

Successive convex approximations to cardinality-constrained convex programs: a piecewise-linear DC approach

Xiaojin Zheng · Xiaoling Sun · Duan Li · Jie Sun

Received: 9 February 2013 / Published online: 9 July 2013
© Springer Science+Business Media New York 2013

Abstract In this paper we consider cardinality-constrained convex programs that minimize a convex function subject to a cardinality constraint and other linear constraints. This class of problems has found many applications, including portfolio selection, subset selection and compressed sensing. We propose a successive convex approximation method for this class of problems in which the cardinality function is first approximated by a piecewise linear DC function (difference of two convex functions) and a sequence of convex subproblems is then constructed by successively linearizing the concave terms of the DC function. Under some mild assumptions, we establish that any accumulation point of the sequence generated by the method is a KKT point of the DC approximation problem. We show that the basic algorithm can

Dedicated to Professor Masao Fukushima in celebration of his 65th birthday.

This research was supported by the NSFC grants 10971034, 11101092 and the Joint NSFC/RGC grant 71061160506, and by Hong Kong RGC Grants CUHK414610 and 2050524.

X. Zheng

School of Economics and Management, Tongji University, Shanghai 200092, P.R. China
e-mail: xjzheng@tongji.edu.cn

X. Sun (✉)

Department of Management Science, School of Management, Fudan University, Shanghai 200433, P.R. China
e-mail: xls@fudan.edu.cn

D. Li

Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
e-mail: dli@se.cuhk.edu.hk

J. Sun

Department of Decision Sciences and Risk Management Institute, National University of Singapore, Singapore 117245, Singapore
e-mail: jsun@nus.edu.sg

be refined by adding strengthening cuts in the subproblems. Finally, we report some preliminary computational results on cardinality-constrained portfolio selection problems.

Keywords Convex programs · Cardinality constraint · DC approximation · Successive convex approximation · Strengthening cuts · Portfolio selection

1 Introduction

Cardinality constraint is often encountered in optimization models of real-world applications when the decision variables have to be *sparse* or the number of nonzero variables is required to be less than the total number of the decision variables. A general form of cardinality-constrained convex program can be expressed as

$$\begin{aligned} (P) \quad & \min f(x) \\ \text{s.t.} \quad & g(x) \leq b, \\ & \text{card}(x) \leq K, \end{aligned}$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a differentiable convex function, $g(x) = (g_1(x), \dots, g_m(x))^T$, $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$, are differentiable convex functions, $b \in \mathbb{R}^m$, and $\text{card}(x) \leq K$ is called the *cardinality constraint*, where $\text{card}(x)$ is the number of nonzero variables of x and $1 \leq K < n$ is a given integer. In the absence of the cardinality constraint, problem (P) reduces to a conventional convex program that can be solved efficiently by nonlinear programming methods (see, e.g., [29]).

It has been shown that problem (P) is in general NP-hard due to the presence of the cardinality constraint (see [3, 31]). The computational difficulty of problem (P) stems from the combinatorial nature of the cardinality constraint. In fact, by introducing a 0-1 variable y_i to enforce $x_i = 0$ or $x_i \neq 0$, problem (P) can be reformulated as the following mixed-integer 0-1 convex program:

$$\begin{aligned} (MIP) \quad & \min f(x) \\ \text{s.t.} \quad & g(x) \leq b, \\ & e^T y \leq K, \quad y \in \{0, 1\}^n \\ & l_i y_i \leq x_i \leq u_i y_i, \quad i = 1, \dots, n, \end{aligned}$$

where e is the all one column vector and l_i and u_i are the lower bound and upper bound of x_i , respectively. In practice, when there are no explicit lower and upper bounds for the set $X = \{x \in \mathbb{R}^n \mid g(x) \leq b\}$, we can set $l_i = -M$ and $u_i = M$ ($i = 1, \dots, n$) for a sufficiently large M .

In the applications of cardinality-constrained convex programs, the following quadratic case of (P) is of particular interest:

$$\begin{aligned}
 (QP) \quad & \min f(x) := x^T Qx + c^T x \\
 \text{s.t.} \quad & Ax \leq b, \\
 & \text{card}(x) \leq K.
 \end{aligned}$$

For instance, the cardinality-constrained or limited diversified mean-variance portfolio selection model is a special case of (QP) where the cardinality constraint limits the total number of different assets in the optimal portfolio (see, e.g., [3, 4, 8, 21, 31, 35]). Recently, the topic of finding sparse solutions to a system of linear equations has attracted much attention in subset selection problems of multivariate linear regression (see [1, 28]) and compressed sensing problems of signal processing (see [6] and the references therein). The optimization model of such problems is a special case of (QP) where $f(x) = \|Ax - b\|_2^2$ and the general linear constraints $Ax \leq b$ are absent.

Until recently, the studies of cardinality-constrained convex programs have focused mainly on the quadratic case (QP) . Solution methods in the literature for (QP) can be classified into two main categories: exact methods and heuristic methods. Most of the exact methods are based on tackling the mixed-integer quadratic program reformulation of (QP) and are of branch-and-bound framework using various relaxations and bounding techniques (see, e.g., [2, 3, 5, 23, 31, 33]). An MIQCQP reformulation is derived in [13] for a class of cardinality-constrained portfolio selection problems where the assets returns are driven by factor models. In [36], an SDP approach is proposed for finding the “best” diagonal decomposition in the perspective reformulation of quadratic program with cardinality and minimum threshold constraints. Recently, a novel geometric approach is proposed in [17] for minimizing a quadratic function subject to a cardinality constraint. Based on this geometric approach, a branch-and-bound method is then developed in [17] for solving cardinality-constrained portfolio selection problems. Heuristic methods for (QP) are typically metaheuristic approaches based on genetic algorithms, tabu search and simulated annealing. A detailed literature review on metaheuristic methods for portfolio selection with discrete features can be found in [8] and [35]. Heuristic methods and local search methods for portfolio selection models with cardinality constraint and minimum buy-in threshold constraints have been also studied by many other authors in the context of limited-diversification, small portfolios and empirical study for comparing different portfolio selection models with real features (see, e.g., [4, 21, 22, 27]).

It is often convenient to define the cardinality function $\text{card}(x)$ as (quasi) ℓ_0 -norm $\|x\|_0$, which is the limit of the ℓ_p -norm $\|x\|_p$ as p tends to zero. The ℓ_1 -norm approximation has been a popular method for finding sparse solutions to linear system (see [6] and the references therein). Replacing the cardinality constraint by the ℓ_1 -norm constraint $\|x\|_1 \leq K$, we can obtain a convex approximation of (P) :

$$\begin{aligned}
 (P_1) \quad & \min f(x) \\
 \text{s.t.} \quad & g(x) \leq b, \\
 & \|x\|_1 \leq K.
 \end{aligned}$$

In contrast to its successful application in sparse solution to linear system, the above ℓ_1 -norm approximation problem, however, does not necessarily produce solutions with desired sparsity due to the presence of the general constraints $g(x) \leq b$.

