

Global Error Bounds for Convex Conic Problems

Shuzhong Zhang
Econometric Institute
Erasmus University
P.O. Box 1738
3000 DR Rotterdam
The Netherlands

July, 1998

ABSTRACT

In this paper Lipschitzian type error bounds are derived for general convex conic problems under various regularity conditions. Specifically, it is shown that if the recession directions satisfy Slater's condition then a global Lipschitzian type error bound holds. Alternatively, if the feasible region is bounded, then the ordinary Slater condition guarantees a global Lipschitzian type error bound. These can be considered as generalizations of previously known results for inequality systems. Moreover, some of the results are also generalized to the intersection of multiple cones. Under Slater's condition alone, a global Lipschitzian type error bound may not hold. However, it is shown that such an error bound holds for a specific region. For linear systems we show that the constant involved in Hoffman's error bound can be estimated by the so-called condition number for linear programming.

Key words: Error bound, convex conic problems, LMIs, condition number.

AMS subject classification: 52A20, 65G99, 90C31.

1 Introduction

In optimization theory it is often desirable to measure the distance to the solution set from a certain given point. In general, this distance can be difficult to assess, since one may not have a complete knowledge about the solution set. However, if the form of the solution set is explicitly given, then in some cases it is possible to estimate the distance to the solution set by the so-called *constraint violation* which is computable. This kind of estimation is termed *error bound relation*. The first such result was obtained by Hoffman [7] for systems of linear equalities and inequalities. We shall discuss Hoffman's error bound in the paper too. A recent extensive survey on various types of error bound results can be found in Pang [15].

Most papers discussing error bound results assume that the solution set is given by equations and inequalities, e.g.

$$S = \{x \mid f_i(x) = 0 \text{ for } i = 1, \dots, m \text{ and } g_j(x) \leq 0 \text{ for } j = 1, \dots, l\}.$$

For a given point x the amount of constraint violation can be measured as the following quantity

$$v(x) = \|f(x)\| + \|(g(x))_+\|$$

where $f(x) = (f_1(x), \dots, f_m(x))$ and $(g(x))_+ = ((g_1(x))_+, \dots, (g_l(x))_+)$ with the notation $(y)_+ = \max(y, 0)$.

A measure for constraint violation is similar to a penalty function in the sense that it takes positive value for points outside the set, and zero otherwise. Note that a measure for constraint violation should be easy computable, such as the case for the above defined function $v(x)$. Hoffman's lemma [7] states that if $S \neq \emptyset$, and f_i and g_j are all affine linear functions, then there is a positive constant $\kappa > 0$ such that

$$\text{dist}(x, S) \leq \kappa v(x) \tag{1.1}$$

for all $x \in \mathfrak{R}^n$. This means that the distance to S is of the same magnitude as $v(x)$. Such a relation is known as a Lipschitzian type error bound.

In the case that f_i and g_j are not linear, the above inequality (1.1) does not hold in general. Early results concerning nonlinear functions are due to Robinson [17] and Mangasarian [13]. Robinson [17] showed that for inequality systems if all functions are convex and differentiable, S is bounded and the Slater condition holds, i.e. there is a \hat{x} such that $g_j(\hat{x}) < 0$ for all j , then relation (1.1) holds. Mangasarian [13] removed the assumption that S is bounded by assuming an additional asymptotic constraint qualification condition, which however can be difficult to verify in general.

In this paper we consider the following convex conic set:

$$\mathcal{F} = (b + \mathcal{L}) \cap \mathcal{K} \tag{1.2}$$

where $b \in \mathfrak{R}^n$, \mathcal{L} is a subspace of \mathfrak{R}^n and $\mathcal{K} \subseteq \mathfrak{R}^n$ is a closed convex cone. Polynomial-time interior-point algorithms for solving convex optimization problems with convex conic feasible set were introduced in a systematic manner by Nesterov and Nemirovskii [14]. It turns out that many important classes of optimization problems, such as linear programming and semidefinite programming, can be cast in this form. The focus of this paper is to discuss how error bound type relation can be established for such problems. Throughout this paper we make the following assumption:

Assumption 1 $\mathcal{F} \neq \emptyset$.

The organization of the paper is as follows. In the next section we prove that with a proper definition of constraint violation a Lipschitzian type error bound (1.1) can be established for general convex conic problems, under various conditions on the relations between \mathcal{L} and \mathcal{K} , including Slater type conditions. In Section 3 we discuss a link between the constant in Hoffman's error bound and the so-called condition number for linear programming. Finally, we conclude the paper in Section 4.

We use the following notation in this paper. Matrices are denoted by capital letters, e.g. X . For symmetric matrix X , $\lambda_{\max}(X)$ indicates the maximum eigenvalue of X , and $\lambda_{\min}(X)$ the minimum eigenvalue of X . We denote n -dimensional Euclidean space by \mathfrak{R}^n and its nonnegative quadrant by \mathfrak{R}_+^n . The space of all symmetric n by n matrices is denoted by $\mathcal{S}^{n \times n}$ and the cone of all symmetric positive semidefinite n by n matrices by $\mathcal{S}_+^{n \times n}$. Vector e represents a vector of all ones with appropriate dimension. For a vector $v \in \mathfrak{R}^n$, we use the capitalized letter V to denote the diagonal matrix which takes v as its diagonal elements. For two vectors $x \in \mathfrak{R}^n$ and $y \in \mathfrak{R}^n$ we write $xy \in \mathfrak{R}^n$ as the component-wise Hadamard product. We use the Euclidean norm for vectors and the spectral norm for matrices. A vector $a \geq 0$ means that each component of a is nonnegative, and $X \succeq 0$ indicates that X is positive semidefinite.

2 Convex conic systems

Consider the convex conic set (1.2). For convenience we further assume that \mathcal{K} is a pointed and solid cone, i.e. $\mathcal{K} \cap (-\mathcal{K}) = \{0\}$ and $\dim \mathcal{K} = n$.

The dual of \mathcal{K} is

$$\mathcal{K}^* = \{x \mid x^T y \geq 0 \text{ for all } y \in \mathcal{K}\}.$$

Since \mathcal{K} is pointed and solid, \mathcal{K}^* too is a closed, convex, pointed and solid cone.

An immediate next question is: How can we define a constraint violation function for \mathcal{F} ? For this purpose we note the following lemma due to Moreau (see Theorem 31.5 in [18]).

