

# Recent Advances in Mathematical Programming with Semi-continuous Variables and Cardinality Constraint

Xiaoling Sun · Xiaojin Zheng · Duan Li

Received: 15 December 2012 / Revised: 19 February 2013 / Accepted: 26 February 2013 /  
Published online: 14 March 2013  
© Operations Research Society of China, Periodicals Agency of Shanghai University, and  
Springer-Verlag Berlin Heidelberg 2013

**Abstract** Mathematical programming problems with semi-continuous variables and cardinality constraint have many applications, including production planning, portfolio selection, compressed sensing and subset selection in regression. This class of problems can be modeled as mixed-integer programs with special structures and are in general NP-hard. In the past few years, based on new reformulations, approximation and relaxation techniques, promising exact and approximate methods have been developed. We survey in this paper these recent developments for this challenging class of mathematical programming problems.

**Keywords** Semi-continuous variables · Cardinality and sparsity constraint · Mixed-integer 0–1 quadratic programming · Perspective reformulation · Lagrangian decomposition · Approximate methods

---

This research was supported by the National Natural Science Foundation of China grants (Nos. 11101092, 10971034) and the Joint National Natural Science Foundation of China/Research Grants Council of Hong Kong grant (No. 71061160506), and by the Research Grants Council of Hong Kong grants (Nos. CUHK414808 and CUHK414610).

X.L. Sun (✉)

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

X.J. Zheng

School of Economics and Management, Tongji University, Shanghai 200092, P.R. China  
e-mail: [xjzheng@tongji.edu.cn](mailto:xjzheng@tongji.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](mailto:dli@se.cuhk.edu.hk)

### 1 Introduction

In many real-world applications of mathematical programming models, the continuous decision variables,  $x_i, i = 1, \dots, n$ , have to be confined in a disconnected set,  $\{0\} \cup [l_i, u_i], i = 1, \dots, n$ , and the number of nonzero variables of  $x_i$  has to be upper bounded by a certain number  $K < n$ . In general, a variable  $x_i \in \{0\} \cup [l_i, u_i]$  is referred to as *semi-continuous* variable, where  $0 \in [l_i, u_i]$  is also allowed, and the constraint for controlling the number of nonzero variables is referred to as cardinality constraint. In the production planning, the semi-continuous variables can be used to describe the state of a production process that is either turned off (inactive), hence nothing is produced, or turned on (active) such that the amount of the production has to lie in certain interval due to managerial and technological considerations. In portfolio selection optimization models, the semi-continuous variables are closely related to the so called *minimum buy-in threshold* which prevents the investors from holding a small position for assets, while the cardinality constraint limits the total number of different assets in the optimal portfolio due to the transaction cost and managerial concerns. Cardinality constraint is particularly important in portfolio management using index tracking strategy where a market benchmark index is tracked by a small group of assets.

The concepts of semi-continuous variables and cardinality constraint can be generalized to a decision vector of the form  $x = (x_1, \dots, x_p)^T \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}$  with  $n_1 + \dots + n_p = n$ . The sub-vector  $x_i \in \mathbb{R}^{n_i}$  is called semi-continuous if

$$x_i \in \mathcal{P}_i = \{x_i \in \mathbb{R}^{n_i} \mid A_i x_i \leq b_i y_i, y_i \in \{0, 1\}\}, \tag{1}$$

where  $A_i \in \mathbb{R}^{m_i \times n_i}, b_i \in \mathbb{R}^{m_i}$  and  $A_i x_i \leq 0$  implies  $x_i = 0$ . For  $x = (x_1, \dots, x_p)^T \in \mathbb{R}^n$ , the cardinality constraint for  $x$  can be expressed as

$$\text{card}(x) \leq K, \tag{2}$$

where  $\text{card}(x)$  is defined as the number of  $i \in \{1, \dots, p\}$  such that  $x_i \neq 0$ . Using the binary variable  $y_i \in \{0, 1\}$  in (1), the cardinality constraint (2) is equivalent to  $e^T y \leq K$ , where  $e$  is the column vector of all ones and  $y \in \{0, 1\}^p$ .

A general formulation of mathematical programming with semi-continuous variables and cardinality constraint can be expressed as the following mixed-integer programming:

$$\begin{aligned} \text{(P)} \quad & \min F(x, y, z) \\ & \text{s.t. } (x, y, z) \in \Omega, \\ & e^T y \leq K, \\ & A_i x_i \leq b_i y_i, \quad y_i \in \{0, 1\}, i = 1, \dots, p, \end{aligned}$$

where  $F(x, y, z): (\mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_p}) \times \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$  is a convex function, the set  $\Omega \subseteq \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$  represents the general constraints for  $(x, y, z)$ , and  $K$  is an integer with  $1 \leq K \leq p$ . Problem (P) is, in general, NP-hard as testing the feasibility of (P) with linear constraints is already NP-complete (see [7]). We point out that there

are polynomially solvable cases of (P); for instance, a class of polynomially solvable cardinality-constrained quadratic optimization problems was identified in [28].

In this paper, we survey some of the recent advances in optimization problems with semi-continuous variables and cardinality constraint. This survey is motivated by continuously increasing interests in problem (P) in recent years from operations research, management science, finance engineering and engineering communities. In the sequel, we only consider the situation when the objective function of (P) has the following separable form:

$$F(x, y, z) = f(x) + c^T y + g(z), \quad (3)$$

where  $f(x)$  and  $g(z)$  are convex cost functions of  $x$  and  $z$ , and  $c \in \mathfrak{R}^n$  is the vector of fixed cost coefficients associated with the semi-continuous variables.

In the next section, we describe some examples of problem (P) arising from different real-world applications. One of the efficient techniques for improving the continuous relaxation of (P) is the construction of convex envelope of a univariate function over semi-continuous variable using perspective function. In Sect. 3, we describe perspective reformulation of problem (P) when  $f(x)$  in (3) is a separable function. In particular, we discuss two tractable perspective reformulations: SOCP reformulation and perspective cut reformulation. A dual method for deriving the perspective reformulation is also presented. In Sect. 4, we focus on the quadratic case of (P). We give an SDP approach for computing the “best” diagonal decomposition of quadratic objective function. Specialized branch-and-bound and branch-and-cut methods for the quadratic case of (P) are also discussed. In Sect. 5, we discuss various approximate methods and techniques for dealing with the cardinality or sparse constraint. Finally, we give some concluding remarks in Sect. 6.

## 2 Examples of Applications

In this section, we describe some examples of problem (P) arising from portfolio selection, compressed sensing and subset selection, quadratic uncapacitated facility location problem and unit commitment problem in power system.

*Example 1 (Portfolio selection)* Suppose there are  $n$  risky assets in the financial market with random return vector  $R = (R_1, \dots, R_n)^T$ . The expected return vector and the covariance matrix of  $R$  are given as  $\mu$  and  $Q$ , respectively. A mean-variance portfolio selection model with cardinality and minimum buy-in threshold can be modeled as

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

where  $x_i \in \mathfrak{R}$  represents the proportion of the total capital invested in the  $i$ th asset,  $\rho$  is a prescribed return level set by the investor,  $\alpha_i \in (0, 1)$  represents the minimum transaction amount of the  $i$ th asset, and  $u_i \in (\alpha_i, 1]$  is the maximum position of the  $i$ th asset.

Portfolio selection problems with cardinality and minimum threshold constraints have been studied by many researchers (see, e.g., [6, 7, 9, 11, 16, 29, 37, 44]). The model (MV) has been also investigated in finance literature in the context of limited-diversification, small portfolios and portfolio selection models with real features (see, e.g., [8, 33, 34, 42, 46]). In the literature, there are also heuristic procedures based on metaheuristics such as genetic algorithms, tabu search, and simulated annealing for problem (MV) (see, e.g., [12, 15, 20, 40, 43]). However, these metaheuristics do not guarantee to find the optimal or even a satisfactory near-optimal solution of (MV).

*Example 2* (Compressed sensing and subset selection) Compressed sensing is an important problem in signal processing (see, e.g., [10] and the references therein). The problem can be formulated as

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

where  $A \in \mathfrak{R}^{m \times n}$  is a data matrix,  $b \in \mathfrak{R}^m$  is an observation vector, and  $1 \leq K \leq n$  is an integer for controlling the sparsity of the solution. In the compressed sensing problem, it is often assumed that  $m < n$ .