Several nonconvex approximations to ℓ_0 -norm have been recently studied in the context of ℓ_2 - ℓ_0 minimization problem defined by

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2 + \lambda \|x\|_0,$$

where $\lambda \in (0, +\infty)$. Replacing $\lambda \|x\|_0$ by a nonconvex and nonsmooth regularization term $\lambda \|x\|_p^p$ with $p \in (0, 1)$ results in the ℓ_2 - ℓ_p minimization problem. The complexity and the lower bound theory of nonzero entries of ℓ_2 - ℓ_p minimization were investigated in [9–11]. It was shown in [18] that the ℓ_2 - ℓ_p minimization with $p \in (0, 1)$ is NP-hard. An interior-point potential reduction algorithm was proposed in [18] to search for a local solution of ℓ_2 - ℓ_p minimization. Similar nonconvex approximations to ℓ_0 -norm have been used in sparse generalized eigenvalue problem ([32]), feature selection ([34]), sparse signal recovery ([7]) and matrix rank minimization ([14]). In [26], the following exponential regularization term

$$\lambda \sum_{i=1}^n (1 - \exp(-\alpha |x_i|)), \quad \alpha > 0,$$

is suggested to seek for solutions of a concave minimization over a polyhedral set with a minimum number of nonzero components. A successive linearization algorithm was proposed in [26] to find a stationary point of the regularized minimization problem over the polyhedral set. To the best of our knowledge, the use of the above mentioned nonconvex approximations to $\|x\|_0$ has not yet been investigated in the context of the general cardinality-constrained convex program (P) in the literature. The ℓ_p -(quasi) norm approximation and the exponential approximation in [26] have been only used in the regularized models of ℓ_0 minimization problems. Recently, a novel penalty decomposition (PD) method was proposed in [25] for the general case of (P) . This PD method framework was also used in [24] for rank minimization. Alternating direction method or block coordinate descent method was utilized in [25] and [24] to find a local optimal solution of the problem.

In this paper, we focus on solving the cardinality-constrained convex program directly, without appealing to regularization terms. We propose a successive convex approximation (SCA) method for solving problem (P) . This method is based on a new piecewise linear DC approximation of the cardinality function $\text{card}(x)$ or $\|x\|_0$. A prominent feature of this piecewise linear DC approximation lies in its polyhedral properties which can be exploited to construct tighter convex subproblems using strengthening cuts when linearization method is used to derive convex approximation. We present a basic iterative method in which a sequence of convex subproblems are solved successively. Under some mild assumptions, we establish that any accumulation point of the sequence generated by the method is a KKT point of the DC approximation problem. This basic SCA method can be further refined and improved by adding strengthening cuts derived from the piecewise linear DC approximation to $\|x\|_0 \leq K$. We report computational results on test problems of portfolio selection

which show that our method is promising for finding feasible solutions of good quality. In particular, the proposed method is competitive with the PD method of [25] and the two versions of the successive convex approximation methods using ℓ_p -norm approximation and the exponential approximation of [26], respectively, in terms of the quality of the feasible solutions obtained within reasonable computation time.

The rest of the paper is organized as follows. In Sect. 2, we derive a piecewise linear DC approximation to the cardinality function and establish some technical results on the DC approximation. In Sect. 3, we first present the basic successive convex approximation method and establish its convergence. We then describe a refined method by adding n strengthening cuts into the subproblems. In Sect. 4, we carry out numerical experiments to evaluate the performance of the algorithm for solving test problems of cardinality-constrained portfolio selection problems. Finally, we give some concluding remarks in Sect. 5.

2 Piecewise linear DC approximation to $\|x\|_0$

In this section, we derive a new piecewise linear DC approximation to the cardinality function $\text{card}(x)$, or equivalently, the ℓ_0 -norm $\|x\|_0$. We also present some basic properties of this DC approximation.

We first notice that

$$\|x\|_0 = \sum_{i=1}^n \text{sign}(|x_i|), \quad (1)$$

where $\text{sign}(z)$ denotes the sign function of $z \in \Re$ which is discontinuous at 0. Consider the following piecewise linear approximation of $\text{sign}(|z|)$:

$$\psi(z, t) = \min \left\{ 1, \frac{1}{t}|z| \right\}, \quad (2)$$

where $t > 0$ is a parameter. Figure 1 illustrates the graphs of functions $y = \text{sign}(|z|)$ and $y = \psi(z, t)$.

We see that function $\psi(z, t)$ can be also expressed as

$$\psi(z, t) = \frac{1}{t}|z| - \frac{1}{t}[(z - t)^+ + (-z - t)^+] = \frac{1}{t}[d(z, 0) - d(z, t)],$$

where $a^+ = \max(a, 0)$ and $d(z, t) = (z - t)^+ + (-z - t)^+$. Since $d(z, t)$ is a convex function of z , $\psi(z, t)$ is a DC function (difference of two convex functions) of z . Using $\psi(z, t)$, we can construct the following piecewise linear underestimation of the ℓ_0 -norm function $\|x\|_0$ for $x \in \Re^n$:

$$\phi(x, t) = \sum_{i=1}^n \psi(x_i, t) = \frac{1}{t}\|x\|_1 - \frac{1}{t}h(x, t), \quad (3)$$

where $h(x, t) = \sum_{i=1}^n d(x_i, t)$. We see that $\phi(x, t)$ is a nonsmooth piecewise linear DC function of x .

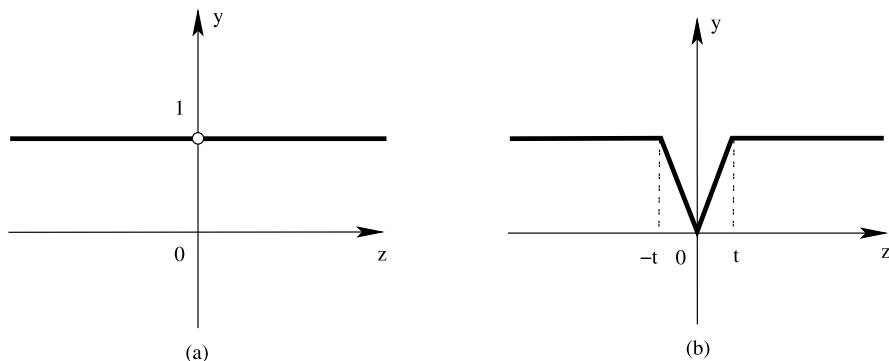


Fig. 1 (a) function $y = \text{sign}(|z|)$; (b) function $y = \psi(z, t)$

The following lemma summarizes some basic properties of $\phi(x, t)$ and can be easily proved.

Lemma 1

- (i) For any $t > 0$, $\phi(x, t)$ is a piecewise linear underestimation of $\|x\|_0$, i.e., $\phi(x, t) \leq \|x\|_0$, $\forall x \in \mathbb{R}^n$, and $\phi(x, t)$ is a non-increasing function of t .
- (ii) For any fixed $x \in \mathbb{R}^n$, it holds that

$$\lim_{t \rightarrow 0^+} \phi(x, t) = \|x\|_0.$$

- (iii) For any fixed $t > 0$, the subgradient of the convex function $h(x, t)$ with respect to x is $\partial h(x, t) = \{(\xi_1, \dots, \xi_n)^T \mid \xi_i \in \partial d(x_i, t), i = 1, \dots, n\}$, where

$$\partial d(x_i, t) = \begin{cases} -1, & x_i \in (-\infty, -t), \\ [-1, 0], & x_i = -t, \\ 0, & x_i \in (-t, t), \\ [0, 1], & x_i = t, \\ 1, & x_i \in (t, \infty). \end{cases} \quad (4)$$

- (iv) For any fixed $t > 0$, $\phi(x, t)$ is a continuous nonconvex function of $x \in \mathbb{R}^n$ and its Clarke's generalized gradient (see [12]) with respect to x is $\partial \phi(x, t) = \{(\xi_1, \dots, \xi_n)^T \mid \xi_i \in \partial \psi(x_i, t), i = 1, \dots, n\}$, where

$$\partial \psi(x_i, t) = \begin{cases} 0, & x_i \in (-\infty, -t), \\ [-1/t, 0], & x_i = -t, \\ -1/t, & x_i \in (-t, 0), \\ [-1/t, 1/t], & x_i = 0, \\ 1/t, & x_i \in (0, t), \\ [0, 1/t], & x_i = t, \\ 0, & x_i \in (t, \infty). \end{cases} \quad (5)$$

