İstanbul Üniversitesi İşletme Fakültesi Dergisi
Istanbul University Journal of the School of Business Administration
Cilt/Vol:40, Sayı/No:2, 2011, 124-137
ISSN: 1303-1732 – www.ifdergisi.org © 2011

Approximation, reformulation and convex techniques for cardinality optimization problems

Mohammad J. Abdi [1]
School of Mathematics
University of Birmingham, United Kingdom

Yunbin Zhao [2]
School of Mathematics
University of Birmingham, United Kingdom

Abstract

The cardinality minimization problem (CMP) is finding a vector with minimum cardinality, which satisfies certain linear (or non-linear) constraints. This problem is closely related to the so-called compressive sensing problems. In this paper we survey and study different approximation, reformulation and convex relaxations for both cardinality constraint problems and cardinality minimization problems, and discuss how to add a penalty function to the objective in order to get a reformulation/approximation model of the original problems, instead of simply dropping the rank constraint. By reformulation techniques, under some mild condition we may either transform the problem to a mixed integer linear program (MILP) or the so-called bilevel SDP problems. We also point out that a continuous approximation of cardinality functions can enable us to apply majorization method to extract proper weights for the (re)weighted *l1* algorithms.

Keywords: Cardinality optimization problem, l_1-minimization, compressive sensing, convex optimization, (re)weighted l_1- minimization, Lagrangian relaxation.

Küme eleman sayılarının (cardinality) optimizasyon problemlerine yönelik yakınsak, yeniden formüle etmeli ve dışbükey teknikler

Özet

Küme eleman sayılarının minimizasyon problemi, belirli doğrusal (veya doğrusal olmayan) kısıtları karşılayan minimum küme eleman sayısını içeren bir vektör bulmakla ilgilidir. Problem, başınç algılama problemi olarak da anılan problemle yakından ilişkilidir. Bu çalışmada, küme eleman kısıt problemleri ve küme eleman sayılarının minimizasyon problemleri için çeşitli yakınsak, yeniden formüle etme ve dışbükey gevşetmeler yer almakta ve yalnızca rank kısıtını dışlamaktan çok orijinal problemin yeniden formüle edilmesi/yakınsanması için amaca nasıl bir ceza fonksiyonu ekleneceğini tartışılmaktadır. Yeniden formüle etme teknikleri ile bazı hafif koşullarda, problem, ya karma tam sayılı doğrusal programlama ya da iki kademeli yarı tanımlı programlama problemlerine dönüştürülebilir. Küme eleman sayısı fonksiyonlarının sürekli yakınsanması, *l1* algoritmalarının (yeniden) ağırlıklandırılarak uygun ağırlıklarının belirlenmesi amacıyla majorlaştırma yönteminin uygulanmasına izin verir.

Anahtar Sözcükler: *Küme eleman sayılarının (cardinality) optimizasyon problemi,* l_1-minimizasyonu, basınçlı algılama, dışbükey optimizasyon, (yeniden) ağırlıklandırılmış l_1-minimizasyonu, Lagrangian gevşetme.

[1] *ab-dimj@maths.bham.ac.uk (M. J. Abdi)*

[2] *y.zhao.2@bham.ac.uk (Y. Zhao)*

1. Introduction

The general cardinality minimization problem (CMP) over a convex set *C*, and cardinality constraint problem can be cast repsectively as

$$\text{Minimize } \{\text{Card } (x) : x \in C\}, \tag{1}$$

and

$$\text{Minimize } \{f(x) : \text{Card } (x) \le \tau,\ x \in C\}, \tag{2}$$

The set *C* can be also non-convex in some situations. So CMP is to maximize the number of zero components or equivalently to minimize the number of non-zero components of a vector satisfying certain constraints. In another word, CMP is looking for the sparest vector in a given feasible set or looking for the simplest model for describing or fitting a certain phenomena. The card function, *card*(x), can be expressed as l_0 norm. While l_0 is not a norm, we can still call it l_0 'norm', due to the following fact.

$$\|x\|_0 = \lim_{p\to 0}\|x\|_0 = \lim_{p\to 0}\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}} = Card(x).$$

l_0 norm is a non-convex, non-smooth and integer valued function, and the optimization problems with `card' objective or constraints are known as NP-hard problems [18, 28], and thus CMPs are not computationally tractable in general.