In multivariate linear regression, we are given  $m$  observed data points  $(a^i, b^i)$  with  $a^i \in \mathfrak{R}^n$  and  $b^i \in \mathfrak{R}$ . The goal is to minimize the least square measure of  $\sum_{i=1}^m (a_i^T x - b_i)^2$  with only a subset of the prediction variables in  $x$ . This subset selection problem then has the same form as (CS) (see [3, 6, 41]). In contrast with the case of compressed sensing, the number of data in subset selection is often much larger than the dimension of the data ( $m > n$ ).

In practice, we can always impose lower bound and upper bound on  $x$ , i.e.,  $-l_i \leq x_i \leq u_i$ ,  $i = 1, \dots, n$ , for some sufficiently large positive numbers  $l_i$  and  $u_i$ . Thus, (CS) is a special case of (P).

*Example 3* (Separable quadratic UFL) Given a set of customers  $N = \{1, \dots, n\}$  and a set of facilities  $M = \{1, \dots, m\}$ . A fixed cost  $c_i$  occurs if facility  $i \in M$  is opened and the total number of opened facilities should be no more than  $1 \leq K \leq m$ . All customers have unit demand that can be satisfied from the opened facilities. Let  $y_i \in \{0, 1\}$  indicate whether or not facility  $i$  is opened, and  $x_{ij}$  denote the fraction of demand of customer  $j$  satisfied from facility  $i$ . The transportation cost is defined by  $q_{ij}x_{ij}^2$ . The separable quadratic uncapacitated facility location problem [31] can be then formulated as

$$\text{(SQUFL)} \quad \min \sum_{i=1}^m \sum_{j=1}^n q_{ij}x_{ij}^2 + \sum_{i=1}^m c_i y_i$$

$$\begin{aligned} \text{s.t. } & 0 \leq x_{ij} \leq y_i, \quad i = 1, \dots, m, j = 1, \dots, n, \\ & e^T y \leq K, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, m, \\ & \sum_{i=1}^m x_{ij} = 1, \quad j = 1, \dots, n. \end{aligned}$$

The problem (SQUFL) is a generalization of the classical linear uncapacitated facility location problem in combinatorial optimization. We see that the decision vector can be written as  $x = (x_1, \dots, x_m) \in \mathfrak{H}^{mn}$ , where  $x_i = (x_{i1}, \dots, x_{in})^T$  for  $i = 1, \dots, m$ . Since  $0 \leq x_i \leq y_i e$ ,  $x_i$  is semi-continuous for  $i = 1, \dots, m$ .

*Example 4* (Unit commitment problem) This is a problem arising from electrical power production (see [21, 25]). Given a set  $I$  of thermal generating units, for each  $i \in I$ , the unit is either turned off or turned on with the power output lying between  $p_{\min}^i$  and  $p_{\max}^i$ . The cost function of the  $i$ th unit is defined by a quadratic function of the power output  $p$ :  $f^i(p) = a^i p^2 + b^i p + c^i$ ,  $i \in I$ . The power demand over time period  $t \in T$  is  $d_t$ , where  $T$  is the set of time periods in the planning horizon. The unit commitment problem is to generate power to meet the demand while minimizing the total cost. The operation of thermal units has to satisfy certain minimum up- and down-time constraints: whenever a unit is turned on or turned off it must remain committed or decommitted for a certain time. Let  $u_{it} \in \{0, 1\}$  represent the status of commitment of unit  $i$  at time period  $t$ , and  $x_{it} \geq 0$  the power output of unit  $i$  at time period  $t$ . Let  $U$  denote the set of  $u_{it}$  satisfying the minimum up- and down-time constraints. The basic unit commitment problem can be then formulated as a mixed-integer separable quadratic problem:

$$\begin{aligned} \text{(UC)} \quad \min & \sum_{i \in I} \sum_{t \in T} a^i x_{it}^2 + b^i x_{it} + c^i y_{it} \\ \text{s.t. } & \sum_{i \in I} x_{it} = d_t, \quad t \in T, \\ & p_{\min}^i u_{it} \leq x_{it} \leq p_{\max}^i u_{it}, \quad i \in I, t \in T, \\ & u = (u_{it}) \in U, \quad u_{it} \in \{0, 1\}, \quad i \in I, t \in T. \end{aligned}$$

It is clear that (UC) is a special case of (P) with only semi-continuous variables. More realistic constraints can be further attached to the above basic formulation (see [25]).

### 3 Perspective Reformulations for Mathematical Programming with Semi-continuous Variables

In this section, we consider the following special case of (P):

$$\text{(P}_s) \quad \min \sum_{i=1}^n f_i(x_i) + c^T y + g(z)$$

$$\begin{aligned} \text{s.t. } & Ax + By + Cz \leq b, \\ & l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, n, \end{aligned}$$

where  $f_i$  ( $i = 1, \dots, n$ ) are univariate convex functions,  $A$ ,  $B$ ,  $C$  are matrices with appropriate dimensions. Note that the cardinality constraint  $\sum_{i=1}^n y_i \leq K$  can be included in the linear constraints of  $(x, y, z)$ . A novel reformulation technique called *perspective reformulation* was proposed by Frangioni and Gentile [21–23] (see also [31]). This reformulation method is based on constructing convex envelope of a univariate function over the semi-continuous variable and gives rise to much more efficient mixed-integer program reformulation than the standard formulation ( $P_s$ ).

### 3.1 Convex Envelope and Perspective Reformulation

Consider the following one-dimensional minimization over a semi-continuous variable:

$$\begin{aligned} \min & h(s) + dt \\ \text{s.t. } & \alpha t \leq s \leq \beta t, \quad t \in \{0, 1\}, \end{aligned} \quad (4)$$

where  $h : \Re \rightarrow \Re$  is a convex function. This problem can be equivalently restated as

$$\min \{ \hat{h}(s, t) \mid (s, t) \in \Re^2 \}, \quad (5)$$

where

$$\hat{h}(s, t) = \begin{cases} 0, & \text{if } s = t = 0, \\ h(s) + d, & \text{if } s \in [\alpha, \beta], t = 1, \\ +\infty, & \text{otherwise.} \end{cases} \quad (6)$$

The convex envelope of  $\hat{h}(s, t)$  is defined by

$$\overline{\text{co}}(\hat{h})(s, t) = \inf \{ v \mid (s, t, v) \in \text{conv}(\text{epi}(\hat{h})) \},$$

where  $\text{epi}(\hat{h})$  is the epigraph defined by  $\text{epi}(\hat{h}) = \{(s, t, v) \mid v \geq \hat{h}(s, t)\}$ . By (6), the convex hull of  $\text{epi}(\hat{h})$  consists of all the points of the following form:

$$(1 - \theta)(0, 0, \bar{w}) + \theta(\bar{s}, 1, \bar{v}) = (\theta\bar{s}, \theta, (1 - \theta)\bar{w} + \theta\bar{v}),$$

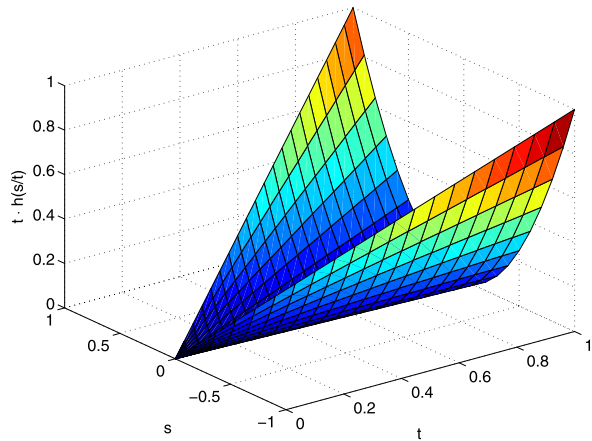
where  $\bar{w} \geq 0$ ,  $\bar{v} \geq h(\bar{s}) + d$ ,  $\bar{s} \in [\alpha, \beta]$  and  $\theta \in [0, 1]$ . Letting  $t = \theta$  and  $s = \theta\bar{s} = t\bar{s}$ , we have  $\bar{s} = s/t$  and  $\alpha t \leq s \leq \beta t$ . Since  $\bar{w} \geq 0$  and  $\bar{v} \geq h(\bar{s}) + d$ , we have

$$(1 - \theta)\bar{w} + \theta\bar{v} \geq \theta(h(\bar{s}) + d) = t \cdot h(s/t) + dt$$

for  $t \in (0, 1]$ . Thus,

$$\overline{\text{co}}(\hat{h})(s, t) = \begin{cases} 0, & \text{if } s = t = 0, \\ t \cdot h(s/t) + dt, & \text{if } \alpha t \leq s \leq \beta t, t \in (0, 1], \\ +\infty, & \text{otherwise.} \end{cases}$$

**Fig. 1** The perspective function of  $h(s) = s^2$  on  $[-1, 1]$



This function is continuous at  $(0, 0)$  by defining  $0/0 := 0$ . The function  $t \cdot h(s/t)$  is called *perspective function* of  $h(s)$  in convex analysis [32], see Fig. 1. Replacing  $\hat{h}(s, t)$  in (5) by  $\overline{co}(\hat{h})(s, t)$ , we obtain the following equivalent form of (4):

$$\begin{aligned} \min \psi(s, t) &:= t \cdot h(s/t) + dt \\ \text{s.t. } \alpha t &\leq s \leq \beta t, \quad t \in \{0, 1\}. \end{aligned} \tag{7}$$

Problem (7) is hence called the perspective reformulation of problem (4). Its continuous relaxation is

$$\begin{aligned} \min \psi(s, t) &:= t \cdot h(s/t) + dt \\ \text{s.t. } \alpha t &\leq s \leq \beta t, \quad t \in [0, 1]. \end{aligned} \tag{8}$$

Since the objective function of (8) is the convex envelope of  $h(s) + dt$ , (8) is tighter than the direct continuous relaxation of (4). Actually, since there are no other constraints in (4), the optimal values of (4) and (8) are the same.