We now consider the DC approximation to problem (P). Replacing the ℓ_0 -norm (cardinality) function in (P) by the above DC approximation $\phi(x, t)$, we obtain the

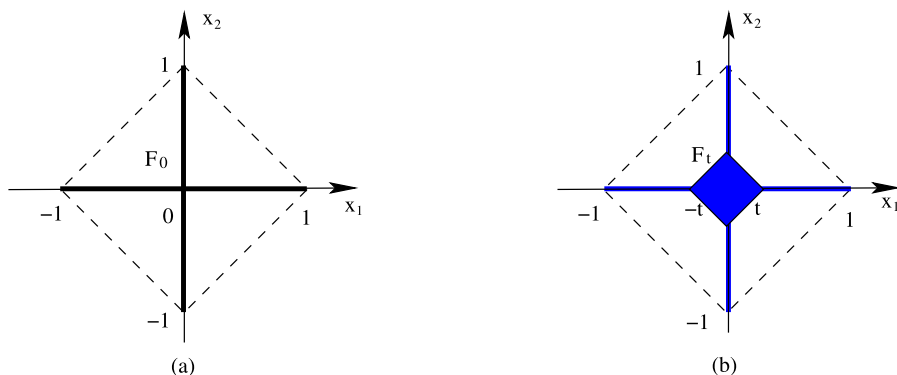


Fig. 2 Illustration of the piecewise linear approximation: (a) Set F_0 ; (b) Set F_t

following problem:

$$\begin{aligned} (P_t) \quad & \min f(x) \\ \text{s.t.} \quad & g(x) \leq b, \\ & \phi(x, t) \leq K. \end{aligned}$$

Since $\phi(x, t) \leq \|x\|_0$ for any $x \in \mathbb{R}^n$ and $t > 0$, problem (P_t) is a relaxation of (P) . Let F_0 and F_t denote the feasible sets of problem (P) and (P_t) , respectively, i.e.,

$$F_0 = \{x \in \mathbb{R}^n \mid g(x) \leq b, \|x\|_0 \leq K\}, \quad F_t = \{x \in \mathbb{R}^n \mid g(x) \leq b, \phi(x, t) \leq K\}.$$

It is easy to see that F_0 and F_t are closed sets in \mathbb{R}^n and $F_0 \subseteq F_t$ for any $t > 0$. In the sequel, we always assume that $F_0 \neq \emptyset$ so that $F_t \neq \emptyset$ for any $t > 0$.

Example 1 To illustrate the feasible sets of (P) and (P_t) , let us consider a small example of (P) where $F_0 = \{x \in [-1, 1]^2 \mid \|x\|_0 \leq 1\}$. The feasible set of (P_t) is $F_t = \{x \in [-1, 1]^2 \mid \phi(x, t) \leq 1\}$ with $0 < t < 1$. Figure 2 illustrates the two feasible sets, from which we can see that F_0 is the union of the two cross lines and F_t is the union of the small diamond and the four attached shorter lines.

Let x^* be a local optimal solution to (P_t) . By Proposition 2.3.3 and Theorem 6.1.1 in [12], under certain constraint qualification, there exist $\lambda^* \in \mathbb{R}_+^m$ and $\mu^* \in \mathbb{R}_+$ such that

$$0 \in \nabla f(x^*) + \nabla g(x^*)\lambda^* + \mu^* \partial \phi(x^*, t), \quad (6)$$

$$\lambda_i^* (g_i(x^*) - b_i) = 0, \quad i = 1, \dots, m, \quad (7)$$

$$\mu^* (\phi(x^*, t) - K) = 0, \quad (8)$$

where $\nabla g(x^*) \in \mathbb{R}^{n \times m}$ is the Jacobian of $g(x)$ at x^* . We call a point x^* satisfying conditions (6)–(8) as a *KKT point* of problem (P_t) .

Let $v(\cdot)$ denote the (global) optimal value of problem (\cdot) . The following theorem shows the relationship between (P_t) and (P) .

Theorem 1

- (i) F_t is a non-decreasing set-valued function of $t > 0$ and $\lim_{t \rightarrow 0^+} F_t = F_0$;
(ii) $v(P_t)$ is a non-increasing function of $t > 0$ and $\lim_{t \rightarrow 0^+} v(P_t) = v(P)$.

Proof (i) Since $\phi(x, t)$ is non-increasing with respect to t , we have $\phi(x, t_2) \leq \phi(x, t_1)$ for any $0 < t_1 \leq t_2$, which in turn implies that $F_0 \subset F_{t_1} \subseteq F_{t_2}$. It follows that $\lim_{t \rightarrow 0^+} F_t$ exists (see, e.g., [30]) and $\lim_{t \rightarrow 0^+} F_t \supseteq F_0$. Next, we prove that $\lim_{t \rightarrow 0^+} F_t \subseteq F_0$. For any $x \in \lim_{t \rightarrow 0^+} F_t$, there exist $t_k \rightarrow 0^+$ and $x^k \in F_{t_k}$ such that $x^k \rightarrow x$. Since $x^k \in F_{t_k}$, we have $x^k \in X$ and $\phi(x^k, t_k) \leq K$. Taking $k \rightarrow \infty$ and using Lemma 1 (ii), we have $x \in X$ and $\|x\|_0 \leq K$. Therefore, $x \in F_0$ and $\lim_{t \rightarrow 0^+} F_t \subseteq F_0$.

(ii) For any set $\Omega \in \mathfrak{R}^n$, define $I_\Omega(x)$ as follows: $I_\Omega(x) = 0$ if $x \in \Omega$ and $I_\Omega(x) = +\infty$ if $x \notin \Omega$. Let $\theta_t(x) = f(x) + I_{F_t}(x)$ and $\theta_0(x) = f(x) + I_{F_0}(x)$. Since $f(x)$ is a continuous function, $\theta_t(x)$ and $\theta_0(x)$ are lower semi-continuous and proper functions. It can be easily verified that $\theta_t(x)$ epi-converges to $\theta_0(x)$ as $t \rightarrow 0^+$. Thus, it follows from (i) and Proposition 7.4 in [30] that $I_{F_t}(x)$ epi-converges to $I_{F_0}(\cdot)$ as $t \rightarrow 0^+$. Therefore, by Theorem 7.33 in [30], we have $\lim_{t \rightarrow 0^+} v(P_t) = v(P)$. \square

3 Successive convex approximation methods

In this section, we first describe a basic successive convex approximation method for the DC approximation problem (P_t) by constructing a sequence of convex subproblems. We then establish the convergence of the basic method to a KKT point of (P_t) . To improve the approximation effects of the subproblems to the original problem (P) , we further refine the basic method by adding a group of strengthening cuts in the subproblems.

