文章编号:2095-6134(2015)05-0577-05

A further discussion on the conservatism of robust linear optimization problems*

LIU Pengfei, YANG Wenguo†

(*School of Mathematical Sciences*, *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*;
Key Laboratory of Big Data Mining and Knowledge Management, *Chinese Academy of Sciences*, *Beijing* 100049, *China*)

(Received 19 January 2015; Revised 14 April 2015)

Liu P F, Yang W G. A further discussion on the conservatism of robust linear optimization problems[J]. Journal of University of Chinese Academy of Sciences, 2015,32(5):577-581.

Abstract The conservatism is an important indicator for measuring a robust approach. In the process of our previous research for the conservatism of robust linear programming problems, we have found that k is a critical parameter to depict the conservatism of robust linear programming problems, where k is the number of nonzero components in optimal solution of the extremely conservative robust linear programming problems. In this paper we give the distribution and expectation of k through analyzing the probability that any basic solutions are the optimal solutions of the extremely conservative robust linear programming problems.

Key words robust approach; conservatism; linear programming; distribution

CLC number: O221.1 **Document code**: A **doi**: 10.7523/j.issn.2095-6134.2015.05.001

对鲁棒线性规划保守性的进一步讨论

刘鹏飞，杨文国

(中国科学院大学数学科学学院, 北京 100049; 中国科学院大数据挖掘与知识管理重点实验室, 北京 100049)

摘　要 保守性是衡量鲁棒优化模型好坏的重要指标,也是研究鲁棒优化方法的一个关键问题.在先前关于鲁棒线性优化保守性的研究中,我们发现,线性规划最优解中非零分量的数目 k 是刻画鲁棒线性规划模型保守性的一个重要参数.本文通过分析基解是鲁棒线性规划问题最优解的概率,给出了参数 k 的概率分布和数学期望.

关键词 鲁棒方法;保守性;线性规划;分布

In recent years, the robust method has been widely used in numerous fields and many models have been developed. However, the conservatism of robust approaches has been a controversial issue

* Supported by National 973 Plan Project(2011CB706900), 863 Plan Project(2011AA01A102), NSFC(71171189, 11331012, 71271204, and 11101420), the "Strategic Priority Research Program" of Chinese Academy of Sciences (XDA06010302), and the Open Preject of Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences and Huawei Technology Co., Ltd.

†Corresponding author, E-mail: yangwg@ucas.ac.cn

since they were proposed. In order to reduce degree of the conservatism, numerous formulations were developed[1-7]. One of the most influential formulations is the one proposed by Bertsimas and Sim[3]. They considered the following nominal linear programming problem:

$$\begin{aligned} \max \quad & c'x \\ \text{s.t} \quad & \sum_j a_{ij} x_j \leqslant b_i \\ & x \geqslant 0. \end{aligned}$$

Each entry a_{ij} is modeled as a symmetric and bounded random variable, and a_{ij} takes values in $[a_{ij}-\hat{a}_{ij}, a_{ij}+\hat{a}_{ij}]$, where $\hat{a}_{ij} \geqslant 0$. Besides, a limit is applied to the number of coefficients that are allowed to change simultaneously. The model of Bertsimas and Sim is:

$$\begin{aligned} \max \quad & c'x \\ \text{s.t} \quad & \sum_j a_{ij} x_j + \\ & \max_{\{S_i \cup t_i | S_i \subseteq J, |S_i| = \lfloor \Gamma \rfloor, t_i \in J \backslash S_i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} x_j + \right. \\ & \left. (\Gamma - \lfloor \Gamma \rfloor) \hat{a}_{it_i} x_{t_i} \right\} \leqslant b_i \quad \forall i \\ & x \geqslant 0. \end{aligned} \tag{1}$$

Here, J is the set of the indices, i. e. $J = \{1, 2, \cdots, n\}$, where n is the dimension of x. As mentioned in Ref. [3], model (1) gives full control of the degree of the conservatism associated with the constraints. When Γ progressively increases from zero to n, the degree of the conservatism of the above approach is expected to increase gradually. When Γ progressively increases to n, we get the extremely conservative robust approach:

$$\begin{aligned} \max \quad & c'x \\ \text{s.t} \quad & \sum_j (a_{ij} + \hat{a}_{ij}) x_j \leqslant b_i \quad \forall i \\ & x \geqslant 0. \end{aligned} \tag{2}$$

However, Liu and Yang[1] pointed out that model (1) may become extremely conservative even when Γ is far less than n. Moreover, it is showed that model (1) does not reach the extremely conservative state when Γ is less than k, where k is the number of nonzero components of the optimal solution of the extremely conservative robust approach.

