Successive Convex Approximations to Cardinality-Constrained Quadratic Programs: A DC Approach

Xiaojin Zheng
School of Economics and Management, Tongji University, Shanghai 200092, P. R. China, xjzheng@tongji.edu.cn

Xiaoling Sun
Department of Management Science, School of Management, Fudan University, Shanghai 200433, P. R. China, xls@fudan.edu.cn

Duan Li
Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong, dli@se.cuhk.edu.hk

Jie Sun
Department of Decision Sciences and Risk Management Institute, National University of Singapore, Singapore 117245, jsun@nus.edu.sg

In this paper we consider a cardinality-constrained quadratic program that minimizes a convex quadratic function subject to a cardinality constraint and 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 are 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 be refined by adding valid inequalities in the subproblems. We report some preliminary computational results which show that our method is promising for finding good suboptimal solutions and is competitive with some other local solution methods for cardinality-constrained quadratic programs.

Key words: Quadratic program; cardinality constraint; DC approximation; successive convex approximation; valid inequalities

---

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 cardinality-constrained quadratic program can be expressed as

$$(P) \qquad \begin{aligned} &\min\ x^T Qx + c^T x \\ &\text{s.t. } \mathrm{card}(x) \le K, \\ &\quad\ \ x \in X, \end{aligned}$$

where Q is an $n \times n$ positive semidefinite matrix, $c \in \Re^n$, $\mathrm{card}(x)$ is the number of nonzero variables of x, $0 < K < n$ is an integer, and X is a bounded polyhedral set in the form of $X = \{x \in \Re^n \mid Ax \le b\}$. In the absence of the cardinality constraint $\mathrm{card}(x) \le K$, problem (P) reduces to a conventional convex quadratic program that can be solved efficiently (see, e.g., Nocedal and Wright (1999)). It has been shown that problem (P) is in general NP-hard due to the presence of the cardinality constraint (see Bienstock (1996); Shaw et al. (2008)).

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 standard mixed-integer 0-1 quadratic program (MIQP):

$$\begin{aligned} &\min\ x^T Qx + c^T x \\ &\text{s.t. } e^T y \le K,\ y \in \{0,1\}^n, \\ &\quad\ \ l_i y_i \le x_i \le u_i y_i,\ i = 1, \ldots, n, \\ &\quad\ \ x \in X, \end{aligned}$$

where $e \in \Re^n$ is the all one column vector, and l and u are the lower bound and upper bound vectors of X, respectively.

Many optimization problems can be formulated as cardinality-constrained quadratic programs. For instance, the cardinality-constrained or limited diversified mean-variance portfolio selection model takes the form of (P) where the cardinality constraint limits the total number of different assets in the optimal portfolio (see, e.g., Bienstock (1996); Blog et al. (1983); Chang et al. (2000); Jacob (1974); Shaw et al. (2008); Woodside-Oriakhi et al. (2011)).

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 Arthanari and Dodge (1993); Miller (2002)) and compressed sensing problems of signal processing (see Bruckstein et al. (2009) and the references therein). The optimization model of such problems is a special case of (P) where $q(x) = \|Ax - b\|_2^2$ and the general linear constraint $x \in X$ is absent.

Solution methods in the literature for problem (P) can be classified into two main categories: exact methods and heuristic methods. We also call a heuristic method as a *local method* if certain convergence to some suboptimal or local solution can be established for this heuristic method. Most of the exact methods for (P) are of branch-and-bound framework based on various relaxations and bounding techniques. Bienstock (1996) developed a branch-and-cut method for solving the MIQP reformulation of (P) where four types of cutting planes for the MIQP reformulation of (P) were considered. Li et al. (2006) proposed an exact solution method for cardinality constrained mean-variance models under round lot constraints and concave transaction costs. The method of Li et al. (2006) is based on a Lagrangian relaxation scheme and contour-domain cut branching rules. Shaw et al. (2008) presented a branch-and-bound method for cardinality constrained mean-variance portfolio problems, where the asset returns are driven by a factor model. Subgradient method is employed in Shaw et al. (2008) to compute a Lagrangian bound at each node of the search tree. Vielma et al. (2008) developed a branch-and-bound method for the MIQP reformulation of (P) based on a lifted polyhedral relaxation of conic quadratic constraints. Bertsimas and Shioda (2009) presented a specialized branch-and-bound method for (P) where a convex quadratic programming relaxation at each node is solved via Lemke's pivoting algorithm. Bonami and Lejeune (2009) proposed a branch-and-bound method for the mean-variance model under stochastic and integer constraints, including cardinality and minimum threshold constraints.

Another line of research on exact methods for (P) is the development of tight MIQP reformulations of (P). Frangioni and Gentile (2006, 2007) derived a novel perspective reformulation for quadratic programs with semi-continuous variables $x_i \in \{0\} \cup [\alpha_i, u_i]$ for $i = 1, \ldots, n$ (see also Frangioni and Gentile (2009); Günlük and Linderoth (2010)). The perspective reformulation can be easily extended to deal with cardinality constraint in (P). It can be shown that the perspective reformulation of (P) is more efficient than the standard MIQP reformulation in the sense that the continuous relaxation of perspective reformulation is tighter than that of the standard MIQP reformulation. Cui et al. (2012) investigated a class of cardinality-constrained portfolio selection problems where the assets returns are driven by factor models. A second-order cone program relaxation and an MIQCQP reformulation were derived in Cui et al. (2012) using the Lagrangian decomposition scheme similar to that of Shaw et al. (2008) and the special structure of the factor models. Zheng et al. (2011) proposed an SDP approach for finding the "best" diagonal decomposition in the perspective reformulation of quadratic program with cardinality and minimum threshold constraints.

Heuristic methods for (P) are typically metaheuristic approaches based on genetic algorithms, tabu search and simulated annealing. Chang et al. (2000) and Woodside-Oriakhi et al. (2011) proposed several metaheuristic methods for finding the efficient frontier of problem (P). A detailed literature review on metaheuristic methods for portfolio selection with discrete features can be found in Chang et al. (2000) and Woodside-Oriakhi et al. (2011). Murray and Shek (2011) proposed an efficient heuristic method for problem (P) which explores the structure of factor-model of the asset returns and utilizes the clustering techniques to identify subsets of similar assets. Heuristic methods and local search methods for portfolio selection models with cardinality constraints 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., Blog et al. (1983); Jacob (1974); Jobst et al. (2001); Maringer and Kellerer (2003)).

It is often convenient to define the cardinality function $\mathrm{card}(x)$ as 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 systems (see Bruckstein et al. (2009) 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):

$$(P_1) \qquad \begin{aligned} \min\ & q(x) := x^T Qx + c^T x \\ \text{s.t. } & \|x\|_1 \leq K, \\ & x \in X, \end{aligned}$$

which can be further reduced to a convex quadratic program with linear constraints. In contrast to its successful application in sparse solution to linear system, the ℓ_1-norm approximation problem (P_1), however, does not often produce solutions with desired sparsity, as witnessed in our computational results.