3.1 Basic successive convex approximation method

Let y be a feasible solution to (P_t) . Recall that $\phi(x, t) = \frac{1}{t}\|x\|_1 - \frac{1}{t}h(x, t)$. Let $\xi \in \partial h(y, t)$, where $\partial h(y, t)$ is defined in (4). By the convexity of $h(\cdot, t)$, we have $h(x, t) \geq h(y, t) + \xi^T(x - y)$ for all x . Thus, we have the following convex overestimation of $\phi(x, t)$ at y :

$$\phi(x, t) \leq u(x, y, \xi, t) := \frac{1}{t}\|x\|_1 - \frac{1}{t}[h(y, t) + \xi^T(x - y)], \quad \forall x \in \mathfrak{R}^n. \quad (9)$$

Using (9), the following convex subproblem can be constructed at y :

$$\begin{aligned} (P_t(y, \xi)) \quad & \min f(x) \\ \text{s.t.} \quad & g(x) \leq b, \\ & u(x, y, \xi, t) \leq K. \end{aligned}$$

The feasible set of $(P_t(y, \xi))$ is an inner approximation to problem (P_t) . The nonsmoothness of the term $\|x\|_1$ in function $u(x, y, \xi, t)$ can be eliminated by introducing a variable z_i for each $|x_i|$. The resulting problem is the following convex program:

$$\begin{aligned} (CP_t(y, \xi)) \quad & \min f(x) \\ \text{s.t.} \quad & g(x) \leq b, \\ & \frac{1}{t}e^T z - \frac{1}{t}[h(y, t) + \xi^T(x - y)] \leq K, \\ & -z_i \leq x_i \leq z_i, \quad i = 1, \dots, n. \end{aligned}$$

Lemma 2 *If (x^*, \bar{z}) solves $(CP_t(y, \xi))$, then (x^*, z^*) with $z_i^* = |x_i^*|$ ($i = 1, \dots, n$) also solves $(CP_t(y, \xi))$ and x^* solves $(P_t(y, \xi))$.*

Proof We first note that $-z_i \leq x_i \leq z_i$ is equivalent to $|x_i| \leq z_i$. If (x^*, \bar{z}) solves $(CP_t(y, \xi))$, then the solution (x^*, z^*) with $z_i^* = |x_i^*|$ ($i = 1, \dots, n$) is also feasible to $(CP_t(y, \xi))$ and has the same objective value as (x^*, \bar{z}) . Thus, (x^*, z^*) is an optimal solution to $(CP_t(y, \xi))$ and x^* is an optimal solution to $(P_t(y, \xi))$. \square

The above lemma suggests that we can always obtain an optimal solution (x^*, z^*) to $(CP_t(y, \xi))$ with $z_i^* = |x_i^*|$ by solving $(CP_t(y, \xi))$ and setting $z_i^* := |x_i^*|$ if necessary.

Let $F_t(y, \xi)$ be the feasible sets of problem $(P_t(y, \xi))$. Since $y \in F_t(y, \xi)$, we have $F_t(y, \xi) \neq \emptyset$. Also, by (9), it holds that $F_t(y, \xi) \subseteq F_t$ and hence $v(P_t(y, \xi)) \geq v(P_t)$. Therefore, $(P_t(y, \xi))$, or equivalently, $(CP_t(y, \xi))$, is a convex inner approximation of (P_t) .

In the sequel, we always assume that a certain constraint qualification holds to guarantee the existence of a KKT point of $(CP_t(y, \xi))$. This is not a restrictive assumption and can be always satisfied when $g_i(x)$ ($i = 1, \dots, m$) are linear or quadratic functions.

The following lemma gives a sufficient condition for an optimal solution of $(CP_t(y, \xi))$ to be a KKT point of (P_t) .

Lemma 3 *Let $\bar{y} \in F_t$. If (\bar{y}, \bar{z}) with $\bar{z}_i = |\bar{y}_i|$ ($i = 1, \dots, n$) solves $(CP_t(\bar{y}, \bar{\xi}))$ for some $\bar{\xi} \in \partial h(\bar{y}, t)$, then \bar{y} is a KKT point of (P_t) .*

Proof By KKT conditions, there exist multipliers $(\bar{\lambda}, \bar{\alpha}, \bar{\beta}, \bar{\mu}) \geq 0$ such that

$$\nabla f(\bar{y}) + \nabla g(\bar{y})\bar{\lambda} + \bar{\alpha} - \bar{\beta} - \frac{1}{t}\bar{\mu}\bar{\xi} = 0, \quad (10)$$

$$\frac{1}{t}\bar{\mu}e - \bar{\alpha} - \bar{\beta} = 0, \quad (11)$$

$$\bar{\lambda}_i(g_i(\bar{y}) - b_i) = 0, \quad i = 1, \dots, m, \quad (12)$$

$$\bar{\mu} \left[\frac{1}{t}e^T \bar{z} - \frac{1}{t}h(\bar{y}, t) - K \right] = 0, \quad (13)$$

$$\bar{\alpha}_i(\bar{z}_i - \bar{y}_i) = 0, \quad \bar{\beta}_i(\bar{z}_i + \bar{y}_i) = 0, \quad i = 1, \dots, n. \quad (14)$$

We now show that $\bar{\alpha} - \bar{\beta} - (1/t)\bar{\mu}\bar{\xi} \in \bar{\mu}\partial\phi(\bar{y}, t)$. In fact, since $\bar{z}_i = |\bar{y}_i|$ for $i = 1, \dots, n$, (14) implies that $\bar{y}_i > 0 \Rightarrow \bar{\beta}_i = 0$ and $\bar{y}_i < 0 \Rightarrow \bar{\alpha}_i = 0$. Thus, we deduce from (11) that $\bar{\alpha}_i = (1/t)\bar{\mu}$ when $\bar{y}_i > 0$ and $\bar{\beta}_i = (1/t)\bar{\mu}$ when $\bar{y}_i < 0$. When $\bar{y}_i = 0$, (11), together with $\bar{\alpha}_i \geq 0$ and $\bar{\beta}_i \geq 0$, implies $\bar{\alpha}_i - \bar{\beta}_i \in [-(1/t)\bar{\mu}, (1/t)\bar{\mu}]$. It then follows from (5) that $\bar{\alpha} - \bar{\beta} - (1/t)\bar{\mu}\bar{\xi} \in \bar{\mu}\partial\phi(\bar{y}, t)$. Thus, we infer from (10), (12) and (13) that \bar{y} satisfies the KKT conditions (6)–(8) of (P_t) . \square

The main idea of the successive convex approximation (SCA) method is to generate a sequence of solutions $\{x^k\} \subset F_t$ by successively solving the convex conservative approximation subproblems $(P_t(x^k, \xi^k))$ via $(CP_t(x^k, \xi^k))$, where x^0 is an initial feasible solution of (P_t) and x^{k+1} is an optimal solution of $(P_t(x^k, \xi^k))$. A detailed description of the basic successive convex approximation method is given as follows.

Algorithm 1 (Basic SCA Method for (P_t))

- Step 0. Choose a small $t > 0$ and a stopping parameter $\epsilon > 0$. Choose $x^0 \in F_t$ and $\xi^0 \in \partial h(x^0, t)$. Set $k := 0$.
 Step 1. Solve the convex subproblem $(CP_t(x^k, \xi^k))$, where $\xi^k \in \partial h(x^k, t)$. Let (x^{k+1}, z^{k+1}) be an optimal solution to $(CP_t(x^k, \xi^k))$.
 Step 2. If $\|x^{k+1} - x^k\| \leq \epsilon$, stop.
 Step 3. Set $k := k + 1$ and go to Step 1.

A sequential convex algorithm was proposed in [19] for solving a joint chance constrained program, where the joint chance constraint is rewritten as a DC constraint $g_1(x) - g_2(x) \leq 0$ with g_i ($i = 1, 2$) being continuously differentiable convex functions. Algorithm 1 can be viewed as a generalization of the method in [19] for solving problem (P_t) which has a nonsmooth DC constraint $\phi(x, t) \leq K$.

We now establish the convergence of Algorithm 1 to a KKT point of (P_t) . For the convenience of analysis, suppose that $\xi^k \in \partial h(x^k, t)$ in Step 0 and Step 1 is determined by

$$\xi_i^k = \begin{cases} -1, & x_i^k < -t_i, \\ -0.5, & x_i^k = -t_i, \\ 0, & -t_i < x_i^k < t_i, \\ 0.5, & x_i^k = t_i, \\ 1, & x_i^k > t_i, \end{cases} \quad (15)$$

for $i = 1, \dots, n$.