Now, we apply the one-dimensional perspective reformulation (7) to problem  $(P_s)$  which has a separable objective function. The resulting perspective reformulation is

$$\begin{aligned} (\text{PR}_s) \quad \min \quad & \sum_{i=1}^n y_i f_i(x_i/y_i) + c^T y + g(z) \\ \text{s.t.} \quad & Ax + By + Cz \leq b, \\ & l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, n. \end{aligned}$$

The problem  $(\text{PR}_s)$  is more efficient than  $(P_s)$  in the sense that the continuous relaxation of  $(\text{PR}_s)$  is tighter than that of  $(P_s)$ .

The above perspective reformulation technique can be extended to problem which has an additional nonseparable term  $p(x)$  in the objective function. In fact, we can introduce a copy constraint  $x = \tilde{z}$  and rewrite the objective function as  $\sum_{i=1}^n f_i(x_i) + c^T y + g(z) + p(\tilde{z})$ . The perspective reformulation is then applicable to the resulting

problem. In general, we can use perspective reformulation for minimization of any convex function over semi-continuous variables provided that the objective function has a separable term  $\sum_{i=1}^n f_i(x_i)$ .

### 3.2 A Dual Approach for Perspective Reformulation

In this subsection, we describe a dual approach for deriving the perspective reformulation (PR<sub>s</sub>). Without loss of generality, we consider a simple version of (P<sub>s</sub>):

$$\begin{aligned} \min \quad & \sum_{i=1}^n f_i(x_i) + c^T y \\ \text{s.t.} \quad & Ax \leq b, \\ & l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, n. \end{aligned} \quad (9)$$

Dualizing the first constraint of (9) yields the following Lagrangian relaxation:

$$\begin{aligned} d(\lambda) = \min \quad & \sum_{i=1}^n f_i(x_i) + c^T y + \lambda^T (Ax - b) \\ \text{s.t.} \quad & l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, n. \end{aligned}$$

The Lagrangian dual of problem (9) then takes the following form,

$$(D_s) \quad \max_{\lambda \geq 0} d(\lambda).$$

Next, we show that the dual problem (D<sub>s</sub>) is equivalent to the continuous relaxation of (PR<sub>s</sub>). For each  $i = 1, \dots, n$ , let  $q_i = \min_{l_i \leq x_i \leq u_i} (f_i(x_i) + \lambda^T a_i x_i)$ , where  $a_i$  is the  $i$ th column of  $A$ . Then,

$$d(\lambda) = -\lambda^T b + \sum_{i=1}^n \min(0, q_i + c_i). \quad (10)$$

Since  $f_i$  is convex, by strong duality, we have

$$\begin{aligned} q_i &= \max_{(\zeta_i, \eta_i) \geq 0} \min_{x_i \in \mathcal{D}_i} [f_i(x_i) + \lambda^T a_i x_i + \eta_i (x_i - u_i) + \zeta_i (l_i - x_i)] \\ &= \max_{(\zeta_i, \eta_i) \geq 0} r_i(\zeta_i, \eta_i, \lambda), \end{aligned} \quad (11)$$

where  $\zeta_i$  and  $\eta_i$  are the multipliers for the inequalities  $l_i \leq x_i$  and  $x_i \leq u_i$ , respectively, and  $r_i(\zeta_i, \eta_i, \lambda)$  is the corresponding dual function. It then follows from (10) and (11) that

$$v(D_s) = \max_{\lambda \geq 0} \left[ -\lambda^T b + \sum_{i=1}^n \min(0, q_i + c_i) \right]$$



$$\begin{aligned}
 &= \max_{\lambda \geq 0} -\lambda^T b - \sum_{i=1}^n s_i \\
 &\quad \text{s.t.} \quad -s_i \leq 0, \quad -s_i \leq q_i + c_i, \quad i = 1, \dots, n, \\
 &= \max_{\lambda \geq 0} -\lambda^T b - \sum_{i=1}^n s_i \\
 &\quad \text{s.t.} \quad -s_i \leq 0, \quad -s_i - c_i \leq \max_{(\zeta_i, \eta_i) \geq 0} r_i(\zeta_i, \eta_i, \lambda), \quad i = 1, \dots, n, \\
 &= \max_{\lambda \geq 0} -\lambda^T b - \sum_{i=1}^n s_i \\
 &\quad \text{s.t.} \quad -s_i - c_i \leq r_i(\zeta_i, \eta_i, \lambda), \quad i = 1, \dots, n, \\
 &\quad (\lambda, s, \zeta, \eta) \geq 0.
 \end{aligned} \tag{12}$$

Note that problem (12) is a convex program since  $r_i(\zeta_i, \eta_i, \lambda)$  is a concave function. It is obvious that the Slater condition holds for (12). Dualizing the first constraint in problem (12) with multiplier  $y_i \geq 0$  and using strong duality, we have

$$\begin{aligned}
 v(D_s) &= \min_{y \geq 0} \max_{(\lambda, s, \zeta, \eta) \geq 0} \left\{ -\lambda^T b + \sum_{i=1}^n [-s_i + y_i(s_i + c_i + r_i(\zeta_i, \eta_i, \lambda))] \right\} \\
 &= \min_{y \in [0, 1]^n} \max_{(\lambda, \zeta, \eta) \geq 0} \left\{ -\lambda^T b + c^T y + \sum_{i=1}^n y_i r_i(\zeta_i, \eta_i, \lambda) \right\} \\
 &= \min_{y \in [0, 1]^n} \left\{ c^T y + \max_{(\lambda, \zeta, \eta) \geq 0} \min_{x \in \mathcal{R}^n} \left[ \lambda^T \left( \sum_{i=1}^n y_i a_i x_i - b \right) \right. \right. \\
 &\quad \left. \left. + \sum_{i=1}^n (y_i f_i(x_i) + y_i \eta_i(x_i - u_i) + y_i \zeta_i(l_i - x_i)) \right] \right\}.
 \end{aligned}$$

Applying again the strong duality to the inner maximization of the above problem, we obtain

$$\begin{aligned}
 v(D_s) &= \min \sum_{i=1}^n y_i f_i(x_i) + c^T y \\
 &\quad \text{s.t.} \quad \sum_{i=1}^n a_i y_i x_i \leq b, \\
 &\quad l_i y_i \leq y_i x_i \leq u_i y_i, \quad i = 1, \dots, n, \\
 &\quad y \in [0, 1]^n.
 \end{aligned}$$

Letting  $z_i = y_i x_i$  makes  $x_i = z_i / y_i$ . Substituting  $x_i = z_i / y_i$  into the above problem yields exactly the continuous relaxation of (PR<sub>s</sub>) for problem (9).

The above discussion reveals that the lower bound provided by the continuous relaxation of the perspective reformulation (PR<sub>s</sub>) is the same as the dual bound of (P<sub>s</sub>), which is usually much tighter than the direct continuous relaxation of the original problem (P<sub>s</sub>).

### 3.3 SOCP Reformulation and Perspective Cut Reformulation

Although the perspective reformulation (PR<sub>s</sub>) is, in theory, tighter than the original mixed-integer formulation, the nonlinear term  $y_i f_i(x_i/y_i)$  makes the objective more nonlinear and intractable even for simple nonlinear functions such as quadratic functions. In this subsection, we describe two tractable reformulations derived from the basic perspective reformulation.