Several nonconvex approximations to ℓ_0-norm have been recently studied in the literature. The complexity and the lower bound theory of nonzero entries of ℓ_2-ℓ_p minimization were investigated in Chen et al. (2009) and Chen et al. (2011). It was shown in Ge et al. (2011) that the ℓ_p minimization ($0 < p < 1$) is NP-hard. An interior-point potential reduction algorithm was then proposed in Ge et al. (2011) to search for a local solution of ℓ_p minimization. Similar nonconvex approximations to ℓ_0-norm have been used in sparse generalized eigenvalue problem (Sriperumbudur et al. (2011)), feature selection (Weston et al. (2003)), sparse signal recovery (Candes et al. (2008)) and matrix rank minimization (Fazel et al. (2003)). A local deterministic optimization approach based on DC programming was developed in Gulpinar et al. (2010) for worst-case mean-variance portfolio selection problems with cardinality constraint. Gulpinar et al. (2010) first formulated the MIQP form of (P) as a DC program by using a penalty term $\sum_{i=1}^n y_i(1-y_i)$ for 0-1 variables $y \in \{0,1\}^n$ and then applied the DC algorithm (see Le Thi and Pham Dinh (2005); Pham Dinh and Le Thi (1997)) to find a local optimal solution of the problem. A linear program is solved at each iteration of this DC algorithm. Recently, Lu and Zhang (2012) proposed a novel penalty decomposition (PD) method for ℓ_0-norm minimization problem which minimizes a general nonlinear convex function subject to $\|x\|_0 \le K$. This PD method framework was also used in Lu and Zhang (2010) for rank minimization. Alternating direction method or block coordinate descent method was utilized in Lu and Zhang (2012) and Lu and Zhang (2010) to find a local optimal solution of the problems.

In this paper 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 when linearization method is used to derive convex approximation. We present a basic iteration method in which a sequence of convex quadratic 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 valid inequalities derived from the piecewise linear DC approximation to $\|x\|_0 \le K$. We report computational results of the method when applied to test problems of portfolio selection, subset selection and compressed sensing. The computational results show that our method is promising for finding good suboptimal solutions. In particular, our method can often improve the feasible solutions found by ℓ_1-norm approximation and is competitive with the PD method of Lu and Zhang (2012) in terms of the computation time and the quality of the feasible solutions obtained.

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

2. Piecewise Linear DC Approximation to $\|x\|_0$

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

We first notice that

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

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

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

where $t > 0$ is a parameter. Fig. 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

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

where $a^+ = \max(a,0)$ and $h(z,t) = (z-t)^+ + (-z-t)^+$. Since $h(z,t)$ 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 $x \in \Re^n$:

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

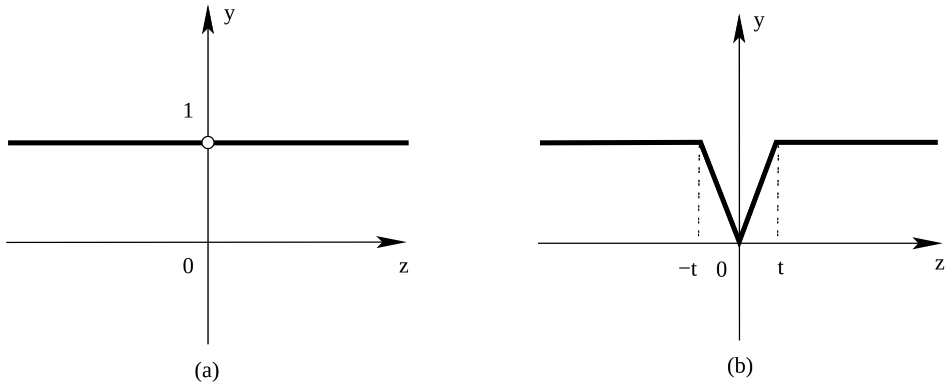

Figure 1: (a) function $y = \text{sign}(|z|)$; (b) function $y = \psi(z,t)$

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

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., $\phi(x,t) \le \|x\|_0$, $\forall x \in \Re^n$, and $\phi(x,t)$ is a non-increasing function of t.*

(ii) *For any fixed $x \in \Re^n$, it holds*

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

(iii) *The subgradient of the convex function $g(x,t)$ is $\partial g(x,t) = \{(\xi_1, \ldots, \xi_n)^T \mid \xi_i \in \partial h(x_i,t),\ i = 1, \ldots, n\}$, where*

$$\partial h(x_i,t) = \begin{cases} -1, & x_i \in (-\infty, -t), \\ [-1,0], & x_i = -t, \\ 0, & x_i \in (-t,t), \\ [0,1], & x_i = t, \\ 1, & x_i \in (t,\infty). \end{cases} \tag{3}$$

(iv) *$\phi(x,t)$ is a continuous nonconvex function of $x \in \Re^n$ and its Clarke's generalized gradient (see Clarke (1983)) is $\partial \phi(x,t) = \{(\xi_1, \ldots, \xi_n)^T \mid \xi_i \in \partial \psi(x_i,t),\ i = 1, \ldots, n\}$, where*

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

It is worth pointing out that $\phi(x,t)$ does not uniformly converge to $\|x\|_0$ on X even if X is a convex compact set.

We now consider the DC approximation to problem (P). Replacing the ℓ_0-norm (cardinality) constraint in (P) by the above DC approximation $\phi(x,t)$, we obtain the following quadratic program with a DC constraint:

$$(P_t) \qquad \min \ x^T Q x + c^T x$$
$$\text{s.t. } \phi(x,t) \le K,$$
$$Ax \le b.$$

Since $\phi(x,t) \le \|x\|_0$ for any $x \in \Re^n$ 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.

$$F_0 = \{x \in \Re^n \mid \|x\|_0 \le K, \ Ax \le b\}, \ F_t = \{x \in \Re^n \mid \phi(x,t) \le K, \ Ax \le b\}.$$

It is easy to see that F_0 and F_t are closed sets in $\Re^n$ and $F_0 \subseteq F_t$ 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 $F_0 = \{x \in [-1,1]^2 \mid \|x\|_0 \le 1\}$. The feasible set of (P_t) is $F_t = \{x \in [-1,1]^2 \mid \phi(x,t) \le 1\}$ with $0 < t < 1$. Fig. 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. The diamond area inside the dashed lines is the ℓ_1-norm approximation of F_0.

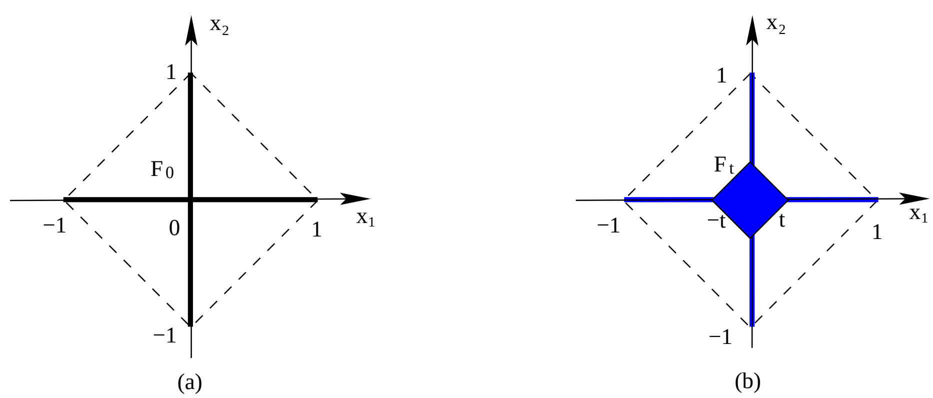

Figure 2: (a) Set F_0; (b) Set F_t

Let x^* be a local optimal solution to (P_t). By Proposition 2.3.3 and Theorem 6.1.1 in Clarke (1983), there exist $\lambda^* \in \Re^m_+$ and $\mu^* \in \Re_+$ such that

$$0 \in 2Qx^* + c + A^T\lambda^* + \mu^*\partial\phi(x^*, t), \tag{5}$$

$$(\lambda^*)^T(Ax^* - b) = 0, \tag{6}$$

$$\mu^*(\phi(x^*, t) - K) = 0. \tag{7}$$

We call a point x^* satisfying conditions (5)-(7) as a *KKT point* of problem (P_t).

The following theorem shows the relationship between (P_t) and (P).