Lemma 2.1 For any $x \in \mathfrak{R}^n$ there is a unique $x_p \in \mathcal{K}$ and $x_d \in \mathcal{K}^*$ such that $x = x_p - x_d$ and $x_p^T x_d = 0$.

In fact, x_p is simply the projection of x onto \mathcal{K} and $\|x_d\|$ measures the distance from x to \mathcal{K} . A natural definition for the constraint violation for \mathcal{F} is now in order:

Definition 2.1 For any $x \in \mathfrak{R}^n$ define

$$v(x; \mathcal{F}) := \text{dist}(x, b + \mathcal{L}) + \|x_d\|$$

as the constraint violation function for \mathcal{F} .

It is readily seen that $v(x, \mathcal{F}) = 0$ iff $x \in \mathcal{F}$.

It is, however, not immediately clear how the function $v(x; \mathcal{F})$ can be computed. Below we shall see some examples in which this function is explicitly derived. First we consider the case $\mathcal{K} = \mathfrak{R}_+^n$, the nonnegative quadrant of \mathfrak{R}^n . Clearly, $x = x_+ + x_-$ where $x_+ = ((x_1)_+, \dots, (x_n)_+)$ and $x_- = (-(-x_1)_+, \dots, -(-x_n)_+)$. Obviously, $x_+ \in \mathcal{K}$ and $-x_- \in \mathcal{K}$ and $x_+^T x_- = 0$. Therefore, $\|x_d\| = \|(x)_-\|$, which is exactly the usual definition of the violation for nonnegativity constraints.

Another example is $\mathcal{K} = \mathcal{S}_+^{n \times n}$, the cone of n by n symmetric positive semidefinite matrices. Consider a given n by n symmetric matrix X . Following Lemma 2.1 we know that there is unique positive semidefinite matrices X_p and X_d such that $X = X_p - X_d$ and $\text{tr } X_p X_d = 0$. Matrices X_p and X_d can be computed as follows. Let $X = Q\Lambda Q^T$ with Q an orthonormal matrix and Λ is a diagonal matrix with eigenvalues of X as its components. Splitting $\Lambda = \Lambda_+ + \Lambda_-$, where Λ_+ and Λ_- denote the nonnegative and nonpositive parts of Λ respectively, it follows that $X_p = Q\Lambda_+ Q^T$, $X_d = -Q\Lambda_- Q^T$ and $X_p X_d = 0$. Hence, $\|X_d\| = \max(0, -\lambda_{\min}(X))$, where $\lambda_{\min}(X)$ denotes the minimum eigenvalue of X .

Finally, we consider another popular convex cone: the second order cone $\mathcal{K} \in \mathfrak{R}^{n+1}$ defined as

$$\mathcal{K} = \{(x_0, x) \mid x \in \mathfrak{R}^n \text{ and } x_0 \geq \|x\|\}.$$

It can be shown that in this case

$$\|x_d\| = (\|x\| - x_0)_+ / \sqrt{2}.$$

In general, Definition 2.1 is only related to the geometry of the object under consideration.

Consider now an arbitrary point $z \in \mathfrak{R}^n$. Assume that $z \notin \mathcal{F}$. The following problem yields a unique point in \mathcal{F} with the shortest Euclidean distance to z :

$$\begin{aligned} \text{(Proj)} \quad & \text{minimize} && \frac{1}{2} \|x - z\|^2 \\ & \text{subject to} && x \in b + \mathcal{L} \\ & && x \in \mathcal{K}. \end{aligned}$$

Let this optimal solution be \bar{x} . The Karush-Kuhn-Tucker optimality condition for (Proj) is given as follows:

$$(\text{KKT}) \begin{cases} \bar{x} - z + \lambda - \mu = 0 \\ \bar{x}^T \mu = 0 \\ \bar{x} \in (b + \mathcal{L}) \cap \mathcal{K} \\ \mu \in \mathcal{K}^* \\ \lambda \in \mathcal{L}^\perp. \end{cases} \quad (2.1)$$

Hence,

$$\begin{aligned} \|\bar{x} - z\|^2 &= (\bar{x} - z)^T (\bar{x} - z) \\ &= (\bar{x} - z)^T (\mu - \lambda) \\ &= -(z_p - z_d)^T \mu + (z - \bar{x})^T \lambda \\ &\leq z_d^T \mu + (z - \bar{x})^T \lambda \\ &\leq \|z_d\| \|\mu\| + (z - \bar{x})^T \lambda, \end{aligned} \quad (2.2)$$

where the first inequality follows from the fact that $z_p \in \mathcal{K}$ and $\mu \in \mathcal{K}^*$.

Let the projection of z onto the affine subspace $b + \mathcal{L}$ be z_l . Then,

$$\begin{aligned} (z - \bar{x})^T \lambda &= (z - z_l + z_l - \bar{x})^T \lambda \\ &= (z - z_l)^T \lambda \\ &\leq \|z - z_l\| \|\lambda\| \\ &= \text{dist}(z, b + \mathcal{L}) \|\lambda\|. \end{aligned}$$

Substitute this relation into (2.2) we obtain

$$(\text{dist}(z, \mathcal{F}))^2 = \|\bar{x} - z\|^2 \leq \|z_d\| \|\mu\| + \text{dist}(z, b + \mathcal{L}) \|\lambda\|. \quad (2.3)$$

In Section 3 we shall discuss how to further bound the errors when \mathcal{K} is a polyhedral cone, which is the situation when the original Hoffman lemma applies. In the rest of this section we assume that \mathcal{K} is a general convex cone. In addition to this we assume that the Slater condition is satisfied, i.e.

Assumption 2 $(b + \mathcal{L}) \cap \text{int } \mathcal{K} \neq \emptyset$.

The following lemma is well-known; see e.g., Duffin [5], Borwein and Wolkowicz [2], Luo, Sturm and Zhang [11], Nesterov and Nemirovskii [14] and Sturm [21]. For completeness we provide here a short proof.

Lemma 2.2 *Suppose that Assumption 2 holds. Then, for any $y \in \mathcal{L}^\perp \cap \mathcal{K}^*$ with $y \neq 0$ it must follow that $b^T y > 0$.*

Proof. Suppose, for the sake of deriving a contradiction, that there is $y \neq 0$ such that $y \in \mathcal{L}^\perp \cap \mathcal{K}^*$ and $b^T y \leq 0$.