As stated above, when $\Gamma \geqslant k$, model (1) may reach the extremely conservative state; when $\Gamma < k$, model (1) does not reach the extremely conservative state. That is, k is an important parameter to analyze the conservativeness of model (1). To some degree, k is a good indicator of the conservativeness of model (1). Based on the study of the distribution of k, we can have a further understanding of the conservativeness of model (1) on the whole. In this paper we give the distribution and expectation of k under two assumptions which are widely used in the research of linear programming problems.

1 Preliminaries

First we should have a review of the definitions corresponding to linear programming which are necessary for the following statement. All of our discussions are under the assumption that the rows of the constraint matrix are less than the columns. The standard form of linear programming is

$$\begin{aligned} \max \quad & c'x \\ \text{s.t} \quad & \sum_{j=1}^{n} a_{ij} x_j = b_i, i = 1, \cdots, m \\ & x \geqslant 0. \end{aligned}$$

Definition 1.1 If we set $(n-m)$ variables equal to zero and then solve the m equations for the remaining variables, the resulting solution, if unique, is called a basic solution.

Definition 1.2 The $(n-m)$ variables which are set equal to zero are known as nonbasic variables. The remaining m variables are called basic variables. If a basic solution satisfies nonnegative condition, it is called a basic feasible solution.

Definition 1.3 The basic variables with a value of zero are called degenerate, and the same term is applied to the corresponding basic feasible solution.

$$\begin{aligned} \max \quad & c'x \\ \text{s.t} \quad & \sum_j a_{ij} x_j \leqslant b_i \quad \forall i = 1, 2, \cdots, n \\ & x \geqslant 0. \end{aligned} \tag{3}$$

Definition 1.4 The constraints of the form

$Ax(\leqslant \text{ or } \geqslant)b$ are called matrix constraints, and the constraints of the form $x(\leqslant \text{ or } \geqslant)0$ are called sign constraints. For a basic solution $\hat{x}$, if the right side of certain matrix constraints (or sign constraints) is equal to the left at $\hat{x}$, the constraints are called efficient at $\hat{x}$.

For example, consider the following linear programming. The matrix constraints are $x_1 + x_2 \leqslant 5$ and $x_1 + 2x_2 \leqslant 7$, the sign constraints are $x_1 \geqslant 0$ and $x_2 \geqslant 0$. For the solution $(5,0)$, the efficient constraints are $x_1 + x_2 \leqslant 5$ and $x_2 \geqslant 0$.

$$\begin{aligned} \max \quad & 2x_1 + x_2 \\ \text{s.t} \quad & x_1 + x_2 \leqslant 5 \\ & x_1 + 2x_2 \leqslant 7 \\ & x_1 \geqslant 0, x_2 \geqslant 0. \end{aligned}$$

Lemma 1.1 If $\hat{x}$ is a non-degenerate basic solution of model (3), there are n and only n efficient constraints at $\hat{x}$ among all the matrix constraints and sign constraints of model (3). Besides, the coefficient vectors of the n efficient constraints are linearly independent.

Proof See Ref. [8]. □

Consider the above example. The solution $(5,0)$ is a non-degenerate basic solution, *and* there are two and only two efficient constraints at $(5,0)$, i.e. $x_1 + x_2 \leqslant 5$ and $x_2 \geqslant 0$. Obviously, coefficient vectors $(1,1)$ and $(0,1)$ are linearly independent. Besides, it is easy to see that $(5,0)$ is the optimal solution and the target vector $(2,1) = 2\times(1,1) - (0,1)$, which is an example of the following lemma.

Lemma 1.2 Assume the target vector $c \neq 0$. Let $\tilde{x}$ be a nondegenerate basic feasible solution of model (3). Let $[a_{j_1}, a_{j_2}, \cdots, a_{j_k}, I_{j_{k+1}}, I_{j_{k+2}}, \cdots, I_{j_n}]$ be the rows which are coefficient vectors of the efficient constraints at $\tilde{x}$, where a_j and I_j are the ith row of A and the ith row of unit matrix, respectively.

i) The necessary and sufficient condition that $\tilde{x}$ is the optimal solution of model (3) is that there exists a $\boldsymbol{\lambda} \geqslant \mathbf{0}$ such that $c = \lambda_1 a_{j_1} + \lambda_2 a_{j_2} + \cdots + \lambda_k a_{j_k} - \lambda_{k+1} I_{j_{k+1}} - \lambda_{k+2} I_{j_{k+2}} - \cdots - \lambda_n I_{j_n}$.

ii) If $\tilde{x}$ is the unique optimal solution of model (3), $\boldsymbol{\lambda} > 0$.

