A priori bounds on the condition numbers in interior-point methods

Florian Jarre, Mathematisches Institut,
Heinrich-Heine Universität Düsseldorf, Germany.

Abstract

Interior-point methods are known to be sensitive to ill-conditioning and to scaling of the data. This paper presents new asymptotically sharp bounds on the condition numbers of the linear systems at each iteration of an interior-point method for solving linear or semidefinite programs and discusses a stopping test which leads to a problem-independent "a priori" bound on the condition numbers.

Key words: Condition number, interior-point method, stopping test.

August 7, 2015

1 Introduction

Asymptotically sharp bounds on the condition number of the linear systems in interior-point methods are presented. These bounds can be reduced by a preprocessing that is applicable for dense problems. To motivate the bounds, the preprocessing is discussed first followed by a statement of the bounds with and without preprocessing. As a starting point, linear programs are considered in Section 2. The extension to semidefinite programs is addressed in Section 3.

1.1 Notation

The components of a vector $x \in \mathbb{R}^n$ are denoted by x_i. Inequalities such as $x \geq 0$ are understood componentwise. The Euclidean norm of a vector $x \in \mathbb{R}^n$ is denoted by $\|x\|$ and $\mathrm{Diag}(x)$ denotes the $n \times n$ diagonal matrix with diagonal entries x_i for $1 \leq i \leq n$. The identity matrix is denoted by I, its dimension being evident from the context. The space of symmetric $n \times n$-matrices is denoted by $\mathcal{S}^n$. The scalar product of two symmetric matrices X, S is given by $X \bullet S := \mathrm{trace}(XS)$ inducing the Frobenius-norm and $X \succeq 0$ ($X \succeq 0$) indicates that X is symmetric and positive semidefinite (positive definite). The condition number of a matrix is always measured with respect to the 2-norm.

2 Linear programming

Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ be given. Consider a linear program in standard form,

$$(LP) \qquad \min\ c^T x \mid Ax = b, \quad x \geq 0.$$

At each iteration of an interior-point algorithm with an infeasible starting point, given some iterate x, y, s with $x, s > 0$, systems of the following form (or a very similar form) for finding a correction $\Delta x, \Delta y, \Delta s$ are considered:

$$\begin{array}{rcl} A\Delta x &=& b - Ax \\ A^T\Delta y + \Delta s &=& c - A^T y - s \\ S\Delta x + X\Delta s &=& \mu_+ e - Xs. \end{array} \tag{1}$$

Here, $S = \mathrm{Diag}(s)$, $X = \mathrm{Diag}(x)$, $e = (1, \ldots, 1)^T$, and μ_+ is chosen as a parameter satisfying $\mu_+ \leq \mu := \frac{x^T s}{n}$.

Solving the second line for Δs and inserting both into the first row leads to a linear system of the form

$$AD^2A^T\Delta y = rhs \tag{2}$$

for some right hand side rhs, where $D^2 = XS^{-1}$ see e.g. [4].

2.1 Preprocessing of dense LPs

For problems with a dense data matrix A the following preprocessing lends itself to reduce the dependence of interior-point algorithms on the scaling of the data of the linear program:

Prior to solving problem (LP), a Cholesky factor L with $LL^T = AA^T$ is computed and $[A, b]$ is replaced with $L^{-1}[A, b]$. (After this transformation A has orthonormal rows.) Then, c is replaced with $c - A^T Ac$, and b and c are scaled to Euclidean norm 1. This scaling implies that the approximate solution generated by the interior-point algorithm needs to be rescaled to fit the original problem. (For $b = 0$ or $c = 0$ one may consider modified problems with $b = Ae$ or $c = e$.)

2.2 A geometric stopping test

A point (x, y, s) is primal and dual optimal, if x and (y, s) satisfy the linear equations, if $x, s \geq 0$, and if x is orthogonal to s (with respect to the scalar product associated with the Euclidean norm). By the above preprocessing, the origin has Euclidean distance 1 from the set $\{x \mid Ax = b\}$, and from the set $\{(y, s) \mid A^T y + s = c\}$.

A natural starting point in this setting is the point $y = 0$, $x = s = e$ for which the angle between x and s is zero. An infeasible-starting-point interior-point method then strives at increasing this angle to $\pi/2$ while simultaneously reducing $\|Ax - b\|$ and $A^T y + s - c\|$ and maintaining $x, s > 0$. The Euclidean setting of the optimality conditions and of this starting point suggests the following geometric stopping criterion:

Geometric stopping test:

1. Require the Euclidean distance to the primal-dual equality constraints to be at most ϵ, i.e.,

$$\|Ax - b\| \le \epsilon \quad \text{and} \quad \|A^T y + s - c\| \le \epsilon,$$

2. Require the cosine of the angle between x and s to be less or equal to ϵ,

$$x^T s \le \epsilon \|x\| \|s\|.$$

For small $\epsilon > 0$ the second inequality approximately states that the angle between x and s is $\frac{\pi}{2} - \alpha$ where $0 \le \alpha \le \epsilon$.