Theorem 2 Let $\epsilon = 0$ and $\{x^k\}$ be a sequence of solutions generated by Algorithm 1.

- (i) If the algorithm stops when $x^{k+1} = x^k$, then x^k is a KKT point of problem (P_t) .
 (ii) Any accumulation point of $\{x^k\}$ is a KKT point of (P_t) .

Proof (i) If the algorithm stops when $x^{k+1} = x^k$, then (x^k, z^k) with $z_i^k = |x_i^k|$ solves $(CP_t(x^k, \xi^k))$. By Lemma 3, x^k is a KKT point of problem (P_t) .

(ii) Since x^{k+1} solves $(P_t(x^k, \xi^k))$ and x^k is feasible to $(P_t(x^k, \xi^k))$, it holds that $v(P_t) \leq f(x^{k+1}) \leq f(x^k)$. So $\{f(x^k)\}$ is non-increasing and convergent with $\lim_{k \rightarrow \infty} f(x^k) = \inf_k f(x^k) \geq v(P_t)$.

Let \bar{x} be an accumulation point of $\{x^k\}$. Then, there exists a subsequence $\{x^{k_j}\} \subset \{x^k\}$ such that $x^{k_j} \rightarrow \bar{x}$. Since $x^k \in F_t$, we have $\bar{x} \in F_t$ and $f(\bar{x}) = \lim_{k \rightarrow \infty} f(x^k) = \inf_k f(x^k)$.

Let $\bar{z}_i = |\bar{x}_i|$ for $i = 1, \dots, n$. In the following, we prove that (\bar{x}, \bar{z}) is an optimal solution to $(CP_t(\bar{x}, \bar{\xi}))$ for some $\bar{\xi} \in \partial h(\bar{x}, t)$ and hence Lemma 3 implies that \bar{x} is a KKT point of (P_t) . We first show that there exists a sufficiently large j such that the feasible set of $(CP_t(x^{k_j}, \xi^{k_j}))$ is the same as that of $(CP_t(\bar{x}, \bar{\xi}))$ for some $\bar{\xi} \in \partial h(\bar{x}, t)$. To this end, we consider the linearized constraint in $(CP_t(y, \xi))$:

$$\frac{1}{t} e^T z - \frac{1}{t} [h(y, t) + \xi^T (x - y)] \leq K, \quad (16)$$

where $h(y, t) = \sum_{i=1}^n d(y_i, t)$ with $d(y_i, t) = (y_i - t)^+ + (-y_i - t)^+$ and $\xi \in \partial h(y, t)$. Let

$$p(x_i, y_i, \xi_i) = d(y_i, t) + \xi_i (x_i - y_i),$$

where $\xi_i \in \partial d(y_i, t)$. Then, (16) can be expressed as

$$\frac{1}{t} e^T z - \frac{1}{t} \sum_{i=1}^n p(x_i, y_i, \xi_i) \leq K. \quad (17)$$

For each $i = 1, \dots, n$, consider the following two cases of \bar{x}_i :

- (a) $\bar{x}_i \in (-\infty, -t) \cup (-t, t) \cup (t, +\infty)$;
- (b) $\bar{x}_i = t$ or $\bar{x}_i = -t$.

In case (a), since $x_i^{k_j} \rightarrow \bar{x}_i$, when j is sufficiently large, $x_i^{k_j}$ lies in the same open interval as \bar{x}_i does. Thus, by (15), one of the following three subcases happens:

- (1) $\bar{x}_i, x_i^{k_j} \in (-\infty, -t)$, $\xi_i^{k_j} = -1 = \bar{\xi}_i \in \partial d(\bar{x}_i, t)$, and

$$p(x_i, x_i^{k_j}, \xi_i^{k_j}) = (-x^{k_j} - t) - (x_i - x_i^{k_j}) = -x_i - t = p(x_i, \bar{x}_i, \bar{\xi}_i).$$

- (2) $\bar{x}_i, x_i^{k_j} \in (-t, t)$, $\xi_i^{k_j} = 0 = \bar{\xi}_i \in \partial d(\bar{x}_i, t)$, and

$$p(x_i, x_i^{k_j}, \xi_i^{k_j}) = 0 = p(x_i, \bar{x}_i, \bar{\xi}_i).$$

- (3) $\bar{x}_i, x_i^{k_j} \in (t, +\infty)$, $\xi_i^{k_j} = 1 = \bar{\xi}_i \in \partial d(\bar{x}_i, t)$, and

$$p(x_i, x_i^{k_j}, \xi_i^{k_j}) = (x_i^{k_j} - t) + (x_i - x_i^{k_j}) = x_i - t = p(x_i, \bar{x}_i, \bar{\xi}_i).$$

In case (b), when j is sufficiently large, then $x_i^{k_j} > 0$ if $\bar{x}_i = t$ or $x_i^{k_j} < 0$ if $\bar{x}_i = -t$. When $\bar{x}_i = t$, by (15), we have $\xi_i^{k_j} = 0$ if $0 < x_i^{k_j} < t$, $\xi_i^{k_j} = 0.5$ if $x_i^{k_j} = t$, and

$\xi_i^{k_j} = 1$ if $x_i^{k_j} > t$. Consequently, we have

$$p(x_i, x_i^{k_j}, \xi_i^{k_j}) = \begin{cases} 0, & 0 < x_i^{k_j} < t, \\ 0.5(x_i - x_i^{k_j}) = 0.5(x_i - t), & x_i^{k_j} = t, \\ (x_i^{k_j} - t) + (x_i - x_i^{k_j}) = x_i - t, & x_i^{k_j} > t. \end{cases} \quad (18)$$

On the other hand, when $\bar{x}_i = t$, we have $p(x_i, \bar{x}_i, \bar{\xi}_i) = \bar{\xi}_i(x_i - t)$, where $\bar{\xi}_i \in \partial d(\bar{x}_i, t) = [0, 1]$. Thus, we have from (18) that

$$p(x_i, x^{k_j}, \xi_i^{k_j}) = p(x_i, \bar{x}_i, \bar{\xi}_i), \quad \text{for some } \bar{\xi}_i \in \partial d(\bar{x}_i, t). \quad (19)$$

Similarly, we can show that (19) holds when $\bar{x}_i = -t$.

In summary, we have shown that (19) holds for each $i = 1, \dots, n$ when j is sufficiently large. Therefore, we deduce from (17) that the feasible set of problem $(CP_t(x^{k_j}, \xi^{k_j}))$ is identical to that of problem $(CP_t(\bar{x}, \bar{\xi}))$ for some $\bar{\xi} \in \partial h(\bar{x}, t)$, when j is sufficiently large. Since (x^{k_j+1}, z^{k_j+1}) is an optimal solution to $(CP_t(x^{k_j}, \xi^{k_j}))$, it is also an optimal solution to $(CP_t(\bar{x}, \bar{\xi}))$. Also, since (\bar{x}, \bar{z}) is a feasible solution to $(CP_t(\bar{x}, \bar{\xi}))$, we have $f(x^{k_j+1}) \leq f(\bar{x})$, so it must hold $f(x^{k_j+1}) = f(\bar{x})$ by the fact $f(\bar{x}) = \inf_k f(x_k)$. Therefore, (\bar{x}, \bar{z}) is an optimal solution to $(CP_t(\bar{x}, \bar{\xi}))$ for some $\bar{\xi} \in \partial h(\bar{x}, t)$. This completes the proof of the theorem. \square

3.2 Refined SCA method with strengthening cuts

In this subsection, we construct a group of strengthening cuts for improving the convex approximation problem $(CP_t(y, \xi))$. Recall that $\phi(x, t) = \sum_{i=1}^n \psi(x_i, t)$ and the convex overestimation of $\phi(x, t)$ is

$$u(x, y, \xi, t) = \sum_{i=1}^n w(x_i, y_i, \xi_i, t),$$

where $w(x_i, y_i, \xi_i, t) = \frac{1}{t}|x_i| - \frac{1}{t}[d(y_i, t) + \xi_i(x_i - y_i)]$ is the convex overestimation of $\psi(x_i, t)$.