The second-order cone programming (SOCP) reformulation [1, 31, 45] is obtained by introducing an additional variable  $\phi_i = y_i f_i(x_i/y_i)$  for each  $i$ . The problem (PR<sub>s</sub>) can be then restated as

$$\begin{aligned}
 (\text{SOCP}_s) \quad & \min \sum_{i=1}^n \phi_i + c^T y + g(z) \\
 \text{s.t.} \quad & Ax + By + Cz \leq b, \\
 & \phi_i \geq y_i f_i(x_i/y_i), \quad i = 1, \dots, n, \\
 & l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, n.
 \end{aligned}$$

If the constraints  $\phi_i \geq y_i f_i(x_i/y_i)$  is SOCP-representable, then the continuous relaxation of problem (SOCP<sub>s</sub>) is an SOCP problem that can be solved efficiently by interior-point methods [2, 5]. For instance, consider the quadratic case where  $f_i(x_i) = a_i x_i^2 + b_i x_i$  with  $a_i > 0$ . Then,  $y_i f_i(x_i/y_i) = a_i x_i^2/y_i + b_i x_i$ . Problem (PR<sub>s</sub>) can thus be equivalently rewritten as

$$\begin{aligned}
 \min \quad & \sum_{i=1}^n (\phi_i + b_i x_i) + c^T y + g(z) \\
 \text{s.t.} \quad & Ax + By + Cz \leq b, \\
 & \phi_i y_i \geq a_i x_i^2, \quad \phi_i \geq 0, \quad i = 1, \dots, n, \\
 & l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, n.
 \end{aligned} \tag{13}$$

It is easy to see that  $\phi_i y_i \geq a_i x_i^2$ ,  $\phi_i \geq 0$ ,  $y_i \geq 0$  can be represented by the following SOCP constraint:

$$\left\| \begin{array}{c} \sqrt{a_i} x_i \\ \frac{\phi_i - y_i}{2} \end{array} \right\| \leq \frac{\phi_i + y_i}{2}, \quad i = 1, \dots, n.$$

Thus, problem (13) is a mixed-integer SOCP reformulation of (P<sub>s</sub>). The standard MIQP solvers such as CPLEX can be used to solve (13). Computational results in [21, 31, 48] showed that the SOCP reformulation is more efficient than the standard MIQP formulation (P<sub>s</sub>).

An alternative way in deriving tractable reformulation from  $(PR_s)$  is the perspective cut reformulation proposed in [21–23]. Recall that the convex envelope of the extended univariate function  $\hat{h}(s, t)$  defined in (6) is given by  $\psi(s, t) = t \cdot h(s/t) + dt$  when  $\alpha t \leq s \leq \beta t$  and  $t \in (0, 1]$ . Assume that  $h$  is differentiable. Then,

$$\nabla \psi(s, t) = \left[ h'(s/t), h(s/t) - \frac{s}{t}h'(s/t) + d \right]^T.$$

Note that  $\nabla \psi(s, t)$  only depends on  $s/t$  and hence is constant on the line  $\{(\bar{s}t, t) \mid t \in [0, 1]\}$  for any  $\bar{s} \in [\alpha, \beta]$ . Recall from the convex analysis that the convex function  $\psi(s, t)$  is fully characterized by its tangent planes. For a given point  $(\bar{s}, \bar{t}) \in \text{dom}(\hat{h})$ , the tangent plane of  $\psi$  is

$$v = \psi(\bar{s}, \bar{t}) + \nabla^T \psi(\bar{s}, \bar{t})[(s, t) - (\bar{s}, \bar{t})].$$

Since  $\nabla \psi(s, t)$  is constant on  $(\bar{s}t, t)$ , we only need to consider the tangent planes at  $(\bar{s}, 1)$  for  $\bar{s} \in [\alpha, \beta]$ . The epigraph  $\text{epi}(\hat{h})$  can be then represented by the following infinite many linear inequalities:

$$v \geq h(\bar{s}) + d + h'(\bar{s})(s - \bar{s}) + [h(\bar{s}) + d - h'(\bar{s})\bar{s}](t - 1), \quad \forall \bar{s} \in [\alpha, \beta],$$

which can be simplified to

$$v \geq h(\bar{s})s + [h(\bar{s}) + d - h'(\bar{s})\bar{s}]t, \quad \forall \bar{s} \in [\alpha, \beta]. \tag{14}$$

The inequalities (14) are called *perspective cuts* [21]. Applying this “linearized” representation of  $\text{epi}(\hat{h})$  to the perspective reformulation  $(PR_s)$ , we obtain the following perspective cut (P/C) reformulation of  $(P_s)$ :

$$\begin{aligned} (PC_s) \quad & \min \sum_{i=1}^n v_i + g(z) \\ \text{s.t.} \quad & v_i \geq f'_i(\bar{x}_i)x_i + [f_i(\bar{x}_i) + c_i - f'_i(\bar{x}_i)\bar{x}_i]y_i, \quad \forall \bar{x}_i \in [l_i, u_i], \quad \forall i, \\ & Ax + By + Cz \leq b, \\ & l_i y_i \leq x_i \leq u_i y_i, \quad y_i \in \{0, 1\}, \quad i = 1, \dots, n. \end{aligned}$$

Problem  $(PC_s)$  is a semi-infinite mixed-integer linear programming problem, which cannot be solved directly. Nevertheless, “localized” subproblems of  $(PC_s)$  with a small finite subset of perspective cuts can be embedded in a branch-and-cut framework, where the violated perspective cuts with  $\bar{x}_i = x_i^*/y_i^*$  are added at each node when  $y_i^*$  is fractional in the optimal solution  $(x^*, y^*, z^*, v^*)$  of the continuous subproblem of the current node. The above solution scheme can be either implemented by tailor-made branch-and-cut method (see [21, 22]) or by means of `cutcallback` procedures in `CPLEX` (see [23]).

Computational results in [23] showed that the P/C reformulation is more efficient than the SOCP reformulation for problems with only semi-continuous variables. As mentioned in [31], the SOCP reformulation has the advantage featured with a simple

and straightforward implementation using optimization modeling languages, without appealing to the specialized branch-and-cut procedures. Moreover, the numerical results in [48] also suggested that the SOCP reformulation can be more efficient than the P/C reformulation when cardinality constraint is present with small cardinality  $K$ .

A main drawback of the SOCP reformulation is the introduction of additional variables and constraints that makes the continuous relaxation to be more complex and time-consuming to solve. Recently, Frangioni et al. [24] proposed a projected perspective reformulation for problem (P<sub>s</sub>) with quadratic function  $f_i(x_i) = a_i x_i^2 + b_i x_i$ . This reformulation is based on a piecewise-quadratic description of  $(1/y_i)a_i x_i^2 + b_i x_i + c_i y_i$ , which is the convex envelope of  $f_i(x_i) + c_i y_i$  over the semi-continuous variable. Consequently, the projected perspective reformulation in [24] is a mixed-integer piecewise-quadratic program whose continuous relaxation has roughly the same size of the standard continuous relaxation.

### 4 Quadratic Programming with Semi-continuous Variables and Cardinality Constraints

In this section, we focus on quadratic programming with semi-continuous variables and cardinality constraint. This problem is a special case of (P) with convex objective quadratic function and linear constraints:

$$\begin{aligned} \text{(CCQP)} \quad & \min q(x) := x^T Qx + c^T x \\ & \text{s.t. } Ax \leq b, \end{aligned} \tag{15}$$

$$e^T y \leq K, \tag{16}$$

$$\alpha_i y_i \leq x_i \leq u_i y_i, \quad i = 1, \dots, n, \tag{17}$$

$$y \in \{0, 1\}^n, \tag{18}$$

where  $Q$  is an  $n \times n$  positive semidefinite matrix,  $c \in \mathfrak{R}^n$ ,  $A \in \mathfrak{R}^{m \times n}$ ,  $b \in \mathfrak{R}^m$ ,  $K$  is an integer satisfying  $1 \leq K \leq n$ ,  $0 < \alpha_i < u_i$ . Problem (CCQP) is an important case of (P) because quadratic objective functions are used in most applications of (P), as seen in Sect. 2.