Typically, infeasible interior-point methods simultaneously reduce μ and the residuals of the primal-dual equations. By limiting the reduction of μ in case the norms of the residuals $\|Ax - b\|$ or $\|A^T y + s - c\|$ are large, the algorithm can be tuned such that the stop test will be triggered by the second inequality. Below, this situation is considered, and in particular, iterates are considered for which the stop test is not yet strictly satisfied, more precisely, where

$$x^T s \ge \epsilon \|x\| \|s\|.$$

The geometric stopping test is not only motivated by the Euclidean setting of the optimality conditions, it also allows to bound the condition number of the linear systems (2) to be solved at each iteration[1]. This is the subject of the next section.

2.3 Bounds in the dense case

Many implementations of interior-point algorithms generate iterates in a "$-\infty$-norm" neighborhood of the central path, i.e. $x_i s_i \ge \sigma\mu$ for $1 \le i \le n$ where $\mu := x^T s/n$ and $\sigma \in (0,1)$ defines the "width" of the neighborhood. For problems that are preprocessed as in Section 2.1 and for interior-point algorithms that generate iterates in such a "$-\infty$-norm" neighborhood, and with a stopping test as in Section 2.2, the following lemma applies to all iterates possibly generated by the algorithm:

Lemma 1: Assume that A has orthonormal rows. Let $x, s > 0$ and $\sigma \in (0,1]$ be given with $x_i s_i \ge \sigma\mu$ for $1 \le i \le n$ where $\mu := x^T s/n$. Assume further that $\epsilon > 0$ is given with $x^T s \ge \epsilon \|x\| \|s\|$. Then, the condition number of the matrix AD^2A^T in (2) is bounded by

$$\text{cond}(AD^2A^T) \le \frac{n^2}{\sigma^2 \epsilon^2}.$$

This bound is asymptotically sharp in the sense that for a certain class of matrices A it cannot be improved by any constant factor when $\min\{n, 1/\epsilon\} \to \infty$.

[1] In case of poorly scaled problems (and for problems that do not have a finite optimal solution) the geometric stopping test implies, however, that the stopping test may be satisfied before the duality gap is below a desired threshold. In this case, conditions such as Lemma 5.3.5 in [2] state that if (LP) has an optimal solution (x^*, y^*, s^*), then $\|(x^*, s^*)\|$ must be rather large.

Proof: First observe that (by assumption on A)

$$\min_{\|y\|=1} y^T ADA^T y \geq \min_{\|x\|=1} x^T Dx \quad \text{and} \quad \max_{\|y\|=1} y^T ADA^T y \leq \max_{\|x\|=1} x^T Dx$$

so that $\text{cond}(ADA^T) \leq \text{cond}(D)$. For $1 \leq i \leq n$ it follows from

$$\sigma\mu \leq x_i s_i \leq x^T s = n\mu$$

that

$$x_i s_i^{-1} = x_i^2 x_i^{-1} s_i^{-1} \leq \frac{x_i^2}{\sigma\mu} \leq \frac{\|x\|^2}{\sigma\mu}$$

and

$$x_i s_i^{-1} = s_i^{-2} x_i s_i \geq \sigma\mu s_i^{-2} \geq \frac{\sigma\mu}{\|s\|^2},$$

hence,

$$\text{cond}(D) \leq \frac{\|x\|^2}{\sigma\mu} / \frac{\sigma\mu}{\|s\|^2} = \frac{\|x\|^2\|s\|^2}{\sigma^2\mu^2} \leq \left(\frac{x^T s}{\epsilon}\right)^2 \frac{1}{\sigma^2\mu^2} = \frac{n^2}{\sigma^2\epsilon^2}.$$

To verify that this bound is asymptotically sharp let $\alpha \geq 1$ and

$$x := \begin{pmatrix} \sigma\sqrt{\epsilon}/\alpha \\ \alpha/\sqrt{\epsilon} \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad s := \begin{pmatrix} \alpha/\sqrt{\epsilon} \\ \sigma\sqrt{\epsilon}/\alpha \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \text{and} \quad A := \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix}.$$

Here, since $\mu = \frac{n-2}{n} + \frac{2\sigma}{n} \leq 1$, the bound $x_i s_i \geq \sigma\mu$ is satisfied. The other assumption to be satisfied is $x^T s \geq \epsilon\|x\|\|s\|$. Since $\|x\| = \|s\|$, this is equivalent to

$$2\sigma + n - 2 \geq \epsilon(\frac{\alpha^2}{\epsilon} + \frac{\sigma^2\epsilon}{\alpha^2} + n - 2)$$

which in turn is implied, if

$$(n-2)(1-\epsilon) \geq \alpha^2 \tag{3}$$

holds true. Defining $\alpha := \sqrt{n}(1-\rho)$ for some number $\rho > 0$, it is evident that ρ can be chosen arbitrarily close to zero without violating (3) if $\min\{n, 1/\epsilon\}$ is sufficiently large.