These kinds of problems have many applications in such areas as finance [22, 7, 25], signal processing and control [27, 16, 21], statistics and principal component analysis [9, 29, 33, 24], compressive sensing [1, 11, 4], etc. Due to the NP-hardness of CMP, the aim of this paper is to survey and introduce different SDP relaxations/approximations of CMPs.

This paper is organized as follows. In section 2, we consider cardinality constraint problems, and discuss SDP relaxation methods for these problems based on Shor's lemma and duality methods. Also we show how this problem can be cast as a bilevel SDP problem which was first pointed out by Y.B.Zhao [31]. In section 3, we review various existing methods for solving CMPs under linear constraint, and as an example of weighted l_1 techniques we introduce a continuous approximation of cardinality function and then apply linearization methods (majorization minimization methods) to solve the problem iteratively. In section 4, we study CMPs under nonlinear non-convex constraints, and show how to find an approximate solution for these problems using reformulation techniques and Lagrangian duality methods. We also explain how the dual problem can be reduced to a semidefinite problem. In section 5, we discuss CMP under 0-1 vectors, and explain how to reformulate these problems by adding certain penalty instead of dropping the rank constraint.

2. Cardinality Constraint Problems

Let us first start with a general cardinality constraint problem. A general cardinality constraint problem is of the form (2) where f (x) and C are convex. Card (x) $\le \tau$ is not a convex constraint, so we try to relax this constraint using semidefinite relaxation. Before doing so, we first note that norms are equivalent in finite dimensional spaces in the following sense: Suppose $\|\bullet\|_\mu$, $\|\bullet\|_\nu$ are norms on R^n. Then there exist scalars a,b ≥ 0, such that $a\|x\|_\mu \le \|x\|_\nu \le b\|x\|_\mu$, $\forall \in R^n$.

For example we have

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\, \|x\|_2.$$

To be more precise note that in general case for a polytope norm defined by max $|a^T x|$, $i = 1, \ldots, n$, we have

$$\frac{1}{\sqrt{\beta}}\sqrt{x^T A_x} \leq \max_{i=1,\ldots,n} \left|a_i^T x\right| \leq \sqrt{\beta}\sqrt{x^T Ax}, \forall A \geq 0, x \in \mathrm{R}^n,$$

where β is a constant.

Proposition 1. [9] The cardinality constraint Card (x) ≤ τ can be relaxed as the following inequality constraint

$$1^T |X| 1 \leq \tau tr(x), \begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \geq 0.$$

Proof. For any given vector x = ($x_1 \ldots x_n$)^T ≠ 0 obviously we have

$$0 \leq \frac{|x_i|}{\sqrt{x_1^2 + \ldots + x_n^2}} \leq 1, i = 1, \ldots, n,$$

and hence

$$\sum_{i=1}^{n} \frac{|x_i|}{\sqrt{x_1^2 + \ldots + x_n^2}} \leq Card(x) \leq \tau,$$

i.e.

$$\frac{\sum_{i=1}^{n} |x_i|}{\sqrt{x_1^2 + \ldots + x_n^2}} = \frac{\|x\|_1}{\|x\|_2} \leq Card(x) \leq \tau. \quad (3)$$

Note that the cardinality of the vector x ($x_1, \ldots, x_n$) is equal to that of the vector $|x|$ = ($|x_1|, \ldots, |x_n|$), so by the fact that *Card*($|x|$)= *Card*(x) ≤ τ, we define the vector ψ = ($\psi_1, \ldots \psi_n$) where for every i, $\psi_i = 1$ if $x_i \neq 0$; otherwise $\psi_i = 0$. By Cauchy-Schwartz inequality, i.e.,

$\left|\langle |x|, |\psi\rangle\right|^2 \leq \left|\langle |x|, |x|\rangle . \langle \psi, \psi\rangle\right|$. and noting that *Card*(x) ≤ τ, we have

$$\left||x|_1 + \ldots + |x|_n\right|^2 \leq \tau\left(|x|_1^2 + \ldots + |x|_n^2\right).$$

Therefore we have

$$\|x\|_1 \leq \sqrt{\tau}\, \|x\|_2 \quad (4)$$

In what follows, we use semidefinite relaxation methods [18]. Consider the matrix, $X = xx^T$, i.e.,

$$X = \begin{pmatrix} x_1^2 & x_1x_2 \cdots & x_1x_n \\ x_2x_1 & x_2^2 \quad \cdots & x_2x_n \\ \vdots & \vdots \quad \ddots & \vdots \\ x_nx_1 & x_nx_2 \cdots & x_n^2 \end{pmatrix}$$

Then (4) can be written as the following convex inequality (see e.g. [9])

$$1^T|x| \ 1 \leq \tau tr(X),$$

where $|x|$ denotes the element-wise absolute value of the matrix X. While the constraint $X = xx^T$ is not convex, it can be relaxed to $X \geq xx^T$, which can be written as

$$\begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \geq 0 \Leftrightarrow \mathrm{X} - xx^T \geq 0, \text{Rank (X)} = 1.$$

by applying Schur's lemma [17, 23, 15]. The proof is complete.