Consider the hyperplane

$$H_y = \{x \mid y^T x = 0\}.$$

For any $x \in b + \mathcal{L}$ we have $y^T x = b^T y \leq 0$, while for any $x \in \mathcal{K}$, since $y \in \mathcal{K}^*$ we have $y^T x \geq 0$. This means that H_y separates $b + \mathcal{L}$ and \mathcal{K} , yielding a contradiction to the fact that $b + \mathcal{L}$ intersects with the interior of \mathcal{K} . □

For fixed \bar{x} we consider again the system (KKT) in terms of μ and λ . After some re-arrangements this yields

$$\begin{cases} \mu - \lambda = \bar{x} - z \\ \bar{x}^T \mu = 0 \\ \mu \in \mathcal{K}^* \\ \lambda \in \mathcal{L}^\perp. \end{cases} \quad (2.4)$$

Define

$$\bar{\mathcal{K}}^* = \mathcal{K}^* \cap \{\mu \mid \bar{x}^T \mu = 0\}$$

which is a closed convex cone as well.

Note that $\bar{x} = 0$ is a trivial case and is omitted in our proof. In many applications, $0 \notin \mathcal{L}$ and so $\bar{x} \neq 0$.

We shall mention another easy case, i.e. \bar{x} lies in the interior of \mathcal{K} then $\bar{\mathcal{K}} = \{0\}$. In this case $\mu = 0$ and $\lambda = z - \bar{x}$, and therefore

$$\text{dist}(z, \mathcal{F}) \leq \text{dist}(z, b + \mathcal{L})$$

due to (2.3). In what remains we shall only concentrate on the situation when $\bar{x} \notin \text{int } \mathcal{K}$.

Remark that for $\bar{x} \in \mathcal{K}$, the cone $\bar{\mathcal{K}}^*$ is known as a *face* of \mathcal{K}^* .

The condition (2.4) is equivalent to

$$\mu \in (\bar{x} - z + \mathcal{L}^\perp) \cap \bar{\mathcal{K}}^*.$$

Lemma 2.3 *If Assumption 2 holds then $\mathcal{L}^\perp \cap \bar{\mathcal{K}}^* = \{0\}$.*

Proof. Suppose for contradiction that there is $y \neq 0$ and $y \in \mathcal{L}^\perp \cap \bar{\mathcal{K}}^*$. This means that

$$y \in \mathcal{L}^\perp \cap \mathcal{K}^* \text{ and } \bar{x}^T y = 0.$$

However, $\bar{x} \in b + \mathcal{L}$ and $y \in \mathcal{L}^\perp$, and so

$$0 = \bar{x}^T y = b^T y$$

which is impossible due to Lemma 2.2. □

Now we define the minimum angle between \mathcal{L}^\perp and $\bar{\mathcal{K}}^*$ as

$$\angle(\mathcal{L}^\perp, \bar{\mathcal{K}}^*) := \min\{\arccos(u^T v / (\|u\| \|v\|)) \mid u \in \mathcal{L}^\perp \setminus \{0\}, v \in \bar{\mathcal{K}}^* \setminus \{0\}\}.$$

Note that both \mathcal{L}^\perp and $\bar{\mathcal{K}}^*$ are closed cones. According to Lemma 2.3, it follows that

$$\angle(\mathcal{L}^\perp, \bar{\mathcal{K}}^*) > 0$$

for any $\bar{x} \in (b + \mathcal{L}) \cap \mathcal{K}$.

In order to pursue our analysis further, one of the following two mutually exclusive cases will be considered.

Assumption 3 *The set $\mathcal{F} = (b + \mathcal{L}) \cap \mathcal{K}$ is compact.*

Assumption 4 $\mathcal{L} \cap \text{int } \mathcal{K} \neq \emptyset$.

Let us first consider the situation when Assumption 3 holds. In that case we know that there exists $\theta > 0$ such that for any $\bar{x} \in \mathcal{F}$ we always have

$$\angle(\mathcal{L}^\perp, \bar{\mathcal{K}}^*) \geq \theta > 0.$$

Now take $\mu \in (\bar{x} - z + \mathcal{L}^\perp) \cap \bar{\mathcal{K}}^*$. Let the projection of 0 onto $\bar{x} - z + \mathcal{L}^\perp$ be p .

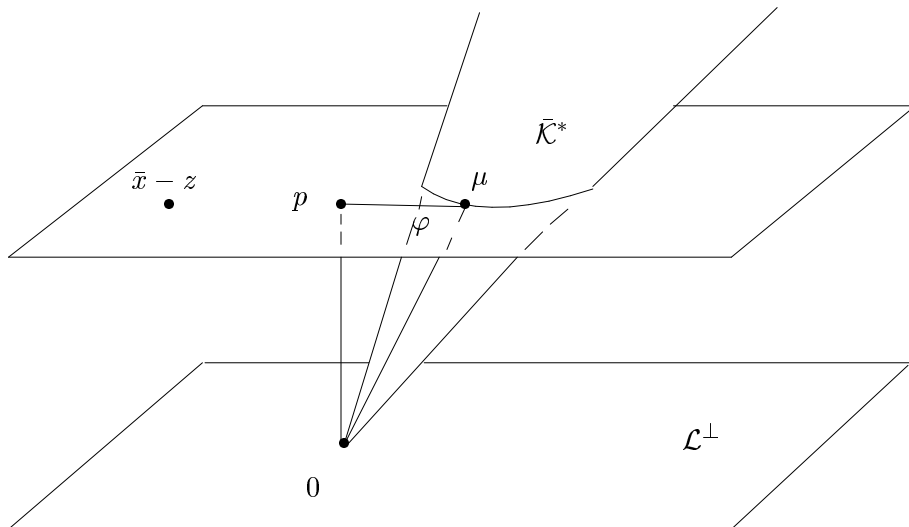


Figure: Subspace \mathcal{L}^\perp and the cone $\bar{\mathcal{K}}^*$.

Let the angle between μ and $\mu - p$ be φ . Clearly, $\theta \leq \varphi \leq \pi/2$. Moreover,

$$\|\mu\| = \|p\|/\sin \varphi \leq \|p\|/\sin \theta \leq \|\bar{x} - z\|/\sin \theta. \quad (2.5)$$

Denote

$$\kappa = 1 + 1/\sin \theta.$$

We are now in a position to prove the following error bound result.

Theorem 2.1 *If Assumption 2 and Assumption 3 hold then*

$$\text{dist}(z, \mathcal{F}) \leq \kappa v(z; \mathcal{F})$$

for all $z \in \mathfrak{R}^n$.