Theorem 1 *Let $v(\cdot)$ denote the optimal value of problem $(\cdot)$. Then*

(i) *F_t is a non-decreasing set-valued function of $t > 0$ and $\lim_{t\to 0^+} F_t = F_0$;*

(ii) *$v(P_t)$ is a non-decreasing function of $t > 0$ and $\lim_{t\to 0^+} v(P_t) = v(P)$.*

Proof. (i) Since $\phi(x, t)$ is non-increasing with respect to t, we have $\phi(x, t_2) \le \phi(x, t_1)$ for any $0 < t_1 \le t_2$, which in turn implies that $F_0 \subset F_{t_1} \subseteq F_{t_2}$. It follows that $\lim_{t\to 0^+} F_t$ exists (see, e.g., Rockafellar and Wets (2004)) and $\lim_{t\to 0^+} F_t \supseteq F_0$. Next, we prove that $\lim_{t\to 0^+} F_t \subseteq F_0$. For any $x \in \lim_{t\to 0^+} F_t$, there exist $t_k \to 0^+$ and $x^k \in F_{t_k}$ such that $x^k \to x$. Since $x^k \in F_{t_k}$, we have $x^k \in X$ and $\phi(x^k, t^k) \le K$. Taking $k \to \infty$ and using Lemma 1 (ii), we have $x \in X$ and $\|x\|_0 \le K$. Therefore, $x \in F_0$ and $\lim_{t\to 0^+} F_t \subseteq F_0$.

(ii) For any set $A \in \Re^n$, define $I_A(x)$ as follows: $I_A(x) = 0$ if $x \in A$ and $I_A(x) = +\infty$ if $x \notin A$. Let $\theta_t(x) = x^TQx + c^Tx + I_{F_t}(x)$ and $\theta_0(x) = x^TQx + c^Tx + I_{F_0}(x)$. Since $x^TQx + c^Tx$ is a continuous function, $\theta_t(\cdot)$ and $\theta_0(\cdot)$ are lower semi-continuous and proper functions. Moreover, $\theta_t(\cdot)$ epi-converges to $\theta_0(\cdot)$ as $t \to 0^+$. It then follows from (i) and Proposition 7.4 in Rockafellar and Wets (2004) that $I_{F_t}(\cdot)$ epi-converges to $I_{F_0}(\cdot)$ as $t \to 0^+$. Therefore, by Theorem 7.33 in Rockafellar and Wets (2004), we have $\lim_{t\to 0^+} v(P_t) = v(P)$. □

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 quadratic subproblems. Convergence to a KKT point of (P_t) is then established. To improve the approximation effects of the subproblems to the original problem (P), we further refine the basic algorithm by including a group of valid inequalities to the convex quadratic subproblems.

3.1 Basic Successive Convex Approximation Method

Let y be a feasible solution to (P_t). Recall that $\phi(x,t) = \frac{1}{t}\|x\|_1 - \frac{1}{t}g(x,t)$. Let $\xi \in \partial g(y,t)$, where $\partial g(y,t)$ is defined in (3). Since $g(x,t)$ is a convex function of x, we have the following convex underestimation of $\phi(x,t)$ at y:

$$\phi(x,t) \geq u(x,y,\xi,t) := \frac{1}{t}\|x\|_1 - \frac{1}{t}[g(y,t) + \xi^T(x-y)], \quad \forall x \in \Re^n. \tag{8}$$

Using (8), the following convex subproblem can be constructed at y:

$$\begin{aligned} (P_t(y,\xi)) \qquad & \min \ q(x) := x^T Q x + c^T x \\ & \text{s.t. } Ax \leq b \\ & \qquad u(x,y,\xi,t) \leq K. \end{aligned}$$

The nonsmoothness of the term $\|x\|_1$ in function $u(x,y,\xi,t)$ can be eliminated by introducing variables $z_i = |x_i|$ for $i = 1,\dots,n$. The resulting problem is the following convex quadratic program:

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

It is easy to see that x^* solves $(P_t(y,\xi))$ if and only if (x^*, z^*) solves $(CP_t(y,\xi))$ with $z_i^* = |x_i^*|$ $(i = 1,\dots,n)$. In fact, for any optimal solution (x^*, z^*) of $(CP_t(y,\xi))$, it must have $z_i^* = |x_i^*|$ since the second constraint of $(CP_t(y,\xi))$ implies that z_i^* should be as small as possible and the third constraint enforces $z_i^* = |x_i^*|$ for $i = 1,\dots,n$.

Let $F_t(y,\xi)$ be the feasible sets of problem $(P_t(y,\xi))$. Then $F_t(y,\xi) \neq \emptyset$ as we always have $y \in F_t(y,\xi)$. Also, by (8), it holds $F_t(y,\xi) \subseteq F_t$ 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 *conservative approximation* of (P_t).

The following lemma shows when the optimal solution to $(CP_t(y,\xi))$ is also an optimal solution to (P_t).

Lemma 2 *Let $\bar{y} \in F_t$. If $(\bar{y}, \bar{z})$ solves $(CP_t(\bar{y}, \bar{\xi}))$ for some $\bar{\xi} \in \partial g(\bar{y}, t)$, then $\bar{y}$ is KKT point of (P_t).*

Proof. We first notice that $(\bar{y}, \bar{z})$ must satisfy $\bar{z}_i = |\bar{y}_i|$ for $i = 1, \ldots, n$. Since $(CP_t(\bar{y}, \bar{\xi}))$ is a convex quadratic problem, by the KKT theorem, there exist multipliers $(\bar{\lambda}, \bar{\alpha}, \bar{\beta}, \bar{\mu}) \geq 0$ such that

$$2Q\bar{y} + c + A^T\bar{\lambda} + \bar{\alpha} - \bar{\beta} - \frac{1}{t}\bar{\mu}\bar{\xi} = 0, \tag{9}$$

$$\frac{1}{t}\bar{\mu}e - \bar{\alpha} - \bar{\beta} = 0, \tag{10}$$

$$\bar{\lambda}^T(A\bar{y} - b) = 0, \tag{11}$$

$$\bar{\mu}[\frac{1}{t}e^T\bar{z} - \frac{1}{t}g(\bar{y}, t) - K] = 0, \tag{12}$$

$$\bar{\alpha}_i(\bar{z}_i - \bar{y}_i) = 0, \ \bar{\beta}_i(\bar{z}_i + \bar{y}_i) = 0, \ i = 1, \ldots, n. \tag{13}$$

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 $i = 1, \ldots, n$, (13) 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 (10) that $\alpha_i = (1/t)\bar{\mu}$ when $\bar{y}_i > 0$ and $\beta_i = (1/t)\bar{\mu}$ when $\bar{y}_i < 0$. When $\bar{y}_i = 0$, (10), 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 (4) that $\bar{\alpha} - \bar{\beta} + (1/t)\bar{\mu}\bar{\xi} \in \bar{\mu}\partial\phi(\bar{y}, t)$. Thus, we infer from (9), (11) and (12) that $\bar{y}$ satisfies the KKT conditions (5)-(7) of (P_t). □

The main idea of the successive convex approximation (SCA) method is to generate a sequence of feasible solutions $\{x^k\}$ 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 to (P_t) and x^{k+1} is an optimal solution to $(P_t(x^k, \xi^k))$. A detailed description of the basic successive convex approximation method is given as follows.

Algorithm 1 (Basic SCA Method for (P))

Step 0. Choose $t > 0$ and a stopping parameter $\epsilon > 0$. Choose $x^0 \in F_t$ and $\xi^0 \in \partial g(x^0, t)$. Set $k := 0$.

Step 1. Solve the convex quadratic subproblem $(CP_t(x^k, \xi^k))$, where $\xi^k \in \partial g(x^k, t)$ is determined by

$$\xi_i^k = \begin{cases} -1, & x_i^k < -t_i, \\ -0.5, & x_i^k = -t_i, \\ 0, & -t_i < x_i^k < t_i, \\ 0.5, & x_i^k = t_i, \\ 1, & x_i^k > t_i, \end{cases} \tag{14}$$

for $i = 1, \ldots, n$. Let (x^{k+1}, z^{k+1}) be an optimal solution to $(CP_t(x^k, \xi^k))$.