By the definition of $\psi(x_i, t)$ in (2), it always holds that $\psi(x_i, t) \leq 1$ ($i = 1, \dots, n$) and thus $\psi(x_i, t) \leq 1$ ($i = 1, \dots, n$) are redundant to the constraints in (P_t) . However, the convex constraints $w(x_i, y_i, \xi_i, t) \leq 1$ ($i = 1, \dots, n$), which are the conservative approximation of $\psi(x_i, t) \leq 1$, provide strengthening cuts to the feasible set of $(P_t(y, \xi))$, which can help to improve the approximation of the subproblem $(P_t(y, \xi))$ to the original problem (P) .

To illustrate the effect of adding strengthening cuts to the feasible set of $(P_t(y, \xi))$, let us consider an example with $F_0 = \{x \in [-1, 1]^3 \mid \|x\|_0 \leq 2\}$. Let $0 < t < 0.5$ and $y = (0.5, t, 0)^T \in F_0$. Take $\xi = (1, 0.5, 0)^T \in \partial h(y, t)$. Then, the feasible set of $(P_t(y, \xi))$ is

$$F_t(y, \xi) = \left\{ x \in [-1, 1]^3 \mid \frac{1}{t}\|x\|_1 - \frac{1}{t}[(x_1 - t) + 0.5(x_2 - t)] \leq 2 \right\}.$$

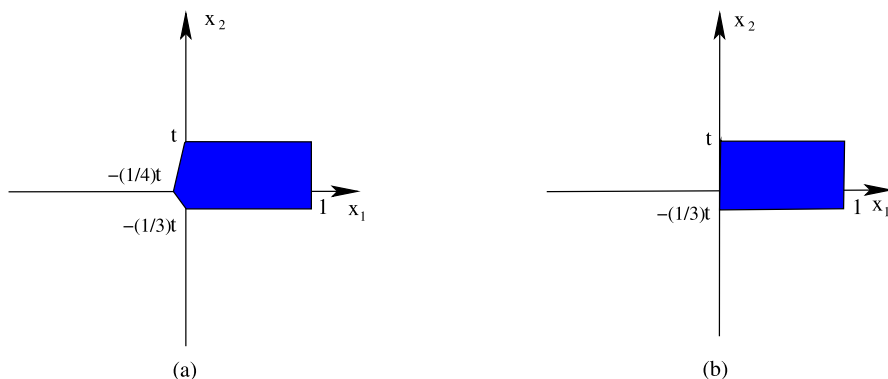


Fig. 3 (a) Projection of set $F_t(y, \xi)$ to L ; (b) Projection of set $F_t^v(y, \xi)$ to L

The strengthening cuts $w(x_i, y_i, \xi_i, t) \leq 1$ ($i = 1, 2, 3$) have the following forms:

$$\frac{1}{t}[|x_1| - (x_1 - t)] \leq 1, \quad \frac{1}{t}[|x_2| - 0.5(x_2 - t)] \leq 1, \quad \frac{1}{t}|x_3| \leq 1.$$

Notice that the first inequality can be reduced to $x_1 \geq 0$ and the second and third inequalities are redundant to $F_t(y, \xi)$. So, the revised feasible set after adding the strengthening cuts is

$$F_t^v(y, \xi) = F_t(y, \xi) \cap \{x \in \mathbb{R}^3 \mid x_1 \geq 0\}.$$

Figure 3 illustrates the projections of sets $F_t(y, t)$ and $F_t^v(y, \xi)$ to the plane $L = \{x \in \mathbb{R}^3 \mid x_3 = 0\}$, respectively, from which we can see clearly that $F_t^v(y, \xi) \subset F_t(y, \xi)$.

Using these n strengthening cuts, the convex approximation subproblem at y can be strengthened to

$$\begin{aligned} (P_t^v(y, \xi)) \quad & \min f(x) \\ \text{s.t.} \quad & g(x) \leq b, \\ & u(x, y, \xi, t) \leq K, \\ & w(x_i, y_i, \xi_i, t) \leq 1, \quad i = 1, \dots, n, \end{aligned}$$

which can be solved via the following convex program:

$$\begin{aligned} (CP_t^v(y, \xi)) \quad & \min f(x) \\ \text{s.t.} \quad & g(x) \leq b, \\ & \frac{1}{t}e^T z - \frac{1}{t}[h(y, t) + \xi^T(x - y)] \leq K, \\ & \frac{1}{t}z_i - \frac{1}{t}[d(y_i, t) + \xi_i(x_i - y_i)] \leq 1, \quad i = 1, \dots, n, \\ & -z_i \leq x_i \leq z_i, \quad i = 1, \dots, n. \end{aligned}$$

Using the above modified subproblem, we have the following refined successive convex approximation method for (P_t) .

Algorithm 2 (Refined SCA Method for (P_t))

- Step 0. Choose a small $t > 0$ and a stopping parameter $\epsilon > 0$. Choose $x^0 \in F_t$ and $\xi^0 \in \partial h(x^0, t)$. Set $k := 0$.
 Step 1. Solve the modified convex subproblem $(CP_t^v(x^k, \xi^k))$, where $\xi^k \in \partial h(x^k, t)$.
 Let (x^{k+1}, z^{k+1}) be an optimal solution to $(CP_t^v(x^k, \xi^k))$. For $i = 1, \dots, n$.
 Step 2. If $\|x^{k+1} - x^k\| \leq \epsilon$, stop.
 Step 3. Set $k := k + 1$ and go to Step 1.

We point out that Algorithms 1 and 2 can be easily extended to the case where the parameter t is dependent on i , i.e., we can use the following general DC function:

$$\phi(x, t) = \sum_{i=1}^n \psi(x_i, t_i) = \sum_{i=1}^n \frac{1}{t_i} |x_i| - \sum_{i=1}^n \frac{1}{t_i} d(x_i, t_i). \quad (20)$$

3.3 Recovery of a feasible solution

We note that Algorithms 1 and 2 tackle the approximation problem (P_t) where $\phi(x, t) \leq K$ is an approximation of $\text{card}(x) \leq K$ for a small $t > 0$. The solution generated by Algorithms 1 and 2 may not satisfy the cardinality constraint $\text{card}(x) \leq K$. Let x^* be a KKT point of (P_t) . A simple heuristic for recovering a feasible solution of (P) from x^* is as follows.

Heuristic 1 Rank the absolute values of x_i^* as: $|x_{i_1}^*| \leq |x_{i_2}^*| \leq \dots \leq |x_{i_n}^*|$. Set $x_{i_j} = 0$ for $j = 1, \dots, n - K$ in (P) . Remove the cardinality constraint and solve the reduced convex program with K variables. If the problem is feasible, then, combined with $x_{i_j} = 0$ for $j = 1, \dots, n - K$, the optimal solution is also feasible to (P) .