Proof. By (2.5) we have

$$\|\mu\| \leq \text{dist}(z, \mathcal{F})/\sin \theta \leq \kappa \text{dist}(z, \mathcal{F}).$$

Using the first equation in (2.4) we also have

$$\|\lambda\| \leq \|\mu\| + \|\bar{x} - z\| \leq (1 + 1/\sin \theta) \text{dist}(z, \mathcal{F}) = \kappa \text{dist}(z, \mathcal{F}).$$

Recall relation (2.3). By the above estimations on $\|\mu\|$ and $\|\lambda\|$, it follows from (2.3) that

$$(\text{dist}(z, \mathcal{F}))^2 \leq \kappa \text{dist}(z, \mathcal{F})(\|z_d\| + \text{dist}(z, b + \mathcal{L}))$$

and consequently

$$\text{dist}(z, \mathcal{F}) \leq \kappa v(z; \mathcal{F}).$$

□

In the other situation, namely if Assumption 4 holds, then a similar result can be shown.

Theorem 2.2 *If Assumption 4 holds, then for any $b \in \mathfrak{R}^n$ we must have $(b + \mathcal{L}) \cap \text{int } \mathcal{K} \neq \emptyset$. Moreover, there is a constant $\kappa > 0$, independent of b , such that*

$$\text{dist}(z, \mathcal{F}) \leq \kappa v(z; \mathcal{F})$$

where $\mathcal{F} = (b + \mathcal{L}) \cap \mathcal{K}$.

Proof. First we show that $(b + \mathcal{L}) \cap \text{int } \mathcal{K} \neq \emptyset$ for all b . Suppose otherwise that there is b with

$$(b + \mathcal{L}) \cap \text{int } \mathcal{K} = \emptyset.$$

Then, there will be a hyperplane separating $b + \mathcal{L}$ and \mathcal{K} , say with $0 \neq y \in \mathfrak{R}^n$ and $c \in \mathfrak{R}$ such that

$$y^T(b + x) \leq c \text{ for all } x \in \mathcal{L}$$

$$y^T x \geq c \text{ for all } x \in \mathcal{K}.$$

Since \mathcal{K} is a closed cone, the above separation implies that $y^T x \geq 0$ for all $x \in \mathcal{K}$ and $c = 0$. Moreover, we also have $y^T x = 0$ for all $x \in \mathcal{L}$. This is in contradiction with the condition $\mathcal{L} \cap \text{int } \mathcal{K} \neq \emptyset$.

Compared with Lemma 2.3, we have now a stronger relation: $\mathcal{L}^\perp \cap \mathcal{K}^* = \{0\}$. This means that the proof of Theorem 2.1 can remain exactly the same, except that now $\theta > 0$ can be taken as the minimum angle between \mathcal{L}^\perp and \mathcal{K}^* which is independent of b . □

We remark that both Theorem 2.1 and Theorem 2.2 easily extend to the case when \mathcal{L} is a closed cone.

Theorem 2.3 *Suppose that \mathcal{K}_1 is a closed convex cone and \mathcal{K}_2 is a closed, convex, solid and pointed cone. Furthermore, suppose that $(b + \mathcal{K}_1) \cap \text{int } \mathcal{K}_2 \neq \emptyset$ and $(b + \mathcal{K}_1) \cap \mathcal{K}_2$ is compact. Then there is a constant $\kappa > 0$ such that*

$$\text{dist}(z, \mathcal{F}) \leq \kappa(\text{dist}(z, b + \mathcal{K}_1) + \text{dist}(z, \mathcal{K}_2))$$

for all $z \in \mathfrak{R}^n$, where $\mathcal{F} = (b + \mathcal{K}_1) \cap \mathcal{K}_2$.

Proof. We follow similar lines as in the proof of Theorem 2.1. Consider

$$\begin{aligned} & \text{minimize} && \frac{1}{2}\|x - z\|^2 \\ & \text{subject to} && x \in b + \mathcal{K}_1 \\ & && x \in \mathcal{K}_2. \end{aligned}$$

Let the optimal solution be \bar{x} . The Karush-Kuhn-Tucker optimality condition yields:

$$\begin{cases} \bar{x} - z - \mu_1 - \mu_2 = 0 \\ (\bar{x} - b)^T \mu_1 = 0 \\ \bar{x}^T \mu_2 = 0 \\ \bar{x} \in (b + \mathcal{K}_1) \cap \mathcal{K}_2 \\ \mu_1 \in \mathcal{K}_1^* \\ \mu_2 \in \mathcal{K}_2^*. \end{cases}$$

Let

$$\bar{\mathcal{K}}_1^* = \mathcal{K}_1^* \cap \{\mu \mid (\bar{x} - b)^T \mu = 0\}$$

and

$$\bar{\mathcal{K}}_2^* = \mathcal{K}_2^* \cap \{\mu \mid \bar{x}^T \mu = 0\}.$$

Both $\bar{\mathcal{K}}_1^*$ and $\bar{\mathcal{K}}_2^*$ are closed convex cones.

Now we claim that

$$(-\bar{\mathcal{K}}_1^*) \cap \bar{\mathcal{K}}_2^* = \{0\}. \quad (2.6)$$

Suppose such is not the case. Then, one should be able to find $\mu \neq 0$ satisfying

$$\begin{cases} \mu \in (-\mathcal{K}_1^*) \cap \mathcal{K}_2^* \\ (\bar{x} - b)^T \mu = 0 \\ \bar{x}^T \mu = 0. \end{cases}$$

Hence, $b^T \mu = 0$. Therefore, $\mu^T(b + x) \leq 0$ for all $x \in \mathcal{K}_1$ and $\mu^T x \geq 0$ for all $x \in \mathcal{K}_2$. This implies that $\{x \mid \mu^T x = 0\}$ separates $b + \mathcal{K}_1$ from \mathcal{K}_2 , contradicting the Slater condition.

Since $-\bar{\mathcal{K}}_1^*$ and $\bar{\mathcal{K}}_2^*$ are closed convex cones and, moreover, $\bar{\mathcal{K}}_2^*$ is a solid pointed cone, we derive from (2.6) that $\bar{\mathcal{K}}_2^*$ can be strictly separated from $-\bar{\mathcal{K}}_1^*$. Due to compactness of \mathcal{F} we may let θ be a positive lower bound on the minimum angle between this separating hyperplane and $\bar{\mathcal{K}}_2^*$. Then we have

$$\|\mu_2\| \leq \|\bar{x} - z\| / \sin \theta$$

and consequently

$$\|\mu_1\| \leq (1 + 1/\sin \theta) \|\bar{x} - z\|.$$

Now,

$$\begin{aligned}
\|\bar{x} - z\|^2 &= (\bar{x} - z)^T (\mu_1 + \mu_2) \\
&= (b - z)^T \mu_1 - z^T \mu_2 \\
&\leq \text{dist}(z, b + \mathcal{K}_1) \|\mu_1\| + \text{dist}(z, \mathcal{K}_2) \|\mu_2\| \\
&\leq (1 + 1/\sin \theta) (\text{dist}(z, b + \mathcal{K}_1) + \text{dist}(z, \mathcal{K}_2)) \|\bar{x} - z\|.
\end{aligned}$$

The desired result thus follows. \square

Similarly, we have the following result, the proof of which is pretty much the same as that of Theorems 2.2 and 2.3 and is omitted here.