Step 2. If $\|x^{k+1} - x^k\| \le \epsilon$, stop.

Step 3. Set $k := k + 1$ and go to Step 1.

A sequential convex algorithm was proposed by Hong et al. (2011) for solving a nonlinear program with a DC constraint $g_1(x) - g_2(x) \le 0$, where g_i $(i = 1, 2)$ are continuously differentiable convex functions. Algorithm 1 can be viewed as a generalization of the method in Hong et al. (2011) for solving problem (P_t) which has a nonsmooth DC constraint $\phi(x, t) \le K$.

We now establish the convergence of Algorithm 1 to a KKT point of (P_t).

Theorem 2 *Let $\epsilon = 0$ and $\{(x^k, z^k)\}$ 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 cluster 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 solves $(CP_t(x^k, \xi^k))$. By Lemma 2, 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 $v(P_t) \le q(x^{k+1}) \le q(x^k)$. So $\{q(x^k)\}$ is non-increasing and convergent with $\lim_{k\to\infty} q(x^k) = \inf_k q(x^k) \ge v(P_t)$.

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

Let $\bar{z}_i = |\bar{x}_i|$ for $i = 1, \ldots, n$. In the following, we prove that $(\bar{x}, \bar{z})$ is an optimal solution to $(CP_t(\bar{x}, \bar{\xi}))$ for some $\bar{\xi} \in \partial g(\bar{x}, t)$ so that by Lemma 2 $\bar{x}$ 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 $\bar{\xi} \in \partial g(\bar{x}, t)$. To this end, we consider the linearized constraint in $(CP_t(y, \xi))$:

$$\frac{1}{t} e^T z - \frac{1}{t}[g(y, t) + \xi^T (x - y)] \le K, \tag{15}$$

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

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

Then, (15) can be expressed as

$$\frac{1}{t} e^T z - \frac{1}{t} \sum_{i=1}^n p(x_i, y_i, \xi) \le K. \tag{16}$$

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

(i) $\bar{x}_i \in (-\infty, -t) \cup (-t, t) \cup (t, +\infty)$;

(ii) $\bar{x}_i = t$ or $\bar{x}_i = -t$.

In case (i), since $x_i^{k_j} \to \bar{x}_i$, when j is sufficiently large , $x_i^{k_j}$ lies in the same open interval as $\bar{x}_i$ does. Thus, by (14), one of the following three subcases happens:

(1) $\bar{x}_i, x_i^{k_j} \in (-\infty, -t)$, $\xi_i^{k_j} = -1 = \bar{\xi}_i \in \partial h(\bar{x}_i, t)$, 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) $\bar{x}_i, x_i^{k_j} \in (-t, t)$, $\xi_i^{k_j} = 0 = \bar{\xi}_i \in \partial h(\bar{x}_i, t)$, and

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

(3) $\bar{x}_i, x_i^{k_j} \in (t, +\infty)$, $\xi_i^{k_j} = 1 = \bar{\xi}_i \in \partial h(\bar{x}_i, t)$, 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 (ii), 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 (14), 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

$$p(x_i, x_i^{k_j}, \xi_i^{k_j}) = \begin{cases} 0 & 0 < x_i^{k_j} < t, \\ 0.5(x_i - x_i^{k_j}) = 0.5(x_i - t), & x_i^{k_j} = t, \\ (x_i^{k_j} - t) + (x_i - x_i^{k_j}) = x_i - t, & x_i^{k_j} > t. \end{cases} \tag{17}$$

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 $\bar{\xi}_i \in \partial h(\bar{x}_i, t) = [0, 1]$. Thus, we have from (17) that

$$p(x_i, x^{k_j}, \xi_i^{k_j}) = p(x_i, \bar{x}_i, \bar{\xi}_i), \quad \text{for some } \bar{\xi}_i \in \partial h(\bar{x}_i, t). \tag{18}$$

Similarly, we can show that (18) holds when $\bar{x}_i = -t$.

In summary, we have shown that (18) holds for each $i = 1, \ldots, n$ when j is sufficiently large. Therefore, we deduce from (16) 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 $\bar{\xi} \in \partial g(\bar{x}, t)$, 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 $(CP_t(\bar{x}, \bar{\xi}))$. Since $(\bar{x}, \bar{z})$ is a feasible solution to $(CP_t(\bar{x}, \bar{\xi}))$, we have $q(x^{k_j+1}) \leq q(\bar{x})$, so it must hold $q(x^{k_j+1}) = q(\bar{x})$ by the fact $q(\bar{x}) = \inf_k q(x_k)$. Therefore, $(\bar{x}, \bar{z})$ is an optimal solution to $(CP_t(\bar{x}, \bar{\xi}))$ for some $\bar{\xi} \in \partial g(\bar{x}, t)$. This completes the proof of the theorem. □

3.2 Refined SCA Method with Valid Inequalities

In the previous subsection, only the convex underestimation (8) is used in the convex subproblem $(P_t(y,\xi))$ to approximate the DC constraint $\phi(x,t) \leq K$. We consider in this subsection a group of valid inequalities for improving the convex approximation problem $(P_t(y,\xi))$.

Recall that $\phi(x,t) = \sum_{i=1} \psi(x_i,t)$ and the convex underestimation of $\phi(x,t)$ is

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

where $w(x_i,y_i,\xi,t) = \sum_{i=1}^{n} \frac{1}{t}|x_i| - \frac{1}{t}[h(y_i,t) + \xi_i(x_i - y_i)]$ is the convex underestimation of $\psi(x_i,t)$.

By the definition of $\psi(x_i,t)$, it always holds $\psi(x_i,t) \leq 1$ $(i = 1,\ldots,n)$ and thus $\psi(x_i,t) \leq 1$ $(i = 1,\ldots,n)$ are redundant to the constraints in (P_t). However, the convex constraints $w(x_i,\xi,y_i,t) \leq 1$ $(i = 1,\ldots,n)$, which are the conservative approximation of $\psi(x_i,t) \leq 1$, provide valid inequalities 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 valid inequalities 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 $\xi = (1,0.5,0)^T \in \partial g(y,t)$. Then, the feasible set of $(P_t(y,\xi))$ is

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

The valid inequalities $w(x_i,\xi,y_i,t) \leq 1$ $(i = 1,2,3)$ have the following forms:

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

Notice that the first valid 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 valid inequalities is

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

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

Using these n valid inequalities, the convex approximation subproblem at y can be strengthened to

$$(P_t^v(y,\xi)) \qquad \min\ x^TQx + c^Tx$$

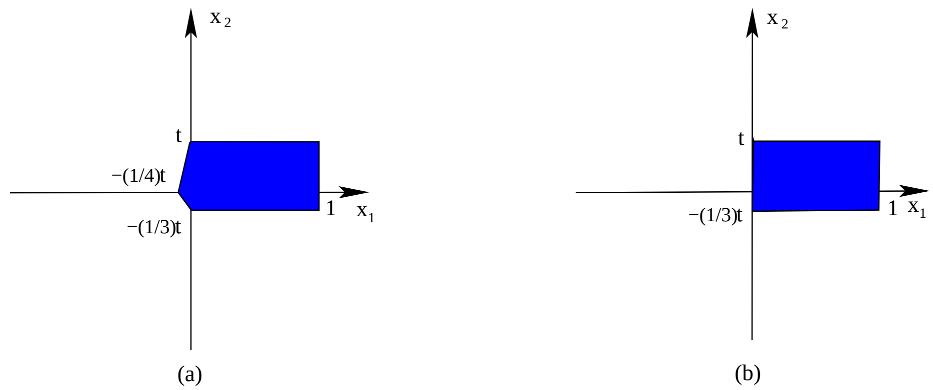

Figure 3: (a) Projection of set $F_t(y,\xi)$ to L; (b) Projection of set $F_t^v(y,\xi)$ to L

$$\begin{aligned} \text{s.t. } & Ax \le b, \\ & u(x,y,\xi,t) \le K, \\ & w(x_i,y_i,\xi_i,t) \le 1, \ i=1,\dots,n, \end{aligned}$$

which is equivalent to a convex quadratic program:

$$\begin{aligned} (CP_t^v(y,\xi)) \qquad & \min \ x^TQx + c^Tx \\ & \text{s.t. } Ax \le b, \\ & \quad \frac{1}{t}e^Tz - \frac{1}{t}[g(y,t) + \xi^T(x-y)] \le K, \\ & \quad \frac{1}{t}z_i - \frac{1}{t}[h(y_i,t) + \xi_i(x_i - y_i)] \le 1, \ i=1,\dots,n, \\ & \quad -x_i \le z_i \le x_i, \quad i=1,\dots,n. \end{aligned}$$

Accordingly, we have the following refined successive convex approximation method for (P).