In [31], Zhao proved that under certain conditions, matrix rank minimization can be formulated as a linear bilevel SDP problem. This motivates the following result.

Proposition 2. If the set C is bounded and defined by linear constraints, *the complexity of the cardinality constrained problem (2) is equivalent to a bilevel SDP problem.*

Proof. From the proof of Proposition 1, one can rewrite the problem (2) as of the form

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{s. t. } x \in C \quad (5) \\ &1^T|X|1 \leq \tau tr(X) \\ &X = x^T x . \end{aligned}$$

Now we can write the problem (8) as the following form

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{s. t. } x \in C \\ &1^T|X|1 \leq \tau tr(X) \quad (6) \\ &\begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \geq 0 \\ &\text{Rank (X)=1,} \end{aligned}$$

which is equivalent to the following bilevel SDP form (see [31])

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{s. t. } x \in C \\ &1^T|X|1 \leq \tau tr(\hat{X}) \end{aligned} \qquad (7)$$

$$\hat{X} = \arg\min_{x \in S^n} \left\{ tr(x) : \begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \geq 0 \right\}.$$

The proof is complete.

The constraint *Rank*(*X*)= 1 is not convex. In order to get a reasonable approximation/relaxization of (2), a simple idea is to drop this constraint. This leads to the following problem

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{s. t. } x \in C \\ &1^T |X| 1 \leq \tau tr(X) \quad (8) \\ &\begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \geq 0, \end{aligned}$$

which can be solved more efficiently than the original problem. Dropping the rank constraint, however, may result in a large gap between the optimal values of the relaxed problem (8) and the original problem. Thus we can use the penalty method instead of dropping the rank constrains to obtain better approximation of the original problem. This idea was first proposed in [31]. This can avoid the the lower level optimization in (7), and yield the following reformulation of (7).

$$\begin{aligned} &\text{Minimize } f(x) \\ &\text{s. t. } x \in C \\ &1^T |X| 1 \leq \tau tr(X) \\ &\begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \geq 0, \end{aligned} \tag{9}$$

where $\xi > 0$ is the penalty parameter which is chosen to be sufficiently large.

A special case of the problem (2) can be cast as

$$\begin{aligned} &\text{Minimize } f(x) = \frac{1}{2} x^T P x + q^T x \\ &\text{s.t. } \mathrm{Ax} \leq \mathrm{b} \\ &Card(x) \leq \tau \\ &0 \leq x_i \leq s_i,\ i = 1, \dots, n, \end{aligned} \tag{10}$$

where P is an $n \times n$ symmetric matrix, $q \in R^n$, $A \in R^{m \times n}$, $b \in R^m$, and $\tau \in N^+$. This problem was studied by Zheng, Sun and Li [32].

As we have seen above, a common way to solve the optimization problems with a cardinality function as an objective or constraint is to relax the cardinality function. We take the specific example above to further demonstrate this approach. First, the problem (10) can be reformulated as the following mixed integer quadratic problem

$$\begin{aligned} \text{Minimize } & f(x) = \frac{1}{2}x^T P x + q^T x \\ \text{s.t. } & \mathrm{Ax} \le \mathrm{b}, \\ & 1^T u \le \tau, u \in \{0,1\}^n, \\ & 0 \le x_i \le s_i u_i, i = 1, \ldots, n, \end{aligned} \tag{11}$$

where 1 still denotes the vector of ones. Note that the constraint $u_i \in \{0, 1\}$, can be written as $u_i^2 - u_i = 0$. Assuming $P \geq 0$, the convex relaxation of the problem above can be achieved by replacing $u_i \in \{0, 1\}$ by $u_i \in [0, 1]$. So it leads to the following problem (see Zheng, Sun and Li [32])

$$\begin{aligned} \text{Minimize } & f(x) = \frac{1}{2}x^T P x + q^T x \\ \text{s.t. } & \mathrm{Ax} \le \mathrm{b}, \\ & 1^T u \le \tau, u \in \{0,1\}^n, \\ & 0 \le x_i \le s_i u_i, i = 1, \ldots, n, \end{aligned} \tag{12}$$