The condition number of ADA^T is then given by $(x_2 s_2^{-1})/(x_1 s_1^{-1}) = \frac{\alpha^4}{\sigma^2\epsilon^2}$ which is arbitrarily close to the given bound. □

The squares in the bound of Lemma 1 are due to the fact that the underlying linear system is a system of normal equations – solving the larger indefinite system (1) instead of (2) generally would be more stable but also computationally more expensive.

Note that the bound of Lemma 1 is independent of μ – but due to the rescaling in Section 2.1 it follows that $\|x\| \geq 1$ and $\|s\| \geq 1$ for any feasible x, s so that $\mu \geq \frac{\epsilon}{n}$ for any feasible x, s satisfying the assumptions of Lemma 1.

2.4 Bound for sparse systems

Generating orthonormal rows of A as in Section 2.1 is not feasible for large scale sparse problems. In this case, the proof of Lemma 1 allows for a simple generalization stated in the next corollary:

Corollary 1: If A does not have orthonormal rows but the remaining assumptions of Lemma 1 are satisfied, generalize the notion of condition number to rectangular matrices by defining $\mathrm{cond}(A)$ as the maximum singular value of A divided by its minimum singular value. Then, the bound of Lemma 1 holds in the following weaker form:

$$\mathrm{cond}(ADA^T) \leq \frac{n^2\,\mathrm{cond}(A)^2}{\sigma^2\epsilon^2}.$$

If all that is known about A is its condition number, then, again, this bound is asymptotically sharp. □

3 Semidefinite programs

Let us consider semidefinite programs of the standard form

$$(SDP) \qquad \min\ C \bullet X \mid \mathcal{A}X = b, \quad X \succeq 0.$$

Here, $\mathcal{A}$ is a linear map: $\mathcal{S}^n \to \mathbb{R}^m$, the i-th component of $\mathcal{A}X$ being given by $A^{(i)} \bullet X$ for some given data matrices $A^{(i)}$. First, we will assume again that $\mathcal{A}$ has orthonormal "rows", i.e. $\|A^{(i)}\|_F = 1$ and $A^{(i)} \bullet A^{(j)} = 0$ for $i \neq j$.

For semidefinite programs, the scaling point W of two matrices $X, S \succ 0$ defines the NT-direction [3]. Here, W is given by

$$W := S^{-1/2}(S^{1/2}XS^{1/2})^{1/2}S^{-1/2}.$$

Throughout this section W will refer to this particular matrix.

Set $P := W^{-1/2}$ and $V := PXP = P^{-1}SP^{-1}$, and for some symmetric matrix Z, define $\mathcal{L}_V(Z) := \frac{1}{2}(VZ + ZV)$. Then, for semidefinite programs, the equivalent of the third block row of (1) is given by the equation

$$\mathcal{L}_V(P\Delta XP) + \mathcal{L}_V(P^{-1}\Delta SP^{-1}) = \mu_+ I - V^2.$$

Multiplying this from left by $(\mathcal{L}_V)^{-1}$ and solving the primal-dual system in an analogous way as in (2) for linear programs leads to a linear system of the form

$$\mathcal{A}\mathcal{D}_W\mathcal{A}^*\Delta y = rhs$$

where $\mathcal{A}^*$ denotes the adjoint of $\mathcal{A}$ and $\mathcal{D}_W$ is defined by the identity "$\mathcal{D}_W[\Delta S] \equiv W\Delta SW$ for all symmetric matrices ΔS".

Lemma 1 can also be generalized to semidefinite programs (SDP). We first consider the case where the iterates are on the central path, i.e. $XS = \mu I$ with $\mu = X \bullet S/n$:

Corollary 2: Assume that $\mathcal{A}$ has orthonormal rows. For (SDP) the bound of Lemma 1 then applies to points on the central path ($\sigma = 1$ in Lemma 1) with $X \bullet S \geq \epsilon\|X\|_F\|S\|_F$ in the form

$$\mathrm{cond}(\mathcal{A}\mathcal{D}_W\mathcal{A}^*) \leq \frac{n^2}{\epsilon^2}.$$