Proof for i), see Ref. [1].

for ii):

Suppose not, assume $\lambda_1 = 0$. Let $G = \{1,2,\cdots,m\}$ and $\hat{G} = \{j_2,\cdots,j_k\}$. Let $G_1 = \{1,2,\cdots,n\}$ and $\hat{G}_1 = \{j_{k+1},j_{k+2},\cdots,j_n\}$. Let $D = \{x \mid a_j x = b_j\ \forall j \in \hat{G}, I_j x = 0\ \forall j \in \hat{G}_1, a_i x \leqslant b_i\ \forall i \in G\setminus\hat{G}, I_i x \geqslant 0\ \forall i \in G_1\setminus\hat{G}_1\}$. Obviously, any $x \in D$ is the feasible solution of model (3). Since there are only $(n-1)$ equality constraints, there are more than one element in D. And we have $c = \lambda_2 a_{j_2} + \cdots + \lambda_k a_{j_k} - \lambda_{k+1} I_{j_{k+1}} - \lambda_{k+2} I_{j_{k+2}} - \cdots - \lambda_n I_{j_n}$. By Lemma 2.2 i), any $x \in D$ is the optimal solution of model (3). This contradicts with the condition that the optimal solution is unique. The statement follows immediately. □

As shown in Ref. [1], Liu and Yang pointed out that model (1) may become extremely conservative even when Γ is far less than n. However, they also give the following result:

Theorem 1.1 Suppose x^* is the optimal solution of model (2) with c as the target vector. Let k be the number of nonzero components in x^*. If $\sum_j |c_j x_j^*| > 0$, $k \geqslant 1$, and $\Gamma \leqslant k-1$, x^* is not the optimal solution of model (1).

Proof See Ref. [1]. □

Theorem 1.1 reveals that k is an important parameter when we analyze the conservativeness of model (1). In the following we will give the distribution and expectation of k.

2 Distribution and expectation of k

To simplify the following exposition, we will use A instead of $\hat{A}$ to represent the constraints matrix of model (2).

Let X be a random variable which represents the number of nonzero components in the optimal solution. We will discuss the probability $p(X=k)$ under the assumption that the probability distribution of (A,b) satisfies the following assumptions[9]:

i) For any fixed (A,b), all sign combinations of the inequalities

$$A(i)\boldsymbol{x} \leqslant (\geqslant)\ b_i \quad i = 1,2,\cdots,m,$$
$$x_j \leqslant (\geqslant) 0 \quad j = 1,2,\cdots,n$$

are equiprobable. Under this assumption, any basic solution will appear in 2^{m+n} instances.

ii) With probability one,

$$\begin{pmatrix} A & b \\ I & 0 \end{pmatrix}$$

are non-degenerate. In other words, with probability one, any basis solutions are non-degenerate.

iii) With probability one, there is a unique optimal solution.

Considering any fixed (A,b,c), there are 2^{m+n} instances in total. We will give the probability that a basic solution is the optimal solution.

Lemma 2.1 Let $\hat{\boldsymbol{x}}$ be a basic solution, then the probability that $\hat{\boldsymbol{x}}$ is the optimal solution is $1/2^{m+n}$.

Proof Let $\hat{\boldsymbol{x}}$ be a basic solution. Obviously, $\hat{\boldsymbol{x}}$ will appear in 2^{m+n} instances. By Lemma 1.1, there are n efficient constraints at $\hat{\boldsymbol{x}}$. Let $[a_{j_1}, a_{j_2}, \cdots, a_{j_k}, I_{j_{k+1}}, I_{j_{k+2}}, \cdots, I_{j_n}]$ be the rows which are coefficient vectors of the efficient constraints at $\hat{\boldsymbol{x}}$. That is, $\hat{\boldsymbol{x}}$ satisfies formulation (4).

$$\begin{cases} a_{j_1}\hat{\boldsymbol{x}} = b_{j_1} \\ \quad\vdots \\ a_{j_k}\hat{\boldsymbol{x}} = b_{j_k} \\ I_{j_{k+1}}\hat{\boldsymbol{x}} = 0 \\ \quad\vdots \\ I_{j_n}\hat{\boldsymbol{x}} = 0. \end{cases} \tag{4}$$