#### 4.1 Diagonal Decompositions

As the quadratic objective function  $q(x)$  is usually nonseparable, the perspective reformulations in Sect. 3 cannot be directly applied to (CCQP). A diagonal decomposition was proposed in [22] to extract separable terms from the quadratic form  $x^T Qx$ . Let  $d \in \mathfrak{R}_+^n$  with  $Q - D \geq 0$ , where  $D = \text{diag}(d)$  denotes the diagonal matrix with  $d$  being the diagonal vector. The quadratic objective function of (P) can be then decomposed as

$$q(x) = x^T (Q - D)x + c^T x + x^T D x. \tag{19}$$

Replacing the separable term  $x^T D x$  with its convex envelope over the semi-continuous variables, which is the sum of the perspective functions of  $d_i x_i^2$  over

$x_i \in \{0\} \cup [\alpha_i, u_i]$  for  $i = 1, \dots, n$ , the perspective reformulation for (CCQP) has the following form:

$$\begin{aligned}
 \text{(PR(d))} \quad & \min x^T(Q - \text{diag}(d))x + c^T x + \sum_{i=1}^n d_i(x_i^2/y_i) \\
 \text{s.t.} \quad & \text{(15), (16), (17), (18)}.
 \end{aligned}$$

The SOCP reformulation of (CCQP) is:

$$\begin{aligned}
 \text{(SOCP(d))} \quad & \min x^T(Q - \text{diag}(d))x + c^T x + \phi^T d \\
 \text{s.t.} \quad & \left\| \frac{x_i}{\frac{\phi_i - y_i}{2}} \right\| \leq \frac{\phi_i + y_i}{2}, \quad i = 1, \dots, n, \tag{20} \\
 & \text{(15), (16), (17), (18)}.
 \end{aligned}$$

On the other hand, using the perspective cuts (14) for quadratic function  $f_i(x_i) = d_i x_i^2$ , we obtain the P/C reformulation of (PR(d)):

$$\begin{aligned}
 \text{(PC(d))} \quad & \min x^T(Q - \text{diag}(d))x + c^T x + \sum_{i=1}^n d_i v_i \\
 \text{s.t.} \quad & v_i \geq 2\bar{x}_i x_i - \bar{x}_i^2 y_i, \quad \forall \bar{x}_i \in [\alpha_i, u_i], \quad i = 1, \dots, n, \\
 & \text{(15), (16), (17), (18)}.
 \end{aligned}$$

A key issue in implementing the SOCP reformulation (SOCP(d)) and the P/C reformulation (PC(d)) is how to choose the parameter vector  $d$ . A natural choice is  $d = (\lambda_{\min} - \epsilon)e$  when  $Q$  is positive definite, where  $\lambda_{\min}$  is the minimum eigenvalue of  $Q$ ,  $\epsilon > 0$  is a sufficiently small scalar and  $e$  is the all one column vector. Frangioni and Gentile [22] suggested to use a heuristic to find a diagonal matrix  $D = \text{diag}(d)$  by solving a simple semidefinite program (SDP):

$$\text{(SDP}_s) \quad \max \{e^T d \mid Q - \text{diag}(d) \succeq 0, d \geq 0\},$$

which we will call the “small” SDP problem. Numerical results in [22] show that this approach favorably compares with the minimum eigenvalue method. A further question arises: How to find a “better”  $d$  in the perspective reformulation?

Zheng et al. [48] proposed an SDP approach for selecting the “best” parameter vector  $d$  in the reformulation (SOCP( $d$ )) such that the continuous relaxation of (SOCP( $d$ )) is as tight as possible.

The continuous relaxation of (SOCP( $d$ )) has the following form:

$$\begin{aligned}
 \overline{\text{(SOCP(d))}} \quad & \min x^T(Q - \text{diag}(d))x + c^T x + \phi^T d \\
 \text{s.t.} \quad & 0 \leq y_i \leq 1, \quad i = 1, \dots, n, \\
 & \text{(15), (16), (17), (20)}.
 \end{aligned}$$

The parameter vector  $d$  corresponding to the tightest continuous relaxation of (SOCP( $d$ )) can be found by solving the following problem:

$$\max\{v(\overline{\text{SOCP}}(d)) \mid d \geq 0, Q - \text{diag}(d) \geq 0\}, \quad (21)$$

where  $v(\cdot)$  denotes the optimal value of problem  $(\cdot)$ . It was shown in [48] that problem (21) is equivalent to the following SDP problem:

$$\begin{aligned} (\text{SDP}_1) \quad & \max -Ks - e^T \pi - \tau \\ \text{s.t.} \quad & \begin{pmatrix} d_i + \mu_i & \frac{1}{2}(c_i - \lambda_i - \beta_i \mu_i) \\ \frac{1}{2}(c_i - \lambda_i - \beta_i \mu_i) & \pi_i + s + \alpha_i u_i \mu_i \end{pmatrix} \geq 0, \quad i = 1, \dots, n, \\ & \begin{pmatrix} Q - \text{diag}(d) & \frac{1}{2}(\lambda + A^T \eta) \\ \frac{1}{2}(\lambda + A^T \eta)^T & -\eta^T b + \tau \end{pmatrix} \geq 0, \\ & (s, \eta, \mu, \pi, d) \in \mathfrak{R}_+ \times \mathfrak{R}_+^m \times \mathfrak{R}_+^n \times \mathfrak{R}_+^n \times \mathfrak{R}_+^n, \quad (\tau, \lambda) \in \mathfrak{R} \times \mathfrak{R}^n, \end{aligned}$$

where  $\beta_i := \alpha_i + u_i$  for  $i = 1, \dots, n$ .

Compared with the “small” SDP formulation (SDP<sub>s</sub>) proposed by [22], the formulation (SDP<sub>1</sub>) has a drawback of having a larger dimension:  $4n + m + 2$  variables in (SDP<sub>1</sub>) compared to only  $n$  variables in (SDP<sub>s</sub>). Also, (SDP<sub>1</sub>) has  $n$  additional  $2 \times 2$  linear matrix inequalities. In spite of the larger size of (SDP<sub>1</sub>), it can still be computed efficiently by the interior-point based methods such as SeDuMi due to its simple structure. The longer time spent on solving the SDP problem (SDP<sub>1</sub>) could be well compensated by the savings in the computation time for the SOCP or P/C reformulations, as witnessed in the computational results in [48].

A Lagrangian decomposition scheme was proposed in [44] for cardinality constrained quadratic program with  $q(x) = x^T (H^T H)x + x^T D x + c^T x$  (without semi-continuous variables). In [44], the dual bound is computed by subgradient method for fixed  $H$  and  $D = \text{diag}(d)$  and is used in a branch-and-bound method. This Lagrangian decomposition method is extended in [48] to give an alternative derivation of the SOCP reformulation for (CCQP) when a fixed diagonal decomposition is used.

## 4.2 Specialized Branch-and-Bound Methods for (CCQP)

In this subsection, we discuss specialized branch-and-bound and branch-and-cut methods for solving (CCQP).

Since the objective function of (CCQP) is convex, a direct implementation of branch-and-bound method is possible: At each node of the search tree, solve a convex quadratic subproblem (CCQP) with all or part of the variables  $(x, y)$  and branch at some  $y_i$  by forcing  $y_i = 0$  or  $y_i = 1$ . This procedure could be not desirable because the number of variables in the continuous relaxation is doubled compared to the number of original decision variables. Bienstock [7] proposed to solve the subproblem with original variables and a surrogate constraint  $\sum_{i=1}^n x_i / u_i \leq K$  which replaces the constraint  $\sum_{i=1}^n y_i \leq K$ . Clearly,  $\sum_{i=1}^n x_i / u_i \leq K$  is valid

to (CCQP) since  $x_i \leq u_i y_i$ . The continuous subproblem then has the following form:

$$\begin{aligned} \min \quad & x^T Q x + c^T x \\ \text{s.t.} \quad & Ax \leq b, \\ & \sum_{i=1}^n x_i / u_i \leq K, \\ & x_i \in [x_i^l, x_i^u], \quad i = 1, \dots, n, \end{aligned}$$

where  $x_i^l$  and  $x_i^u$  are lower and upper bounds of  $x_i$  at the current node; and it is possible  $x_i^l = x_i^u$  when the variable is fixed. Primal feasible method is used in [7] for solving the continuous subproblem. The branching in [7] is done directly on the variable  $x_i$  instead of on the binary variable  $y_i$ : Either the constraint  $x_i \leq 0$  is added when  $x_i$  is branched down or the constraint  $x_i \geq \alpha_i$  is added when  $x_i$  is branched up. Mixed-integer rounding cuts, knapsack cuts and disjunctive cuts for cardinality constraint and semi-continuous variables were investigated in [7]. Recently, strong valid inequalities for semi-continuous knapsack polyhedron and cardinality constraint were derived in [18, 19]. These cuts can be incorporated into branch-and-cut methods.