Algorithm 2 (Refined SCA Method for (P))

Step 0. Choose a small $t > 0$ and a stopping parameter $\epsilon > 0$. Choose $x^0 \in F_t$ and $\xi^0 \in \partial g(x^0,t)$. Set $k := 0$.

Step 1. Solve the convex quadratic subproblem $(CP_t^v(x^k,\xi^k))$, where $\xi^k \in \partial g(x^k,t)$ is determined by (14). Let (x^{k+1},z^{k+1}) be an optimal solution to $(CP_t^v(x^k,\xi^k))$.

Step 2. If $\|x^{k+1} - x^k\| \le \epsilon$, stop.

Step 3. Set $k := k+1$ and go to Step 1.

Similar to Algorithm 1, we can prove that the sequence $\{x^k\}$ generated by Algorithm 2 converges to a KKT point of (P_t).

We point out that the KKT point computed by Algorithms 1 and 2 may not satisfy the cardinality constraint $\|x\|_0 \le K$. A simple heuristic can be used to search a feasible solution to (P) from a KKT point of (P_t). Let $\bar{x}$ be a KKT point of (P_t). Without loss of generality, assume that $|\bar{x}_1| \le |\bar{x}_2| \le \cdots \le |\bar{x}_n|$. We can set $x_i = 0$ for $i = 1, \ldots, n-K$ in (P) and resolve (P) as a reduced convex quadratic program of the variables $x_{n-K+1}, \ldots, x_n$. If the reduced convex quadratic program is feasible, then its optimal solution combined with $x_1 = \cdots = x_{n-K} = 0$ provides a feasible suboptimal solution to (P). For the unconstrained case of (P) where the general linear constraints $Ax \le b$ are absent in (P), this heuristic can always find a feasible suboptimal solution to (P). In general case, however, there is no guarantee for the above heuristic to find a feasible solution of (P). Notice that testing the feasibility of (P) is already NP-complete when A has three rows (see Bienstock (1996)).

We now give an illustrative example to demonstrate Algorithms 1 and 2.

Example 2 Consider the following problem

$$\begin{aligned} \min\, & q(x) := x_1^2 - 4x_1x_2 - 6x_1x_3 + 5x_2^2 + 8x_2x_3 + 17x_3^2 + x_3 \\ \text{s.t. } & \|x\|_0 \le 2, \\ & x \in [-1, 1]^3. \end{aligned}$$

The optimal solution of the example is $x^* = (-0.1875, 0, -0.0625)^T$ with $q(x^*) = -0.0312$.

We first solve the ℓ_1-norm approximation of the example:

$$\begin{aligned} \min\, & x_1^2 - 4x_1x_2 - 6x_1x_3 + 5x_2^2 + 8x_2x_3 + 17x_3^2 + x_3 \\ \text{s.t. } & \|x\|_1 \le 2, \\ & x \in [-1, 1]^3. \end{aligned}$$

The optimal solution of the above problem is $\bar{x}^0 = (-0.8750, -0.2500,\ -0.1250)^T$ with $q(x) = -0.0625$. Setting $x_3 = 0$ and solving the reduced quadratic program, we find a feasible solution $x^0 = (0, 0, 0)^T$ with $q(x^0) = 0$.

We now set $t = 0.01$ and $\epsilon = 10^{-6}$. Starting from x^0, Algorithm 1 stops at $\bar{x}^1 = (0, 0.01, -0.0318)^T$ after 3 iterations. Setting $x_1 = 0$ and resolving the reduced quadratic program, we obtain $x^1 = (0, 0.0290, -0.0362)^T$ with $q(x^1) = -0.0181$. On the other hand, starting from x^0, Algorithm 2 finds $\bar{x}^2 = (-0.01, 0, -0.01)^T$ after 2 iterations. Setting $x_2 = 0$ and resolving the reduced quadratic program, we obtain $x^2 = (-0.1875, 0, -0.0625)^T$ with $q(x^2) = -0.0312$, which is optimal to the example. We see from this example that the heuristic gives a better solution starting from the solution generated by Algorithm 2.

4. Computational Results

In this section, we conduct computational experiments to evaluate the performance of the SCA method (Algorithm2). Three types of test problems are considered in our experiments: portfolio selection problem, subset selection problem and compressed sensing problem. The main purpose of our computational experiments is to test the capability of the SCA method for finding good quality suboptimal solutions. We also compare the SCA method with the ℓ_1-norm approximation and the penalty decomposition (PD) method in Lu and Zhang (2012). Our preliminary numerical results suggest that the proposed successive convex approximation method is promising for generating suboptimal solutions better than those obtained from the ℓ_1-norm approximation and is competitive with the PD method for the three classes of test problems of (P).

4.1 Implementation issues

The SCA method (Algorithm2), the ℓ_1-norm approximation and the PD method are implemented in Matlab 7.9 and run on a PC (2.4G MHZ, 6GB RAM). The Matlab codes in our numerical test are available from `http://my.gl.fudan.edu.cn/teacherhome/xlsun/dcqp`. All the convex quadratic subproblems in our numerical experiments are solved by the QP solver in CPLEX 12.3 (see IBM ILOG CPLEX (2011)) with Matlab interface. In our implementation of the PD method in Lu and Zhang (2012), we use the Matlab code provided by the authors, which is available at: `http://people.math.sfu.ca/∼zhaosong/Codes/PD-sparse`.

In our implementation of the SCA method and the PD method, the initial feasible solutions are obtained by applying the simple heuristic described in the previous section from an optimal solution of the ℓ_1-norm approximation problem (P_1). This simple heuristic is also applied to the solutions computed by the SCA method and the PD method. It turns out that feasible solutions can be obtained by the heuristic for all the three types of test problems in our numerical experiment.

For each test problem, we run Algorithm 2 for different parameter t and record the best feasible solution found as the output of the suboptimal solution of the problem. More specifically, we start from $t = 0.5$ and reduce t by $t := t/3$ for each run of Algorithm 2. This process is terminated either when a feasible solution better than the initial feasible solution is found or when the number of runs reaches 100.

4.2 Portfolio selection problem

In this subsection, we consider cardinality constrained portfolio selection problem. Let μ and Q be the mean and covariance matrix of n risky assets, respectively. The cardinality constrained mean-variance portfolio selection problem can be formulated as

$$
\begin{aligned}
(MV) \qquad & \min \ x^T Q x \\
& \text{s.t} \ \text{card}(x) \le K, \\
& \quad \mu^T x \ge \rho, \\
& \quad e^T x = 1, \ 0 \le x \le u_i, \ i = 1, \ldots, n.
\end{aligned}
$$

To build the test problems of (MV), we use the 90 instances of portfolio selection created by Frangioni and Gentile (2007), 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`.