Note that the constraint $u_i \in [0, 1]$, can be written as $u^2 - u_i \leq 0$. Obviously, the optimal value of the problem (12) is a lower bound for the problem (10). An SDP relaxation for the problem (10) can be obtained as follows: Let $X = xx^T$, and $U = uu^T$ which can be relaxed to $X \geq xx^T$ and $U \geq uu^T$, yielding the relaxed problem

$$\begin{aligned} \text{Minimize } & f(x) = \frac{1}{2}x^T P x + q^T x \\ \text{s.t. } & \mathrm{Ax} \le \mathrm{b}, \\ & 1^T u \le \tau, u \in \{0,1\}^n, \\ & 0 \le x_i \le s_i u_i, i = 1, \ldots, n, \\ & \begin{pmatrix} X & x \\ x^T & 1 \end{pmatrix} \ge 0, \begin{pmatrix} U & u \\ u^T & 1 \end{pmatrix} \ge 0. \end{aligned} \tag{13}$$

Relationship between (12) and (13) was characterized by the following result.

Proposition 4. [32] *Suppose that the feasible set of the problem (12) has an interior point (or a relative interior point, if $Ax \leq b$ includes equality constraint). If $P \geq 0$, then the optimal value of the problems (13) and (12) are equal.*

3. CMP Under Linear Constraints

The Cardinality minimization problem (CMP) with linear constraints, i.e.,

$$\text{Minimize } \{\text{Card } (x): Ax = b\}, \tag{14}$$

where $A \in R^{m \times n}$ is a matrix with $m < n$, has been widely discussed in the field of compressive sensing [5, 20, 26] which deals with the signal processing/recovery which has a wide range of applications in such areas as image processing [19].

The most popular approach for solving (14) (which is NP-hard in general) is to replace the function *card* (x) by its convex envelop $\|x\|_1$ (see e.g. [12]). Hence a relaxation of (14) is as follows:

$$\text{Minimize } \{\|x\|_1: Ax = b\}, \tag{15}$$

which can be also written as

$$\text{Minimize } \{s: \|x\|_1 \leq s, Ax = b\}. \tag{16}$$

Clearly (15) and (16) are linear programming problems. For instance, (16) can be written as the linear program

$$\text{Minimize } x\ 1^T s \tag{17}$$

$$\text{s. t. } -s \leq x \leq s \tag{18}$$

$$Ax = b.$$

Another effective method for solving the problem (14) is to apply weighted l_1 techniques (see e.g. [6]). As an example let us consider the following continuous approximation of *card* (x)

$$Card(x) = \|x\|_0 = \lim_{\in \to 0} \sum_{i=1}^{n} \sin(atan\left(\frac{x_i}{\in}\right)). \tag{19}$$

Hence for a given small ² > 0, an approximation counterpart of (14) is given as follows

$$\text{Minimize } \sum_{i=1}^{n} \sin(atan\left(\frac{x_i}{\in}\right)). \tag{20}$$

$$\text{s. t. } Ax = b. \tag{21}$$

Note that

$$\sin\left(atan\left(\frac{|x_i|}{\in}\right)\right) \leq \sin\left(atan\left(\frac{|y_i|}{\in}\right)\right) + \frac{1}{y_i^2 + \in^2} \cos\left(atan\left(\frac{|y_i|}{\in}\right)\right)\left(|x_i| - |y_i|\right)$$

$$\leq \sin\left(atan\left(\frac{|y_i|}{\in}\right)\right) + \frac{1}{y_i^2 + \in^2}\left(|x_i| - |y_i|\right), \forall x, y.$$

Using linearization techniques (majorization minimization), one obtains the following iterative scheme:

$$x^{(k+1)} = \arg\min_x \{\sum_{i=1}^{n} \frac{|x_i|}{\left(x_i^k\right)^2 + \in^2} : Ax = b\}, \ (22)$$

where $\frac{1}{\left(x_i^{(k)}\right)^2 + \in^2}$ can be interpreted as the weight which forces the nonzero component to be zero if possible. The initial point $x^{(0)}$ can be chosen as the optimal solution of problem (15).