Theorem 2.4 *Suppose that \mathcal{K}_1 is a closed convex cone and \mathcal{K}_2 is a closed, convex, solid and pointed cone. Furthermore, suppose that $\mathcal{K}_1 \cap \text{int } \mathcal{K}_2 \neq \emptyset$. Then, for any $b \in \mathfrak{R}^n$ here is a constant $\kappa > 0$, independent of b , such that*

$$\text{dist}(z, \mathcal{F}) \leq \kappa (\text{dist}(z, b + \mathcal{K}_1) + \text{dist}(z, \mathcal{K}_2))$$

for all $z \in \mathfrak{R}^n$, where $\mathcal{F} = (b + \mathcal{K}_1) \cap \mathcal{K}_2$.

When more than two cones are concerned, a similar result holds under Slater's condition. First we note the following lemma, see e.g. [11].

Lemma 2.4 *Let \mathcal{K} be a convex cone and $\text{int } \mathcal{K} \neq \emptyset$. Then, $x \in \text{int } \mathcal{K}$ if and only if for any $0 \neq \mu \in \mathcal{K}^*$ it holds that $\angle(x, \mu) \geq \theta > 0$.*

Theorem 2.5 *Let \mathcal{K}_i be convex cones, $i = 1, \dots, m$. Suppose that*

$$\bigcap_{i=1}^m \text{int } \mathcal{K}_i \neq \emptyset.$$

Then, there is $\kappa > 0$ such that

$$\text{dist}(z, \bigcap_{i=1}^m \mathcal{K}_i) \leq \kappa \sum_{i=1}^m \text{dist}(z, \mathcal{K}_i)$$

for any $z \in \mathfrak{R}^n$.

Proof. Consider

$$\begin{aligned}
&\text{minimize} && \frac{1}{2} \|x - z\|^2 \\
&\text{subject to} && x \in \mathcal{K}_i, i = 1, \dots, m.
\end{aligned}$$

Hence, for the optimal solution \bar{x} the KKT condition yields

$$\bar{x} - z = \sum_{i=1}^m \mu_i$$

with $\mu_i \in \mathcal{K}_i^*$ and $\bar{x}^T \mu_i = 0$ for $i = 1, \dots, m$.

Let

$$d \in \bigcap_{i=1}^m \text{int } \mathcal{K}_i.$$

By Lemma 2.4 there exists $g_i > 0$ satisfying

$$d^T \mu_i \geq g_i \|\mu_i\|$$

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

Let

$$z = z_{ip} - z_{id}$$

with $z_{ip} \in \mathcal{K}_i$, $z_{id} \in \mathcal{K}_i^*$ and $z_{ip}^T z_{id} = 0$ due to Lemma 2.1. Moreover, $\|z_{id}\| = \text{dist}(z, \mathcal{K}_i)$, $i = 1, \dots, m$.

Therefore,

$$\begin{aligned} \|\bar{x} - z\|^2 &= (\bar{x} - z)^T \sum_{i=1}^m \mu_i \\ &= - \sum_{i=1}^m z^T \mu_i \\ &= - \sum_{i=1}^m (z_{ip} - z_{id})^T \mu_i \\ &\leq \sum_{i=1}^m z_{id}^T \mu_i \\ &\leq \sum_{i=1}^m \|z_{id}\| \|\mu_i\|. \end{aligned}$$

On the other hand, since

$$\|d\| \|\bar{x} - z\| \geq d^T (\bar{x} - z) = \sum_{i=1}^m d^T \mu_i \geq g_i \|\mu_i\|$$

for $i = 1, \dots, m$, it follows that

$$\|\bar{x} - z\|^2 \leq \sum_{i=1}^m \text{dist}(z, \mathcal{K}_i) (\|d\|/g_i) \|\bar{x} - z\|$$

and so by letting

$$\kappa = \max_{i=1, \dots, m} \|d\|/g_i$$

it follows that

$$\text{dist}(z, \cap_{i=1}^m \mathcal{K}_i) \leq \kappa \sum_{i=1}^m \text{dist}(z, \mathcal{K}_i).$$

□

Theorem 2.1 can be viewed as an analogue to Robinson's result for convex inequality systems. In the form of convex inequality systems, Theorem 2.2 can be found in Hu and Wang [9] and Deng and Hu [3]. In particular, Deng and Hu [3] investigated the case when \mathcal{K} is the cone of positive semidefinite matrices. This case is known as *linear matrix inequalities* (LMIs for short). In its optimization version it is also called *semidefinite programming* and has received intensive research attention recently. Sturm [20] mainly investigated error bounds for LMIs in the absence of Slater's condition. In fact, in the context of LMIs, both Theorem 2.1 and Theorem 2.2 also follow from the analysis in [20]. Moreover, an example was given in Sturm [20] showing that Assumption 2 alone cannot guarantee a global Lipschitzian type error bound even for LMIs. Such an error bound is only possible when an additional scaling factor is present.

Below we shall discuss how to derive some conditioned error bound relation for the convex conic problem (1.2) under Assumption 2, without assuming Assumption 3 and Assumption 4.

In this situation the recession cone $\mathcal{L} \cap \mathcal{K}$ must be non-empty and it is not contained in the interior of \mathcal{K} .

For a fixed positive angle $0 < \theta < \pi/2$, consider the following cone

$$\mathcal{C} = \{x \mid \text{the projection of } x \text{ onto } \mathcal{L} \text{ and the cone } \mathcal{L} \cap \mathcal{K} \text{ has an angle at least } \pi/2 + \theta\}.$$

Theorem 2.6 *Suppose that Assumption 2 holds. There exists a constant $\kappa > 0$ such that*

$$\text{dist}(z, \mathcal{F}) \leq \kappa v(z; \mathcal{F}).$$

for all $z \in \mathcal{C}$.