Let $G = \{1,2,\cdots,m\}$ and $\hat{G} = \{j_1, j_2, \cdots, j_k\}$. For any $i \in G\setminus\hat{G}$, $\hat{\boldsymbol{x}}$ belongs to $\{\boldsymbol{x} \mid a_i\boldsymbol{x} \geqslant b_i\}$ or $\{\boldsymbol{x} \mid a_i\boldsymbol{x} \leqslant b_i\}$. By assumption ii) and Lemma 1.1, $\hat{\boldsymbol{x}}$ belongs to only one of them. Let $G_1 = \{1,2,\cdots,n\}$ and $\hat{G}_1 = \{j_{k+1}, j_{k+2}, \cdots, j_n\}$. For any $i \in G_1\setminus\hat{G}_1$, $\hat{\boldsymbol{x}}$ belongs to $\{\boldsymbol{x} \mid I_i\boldsymbol{x} \geqslant 0\}$ or $\{\boldsymbol{x} \mid I_i\boldsymbol{x} \leqslant 0\}$. By assumption ii) and Lemma 1.1, $\hat{\boldsymbol{x}}$ belongs to only one of them. Without loss of generality, suppose $\hat{\boldsymbol{x}}$ satisfies formulation (5). Since $|G\setminus\hat{G}| + |G_1\setminus\hat{G}_1| = m$, the probability that formulation (6) occurs is $1/2^m$.

$$\begin{cases} a_{j_1}\boldsymbol{x} = b_{j_1} \\ \quad\vdots \\ a_{j_k}\boldsymbol{x} = b_{j_k} \\ I_{j_{k+1}}\boldsymbol{x} = 0 \\ \quad\vdots \\ I_{j_n}\boldsymbol{x} = 0 \\ a_i\boldsymbol{x} \leqslant b_i \ \forall i \in G\setminus\hat{G} \\ I_i\boldsymbol{x} \leqslant 0 \ \forall i \in G_1\setminus\hat{G}_1. \end{cases} \tag{5}$$

$$\begin{cases} a_i\boldsymbol{x} \leqslant b_i \ \forall i \in G\setminus\hat{G} \\ I_i\boldsymbol{x} \leqslant 0 \ \forall i \in G_1\setminus\hat{G}_1. \end{cases} \tag{6}$$

By Lemma 1.1, $[a_{j_1}, a_{j_2}, \cdots, a_{j_k}, I_{j_{k+1}}, I_{j_{k+2}}, \cdots, I_{j_n}]$ are linearly independent. Therefore, there exists a unique $\boldsymbol{\lambda}$ such that

$$\boldsymbol{c} = \sum_{i=1}^{k} \lambda_i a_{j_i} - \sum_{i=k+1}^{n} \lambda_i I_{j_i} =$$
$$\sum_{i=1}^{k} |\lambda_i| \frac{\lambda_i}{|\lambda_i|} a_{j_i} - \sum_{i=k+1}^{n} |\lambda_i| \frac{\lambda_i}{|\lambda_i|} I_{j_i},$$

where $\boldsymbol{c}$ is the target vector. Note that $|\lambda_i| > 0$ for any i (by Lemma 1.2 ii) and assumption iii)). By Lemma 1.2 i), $\hat{\boldsymbol{x}}$ is the optimal solution of (7).

$$\begin{aligned} \max \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s. t} \quad & \frac{\lambda_{j_1}}{|\lambda_{j_1}|} a_{j_1}\boldsymbol{x} \leqslant b_{j_1} \\ & \quad\vdots \\ & \frac{\lambda_{j_k}}{|\lambda_{j_k}|} a_{j_k}\boldsymbol{x} \leqslant b_{j_k} \\ & \frac{\lambda_{j_{k+1}}}{|\lambda_{j_{k+1}}|} I_{j_{k+1}}\boldsymbol{x} \leqslant 0 \\ & \quad\vdots \\ & \frac{\lambda_{j_n}}{|\lambda_{j_n}|} I_{j_n}\boldsymbol{x} \leqslant 0 \\ & a_i\boldsymbol{x} \leqslant b_i \quad \forall i \in G\setminus\hat{G} \\ & I_i\boldsymbol{x} \leqslant 0 \quad \forall i \in G_1\setminus\hat{G}_1. \end{aligned} \tag{7}$$

By assumption i), the probability that formulation (8) occurs is $1/2^n$. That is, under the assumption that $\hat{\boldsymbol{x}}$ satisfies formulation (6), the probability that formulation (7) occurs is $1/2^n$. Therefore, under the assumption that $\hat{\boldsymbol{x}}$ satisfies formulation (6), the probability that $\hat{\boldsymbol{x}}$ is the optimal solution is $1/2^n$. Since the probability that formulation (6) occurs is $1/2^m$, the probability that a basic solution is the optimal solution is $1/2^{m+n}$.