Before closing this section, it is worth mentioning that sometimes we are interested in finding a solution with a prescribed cardinality *t*. Such problems can be written as the following feasibility problem:

$$\begin{aligned} &\text{Find} \quad x \\ &\text{s. t.} \ Ax = b \ \ (23) \\ &card(x) \le t. \end{aligned}$$

which can be reformulated as a d.c. programming. In fact, for $x \in R^n$, the problem above is equivalent to the minimization of $(n - t)$ smallest components of x.

Now suppose $S_t(x)$ is defined as the summation over the t largest components of the vector $|x|$ (assume that $|x_1| \ge |x_2 \ge \ldots \ge |x_n||$)

$$S_t(x) = \sum_{i=1}^{t} |x_i|$$

which clearly is a convex function. Hence the problem (23) can be reformulated as

$$\text{Minimize } \{\|x\|_1 - S_t(x) : \text{Ax = b}\}. \qquad (24)$$

which is a d. c. programming problem. This problem can be solved by the cutting plane method, which is a usual approach for solving d. c. problems. However, linearization method can be still used to obtain an approximate solution for the problem.

$$\text{Minimize } \{\nabla (\|x\|_1 - S_t(x))^{\mathrm{T}} x : \text{Ax = b}\}. \qquad (25)$$

This is equivalent to

$$\begin{aligned} &\text{Minimize } (sign(x) - g)^T x \\ &\text{s. t. } Ax = b, \\ &g = \text{Maximize } u^T x \qquad (26) \\ &\text{s. t. } u \in [0, 1] \\ &1^T u = t, \end{aligned}$$

which can be viewed as a special linear bilevel programming problem.

4. CMP with Nonlinear Non-Convex Constraints

In this section, we discuss the CMP with quadratic constraints, i.e., *C* in (1) is of the form

$$C = \{x: b_i x_i^2 - a_i x_i - c_i \leq 0\},\ i = 1, \ldots, n.$$

We assume that the constraint functions are not necessarily convex, i.e., b_i is not necessarily positive. So the problem is NP-hard. In this section, we discuss the approach for the relaxation and/or reformulation of such problems.

By adding a boolean valued slack variable $v = (v_1, \ldots, v_n)^T$ to the problem, the CMP (1), with non-convex quadratic constraints, can be reformulated as

$$\begin{aligned} &Maximize \sum_{i=1}^{n} v_i = 1^T v \\ &\text{s.t. } v_i x_i = 0 \\ &v_i \in \{0,1\}, i = 1,2,\ldots,n \\ &b_i x_i^2 - a_i x_i - c_i \leq 0, i = 1,2,\ldots,n, \end{aligned} \tag{27}$$

A similar reformulation can be found in [8]. We now give a dual formulation of this problem.

Proposition 5. The dual SDP form of the problem above can be written as the following SDP problem

$$\min_{\gamma,\lambda,\mu,\beta} (\gamma : \begin{pmatrix} c(\lambda,\mu,\beta)+\gamma & b(\lambda,\mu,\beta)^T \\ b(\lambda,\mu,\beta) & A(\lambda,\mu,\beta) \end{pmatrix} \geq 0)$$

where $A(\lambda, \mu, \beta)$, $c(\lambda, \mu, \beta)$, $b(\lambda, \mu, \beta)$ are defined in (30), (31).

Proof. We make some small changes to the objective of (27) and rewrite the problem as follows

$$\begin{aligned} &Maximize \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} v \\ x \end{pmatrix} \\ &\text{s.t. } v_i x_{i=0} \\ &v_i \in \{0,1\}, i = 1,2,\ldots,n \\ &b_i x_i^2 - a_i x_i - c_i \leq 0, i = 1,2,\ldots,n, \end{aligned} \tag{28}$$

where $0 \in R^n$ is a column vector with all of its components zeros. The condition $v_i \in \{0, 1\}$ can be relaxed with $v^2 - v_i \leq 0$ which is a convex constraint, producing the following relaxation problem:

$$Maximize - \begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} v \\ x \end{pmatrix}$$

$$\text{s.t. } v_i x_{i=0}$$

$$v_i \in \{0,1\}, i = 1,2,...,n$$

$$b_i x_i^2 - a_i x_i - c_i \leq 0, i = 1,2,...,n,$$

where

$$\begin{pmatrix} v \\ x \end{pmatrix} = \left(v_1,...,v_n,x_1,...,x_n\right)^T, \left(\underbrace{1...1}_{n-times}, \underbrace{0...0}_{n-times}\right).$$