Bertsimas and Shioda [6] proposed a branch-and-bound method where the continuous subproblem at each node is solved by Lemke's pivoting method. One of the advantages of using Lemke's method in a branch-and-bound method is that the optimal solution of the parent node can be used as an initial point of the Lemke's method for solving the continuous subproblem of the current node, because Lemke's method can be started from an infeasible basic solution.

Shaw et al. [44] presented a branch-and-bound method for cardinality constrained mean-variance portfolio problems, where the asset returns are driven by a factor model. The covariance matrix of the asset returns can be then expressed as  $Q = H^T H + D$ , where  $H \in \mathcal{R}^{m \times n}$  and  $D$  is a nonnegative diagonal matrix. Unlike the above two exact methods where a quadratic programming relaxation is solved at each node, the lower bounds in [44] are computed by solving a Lagrangian dual problem via subgradient method. Recently, [27] derived the optimal control law for discrete-time cardinality constrained linear-quadratic control problem, in which the number of time periods where controls can be applied is limited, by using a semi-definite programming solution scheme.

To provide upper bounds in branch-and-bound or branch-and-cut methods, fast heuristics are needed to find feasible solutions of (CCQP) at the root node or some sub-nodes during the branch-and-bound search process. A simple and natural heuristic to generate a feasible solution of (CCQP) is as follows.

**Heuristic 1** ([34]) Let  $x^*$  be the optimal solution of the continuous relaxation. Rank the absolute values of  $x_i^*$  as:  $|x_{i_1}^*| \leq |x_{i_2}^*| \leq \dots \leq |x_{i_n}^*|$ . Resolve the continuous problem with  $x_{i_j} = 0$  for  $j = 1, \dots, n - K$  and  $x_{i_j} \geq \alpha_i$  for  $j = n - K + 1, \dots, n$ . If the problem is feasible, then the optimal solution is feasible to (CCQP).

Another heuristic is based on solving the small mixed-integer subproblem of (CCQP) using branch-and-bound method. The number of variables of the subproblem is  $K + \kappa$ , where  $\kappa$  is a small integer, e.g.,  $\kappa = \lceil n/10 \rceil$ . The heuristic can be described as follows:

**Heuristic 2** ([7]) Let  $x^*$  be the optimal solution of the continuous relaxation. 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 (CCQP). Solve the reduced mixed-integer quadratic program with initial upper bound obtained from Heuristic 1 (or  $+\infty$  if no feasible solution is found by Heuristic 1). Early termination of the branch-and-bound method for the small reduced subproblem 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 fast in finding a reasonably good feasible solution of (CCQP). If the reduced mixed-integer program is infeasible, we can increase  $\kappa$  until the reduced problem is feasible.

## 5 Approximate Methods for Cardinality Constrained Problems

In the literature, for a vector  $x = (x_1, \dots, x_n)^T \in \mathfrak{R}^n$ , the cardinality function  $\text{card}(x)$  is also called (quasi)  $\ell_0$ -norm, denoted by  $\|x\|_0$ , in sparse optimization literature. A general form of cardinality constrained mathematical programming can be formulated as

$$\begin{aligned} (\text{P}_c) \quad & \min f(x) \\ & \text{s.t. } g(x) \leq 0, \quad h(x) = 0, \\ & \|x\|_0 \leq K, \end{aligned}$$

where  $f: \mathfrak{R}^n \rightarrow \mathfrak{R}$ ,  $g: \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ ,  $h: \mathfrak{R}^n \rightarrow \mathfrak{R}^p$  are continuously differentiable functions.  $(\text{P}_c)$  can be regarded as a special case of (P) without explicit semi-continuous variables. An important special case of  $(\text{P}_c)$  is the cardinality constrained quadratic program where  $f$  is quadratic and  $g$  and  $h$  are affine functions. In particular, the following problem of minimizing a quadratic function under cardinality constraint is of interest:

$$(\text{QP}_c) \quad \min \{x^T Qx + c^T x \mid \|x\|_0 \leq K\}.$$

The above formulation includes many applications such as compressed sensing and subset selection. Specialized convex relaxations for  $(\text{QP}_c)$  are discussed in [4]. A closely related problem to the cardinality constrained quadratic program is the  $\ell_0$ -norm minimization problem or sparse solutions of linear equations:

$$\min \{\|x\|_0 \mid Ax = b\}.$$

The reader is referred to [10, 35, 47] for an extensive literature on this problem.

In this section, we describe different inexact methods for  $(\text{P}_c)$  or its special cases. These methods are mainly based on various approximations and relaxations of the



$\ell_0$ -norm function  $\|x\|_0$  except for the penalty decomposition and alternating direction method in [38]. The suboptimal or local solutions obtained from these approximate methods can be used to improve the performance of branch-and-bound methods to find exact solutions of  $(P_c)$ .

### 5.1 $\ell_p$ -Norm Approximation

A popular approach in the literature for dealing with  $\ell_0$ -norm  $\|x\|_0$  is to replace it with the  $\ell_1$ -norm  $\|x\|_1$ . The resulting problem then becomes

$$\begin{aligned} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \quad h(x) = 0, \\ & \|x\|_1 \leq K. \end{aligned} \tag{22}$$

The above problem is a convex relaxation of  $(P_c)$  when  $f$  and  $g$  are convex and  $h$  is affine. The  $\ell_1$ -norm constraint  $\|x\|_1 \leq K$  in problem (22) can be also incorporated into the objective function as a regularized or penalized term, yielding the following convex problem,

$$\begin{aligned} \min & f(x) + \lambda \|x\|_1 \\ \text{s.t.} & g(x) \leq 0, \quad h(x) = 0, \end{aligned} \tag{23}$$

where  $\lambda > 0$  is a regularization parameter. The approach to replace  $\ell_0$ -norm with  $\ell_1$ -norm is called *basis pursuit* [17]. In contrast to its successful applications in sparse solution to linear system, the  $\ell_1$ -norm approximation problem formulation (22) or (23), however, does not necessarily produce solutions with desired sparsity for general cardinality constrained optimization problems. Nonconvex norm  $\ell_p$ -norm ( $0 < p < 1$ ) can be used to replace  $\|x\|_0$  to enforce a stricter sparsity, leading to the following nonconvex approximation of  $(P_c)$ :

$$\begin{aligned} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \quad h(x) = 0, \\ & \|x\|_p \leq K \end{aligned} \tag{24}$$

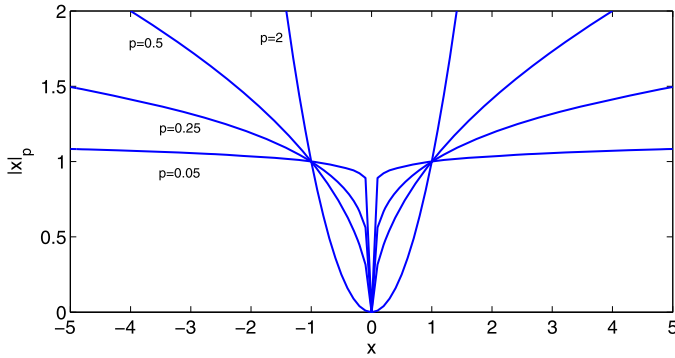
or its regularized problem:

$$\begin{aligned} \min & f(x) + \lambda \|x\|_p^p \\ \text{s.t.} & g(x) \leq 0, \quad h(x) = 0. \end{aligned} \tag{25}$$

The  $\ell_p$ -norm approximation is based on the following property:

$$\lim_{p \rightarrow 0^+} \|x\|_p = \|x\|_0$$

for any fixed  $x$  (see Fig. 2 for the one-dimensional case).



**Fig. 2**  $\ell_p$ -norm functions with different values of  $p$

It is shown in [30] that the  $\ell_p$  minimization:  $\min\{\|x\|_p^p \mid Ax = b\}$  is strongly NP-hard when  $0 < p < 1$ . An interior point method is suggested in [30] for the  $\ell_p$  minimization. Lai and Wang [36] also developed an iterative solution method for the  $\ell_p$  minimization. In [26], it is proved that the  $\ell_0$  and  $\ell_p$  minimization problems with linear equality and inequality constraints are equivalent for some sufficiently small  $p > 0$ . A successive linearization algorithm is proposed in [26] for finding a stationary point of the  $\ell_p$  minimization problem. Chen et al. [13] proved that the unconstrained  $\ell_q$ - $\ell_p$  minimization:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_q^q + \lambda \|x\|_p^p, \quad \text{where } q \geq 1, \lambda > 0, 0 < p < 1,$$

is strongly NP-hard. Lower bounds of the parameter  $\lambda$  in the  $\ell_2$ - $\ell_p$  minimization for achieving the sparsity requirement  $\|x\|_0 \leq K$  are established in [13, 14].