In general cases, however, there is no guarantee for the above heuristic to find a feasible solution of (P) . This is not surprising as testing the feasibility of (P) is already NP-complete when there are three linear inequality constraints in (P) (see [3]). While Heuristic 1 is very fast, the quality of the generated feasible solution could be unsatisfactory. Another heuristic is based on solving a small reduced mixed-integer problem of (P) (see [3]). The number of integer variables of the reduced problem is $K + \kappa$, where κ is a prescribed small integer, e.g., $\kappa = 10$. This heuristic can be described as follows:

Heuristic 2 Rank the absolute values of x_i^* as: $|x_{i_1}^*| \leq |x_{i_2}^*| \leq \dots \leq |x_{i_n}^*|$. Set $x_{i_j} = 0$ for $j = 1, \dots, n - K - \kappa$ in (P) . Solve the reduced mixed-integer convex program with initial upper bound obtained from Heuristic 1 (or $+\infty$ if no feasible solution is found by Heuristic 1). If the reduced mixed-integer convex program is feasible, then, combined with $x_{i_j} = 0$ for $j = 1, \dots, n - K - \kappa$, the optimal solution is also feasible to (P) .

Early termination of the branch-and-bound method for solving the small reduced mixed integer problem in Heuristic 2 can be also done by setting a limit on the maximum time, number of nodes or relative gap. Since $K + \kappa$ is often much less than n in practice, it is expected that Heuristic 2 is efficient in finding a reasonably good feasible solution of (P) . If the reduced mixed-integer convex program is infeasible, we can increase κ until the reduced problem is feasible.

4 Computational results

In this section, we conduct computational experiments to evaluate the performance of the SCA method (Algorithm 2) for problem (P) . The main purpose of our computational experiments is to test the capability of the SCA method, together with Heuristics 1 and 2, for finding good quality solutions of problem (P) when both cardinality and other constraints are present. It is beyond the scope of this paper to compare the proposed method with other specific heuristic methods for sparse solutions of linear system though in theory the DC approximation in (3) could be applicable to any optimization problems involving ℓ_0 -norm.

4.1 Test problems

The test problems in our computational experiments are cardinality constrained mean-variance portfolio selection problems (see [2, 3, 5, 23, 31]). Let μ and Q be the mean and covariance matrix of n risky assets, respectively. The problem can be formulated as

$$\begin{aligned} (MV) \quad & \min f(x) := x^T Q x \\ & \text{s.t. } \text{card}(x) \leq K, \\ & \mu^T x \geq \rho, \\ & \sum_{i=1}^n x_i = 1, \quad 0 \leq x_i \leq u_i, \quad i = 1, \dots, n. \end{aligned}$$

To build the test problems of (MV) , we use the 90 instances of portfolio selection created in [15, 16], 30 instances each for $n = 200, 300$ and 400 . The 30 instances for each n are divided into three subsets, 10 instances in each subset, with different diagonal dominance in matrix Q . The parameters ρ and u_i are uniformly drawn at random from intervals $[0.002, 0.01]$ and $[0.375, 0.425]$, respectively. The data files of these instances are available at: <http://www.di.unipi.it/optimize/Data/MV.html>. For each of the 90 instances, we set $K = 5, 10, 15, 20$, respectively, to build 360 instances of (MV) in total.

4.2 Implementation issues

In our computational experiments, we compare Algorithm 2 with the penalty decomposition (PD) method in [25] and two versions of the successive convex approxi-

mation (SCA) methods using ℓ_p -norm approximation and the exponential approximation in [26], respectively. Using ℓ_p -norm approximation, we can construct an ℓ_p -norm approximation problem to (P):

$$\begin{aligned} \min & f(x) \\ \text{s.t.} & \quad g(x) \leq b, \\ & \quad \|x\|_p \leq K, \end{aligned} \quad (21)$$

where $p > 0$ is a small number. Similarly, we can use the exponential approximation in [26] to construct the following approximation problem to (P):

$$\begin{aligned} \min & f(x) \\ \text{s.t.} & \quad g(x) \leq b, \\ & \quad \sum_{i=1}^n (1 - \exp(-\alpha |x_i|)) \leq K, \end{aligned} \quad (22)$$

where $\alpha > 0$ is a large number. SCA schemes similar to Algorithm 1 can be then applied to problems (21) and (22) to obtain a KKT point of problems (21) and (22), respectively.

In our implementation, the initial feasible solutions of the above methods are obtained by applying Heuristic 1 described in Sect. 3.3 to an optimal solution of the ℓ_1 -norm approximation problem (P_1). When implementing Algorithm 2, we use the DC function defined in (20) in which t_i is set in the following fashion: $t_i = x_{[K]}^0$ for $i \in I_K$ and $t_i = x_{\max}^0$ otherwise, where $x_{[K]}^0$ is the K -th largest variable of the initial feasible solution x^0 , I_K denotes the set of index of the K nonzero variables of x^0 , and x_{\max}^0 is the largest variable of x^0 . Note that for problem (MV), we always have $0 < x_{[K]}^0 \leq x_{\max}^0 < 1$ if $K > 1$. As in the numerical tests of [10, 26], we set $p = \frac{1}{2}$ in (21) and $\alpha = 100$ in (22). For each solution obtained by these methods, we apply Heuristic 2 to recover a feasible solution of (P). In Heuristic 2, we set $\kappa = 10$ for problems with $K = 5, 10, 15$ and $\kappa = 8$ for $K = 20$. The reduced mixed-integer quadratic problems in Heuristic 2 are solved by the MIQP solver in CPLEX 12.3.

The numerical tests have been performed on a PC equipped with Intel Pentium G630 CPU (2.70 GHz) and 4 GB of RAM, running Windows 7 (32-bit). All the convex quadratic subproblems in our numerical experiments are solved by the QP solver in CPLEX 12.3 with Matlab interface (see [20]). In our implementation of the PD method in [25], we use the Matlab code provided by the authors of [25], which is available at: <http://people.math.sfu.ca/~zhaosong/Codes/PD-sparse>.

4.3 Numerical results

To measure the quality of a feasible solution x^* , we use the following relative improvement of the function value of x^* over the solution obtained from the ℓ_1 -norm approximation:

$$\text{rel. imp.} = \frac{f(x_{\ell_1}) - f(x^*)}{\max(1, |f(x_{\ell_1})|)} (\%),$$

Table 1 Comparison results of the four methods for portfolio selection problems

n	K	SCA-DC			SCA- ℓ_p			SCA-exp			PD		
		rel. imp.	time	d	rel. imp.	time	d	rel. imp.	time	d	rel. imp.	time	d
200	5	2.66	1.92	24	1.35	12.60	18	0.79	0.22	16	0.79	29.75	16
300	5	2.02	3.23	25	1.19	23.41	14	0.52	0.29	12	0.52	41.12	12
400	5	2.48	4.91	25	1.40	41.93	19	0.99	0.39	14	0.99	68.09	14
200	10	2.58	9.17	27	2.30	23.23	27	0.95	1.78	19	0.97	29.48	19
300	10	2.77	11.70	25	2.52	40.39	27	0.82	2.11	15	0.78	44.17	15
400	10	3.31	19.98	27	3.03	77.01	28	1.25	2.65	18	1.27	81.38	18
200	15	2.58	85.07	24	1.86	84.94	27	0.85	15.64	13	0.89	43.65	13
300	15	2.92	96.56	25	2.26	105.00	27	1.05	19.50	17	0.97	60.18	17
400	15	3.37	110.34	27	3.07	130.09	29	1.47	15.52	19	1.49	78.17	19
200	20	2.58	73.36	27	1.62	64.66	29	0.98	13.25	17	0.75	43.35	16
300	20	2.32	69.84	23	1.43	86.24	30	0.81	14.61	18	0.73	54.77	16
400	20	3.61	89.68	24	2.34	113.32	28	1.33	14.60	21	1.23	80.20	20

where x_{ℓ_1} is the feasible solution of (P) obtained by applying Heuristic 1 to an optimal solution of the ℓ_1 -norm approximation problem (P_1) .

Table 1 summarizes the numerical results for test problems with $n = 200, 300, 400$ and $K = 5, 10, 15, 20$. Some notations in Table 1 are explained as follows.