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

$$
\text{relative imp.} = \frac{f(x_{\ell_1}) - f(x^*)}{\max(1, |f(x_{\ell_1})|)} (\%), \tag{19}
$$

where $f(x) = x^T Q x$ and x_{ℓ_1} is the feasible solution obtained from the ℓ_1-norm approximation problem (P_1).

Table 1 summarizes the numerical results for test problems with $n = 200, 300, 400$ and $K = 40, 45, 50$. Some notations in Table 1 are explained as follows:

- "SCA" stands for the improved successive convex approximation method (Algorithm 2);

- "PD" stands for the penalty decomposition method in Lu and Zhang (2012);
- "Computing time" (in seconds) and "relative imp." are average results for the 30 instances of each n;
- "d" denotes the number of times the algorithm finds a better suboptimal solution (i.e., relative imp. > 0) for the 30 instances of each n.

We see from Table 1 that the SCA method can often improve the feasible solutions obtained from the ℓ_1-norm approximation, while the PD method fails to make improvement for any of the 90 instances with $K = 40, 45, 50$. It is also observed that the computation time of the SCA method is less than that of the PD method.

Table 1: Comparison results for portfolio selection problems

		SCA			PD		
n	K	Computing time	relative imp.	d	Computing time	relative imp.	d
200	40	2.33	0.73	7	6.68	0.0	0
300	40	4.22	0.20	3	14.01	0.0	0
400	40	7.78	0.28	3	30.25	0.0	0
200	45	2.64	1.16	11	6.28	0.0	0
300	45	4.72	1.03	10	13.07	0.0	0
400	45	9.10	1.34	14	27.80	0.0	0
200	50	3.31	0.77	9	6.03	0.0	0
300	50	5.53	1.13	10	12.27	0.0	0
400	50	10.95	2.05	18	25.89	0.0	0

4.3 Subset selection problem

In this subsection, we consider subset selection problems in multivariate regression (see Arthanari and Dodge (1993); Miller (2002)). For given m data points (a_i, b_i) with $a_i \in \Re^n$ and $b_i \in \Re$, the optimization model for the subset selection problem has the following form:

$$(SS) \qquad \min \ \|Ax - b\|_2^2 \quad \text{s.t. } \operatorname{card}(x) \le K,$$

where $A^T = (a_1, \ldots, a_m)$ and $b \in \Re^m$. Clearly, (SS) is a special form of problem (P) when the general linear constraints are absent. In practice, the number of data points m is usually larger than the data dimension n.

In our test, we generate the data in (SS) in the same fashion as Bertsimas and Shioda (2009). For a fixed n, we generate $m = 3n$ data points (a_i, b_i), $i = 1, \ldots, m$. The elements of a_i are generated from the normal distribution $N(0,1)$ and $b = A\bar{x} + \epsilon$, where $\epsilon = (\epsilon_1, \ldots, \epsilon_m)^T$ and ϵ_i is taken from the normal distribution $N(0,1)$, $i = 1, \ldots, m$. The elements of the vector $\bar{x}$ are from the uniform distribution in $[-1, 1]$ and $n - K$ elements of $\bar{x}$ are randomly set to zero. As in Bertsimas and Shioda (2009), a lower -100 and an upper bound 100 is set for each x_i $(i = 1, \ldots, n)$.

For each $n = 300, 400, 500$ with different cardinality K, we randomly generate 5 instances. The average CPU time and the average relative improvement over the ℓ_1-norm approximation are summarized in Fig. 4 for the SCA method and the PD method, where "relative imp." is defined in a similar way as (19). From Fig. 4, it is clear that the SCA method outperforms the PD method in terms of the quality of the feasible solutions obtained and the CPU time. We see that the average relative improvement of the SCA method over the ℓ_1-norm approximation ranges from 4% to 45% and the relative improvement has a tendency to decrease as the cardinality K increases. We also observe that in some cases the PD method is unable to improve the feasible solution obtained from the ℓ_1-norm approximation.

4.4 Compressed sensing problem

In this subsection, we consider the compressed sensing problem (see Bruckstein et al. (2009) and the references therein) in the following form:

$$
\begin{aligned}
(CS) \qquad & \min \|Ax - b\|_2^2 \\
& \text{s.t. } \operatorname{card}(x) \le K,
\end{aligned}
$$

where $A \in \Re^{m \times n}$ is a data matrix, $b \in \Re^m$ is an observation vector, and $1 \le K \le n$ is an integer for controlling the sparsity of the solution. In the compressed sensing problem, it is often assumed that $m \le n$. As in Lu and Zhang (2012), we consider two data sets for matrix A. In Data Set 1, the elements of A are from the standard Gaussian distribution with its row orthonormalized. In Data Set 2, the elements of A are directly drawn from the standard Gaussian distribution. In both data sets, the elements of $b \in \Re^m$ are from the standard Gaussian distribution. For both data sets, we use two subsets of test problems with $(m, n) = (256, 512)$ and $(m, n) = (256, 1024)$, respectively. For each subset of test problems, we randomly generate 10 instances for different cardinality K.

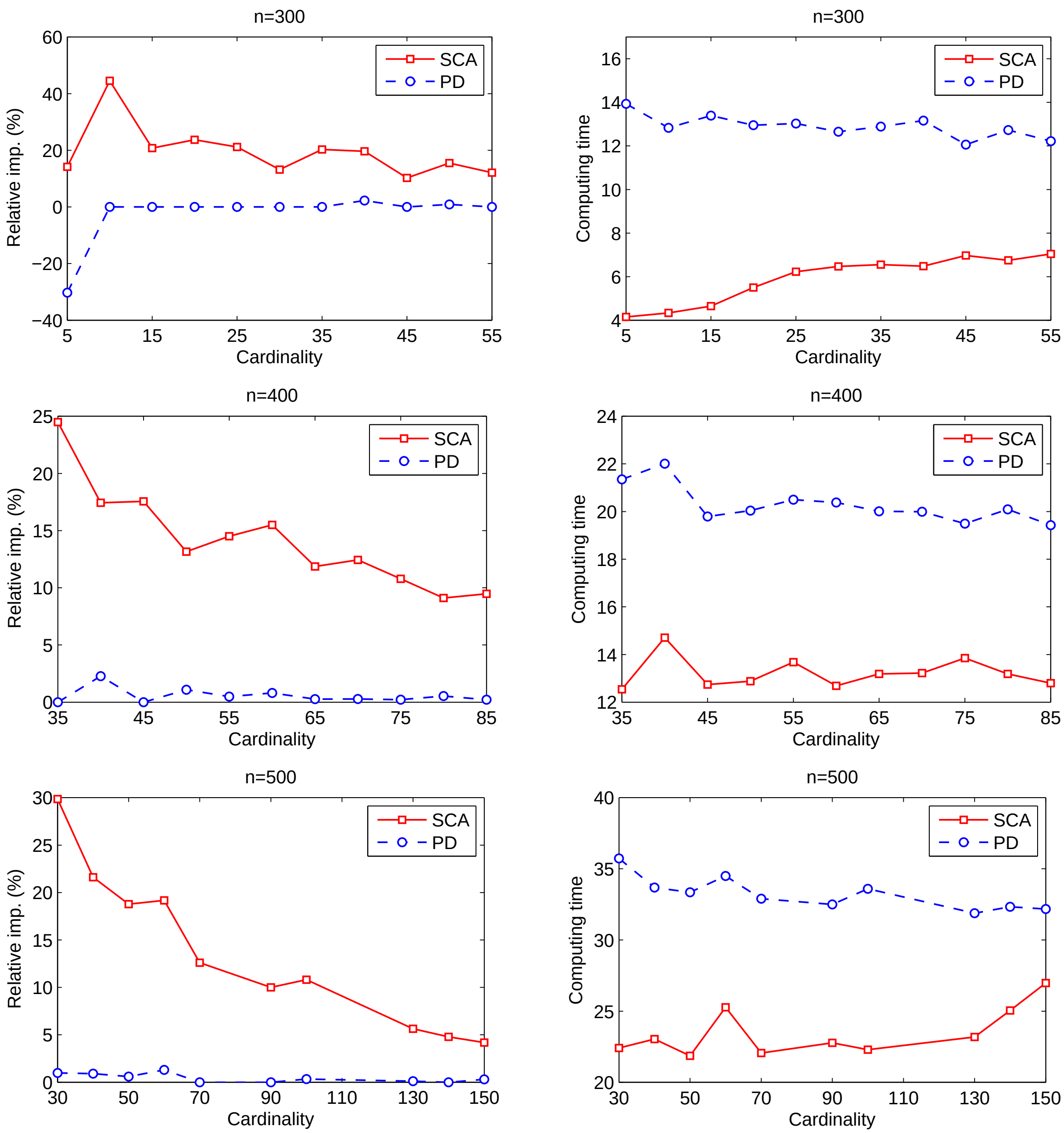

Figure 4: Relative improvement and computing time for subset selection problems with $n = 300, 400, 500$