### 5.2 Mangasarian’s Approximation Method

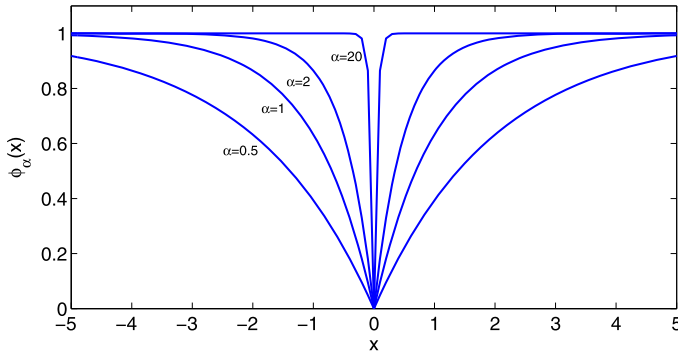
Mangasarian [39] suggested to replace  $\|x\|_0$  by the following exponential function:

$$\phi_\alpha(x) = \sum_{i=1}^n (1 - e^{-\alpha|x_i|})$$

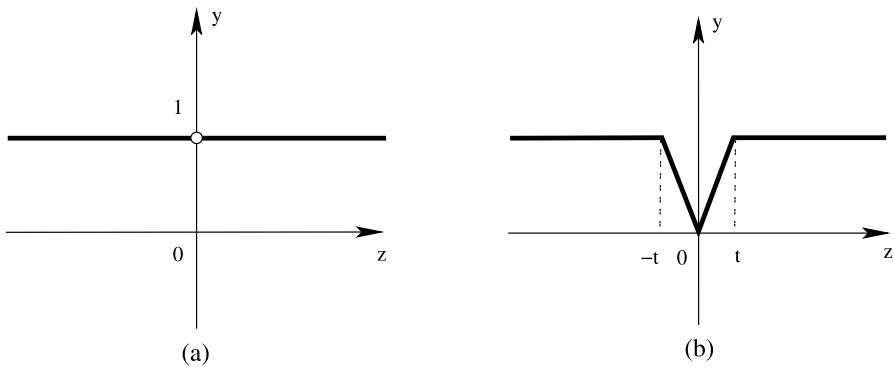
for some  $\alpha > 0$ . It is clear that  $\phi_\alpha(x) \leq \|x\|_0$  and

$$\lim_{\alpha \rightarrow +\infty} \phi_\alpha(x) = \|x\|_0$$

for any fixed  $x$  (see Fig. 3 for the one-dimensional case). It is shown in [39] that the  $\phi_\alpha$  minimization and the  $\ell_0$  minimization over polyhedral set are equivalent for some large  $\alpha > 0$ . A successive linearization algorithm is also proposed in [39] to find a stationary point of the  $\phi_\alpha$  minimization problem.



**Fig. 3** Functions  $1 - e^{-\alpha|x|}$  with different values of  $\alpha$



**Fig. 4** (a) Function  $y = \text{sign}(|z|)$ ; (b) function  $y = w(z, t)$

### 5.3 DC Approximation

In this subsection, we introduce a DC approach [49] to approximate the  $\ell_0$  norm function  $\|x\|_0$ . We first note that

$$\|x\|_0 = \sum_{i=1}^n \text{sign}(|x_i|), \tag{26}$$

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

$$w(z, t) = \min \left\{ 1, \frac{1}{t}|z| \right\}, \tag{27}$$

where  $t > 0$  is a parameter (see Fig. 4). It is easy to see that

$$\lim_{t \rightarrow 0^+} w(z, t) = \text{sign}(|z|)$$

for any fixed  $z$ .

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

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

where  $a^+ = \max(a, 0)$  and  $h(z, t) = (z - t)^+ + (-z - t)^+$ . Since  $h(z, t)$  is a convex function of  $z$ ,  $w(z, t)$  is a DC function (difference of two convex functions) of  $z$ . Using  $w(z, t)$ , we can construct the following piecewise linear underestimation of the  $\ell_0$ -norm function  $\|x\|_0$  for  $x \in \mathfrak{R}^n$ :

$$\psi(x, t) = \sum_{i=1}^n w(x_i, t) = \frac{1}{t} \left( \|x\|_1 - \sum_{i=1}^n h(x_i, t) \right).$$

We see that  $\psi(x, t)$  is a nonsmooth piecewise linear DC function of  $x$  and  $\lim_{t \rightarrow 0^+} \psi(x, t) = \|x\|_0$  for any fixed  $x$ .

A prominent feature of the above 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 of the constraint  $\psi(x, t) \leq K$ . In fact, by the definition of  $w(x_i, t)$ , it always holds  $w(x_i, t) \leq 1$  ( $i = 1, \dots, n$ ). The convex inner approximation of  $w(x_i, t) \leq 1$  at  $y_i$  is

$$\sum_{i=1}^n \frac{1}{t}|x_i| - \frac{1}{t}[h(y_i, t) + \xi_i(x_i - y_i)] \leq 1,$$

where  $\xi_i \in \partial h(y_i, t)$  for  $i = 1, \dots, n$ . The above  $n$  inequalities provide strengthening cuts to the feasible set of the convex subproblems in the linearization method (see [49]).

#### 5.4 Penalty Decomposition and Alternating Direction Method

Alternating direction method (ADM) or block coordinate decent method is a classical method in solving convex programming:  $\min_{(x,y) \in X \times Y} f(x, y)$ . The idea of ADM is quite simple and straightforward: The function  $f(x, y)$  is minimized over  $x$  for fixed  $y$  and minimized over  $y$  for fixed  $x$  alternatively in the hope that the iterative sequence  $\{(x^k, y^k)\}$  eventually converges to the stationary point of the problem. This method is particularly useful when the problem over  $(x, y)$  can be reduced to “easier” subproblems when either  $x$  or  $y$  is fixed.

We now consider the problem  $(P_c)$  with convex  $f$  and  $g$ , which can be rewritten as

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } g(x) \leq 0, \quad h(x) = 0, \\ & \quad x = y, \\ & \quad \|y\|_0 \leq K. \end{aligned}$$

Penalizing the equality constraints  $x = y$ , we obtain

$$\begin{aligned} \min \quad & f(x) + \frac{\rho}{2} \|x - y\|_2^2 \\ \text{s.t.} \quad & g(x) \leq 0, \quad h(x) = 0, \\ & \|y\|_0 \leq K. \end{aligned} \quad (28)$$

We observe that problem (28) reduces to a continuous convex program for fixed  $y$ , and to a minimization of separable quadratic function with cardinality constraint for fixed  $x$ . It turns out the latter problem can be explicitly solved as stated in the following lemma.

**Lemma 1** ([38]) *Consider the problem:*

$$\min \left\{ \sum_{i=1}^n q_i(x_i) \mid \|x\|_0 \leq K, x_i \in X_i, i = 1, \dots, n \right\}, \quad (29)$$

where  $q_i(0) = 0$  ( $i = 1, \dots, n$ ). Let  $q_i^* = \min_{x_i \in X_i} q_i(x_i)$  with minimizer  $x_i = \bar{x}_i^*$ . Let  $\{q_i^*\}_{i=1}^n$  be ranked in an increasing order:  $q_{i_1}^* \leq q_{i_2}^* \leq \dots \leq q_{i_n}^*$ . Then, the optimal solution  $x^*$  of (29) is given by

$$x_{i_k}^* = \begin{cases} \bar{x}_{i_k}^*, & 1 \leq k \leq K, \\ 0, & \text{otherwise.} \end{cases}$$

It is shown in [38] that under some mild conditions, the ADM for (P<sub>c</sub>) converges to a local minimizer of (P<sub>c</sub>). Computational results in [38] show that the ADM is capable of finding feasible solutions of good quality within reasonable computing time. We remark that alternating direction method for (P<sub>c</sub>) can be extended to problem (CCQP) with semi-continuous variables  $x_i \in \{0\} \cup [\alpha_i, u_i]$ , where  $0 < \alpha_i < u_i$ . Finally, we point out that the augmented Lagrangian function  $f(x) + \lambda^T(x - y) + \frac{\rho}{2} \|x - y\|_2^2$  can be also employed in the penalized problem (28).