- “SCA-DC” denotes the refined successive convex approximation method using the DC approximation (Algorithm 2);
- “SCA- ℓ_p ” and “SCA-exp” stand for the two versions of the successive convex approximation methods applied to problem (21) and (22), respectively;
- “PD” denotes the penalty decomposition method in [25];
- “time” (in seconds) is the average total computation time of the algorithm and Heuristic 2 for the 30 instances of each n ;
- “rel. imp.” is the average relative improvement for the 30 instances of each n ;
- “ d ” denotes the number of times the recovered feasible solution is better than the initial feasible solution, i.e., $\text{rel. imp.} > 0$, for the 30 instances of each n .

Comparing the results in Table 1, we observe that among the four methods, SCA-DC generates solutions with the best quality for all instances in terms of the average relative improvement over the ℓ_1 -norm approximation. However, together with the computation time of Heuristic 2, the average total computation time of SCA-DC is larger than that of SCA-exp and PD, but slightly less than that of SCA- ℓ_p for almost all instances. We notice that the computation time of the three versions of the successive convex approximation methods (SCA-DC, SCA- ℓ_p and SCA-exp) tends to increase as the cardinality K increases, while the relative improvements are similar for all K . We also see that SCA- ℓ_p is competitive in terms of the number of times the method beats ℓ_1 -norm approximation, while SCA-exp spends the least computa-

tion time among all the four methods. Overall, it appears that SCA-DC can produce solutions with good quality within reasonable computation time.

5 Concluding remarks

We have presented in this paper a successive convex approximation method for cardinality-constrained convex programming problems. Our method is based on a piecewise linear DC approximation of the cardinality function. In our method, this DC function is successively convexified to construct a sequence of convex subproblems. We have established the convergence of the method to a KKT point of the DC approximation problem. The basic successive convex approximation method can be refined by adding strengthening cuts generated from the separable DC functions. Computational results on test problems of portfolio selection demonstrate that the DC approximation is promising in finding good quality suboptimal solution and is competitive with other existing inexact methods for cardinality constrained problems.

References

1. Athanari, T.S., Dodge, Y.: *Mathematical Programming in Statistics*. Wiley, New York (1993)
2. Bertsimas, D., Shioda, R.: Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.* **43**, 1–22 (2009)
3. Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. *Math. Program.* **74**, 121–140 (1996)
4. Blog, B., van der Hoek, G., Rinnooy Kan, A.H.G., Timmer, G.T.: The optimal selection of small portfolios. *Manag. Sci.* **29**, 792–798 (1983)
5. Bonami, P., Lejeune, M.A.: An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Oper. Res.* **57**, 650–670 (2009)
6. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* **51**, 34–81 (2009)
7. Candes, E., Wakin, M., Boyd, S.: Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.* **14**(5), 877–905 (2008)
8. Chang, T.J., Meade, N., Beasley, J.E., Sharaiha, Y.M.: Heuristics for cardinality constrained portfolio optimisation. *Comput. Oper. Res.* **27**, 1271–1302 (2000)
9. Chen, X., Ge, D., Wang, Z., Ye, Y.: Complexity of unconstrained ℓ_2 - ℓ_p minimization. *Math. Program.* (2011). [arXiv:1105.0638](https://arxiv.org/abs/1105.0638)
10. Chen, X., Xu, F., Ye, Y.: Lower bound theory of nonzero entries in solutions of ℓ_2 - ℓ_p minimization. *SIAM J. Sci. Comput.* **32**, 2832–2852 (2010)
11. Chen, X., Zhou, W.: Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization. Tech. rep, Hong Kong Polytechnic University (2010). Available at <http://www.polyu.edu.hk/ama/staff/xjchen/IRL1-4-8.pdf>
12. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. Wiley, New York (1983)
13. Cui, X.T., Zheng, X.J., Zhu, S.S., Sun, X.L.: Convex relaxations and MIQCQP reformulations for a class of cardinality-constrained portfolio selection problems. *J. Glob. Optim.* (2012). [10.1007/s10898-012-9845-z](https://doi.org/10.1007/s10898-012-9845-z)
14. Fazel, M., Hindi, H., Boyd, S.: Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In: *Proceedings of American Control Conference*, pp. 2156–2162 (2003)
15. Frangioni, A., Gentile, C.: Perspective cuts for a class of convex 0–1 mixed integer programs. *Math. Program.* **106**, 225–236 (2006)
16. Frangioni, A., Gentile, C.: SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Oper. Res. Lett.* **35**, 181–185 (2007)

17. Gao, J., Li, D.: Optimal cardinality constrained portfolio selection. *Oper. Res.* **61**, 745–761 (2013)
18. Ge, D., Jiang, X., Ye, Y.: A note on complexity of ℓ_p minimization. *Math. Program.* **129**, 285–299 (2011)
19. Hong, L., Yang, Y., Zhang, L.: Sequential convex approximations to joint chance constrained programs: a Monte Carlo approach. *Oper. Res.* **59**, 617–630 (2011)
20. IBM ILOG CPLEX: IBM ILOG CPLEX 12.3 User's Manual for CPLEX (2011)
21. Jacob, N.L.: A limited-diversification portfolio selection model for the small investor. *J. Finance* **29**, 847–856 (1974)
22. Jobst, N., Horniman, M., Lucas, C., Mitra, G.: Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. *Quant. Finance* **1**, 489–501 (2001)
23. Li, D., Sun, X.L., Wang, J.: Optimal lot solution to cardinality constrained mean-variance formulation for portfolio selection. *Math. Finance* **16**, 83–101 (2006)
24. Lu, Z., Zhang, Y.: Penalty decomposition methods for rank minimization. Tech. rep, Department of Mathematics, Simon Fraser University (2012). Available at http://www.optimization-online.org/DB_HTML/2010/08/2718.html
25. Lu, Z., Zhang, Y.: Sparse approximation via penalty decomposition methods. Tech. rep, Department of Mathematics, Simon Fraser University (2012). Available at http://www.optimization-online.org/DB_HTML/2010/08/2719.html
26. Mangasarian, O.: Minimum-support solutions of polyhedral concave programs. *Optimization* **45**, 149–162 (1999)
27. Maringer, D., Kellerer, H.: Optimization of cardinality constrained portfolios with a hybrid local search algorithm. *OR Spektrum* **25**, 481–495 (2003)
28. Miller, A.J.: *Subset Selection in Regression*, 2nd edn. Chapman and Hall, London (2002)
29. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer, Berlin (1999)
30. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*. Springer, Berlin (2004)
31. Shaw, D.X., Liu, S., Kopman, L.: Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. *Optim. Methods Softw.* **23**, 411–420 (2008)
32. Sriperumbudur, B., Torres, D., Lanckriet, G.: A majorization-minimization approach to the sparse generalized eigenvalue problem. *Mach. Learn.* **85**, 3–39 (2011)
33. Vielma, J.P., Ahmed, S., Nemhauser, G.L.: A lifted linear programming branch-and-bound algorithm for mixed integer conic quadratic programs. *INFORMS J. Comput.* **20**, 438–450 (2008)
34. Weston, J., Elisseeff, A., Schölkopf, B., Tipping, M.: Use of the zero norm with linear models and kernel methods. *J. Mach. Learn. Res.* **3**, 1439–1461 (2003)
35. Woodside-Oriakhi, M., Lucas, C., Beasley, J.E.: Heuristic algorithms for the cardinality constrained efficient frontier. *Eur. J. Oper. Res.* **213**, 538–550 (2011)
36. Zheng, X.J., Sun, X.L., Li, D.: Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach. Tech. report (2011). Available at http://www.optimization-online.org/DB_HTML/2010/11/2797.html