duality for solving general nonconvex constrained problems," *Optim. Lett.*, vol. 10, no. 8, pp. 1763–1779, 2016.

[41] S. F. Li and A. Gupta, "On dual configuration forces," *J. Elasticity*, vol. 84, no. 1, pp. 13–31, 2006.

[42] Z. P. Lo and B. Bavarian, "Job scheduling on parallel machines using simulated annealing," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, vol. 1, 1991, pp. 391–396.

[43] J. E. Marsden and T. J. R. Hughes, *Mathematical Foundations of Elasticity*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.

[44] R. Mathieu, L. Pittard, and G. Anandalingam, "Genetic algorithm based approach to bi-level linear programming," *Oper. Res.*, vol. 28, no. 1, pp. 1–21, 1994.

[45] J. T. Oden, *An Introduction to Mathematical Modeling*. Hoboken, NJ, USA: Wiley, 2011.

[46] K. Park, K. Lee, and S. Park, "An extended formulation approach to the edge-weighted maximal clique problem," *Eur. J. Oper. Res.*, vol. 95, no. 3, pp. 671–682, 1996.

[47] J. M. W. Rhys, "A selection problem of shared fixed costs and network flows," *Manag. Sci.*, vol. 17, no. 3, pp. 200–207, 1970.

[48] N. Ruan and D. Y. Gao, "Global optimal solution to quadratic discrete programming problem with inequality constraints," in *Canonical Duality Theory: Unified Methodology for Multidisciplinary Study*, D. Y. Gao, V. Latorre, and N. Ruan, Eds. New York, NY, USA: Springer, 2017, pp. 315–338.

[49] S. Ruuska and K. Miettinen, "Constructing evolutionary algorithms for bilevel multiobjective optimization," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2012, pp. 1–7.

[50] A. Sinha, P. Deb, K. Malo, P. Korhonen, and J. Wallenius, "Solving bilevel multi-criterion optimization problems with lower level decision uncertainty," *IEEE Trans. Evol. Comput.*, vol. 20, no. 2, pp. 199–217, Apr. 2016.

[51] H. Stackelberg, *The Theory of the Market Economy*. New York, NY, USA: Oxford Univ. Press, 1952.

[52] G. Strang, *Introduction to Applied Mathematics*. Wellesley, MA, USA: Wellesley–Cambridge Press, 1986.

[53] L. Vicente, G. Savard, and J. Júdice, "Descent approaches for quadratic bilevel programming," *J. Optim. Theory Appl.*, vol. 81, no. 2, pp. 379–399, 1994.

[54] Y. Wang, Y.-C. Jiao, and H. Li, "An evolutionary algorithm for solving nonlinear bilevel programming based on a new constraint-handling scheme," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 32, no. 2, pp. 221–232, May 2005.

[55] Z. B. Wang, S. C. Fang, D. Y. Gao, and W. X. Xing, "Global extremal conditions for multi-integer quadratic programming," *J. Ind. Manag. Optim.*, vol. 4, no. 2, pp. 2–13, 2008.

[56] C. Wilbaut, S. Hanafi, and S. Salh, "A survey of effective heuristics and their application to a variety of knapsack problems," *IMA J. Manag. Math.*, vol. 19, no. 3, pp. 227–244, Jul. 2008.

[57] C. Witzgall, "Mathematical methods of site selection for Electronic Message Systems (EMS)," NBS, Bratislava, Slovakia, Rep., 1975, doi: 10.6028/nbs.ir.75-737.

[58] P. Wu, F. Chu, A. Che, and Y. Zhao, "Dual-objective optimization for lane reservation with residual capacity and budget constraints," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2018.2810114.

[59] Y. M. Xie and G. P. Steven, *Evolutionary Structural Optimization*. London, U.K.: Springer, 1997.

[60] Y. Yu, S. Han, and X. Li, "A robust game-based algorithm for downlink joint resource allocation in hierarchical OFDMA femtocell network system," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2018.2817586.

[61] G. Zhang, J. Lu, and Y. Gao, *Multi-Level Decision Making: Models, Methods and Applications*. Heidelberg, Germany: Springer, 2015.

David Yang Gao received the Ph.D. degree in engineering mechanics and applied mathematics from Tsinghua University, Beijing, China.

He is the Alex Rubinov Professor of Mathematics with Federation University Australia, Mount Helen, VIC, Australia. He has held research and teaching positions in different institutes including MIT, Cambridge, MA, USA, Yale University, New Haven, CT, USA, Harvard University, Cambridge, the University of Michigan, Ann Arbor, MI, USA, and Virginia Tech, Blacksburg, VA, USA. He has published about 15 books and 200 scientific and philosophic papers (about 50% are single authored). His main research contributions include a canonical duality-triality theory, which can be used to model complex phenomena within a unified framework. This theory has been used successfully for solving a large class of nonconvex/nonsmooth problems in nonlinear sciences as well as a series of well-known NP-hard problems in global optimization and computational science. The main part of the canonical duality theory, i.e., the complementary-dual variational principle he proposed in 1997, solved a 50-year open problem in nonlinear elasticity and now is called the Gao principle in literature. This principle plays an important role in nonconvex analysis and computational mechanics. His current research interests include modeling and fundamental principles of duality and triality in multidisciplinary fields of engineering science, mathematics, physics, information technology, natural, and social complex systems.

Prof. Gao is the Founding Editor for Springer book series of Advances in Mechanics and Mathematics. He has served as an Associate Editor for several international journals of applied math, optimization, solid mechanics, dynamical systems, and industrial and management engineering. He is the Founding Vice President and the Secretary General of the International Society of Global Optimization.