The comparison results are summarized in Figs. 5 and 6 where "ℓ_1-norm" stands for the ℓ-norm approximation method and the residual values $\|Ax - b\|_2$ is used to measure the quality of the suboptimal solutions obtained from the three methods. From Fig. 5, we see that both the SCA method and the PD method outperform the ℓ_1-norm approximation in terms of the residual values achieved for different cardinality. The PD method tends to give better residual values than the SCA method for small cardinality whereas the SCA method slightly outperforms the PD method for large cardinality. We observe that when

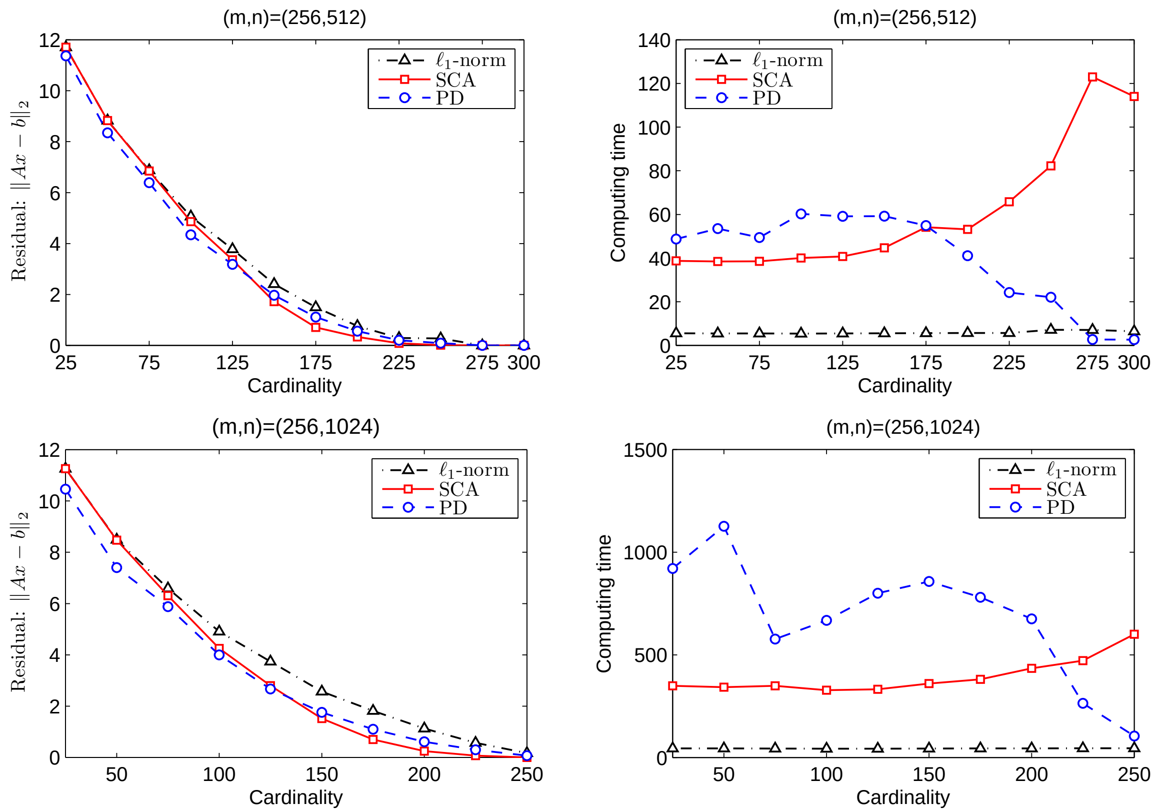

Figure 5: Residual values and computing time for compressed sensing problems from Data Set 1

the cardinality is larger than some threshold value, the three methods all give zero residual values. We also see from Fig. 5 that both the SCA method and the PD method spend much more time than the ℓ_1-norm approximation. This is because the ℓ_1-norm approximation only needs to solve a single convex quadratic program (P_1) while the SCA method and the PD method need to solve a convex quadratic program subproblem at each iteration of the algorithms. Fig. 6 shows the similar pattern as Fig. 5 except that the SCA method can achieve better residual values than the PD method and the ℓ_1-norm approximation for all the test problems. Finally, it is interesting to notice that for both data sets, the computing time of the SCA method tends to increase as the cardinality increases. In particular, the computing time of the SCA method grows rapidly when the cardinality is larger than some threshold value. This seems to suggest that the PD method is more efficient than the SCA method for problems with large cardinality.

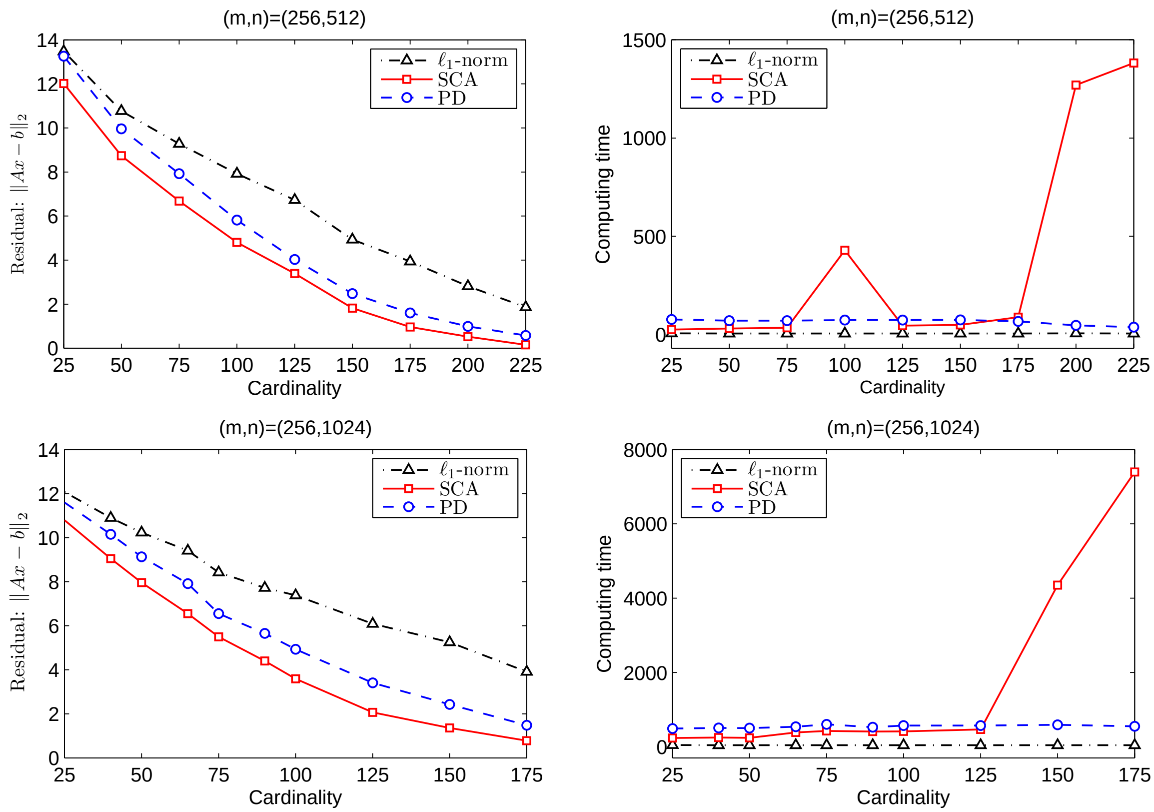

Figure 6: Residual values and computing time for compressed sensing problems from Data Set 2

5. Concluding Remarks

We have presented in this paper a successive convex approximation method for cardinality-constrained quadratic 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 quadratic 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 valid inequalities generated from the separable DC functions. Computational results on test problems of portfolio selection, subset selection and compressed sensing demonstrate that our method is promising in finding good quality suboptimal solution of the original problem and is competitive with the ℓ_1-norm approximation and the PD method.