## 6 Conclusions

We have summarized in this paper some recent advances in mathematical programming with semi-continuous variables and cardinality constraint. Our focuses are mainly on theory and solution techniques that are potentially applicable to problems arising from the real-world applications. On one hand, semi-continuous variables and cardinality constraint have been widely used in modeling real-world optimization problems, leading to increasing interests and demands of efficient solution methods to tackle this kind of discrete constraints. On the other hand, the inherent combinatorial nature makes it very difficult to find the exact solutions of the problems with realistic dimension and data structures. The current literature lacks systematic investigation of the theory and solution methods for mathematical programming with semi-continuous variables and cardinality constraint. Further research efforts are needed to better our understandings of this class of challenging mathematical programming problems.

## References

- [1] Aktürk, M., Atamtürk, A., Gürel, S.: A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Oper. Res. Lett.* **37**, 187–191 (2009)
- [2] Alizadeh, F., Goldfarb, D.: Second-order cone programming. *Math. Program.* **95**, 3–51 (2003)
- [3] Arthanari, T., Dodge, Y.: *Mathematical Programming in Statistics*. Wiley, New York (1993)
- [4] Bach, F., Ahipasaoglu, S., d'Aspremont, A.: Convex relaxations for subset selection (2010) [arXiv:1006.3601](https://arxiv.org/abs/1006.3601)
- [5] Ben-Tal, A., Nemirovski, A.: *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. MPS-SIAM Series on Optimization. SIAM, Philadelphia (2001)
- [6] Bertsimas, D., Shioda, R.: Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.* **43**, 1–22 (2009)
- [7] Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. *Math. Program.* **74**, 121–140 (1996)
- [8] Blog, B., Hoek, G.V.d., Rinnooy Kan, A., Timmer, G.: The optimal selection of small portfolios. *Manag. Sci.* **29**, 792–798 (1983)
- [9] Bonami, P., Lejeune, M.: An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Oper. Res.* **57**, 650–670 (2009)
- [10] 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)
- [11] Cesarone, F., Scozzari, A., Tardella, F.: Efficient algorithms for mean-variance portfolio optimization with hard real-world constraints. *G. Ist. Ital. Attuari* **72**, 37–56 (2009)
- [12] Chang, T., Meade, N., Beasley, J., Sharaiha, Y.: Heuristics for cardinality constrained portfolio optimisation. *Comput. Oper. Res.* **27**, 1271–1302 (2000)
- [13] 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)
- [14] 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)
- [15] Crama, Y., Schyns, M.: Simulated annealing for complex portfolio selection problems. *Eur. J. Oper. Res.* **150**, 546–571 (2003)
- [16] 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). doi:[10.1007/s10898-012-9845-z](https://doi.org/10.1007/s10898-012-9845-z)
- [17] Donoho, D., Huo, X.: Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory* **47**, 2845–2862 (2001)
- [18] de Farias, I., Zhao, M.: A polyhedral study of the semi-continuous knapsack problem. *Math. Program.* (2012). doi:[10.1007/s10107-012-0566-3](https://doi.org/10.1007/s10107-012-0566-3)
- [19] de Farias, I. Jr, Nemhauser, G.: A polyhedral study of the cardinality constrained knapsack problem. *Math. Program.* **96**, 439–467 (2003)
- [20] Fernández, A., Gómez, S.: Portfolio selection using neural networks. *Comput. Oper. Res.* **34**, 1177–1191 (2007)
- [21] Frangioni, A., Gentile, C.: Perspective cuts for a class of convex 0–1 mixed integer programs. *Math. Program.* **106**, 225–236 (2006)
- [22] Frangioni, A., Gentile, C.: SDP diagonalizations and perspective cuts for a class of nonseparable MIQP. *Oper. Res. Lett.* **35**, 181–185 (2007)
- [23] Frangioni, A., Gentile, C.: A computational comparison of reformulations of the perspective relaxation: SOCP vs. cutting planes. *Oper. Res. Lett.* **37**, 206–210 (2009)
- [24] Frangioni, A., Gentile, C., Grande, E., Pacifici, A.: Projected perspective reformulations with applications in design problems. *Oper. Res.* **59**, 1225–1232 (2011)
- [25] Frangioni, A., Gentile, C., Lacalandra, F.: Tighter approximated MILP formulations for unit commitment problems. *IEEE Trans. Power Syst.* **24**, 105–113 (2009)
- [26] Fung, G., Mangasarian, O.: Equivalence of minimal  $l_0$ - and  $l_p$ -norm solutions of linear equalities, inequalities and linear programs for sufficiently small  $p$ . *J. Optim. Theory Appl.* **151**, 1–10 (2011)
- [27] Gao, J.J., Li, D.: Cardinality constrained linear-quadratic optimal control. *IEEE Trans. Autom. Control* **56**, 1936–1941 (2011)
- [28] Gao, J.J., Li, D.: A polynomial case of the cardinality-constrained quadratic optimization problem. *J. Glob. Optim.* (2012). doi:[10.1007/s10898-012-9853-z](https://doi.org/10.1007/s10898-012-9853-z)
- [29] Gao, J.J., Li, D.: Optimal cardinality constrained portfolio selection. *Oper. Res.* (2013), in press

- [30] Ge, D., Jiang, X., Ye, Y.: A note on complexity of  $l_p$  minimization. *Math. Program.* **129**, 285–299 (2011)
- [31] Günlük, O., Linderoth, J.: Perspective reformulations of mixed integer nonlinear programs with indicator variables. *Math. Program.* **124**, 183–205 (2010)
- [32] Hiriart-Urruty, J., Lemaréchal, C.: *Convex Analysis and Minimization Algorithms: Fundamentals*. Springer, New York (1993)
- [33] Jacob, N.: A limited-diversification portfolio selection model for the small investor. *J. Finance* **29**, 847–856 (1974)
- [34] 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)
- [35] Jocar, S., Pfetsch, M.: Exact and approximate sparse solutions of underdetermined linear equations. *SIAM J. Sci. Comput.* **31**, 23–44 (2008)
- [36] Lai, M., Wang, J.: An unconstrained  $\ell_q$  minimization with  $0 < q \leq 1$  for sparse solution of underdetermined linear systems. *SIAM J. Optim.* **21**, 82–101 (2010)
- [37] Li, D., Sun, X., Wang, J.: Optimal lot solution to cardinality constrained mean-variance formulation for portfolio selection. *Math. Finance* **16**, 83–101 (2006)
- [38] 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](http://www.optimization-online.org/DB_HTML/2010/08/2719.html)
- [39] Mangasarian, O.: Minimum-support solutions of polyhedral concave programs. *Optimization* **45**, 149–162 (1999)
- [40] Maringer, D., Kellerer, H.: Optimization of cardinality constrained portfolios with a hybrid local search algorithm. *IEEE Spectr.* **25**, 481–495 (2003)
- [41] Miller, A.: *Subset Selection in Regression*, 2nd edn. Chapman & Hall, London (2002)
- [42] Mitra, G., Ellison, F., Scowcroft, A.: Quadratic programming for portfolio planning: insights into algorithmic and computational issues. Part II: Processing of portfolio planning models with discrete constraints. *J. Asset Manag.* **8**, 249–258 (2007)
- [43] Schaerf, A.: Local search techniques for constrained portfolio selection problems. *Comput. Econ.* **20**, 177–190 (2002)
- [44] Shaw, D., Liu, S., Kopman, L.: Lagrangian relaxation procedure for cardinality-constrained portfolio optimization. *Optim. Methods Softw.* **23**, 411–420 (2008)
- [45] Tawarmalani, M., Sahinidis, N.: Semidefinite relaxations of fractional programs via novel convexification techniques. *J. Glob. Optim.* **20**, 133–154 (2001)
- [46] Xie, J., He, S., Zhang, S.: Randomized portfolio selection, with constraints. *Pac. J. Optim.* **4**, 89–112 (2008)
- [47] Zhao, Y., Li, D.: Reweighted  $\ell_1$ -minimization for sparse solutions to underdetermined linear systems. *SIAM J. Optim.* **22**, 1065–1088 (2012)
- [48] 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. rep., School of Management, Fudan University (2012). Available at: [http://www.optimization-online.org/DB\\_HTML/2010/11/2797.html](http://www.optimization-online.org/DB_HTML/2010/11/2797.html)
- [49] Zheng, X.J., Sun, X.L., Li, D., Sun, J.: Successive convex approximations to cardinality-constrained quadratic programs: a DC approach. Tech. rep., School of Management, Fudan University (2012). Available at: [http://www.optimization-online.org/DB\\_HTML/2012/04/3423.html](http://www.optimization-online.org/DB_HTML/2012/04/3423.html)