Applying Lagrange duality and adding some weight vectors μ, λ, β *yields*

$$L_{v,x}(\mu,\lambda,\beta) = \inf_{(v,x)^T \in R^{2n}} \left(-\begin{pmatrix} 1 \\ 0 \end{pmatrix}^T \begin{pmatrix} v \\ x \end{pmatrix} + \sum_{i=1}^{n} \mu_i \left(v_i^2 - v_i\right) + \sum_{i=1}^{n} \lambda_i v_i x_i + \sum_{i=1}^{n} \beta_i \left(b_i x_i^2 - a_i x_i - c_i\right) \right). \tag{29}$$

Note that

$$L_{v,x}(\mu,\lambda,\beta)$$

$$= \begin{pmatrix} v_1 \\ \vdots \\ v_n \\ x_1 \\ \vdots \\ x_n \end{pmatrix}^T \begin{pmatrix} \mu_1 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \mu_n & 0 & 0 & \lambda_n \\ \lambda_1 & 0 & 0 & \beta_1 b_1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \lambda n & 0 & 0 & \beta_n b_n \end{pmatrix}_{2n \otimes 2n} \begin{pmatrix} v_1 \\ \vdots \\ v_n \\ x_1 \\ \vdots \\ x_n \end{pmatrix} +$$

$$\begin{pmatrix} -1-\mu_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & -1-\mu_n & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_1 \alpha_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_n \alpha_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \\ x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}^T \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -c_1 \\ \vdots \\ -c_1 \end{pmatrix}.$$

Setting

$$A(\mu,\lambda,\beta) = \begin{pmatrix} \mu_1 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \mu_n & 0 & 0 & \lambda_n \\ \lambda_1 & 0 & 0 & \beta_1 b_1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \lambda n & 0 & 0 & \beta_n b_n \end{pmatrix}, c(\mu,\lambda,\beta) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}^T \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -c_1 \\ \vdots \\ -c_1 \end{pmatrix}. \tag{30}$$

$$b(\mu,\lambda,\beta)=\frac{1}{2}\begin{pmatrix} \mu_1 & 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \mu_n & 0 & 0 & \lambda_n \\ \lambda_1 & 0 & 0 & \beta_1 b_1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \lambda n & 0 & 0 & \beta_n b_n \end{pmatrix}, \tag{31}$$

and introducing a new variable y for (v, x), the function Lv,x(μ, λ, β) can be written as

$$Ly(\lambda, \mu, \beta) = y^T A(\lambda, \mu, \beta)y + 2b(\lambda, \mu, \beta)^T y + c(\lambda, \mu, \beta).$$

Assume that λ, μ, β, θ are chosen such that

$$Ly(\lambda, \mu, \beta) - \theta \geq 0, \ \forall y \in R^{2n},$$

then θ is an upper bound for the optimal value of (27). Also from chapter 3 of [2], we have

$$g(y) = y^T Ay + 2b^T y + c - \theta \geq 0 \Leftrightarrow G(y, t) = y^T Ay + 2bt^T y + (c - \theta)t^2 \geq 0.$$

So

$$G(y,t)\geq 0 \Leftrightarrow \begin{pmatrix} c-\theta & b^T \end{pmatrix} \geq 0.$$

Then looking for the best upper bound for the main problem above becomes

$$Max_{\theta,\lambda,\mu,\beta}\left\{\theta : \begin{pmatrix} c-\theta & b^T \\ b & A \end{pmatrix} \geq 0\right\}.$$

Setting θ = −γ yields a relaxation for the original problem

$$Min_{\gamma,\lambda,\mu,\beta}\left\{\gamma : \begin{pmatrix} c-\theta & b^T \\ b & A \end{pmatrix} \geq 0\right\},$$

which is an SDP and can be solved efficiently.

5. CMP with 0-1 variables

In some situations, we are interested in minimizing the cardinality of a boolean vector $x \in R^n$, i.e xi $\in \{0, 1\}$, $i = 1, \ldots, n$. So, we may consider the CMP with 0-1 variables and quadratic constraints:

$$\begin{aligned} &\text{Minimize} \quad Card(x) \\ &\text{s.t.} x^T B_i x - A_i x - b_i \leq 0, i = 1\ldots, m \quad (32) \\ &x \in \{0,1\}, \end{aligned}$$

where $B_i \geq 0$, A_i is a vector with appropriate dimension, b_i is a contant. This problem is also discussed in [13] in which the feasible set is defined by a linear system.