Proof. Observe that if \bar{x} is the projection of z on \mathcal{F} , then it must also be the projection of $z + y$ on \mathcal{F} for any $y \in \mathcal{L}^\perp$. This can be seen as follows. The fact that $\bar{x} \in (b + \mathcal{L}) \cap \mathcal{K}$ is the projection of z is equivalent to the existence of $\lambda \in \mathcal{L}^\perp$ and $\mu \in \mathcal{K}^*$ such that

$$\bar{x} - z = \mu - \lambda \text{ and } \bar{x}^T \mu = 0.$$

(See also (2.1)). Now if z is changed to $z + y$, then we need only to change λ to $\lambda + y \in \mathcal{L}^\perp$ to satisfy the same set of KKT conditions.

Remark also that to prove the theorem it is sufficient to show that, for any $z \in \mathcal{C}$, its projection onto \mathcal{F} is contained in a compact set.

Suppose that the theorem is false and that there is a sequence $\{z^{(k)} \in \mathcal{C} \mid k = 1, 2, \dots\}$, such that the corresponding projection on \mathcal{F} , $\{\bar{x}^{(k)} \in \mathcal{F} \mid k = 1, 2, \dots\}$, is unbounded. Due to the above remarks we have made, we need only to consider the projection of $z^{(k)}$ onto the subspace \mathcal{L} . Without loss of generality, assume that $z^{(k)} \in \mathcal{L} \cap \mathcal{C}$ for all k .

For sufficiently large k we have

$$\begin{aligned} \|z^{(k)} - \bar{x}^{(k)}\|^2 &= \|z^{(k)}\|^2 + \|\bar{x}^{(k)}\|^2 - 2\langle z^{(k)}, \bar{x}^{(k)} \rangle \\ &\geq \|z^{(k)}\|^2 + \|\bar{x}^{(k)}\|^2 + \cos(\theta/2)\|z^{(k)}\|\|\bar{x}^{(k)}\| \\ &> \|z^{(k)} - \bar{x}^{(1)}\|^2 \end{aligned}$$

where the first inequality is because $\bar{x}^{(k)}$ must be pointing towards the cone of recession directions $\mathcal{L} \cap \mathcal{K}$, and the last inequality is due to the fact that $\|\bar{x}^{(k)}\| \rightarrow \infty$. This contradicts to $\bar{x}^{(k)}$ being the closest point in \mathcal{F} to $z^{(k)}$. □

For any given point $z \in \mathfrak{R}^n$, we may decompose $z = z_1 + z_2$ with $z_1 \in \mathcal{L} \cap \mathcal{K}$ and $z_2 \in \mathcal{C}$. The following relation is immediate.

Lemma 2.5 $\text{dist}(z, \mathcal{F}) \leq \text{dist}(z_2, \mathcal{F})$.

Proof. Let the projection of z_2 onto \mathcal{F} be \bar{x}_2 . Then,

$$\text{dist}(z, \mathcal{F}) \leq \|z_2 + z_1 - (\bar{x}_2 + z_1)\| = \text{dist}(z_2, \mathcal{F})$$

where we used the fact that $z_1 \in \mathcal{L} \cap \mathcal{K}$ and so $\bar{x}_2 + z_1 \in \mathcal{F}$. □

Combining Lemma 2.5 and Theorem 2.6 we have

Theorem 2.7 *Suppose that Assumption 2 holds. Then*

$$\text{dist}(z, \mathcal{F}) \leq \kappa v(z_2; \mathcal{F})$$

for all $z \in \mathfrak{R}^n$ with $z = z_1 + z_2$, $z_1 \in \mathcal{L} \cap \mathcal{K}$ and $z_2 \in \mathcal{C}$.

3 Hoffman's error bound and the condition number

In this section we shall discuss error bounds for the linear system $\{x \mid A^T x \leq b\}$ with $A \in \mathfrak{R}^{m \times n}$ and $\text{rank}(A) = m$. This is the setting for which Hoffman's error bound result applies ([7]). Our purpose is to see how the constant in Hoffman's bound is related to other known quantities for the linear system. Previous results on the constant of Hoffman's bound can be found, e.g., in [12, 1, 6, 10].

By introducing a slack $s(x) = b - A^T x$ we confine ourselves to the range space of A^T , i.e.

$$\mathcal{L} = \{s \mid \exists x \in \mathfrak{R}^m : s = A^T x\}.$$

Accordingly, $\mathcal{K} = \mathfrak{R}_+^n$.

For a given $z \in \mathfrak{R}^n$ with $s(z) \notin \mathfrak{R}_+^n$. Let $s(\bar{x}) = b - A^T \bar{x} \in (b + \mathcal{L}) \cap \mathcal{K}$ which minimizes the distance to $s(z)$.

Let

$$K = \{i \mid s(\bar{x})_i > 0\} \text{ and } J = \{1, \dots, n\} \setminus K.$$

Then, for this given $s(\bar{x}) \geq 0$ we can rewrite (2.1) as

$$\begin{cases} A_J \mu_J = AA^T(\bar{x} - z) \\ \mu_K = 0 \\ \mu_J \geq 0. \end{cases} \quad (3.1)$$

As (3.1) is a necessary condition for optimality, it is certain that (3.1) is feasible. What remains to be analyzed is the size of the solution. A key ingredient in our analysis is the following lemma.

Lemma 3.1 *Suppose that A has full row rank. Then,*

$$\chi(A) := \sup\{\|DA^T(ADA^T)^{-1}\| \mid D \text{ diagonal and } D \succ 0\} < \infty.$$

Lemma 3.1 was first shown by Dikin [4] and was used in his convergence analysis for affine scaling methods. Among others, Stewart [19] and Todd [23] rediscovered this result later.

The meaning of Lemma 3.1 can be interpreted as follows. It is well known that $\text{Null}(A) = \{x \mid Ax = 0\}$ and $\text{Range}(A^T) = \{x \mid \exists y \in \mathfrak{R}^m x = A^T y\}$ are orthocomplements to each other. Obviously, for a given positive diagonal matrix D , $\text{Null}(A)$ can only intersect with $D\text{Range}(A^T)$ at the origin, hence there must be a positive angle between them. Lemma 3.1 further states that the minimum angle between $\text{Null}(A)$ and $D\text{Range}(A^T)$ is uniformly bounded from below by a positive constant which is independent of D .