One of the distinctive features of the proposed piecewise linear DC approximation to $\|x\|_0$ is its polyhedral property which can be exploited to construct tighter convex subproblems by linearizing the concave terms in the DC function during the course of iteration. Another contribution of the paper is the investigation of valid inequalities in constructing tighter convex subproblems for solving the DC approximation problem of (P). Our computational experiment suggests that adding valid inequalities derived from the piecewise linear DC approximation can significantly improve the performance of the algorithm.

Finally, we point out that the proposed method can be easily extended to deal with cardinality-constrained problems with a general convex objective function and convex constraints. The only difference is that the convex quadratic subproblem at Step 1 of Algorithms 1 and 2 will become then a general convex subproblem.

Acknowledgments

This research was supported by the NSFC grants 10971034, 11101092 and the Joint NSFC/RGC grant 71061160506, and by the RGC grant of Hong Kong 414207.

References

Arthanari, T. S., Y. Dodge. 1993. *Mathematical Programming in Statistics*. John Wiley & Sons, New York.

Bertsimas, D., R. Shioda. 2009. Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications* **43** 1–22.

Bienstock, D. 1996. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming* **74** 121–140.

Blog, B., G. van der Hoek, A. H. G. Rinnooy Kan, G. T. Timmer. 1983. The optimal selection of small portfolios. *Management Science* **29** 792–798.

Bonami, P., M. A. Lejeune. 2009. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations Research* **57** 650–670.

Bruckstein, A. M., D. L. Donoho, M. Elad. 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review* **51** 34–81.

Candes, E.J., M.B. Wakin, S.P. Boyd. 2008. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications* **14**(5) 877–905.

Chang, T. J., N. Meade, J. E. Beasley, Y. M. Sharaiha. 2000. Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research* **27** 1271–1302.

Chen, X., D. Ge, Z. Wang, Y. Ye. 2011. Complexity of unconstrained ℓ_2-ℓ_p minimization. Tech. rep., Stanford University. ArXiv:1105.0638v1.

Chen, X., F. Xu, Y. Ye. 2009. Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. Tech. rep., The Hong Kong Polytechnic University.

Clarke, F. H. 1983. *Optimization and Nonsmooth Analysis*. Wiley, New York.

Cui, X. T., X. J. Zheng, S. S. Zhu, X. L. Sun. 2012. Convex relaxations and MIQCQP reformulations for a class of cardinality-constrained portfolio selection problems. *Journal of Global Optimization* DOI: 10. 1007/s10898-012-9845-z.

Fazel, M., H. Hindi, S.P. Boyd. 2003. Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. *Proceedings of American Control Conference*. 2156–2162.

Frangioni, A., C. Gentile. 2006. Perspective cuts for a class of convex 0–1 mixed integer programs. *Mathematical Programming* **106** 225–236.

Frangioni, A., C. Gentile. 2007. SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Operations Research Letters* **35** 181–185.

Frangioni, A., C. Gentile. 2009. A computational comparison of reformulations of the perspective relaxation: SOCP vs. cutting planes. *Operations Research Letters* **37** 206–210.

Ge, D., X. Jiang, Y. Ye. 2011. A note on complexity of ℓ_p minimization. *Mathematical Programming* **129** 285–299.

Gulpinar, N., H. A. Le Thi, M. Moeini. 2010. Robust investment strategies with discrete asset choice constraints using DC programming. *Optimization* **59** 45–62.

Günlük, O., J. Linderoth. 2010. Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Mathematical Programming* **124** 183–205.

Hong, L.J., Y. Yang, L.W. Zhang. 2011. Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach. *Operations Research* **59** 617–630.

IBM ILOG CPLEX. 2011. *IBM ILOG CPLEX 12. 3 User's Manual for CPLEX*.

Jacob, N. L. 1974. A limited-diversification portfolio selection model for the small investor. *Journal of Finance* **29** 847–856.

Jobst, N.J., M.D. Horniman, C.A. Lucas, G. Mitra. 2001. Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. *Quantitative Finance* **1** 489–501.

Le Thi, H. A., T. Pham Dinh. 2005. The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research* **133** 23–46.

Li, D., X. L. Sun, J. Wang. 2006. Optimal lot solution to cardinality constrained mean-variance formulation for portfolio selection. *Mathematical Finance* **16** 83–101.

Lu, Z., Y. Zhang. 2010. Penalty decomposition methods for rank minimization. Tech. rep., Department of Mathematics, Simon Fraser University. Available at: http://www.optimization-online.org/DB_HTML/2010/08/2718.html.

Lu, Z., Y. Zhang. 2012. Sparse approximation via penalty decomposition methods. Tech. rep., Department of Mathematics, Simon Fraser University. Available at: http://www.optimization-online.org/DB_HTML/2010/08/2719.html.

Maringer, D., H. Kellerer. 2003. Optimization of cardinality constrained portfolios with a hybrid local search algorithm. *OR Spectrum* **25** 481–495.

Miller, A. J. 2002. *Subset Selection in Regression*. Chapman and Hall. Second edition.

Murray, W., H. Shek. 2011. A local relaxation method for cardinality constrained portfolio optimization problem. Tech. rep., Stanford University. Available at SSRN: http://ssrn.com/abstract=1699527.

Nocedal, J., S. J. Wright. 1999. *Numerical Optimization*. Springer Verlag.

Pham Dinh, T., H. A. Le Thi. 1997. Convex analysis approach to DC programming: Theory, algorithms and applications. *Acta Mathematics Vietnam* **22** 289–355.

Rockafellar, R. T., R. J. B. Wets. 2004. *Variational Analysis*. Springer.

Shaw, D. X., S. Liu, L. Kopman. 2008. Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. *Optimization Methods & Software* **23** 411–420.

Sriperumbudur, B.K., D.A. Torres, G.R.G. Lanckriet. 2011. A majorization-minimization approach to the sparse generalized eigenvalue problem. *Machine learning* **85** 3–39.

Vielma, J. P., S. Ahmed, G. L. Nemhauser. 2008. A lifted linear programming branch-and-bound algorithm for mixed integer conic quadratic programs. *INFORMS Journal of Computing* **20** 438–450.

Weston, J., A. Elisseeff, B. Schölkopf, M. Tipping. 2003. Use of the zero norm with linear models and kernel methods. *The Journal of Machine Learning Research* **3** 1439–1461.

Woodside-Oriakhi, M., C. Lucas, J. E. Beasley. 2011. Heuristic algorithms for the cardinality constrained efficient frontier. *European Journal of Operational Research* **213** 538–550.

Zheng, X. J., X. L. Sun, D. Li. 2011. Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach. Tech. rep. Available at: http://www.optimization-online.org/DB_HTML/2010/11/2797.html.