Define a new variable $z = 2x - 1$. Hence the problem above can be reformulated as

$$\begin{aligned} \text{Minimize} \quad & Card(z+1) \\ \text{subject to} \quad & 0.25(z+1)^T B_i(z+1) - 0.5(A_i(z+1)) - b_i \leq 0,\ i = 1, \ldots, m \\ & z \in \{-1, 1\}. \end{aligned} \tag{33}$$

Now let $Z = zz^T$. By shor's lemma, $Z = zz^T$ is equivalent to $Z \geq zz^T$ and $rank(Z) = 1$. Also note that $tr(Z) = n$. Hence the problem (33) is equivalent to the following problem.

$$\begin{aligned} \text{Minimize} \quad & 1^T z \\ \text{s. t.} \quad & 0.25(z+1)^T B_i(z+1) - 0.5(A_i(z+1)) - b_i \leq 0,\ i = 1, \ldots, m \\ & Tr(Z) = n,\ Rank(Z) = 1. \end{aligned} \tag{34}$$

Since $z \in \{-1, 1\}$, we can remove the constraint $tr(Z) = n$ and replace it by $diag(Z) = 1$.

To relax the constraint $Rank(Z) = 1$, one can replace the *rank* function by the nuclear norm which is certain convex relaxation of the rank function. So we get the following SDP:

$$\begin{aligned} \text{Minimize} \quad & 1^T z \\ \text{s.t.} \quad & 0.25(z+1)^T B_i(z+1) - 0.5(A_i(z+1)) - b_i \leq 0,\ i = 1, \ldots, m \\ & Diag(Z) = 1,\ \|Z\|_* \leq \gamma, \end{aligned} \tag{35}$$

where γ is a constant depends on the bound of $\|Z\|$.

As we mentioned in the previous sections, another simple approach to construct an approximation of the problem (34) is using penalty function as proposed by Zhao [31]. So we obtain the following approximation counterpart of (34):

$$\begin{aligned} \text{Minimize} \quad & 1^T z + \xi\|Z\|_* \\ \text{s.t.} \quad & 0.25(z+1)^T B_i(z+1) - 0.5(A_i(z+1)) - b_i \leq 0,\ i = 1, \ldots, \\ & Z \geq zz^T,\ Z \geq 0,\ Diag(Z) = 1, \end{aligned} \tag{36}$$

where $\xi > 0$ is the penalty parameter.

6. Conclusion

We have discussed different reformulations of cardinality constraint and cardinality minimization problems, and how convex techniques can be used to get approximate counterparts of these NP-hard problems. We have demonstrated that under mild assumption a cardinality constraint problem can be equivalently reformulated as a bilevel SDP problem. This involves how the penalty method can be used to reformulate the problem. Several important specific cases including CMP with linear constraints, CMP with nonlinear non-convex constraint, and CMP with $0-1$ vectors have been discussed.

Kaynakça

[1] R.G. Baraniuk, Compressive Ensing, *IEEE Signal Processing Magazine.*4, 24, 118, (2007).

[2] A. Ben-Tal and A.S. Nemirovski, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. *MPS-SIAM Series on Optimization* (2001).

[3] D.P. Bertsekas, A. Nedic, and A.E. Ozdaglar, Convex Analysis and Optimization. *Athena Scientific*, (2003).

[4] M. Bruglieri, M. Ehrgott, H.W. Hamacher, and F. Maffili, An Annotated Bibliography of Combinatorial Optimization Problems with Fixed Cardinality Constraints. *Discrete Applied Mathematics*, 154, 9, 1344-1357, (2006).

[5] E.J. Cand`es and M.B. Wakin, An Introduction to Compressive Sampling, IEEE *Signal Processing Magazine*, 25, 2, 21-30, (2008).

[6] E.J. Cand`es, M.B. Wakin, and S.P. Boyd, Enhancing Sparsity by Reweighted l1 Minimization, *Journal of Fourier Analysis and Applications*. 14, 5, 877-905, (2008).

[7] T.J. Chang, N. Meade, JE Beasley, and YM Sharaiha, Heuristics for Cardinality Constrained Portfolio Optimisation, *Computers and Operations Research,* 27, 13, 1271-1302, (2000).