To understand this fact we may consider the following example. Let $A = [1, 1]$. Then $\text{Null}(A)$ is simply the line $x_1 + x_2 = 0$. For a given positive diagonal matrix D , $D\text{Range}(A^T)$ is contained

in the first and the third quadrants. The angle between these two subspaces never exceeds $\pi/4$. An important property of the constant $\chi(A)$ is that it reflects an intrinsic, geometric relationship of the spaces. Vavasis and Ye [24] used this constant $\chi(A)$ as a measure of complexity for solving the related linear programming problem. Their results showed that, in a real-number computation model, linear program is solvable in polynomial-time, in terms of total number of basic operations, with respect to the dimension n and the complexity measure $\log \chi(A)$. For problems with integral input data, this result yields the usual polynomiality complexity result for linear programs in terms of the input-length.

Holder, Sturm and Zhang [8] showed that $\chi(A)$ plays an important role in sensitivity analysis for linear programming. Furthermore, Sturm and Zhang [22] extended some of the results in [8] to semidefinite programming. It is known however, that Lemma 3.1 cannot extend to general semidefinite programming for arbitrary invariant scaling of the cone $\mathcal{S}_+^{n \times n}$; see [22].

Fortunately, in analyzing (3.1) we need only to deal with a polyhedral cone. To see how condition number $\chi(A)$ can play a role in error bound analysis, we need to introduce a number of technical lemmas.

First we note the following equivalent definition of $\chi(A)$ for arbitrary matrix A due to Vavasis and Ye [24].

Lemma 3.2 *It holds that*

$$\chi(A) = \sup \left\{ \frac{\|y\|}{\|c\|} \mid y \text{ minimizes } \|D^{1/2}(A^T y - c)\| \text{ for } 0 \neq c \in \Re^n \text{ and } D \text{ positive diagonal} \right\}.$$

For our analysis it is important to know the size of a solution for a linear system. To this end, we note the following two lemmas. Remark that Renegar [16] studied similar problems in a quite general framework using a quantity called *distance to ill-posedness*.

Lemma 3.3 *Suppose that A has full row rank. Further assume that $\{x \mid Ax = b, x > 0\} \neq \emptyset$. Then, there is a solution \bar{x} in $\mathcal{F} = \{x \mid Ax = b, x \geq 0\}$ such that*

$$\|\bar{x}\| \leq \chi(A)\|b\|.$$

Proof. Consider a linear program

$$\begin{aligned} \text{(P)} \quad & \text{minimize} && e^T x \\ & \text{subject to} && Ax = b \\ & && x \geq 0 \end{aligned}$$

and its dual

$$\begin{aligned}
 \text{(D)} \quad & \text{maximize} && b^T y \\
 & \text{subject to} && A^T y + s = e \\
 & && s \geq 0.
 \end{aligned}$$

Both (P) and (D) satisfy Slater's condition. Therefore their respective analytic central paths $\{x(\mu) \mid \mu > 0\}$ and $\{(y(\mu), s(\mu)) \mid \mu > 0\}$ exist, satisfying the following relation:

$$\begin{cases} Ax(\mu) = b \\ A^T y(\mu) + s(\mu) = e \\ x(\mu)s(\mu) = \mu e. \end{cases} \quad (3.2)$$

Multiplying the second equation in (3.2) with $X(\mu)$, the diagonal matrix with $x(\mu)$ as its diagonal components, and applying the first equation in (3.2) we obtain

$$y(\mu) = (AX(\mu)A^T)^{-1}b - \mu(AX(\mu)A^T)^{-1}e.$$

Substituting this into the second equation and finally using the third relation in (3.2) we have

$$x(\mu) = X(\mu)A^T(AX(\mu)A^T)^{-1}b + \mu e - \mu X(\mu)A^T(AX(\mu)A^T)^{-1}Ae.$$

Now we can apply Lemma 3.1 to obtain

$$\|x(0)\| = \|\lim_{\mu \rightarrow 0} x(\mu)\| \leq \chi(A)\|b\|.$$

The lemma is proven. □

Next we shall extend this result to the case when Slater's condition is no longer assumed.

Lemma 3.4 *Suppose that A has full row rank. Further assume that $\{x \mid Ax = b, x \geq 0\} \neq \emptyset$. Then, there is a solution \bar{x} in $\mathcal{F} = \{x \mid Ax = b, x \geq 0\}$ such that*

$$\|\bar{x}\| \leq \chi(A)\|b\|.$$

Proof. Let $\delta > 0$. Consider a perturbed set

$$\mathcal{F}_\delta = \{x \mid Ax = b + \delta Ae, x \geq 0\}.$$

Clearly, \mathcal{F}_δ contains an interior point and therefore Lemma 3.3 can be invoked. Let $x^\delta \in \mathcal{F}_\delta$ and

$$\|x^\delta\| \leq \chi(A)\|b + \delta Ae\|.$$

The set $\{x^\delta \mid 0 < \delta < 1\}$ is bounded. Let x^0 be a cluster point of x^δ as $\delta \rightarrow 0$. Obviously, $x^0 \in \mathcal{F}$ and

$$\|x^0\| \leq \chi(A)\|b\|.$$

□

Next we shall compare the condition number of A and that of its submatrices.

Lemma 3.5 *Let $A = [A_1, A_2]$. Then $\chi(A_1) \leq \chi(A)$.*

Proof. By Lemma 3.2,

$$\chi(A) = \sup\left\{\frac{\|y\|}{\|c\|} \mid y \text{ minimizes } \|D^{1/2}(A^T y - c)\| \text{ for } 0 \neq c \in \mathfrak{R}^n \text{ and } D \text{ positive diagonal}\right\}.$$

Let $c = [c_1^T, c_2^T]^T$ and $D = \text{diag}(D_1, D_2)$ be partitioned in accordance with $A = [A_1, A_2]$.

For fixed $c_1 \neq 0$ and fixed positive diagonal matrix D_1 . Let $c_2 = 0$, and let D_2 be positive diagonal and $D_2 \rightarrow 0$. Clearly, the set of solutions minimizing $\|D^{1/2}(A^T y - c)\|$ converges to the set of solutions minimizing $\|D_1^{1/2}(A_1^T y - c_1)\|$. For given c and D let $y(c, D)$ be a maximum norm solution among solutions which minimize $\|D^{1/2}(A^T y - c)\|$. Define $y(c_1, D_1)$ similarly. It follows that

$$\limsup_{D_2 \rightarrow 0} \|y(c, D)\| \geq \|y(c_1, D_1)\|.$$

As a consequence,

$$\chi(A) \geq \chi(A_1)$$

and so the lemma is proven. □

Applying Lemmas 3.4 and 3.5 to (3.1) we have

$$\|\mu_J\| \leq \chi(A_J)\|(AA^T)(\bar{x} - z)\| \leq \chi(A)\lambda_{\max}(AA^T)\|\bar{x} - z\|. \quad (3.3)$$

Finally we shall give a bound on the constant in Hoffman's error bound for linear systems.