$$\begin{cases} \dfrac{\lambda_{j_1}}{|\lambda_{j_1}|} a_{j_1}\boldsymbol{x} \leqslant b_{j_1} \\ \vdots \\ \dfrac{\lambda_{j_k}}{|\lambda_{j_k}|} a_{j_k}\boldsymbol{x} \leqslant b_{j_k} \\ \dfrac{\lambda_{j_{k+1}}}{|\lambda_{j_{k+1}}|} I_{j_{k+1}}\boldsymbol{x} \leqslant 0 \\ \vdots \\ \dfrac{\lambda_{j_n}}{|\lambda_{j_n}|} I_{j_n}\boldsymbol{x} \leqslant 0. \end{cases} \tag{8}$$

Lemma 2.1 shows that for different basic solutions, the probabilities that they are the optimal solutions are the same, i. e. $1/2^{m+n}$. Considering any fixed (A, b, c), we have the following theorem.

Theorem 2.1 Let X be a random variable which represents the number of nonzero components in the optimal solution, then

$$p(X = k) = \frac{C_m^k C_n^{n-k}}{C_{m+n}^n},$$

$$E(X) = \sum_{k=0}^{m} k \cdot \frac{C_m^k C_n^{n-k}}{C_{m+n}^n} = \frac{mn}{m+n},$$

where $p(X = k)$ is the probability of $X = k$, $E(X)$ is the expectation of X, m is the number of the rows of A, and n is the number of the columns of A.

Proof Let X_1 be a random variable which represents the number of nonzero components in basic solutions. By Lemma 2.1, the probabilities that any basic solutions are the optimal solutions are the same. So the distribution of X_1 is the same as the distribution of X. We will give the distribution of X_1 below.

Let $\hat{\boldsymbol{x}}$ be a basic solution. The total number of basic solutions is C_{m+n}^n. If there are k matrix constraints and $(n-k)$ sign constraints among the efficient constraints at $\hat{\boldsymbol{x}}$, the number of nonzero components in $\boldsymbol{x}$ is k. There are C_m^k cases to choose k constraints to be regarded as efficient constraints among the m matrix constraints. There are C_n^{n-k} cases to choose $(n-k)$ constraints to be regarded as efficient constraints among the n sign constraints. Therefore, the probability that $X_1 = k$ is

$$p(X_1 = k) = \frac{C_m^k C_n^{n-k}}{C_{m+n}^n}.$$

That is,

$$p(X = k) = \frac{C_m^k C_n^{n-k}}{C_{m+n}^n},$$

$$E(X) = \sum_{k=0}^{m} k \frac{C_m^k C_n^{n-k}}{C_{m+n}^n} = \frac{mn}{m+n}. \qquad \square$$

Since Lemma 2.1 and Theorem 2.1 hold for any fixed (A, b, c), Lemma 2.1 and Theorem 2.1 also hold when we regard all the A whose dimensions are $m \times n$ as a space.

References

[1] Liu P F, Yang W G, Guo T D. A discussion on the conservatism of robust linear optimization problems[J/OL]. Eprints for the optimization community. (2014-10)[2015-01-10]. http://www.optimization-online.org/DB_HTML/2014/10/4598.html.

[2] Soyster A L. Technical note-convex programming with set-inclusive constraints and applications to inexact linear programming[J]. Operations Research, 1973, 21(5): 1 154-1 157.

[3] Bertsimas D, Sim M. The price of robustness[J]. Operations Research, 2004, 52(1): 35-53.

[4] Ben-Tal A, Nemirovski A. Robust convex optimization[J]. Mathematics of Operations Research, 1998, 23(4): 769-805.

[5] Ben-Tal A, Nemirovski A. Robust solutions of uncertain linear programs[J]. Operations Research Letters, 1999, 25(1): 1-13.

[6] EI-Ghaoui L, Lebret H. Robust solutions to least-square problems to uncertain data matrices[J]. Sima Journal on Matrix Analysis and Applications, 1997, 18: 1 035-1 064.

[7] El Ghaoui L, Oustry F, Lebret H. Robust solutions to uncertain semidefinite programs[J]. SIAM Journal on Optimization, 1998, 9(1): 33-52.

[8] Hillier F S, Lieberman G J. Introduction to operations research[M]. 9th ed. San Francisco: Mc Graw Hill-Higher Education, 2010: 107-109.

[9] Adler I, Karp R M, Shamir R. A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps[J]. Journal of Complexity, 1987, 3(4): 372-387.