[8] A. d'Aspremont, A Semidefinite Representation for Some Minimum Cardinality Problems, *Arxiv preprint math/0302092.* (2003).

[9] A. d'Aspremont, L.E. Ghaoui, M.I. Jordan, and G.R.G. Lanckriet, A Direct Formulation for Sparse PCA Using Semidefinite Programming, *Arxiv preprint cs/0406021.* (2004).

[10] J. Dattorro, Convex Optimization & Euclidean Distance Geometry, *Meboo Publishing* USA, 2005.

[11] D.L. Donoho, Compressed Sensing, *IEEE Transactions on Information Theory*, 52, 4, 1289-1306, (2006).

[12] M. Fazel, Matrix Rank Minimization with Applications, Ph.D. Thesis, Stanford University, 2002.

[13] M. Fazel, H. Hindi, and S.P. Boyd, Log-Det Heuristic for Matrix Rank Minimization with Applications to Hankel and Euclidean Distance Matrices, American Control Conference, 2003. Proceedings of the 2003, vol. 3, IEEE, 2003, pp. 2156-2162.

[14] M. Fazel, H. Hindi, and S.P. Boyd, Rank Minimization and Applications in System Theory, American Control Conference, 2004.

[15] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins Univ Pr, 1996.

[16] I.F. Gorodnitsky and B.D. Rao, Sparse Signal Reconstruction from Limited Data Using

FOCUSS: A re-weighted minimum norm algorithm, *IEEE Transactions on Signal Processing*, 45, 600-616, (1997).

[17] A.J. Laub, *Matrix Analysis for Scientists and Engineers*, SIAM, 2005.

[18] C. Lemar´echal and F. Oustry, *Semidefinite Relaxations and Lagrangian Duality with Application to Combinatorial Optimization*, Preprint, 1999.

[19] M. Lustig, D. Donoho, and J.M. Pauly, Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging, *Magnetic Resonance in Medicine*. 58, 6, 1182-1195, (2007).

[20] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly, Compressed Sensing MRI, *IEEE Signal Processing Magazine*. 25, 2, 72-82, (2008).

[21] D. Malioutov, M. Cetin, and A.S. Willsky, A Sparse Signal Reconstruction Perspective for Source Localization with Sensor Arrays, *IEEE Transactions on Signa Processing*. 53, 8, 3010-3022, (2005).

[22] D. Maringer and H. Kellerer, Optimization of Cardinality Constrained Portfolios with a Hybrid Local Search Algorithm, *OR Spectrum*. 25, 4, 481-495, (2003).

[23] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra: Solutions Manual*, SIAM, 2000.

[24] B. Moghaddam, Y. Weiss, and S. Avidan, Spectral Bounds for Sparse PCA: Exact and Greedy Algorithms, *Advances in Neural Information Processing Systems*. 18, 915, (2006).

[25] F. Streichert, H. Ulmer, and A. Zell, Evolutionary Algorithms and The Cardinality Con-Strained Portfolio Optimization Problem, Proceedings of the International Conference on Operations Research (OR 2003), Heidleberg, September 3-5, 2003.

[26] M.Y.J. Ting, Signal Processing for Magnetic Resonance Force Microscopy, Ph.D. Thesis, 2006.

[27] Y. Tsaig and D.L. Donoho, Extensions of Compressed Sensing, *Signal processing.* 86, 3, 549-571, (2006).

[28] L. Vandenberghe and S.P. Boyd, Semidefinite Programming, *SIAM Review.* 38, 1, 49-95, (1996).

[29] S. Wold, K. Esbensen, and P. Geladi, Principal Component Analysis, *Chemometrics and Intelligent Laboratory Systems,* 2, 1-3, 37-52, (1987).

[30] H. Xu, C. Caramanis, and S. Mannor, *Robust Regression and Lasso*, Arxiv preprint arXiv: 0811.1790, 2008.

[31] Y.B. Zhao, *Approximation Theory of Matrix Rank Minimization and Its Application to Quadratic Equations*, Arxiv preprint arXiv:1010.0851, 2010.

[32] X. Zheng, X. Sun, and D. Li, *Convex Relaxations and Mixed-integer Quadratic Reformulations for Cardinality Constrained Quadratic Programs*, Preprint, 2010.

[33] H. Zou, T. Hastie, and R. Ibshirani, Sparse Principal Component Analysis, *Journal of Computational and Graphical Statistics.* 15, 2, 265-286, (2006).