Theorem 3.1 *Suppose that $\mathcal{F} = \{x \mid A^T x \leq b\} \neq \emptyset$ and A has full row rank. It holds that*

$$\text{dist}(z, \mathcal{F}) \leq \chi(A)(\text{cond}(AA^T))\|(A^T z - b)_+\|$$

for any $z \in \mathfrak{R}^n$, where $\text{cond}(AA^T) := \lambda_{\max}(AA^T)/\lambda_{\min}(AA^T)$.

Proof. Using (3.1) and (3.3),

$$\|\mu\| = \|\mu_J\| \leq \chi(A)\lambda_{\max}(AA^T)\|\bar{x} - z\|.$$

By (2.3), on one hand we have

$$\begin{aligned} \|s(\bar{x}) - s(z)\|^2 &= (s(\bar{x}) - s(z))^T(-\lambda + \mu) \\ &= (s(\bar{x}) - s(z))^T\mu \\ &= -s(z)^T\mu \\ &= (A^T z - b)^T\mu \\ &\leq (A^T z - b)_+^T\mu \\ &\leq \|(A^T z - b)_+\|\|\mu\| \\ &\leq \|(A^T z - b)_+\|\chi(A)\lambda_{\max}(AA^T)\|\bar{x} - z\|. \end{aligned}$$

On the other hand,

$$\|s(\bar{x}) - s(z)\|^2 \geq \lambda_{\min}(AA^T)\|\bar{x} - z\|^2.$$

Combining these two inequalities, the desired result follows. \square

4 Conclusions

In this paper we discuss error bounds for sets in convex conic form. The notion of constraint violation is extended to this class of problems. For a number of applications the measure of constraint violation is easy computable. We show that under Slater's condition, and additionally, if either the feasible set is bounded or the recession directions satisfy the Slater's condition, then there is a global Lipschitzian type error bound for general convex conic problems. These results can be generalized to the intersection of multiple convex cones, or intersection of two shifted convex cones, one of them being pointed and solid. If only Slater's condition is satisfied without additional assumptions on the feasible region, then a global error bound is impossible as shown by Sturm [20]. In this case, one may still identify a region in which Lipschitzian type error bound holds. Finally, we discuss the bounds in Hoffman's lemma for linear systems. It is shown that such a bound is linked closely with the condition number for linear programming as investigated by Vavasis and Ye [24].

Acknowledgement: I would like to thank Jos Sturm for pointing out an error in an earlier version of the paper.

References

- [1] C. Bergthaller and I. Singer, The distance to a polyhedron, *Linear Algebra and Its Applications* 169 (1992) 111-129.
- [2] J.M. Borwein and H. Wolkowicz, Characterizations of optimality for the abstract convex program with finite dimensional range, *Journal of the Australian Mathematical Analysis and Applications* 83 (1981) 495-530.
- [3] S. Deng and H. Hu, Computable error bounds for semidefinite programming, *Technical Report*, Department of Mathematical Sciences, Northern Illinois University, USA, 1996.
- [4] I.I. Dikin, Iterative solution of problems of linear and quadratic programming, *Soviet Mathematics Doklady* 8 (1967) 674-675.
- [5] R.J. Duffin, Infinite programs, *Linear Inequalities and Related Systems*, editors H.W. Kuhn and A.W. Tucker, Princeton University Press, pp. 157-170, 1956.
- [6] O. Güler, A.J. Hoffman and U.G. Rothblum, Approximations to solutions to systems of linear inequalities, *SIAM Journal of Matrix Analysis and Applications* 16 (1995) 688-696.
- [7] A.J. Hoffman, On approximate solutions of systems of linear inequalities, *Journal of Research of the National Bureau of Standards* 49 (1952) 263-265.
- [8] A. Holder, J.F. Sturm and S. Zhang, Analytic central path, sensitivity analysis and parametric programming, *Report 9801/A*, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1998.
- [9] H. Hu and Q. Wang, On approximate solutions of infinite systems of linear inequalities, *Linear Algebra and Its Applications* 114/115 (1989) 429-438.
- [10] W. Li, The sharp Lipschitz constants for feasible and optimal solutions of a perturbed linear program, *Linear Algebra and Its Applications* 187 (1993) 15-40.
- [11] Z.Q. Luo, J.F. Sturm and S. Zhang, Duality results for conic convex programming, *Report 9719/A*, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1997.
- [12] O.L. Mangasarian, A condition number of linear inequalities and equalities, in: G. Bamber and O. Optiz eds., *Proc. 6th Symp. über Operations Research*, Universität Augsburg, 1981, *Methods of Operations Research* 43 (1981) 3-15.
- [13] O.L. Mangasarian, A condition number for differentiable convex inequalities, *Mathematics of Operations Research* 10 (1985) 474-484.

- [14] Yu. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics, Philadelphia, 1994.
- [15] J.-S. Pang, Error bounds in mathematical programming, *Mathematical Programming* 79 (1997) 299-332.
- [16] J. Renegar, Some perturbation theory for linear programming, *Mathematics Programming* 65 (1994) 73-91.
- [17] S.M. Robinson, An application of error bounds for convex programming in a linear space, *SIAM Journal on Control and Optimization* 13 (1975) 271-273.
- [18] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
- [19] G.W. Stewart, On scaled projections and pseudoinverses, *Linear Algebra and Its Applications* 112 (1989) 189-193.
- [20] J.F. Sturm, Error bounds for linear matrix inequalities, *Technical Report*, Communication Research Laboratory, McMaster University, Canada, 1998.
- [21] J.F. Sturm, Primal-dual interior point approach to semidefinite programming, *Ph.D Thesis Series 156*, Tinbergen Institute, Erasmus University Rotterdam, The Netherlands, 1997.
- [22] J.F. Sturm and S. Zhang, On sensitivity of central solutions in semidefinite programming, *Report 9813/A*, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1998.
- [23] M.J. Todd, A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm, *Operations Research* 38 (1990) 1006-1018.
- [24] S.A. Vavasis and Y. Ye, A primal-dual interior point method whose running time depends only on the constraint matrix, *Mathematical Programming* 74 (1996) 79-120.