Proof: Since $\mathcal{A}$ has orthonormal rows, by the proof of Lemma 1, it suffices to bound the condition number of $\mathcal{D}_W$. For X, S on the central path, the matrix W simplifies to $W = X/\sqrt{\mu}$. Let Λ be a diagonal matrix with the eigenvalues of X on its diagonal. Then, by orthogonal invariance of the Frobenius norm,

$$\|\mathcal{D}_W\| := \max_{\|\Delta S\|_F=1} \|\mathcal{D}_W[\Delta S]\|_F = \max_{\|\Delta S\|_F=1} \frac{\|X\Delta S X\|_F}{\mu} = \max_{\|\Delta S\|_F=1} \frac{\|\Lambda\Delta S\Lambda\|_F}{\mu},$$

so that

$$\|\mathcal{D}_W\|^2 = \max_{\|\Delta S\|_F=1} \frac{1}{\mu^2}\sum_{i,j}\Lambda_{i,i}^2\Lambda_{j,j}^2\Delta S_{i,j}^2 \le \max_{\|\Delta S\|_F=1} \frac{1}{\mu^2}\max_k \Lambda_{k,k}^4 \sum_{i,j}\Delta S_{i,j}^2,$$

and thus, $\|\mathcal{D}_W\| = \frac{1}{\mu}\max_k \Lambda_{k,k}^2$.
Likewise, $\|(\mathcal{D}_W)^{-1}\| = \|\mathcal{D}_{W^{-1}}\| = \mu/\min_k \Lambda_{k,k}^2$, and hence,

$$\text{cond}(\mathcal{D}_W) = \text{cond}(X)^2 \quad (= \text{cond}(W)^2).$$

By assumption, $\epsilon\|X\|_F\|S\|_F \le X \bullet S$, and the relation $XS = \mu I$ further implies

$$\epsilon\|X\|_F\|S\|_F \le X \bullet S = X \bullet (\mu X^{-1}) = n\mu.$$

Replacing S with μX^{-1} on the left-hand side above and dividing by μ yields

$$\epsilon\|X\|_F\|X^{-1}\|_F \le n.$$

Since $\|X\|_F\|X^{-1}\|_F \ge \|X\|_2\|X^{-1}\|_2$ this implies $\text{cond}(X) \le n/\epsilon$ from which the claim follows. □

The above proof can now be generalized to points in a "$-\infty$-norm" neighborhood of the central path. To this end first note that $X^{1/2}SX^{1/2} \succeq \sigma\mu I$ if, and only if, $S^{1/2}XS^{1/2} \succeq \sigma\mu I$ (this follows because $X^{1/2}SX^{1/2}$ and $S^{1/2}XS^{1/2}$ are both similar to XS and thus share the same eigenvalues) so that the matrices X and S are interchangeable in the statement of the next lemma:

Lemma 2: Assume that $\mathcal{A}$ has orthonormal rows. Let $X, S \succ 0$ and $\sigma \in (0, 1]$ be given with $X^{1/2}SX^{1/2} \succeq \sigma\mu I$ where $\mu := X \bullet S/n$. Assume further that $\epsilon > 0$ is given with $X \bullet S \ge \epsilon\|X\|_F\|S\|_F$. Then, the condition number of the matrix $\mathcal{A}\mathcal{D}_W\mathcal{A}^*$ is bounded by

$$\text{cond}(\mathcal{A}\mathcal{D}_W\mathcal{A}^*) \le \frac{n^2}{\sigma^2\epsilon^2}.$$

This bound is asymptotically sharp.

Proof: We begin by quoting a result that follows the statement[2] of Lemma 3.1 in [1]: For W defined as above the following holds true:

$$2(X^{-1} + S)^{-1} \preceq W \preceq (X + S^{-1})/2. \tag{4}$$

By the proof of Corollary 2 it suffices to bound the condition number of W. Denote the minimum and maximum eigenvalues of a symmetric matrix Z by

[2]The reference [1] is self-contained and the proof of this result is easy to follow so that the proof is not reproduced here.

$\lambda_{min}(Z)$ and $\lambda_{max}(Z)$. Observe that W remains invariant when multiplying both, X and S with the same positive constant. We may therefore assume that X and S are scaled such that $\lambda_{max}(X) = \lambda_{max}(S^{-1})$. Let $C := \max(\text{cond}(X), \text{cond}(S))$. Then,

$$X^{-1} \preceq \frac{1}{\lambda_{min}(X)} I = \frac{C}{\lambda_{max}(X)} I \qquad \text{and} \qquad S \preceq \lambda_{max}(S) I = \frac{C}{\lambda_{max}(S^{-1})} I.$$

Thus, the inequality (4) implies

$$\frac{\lambda_{max}(X)}{C} I \preceq W \preceq \lambda_{max}(X) I,$$

so that $\text{cond}(W) \le C$. Since X and S are interchangeable in Lemma 2, we assume without loss of generality that $\text{cond}(W) \le \text{cond}(X)$. Let u be an eigenvector of X to the eigenvalue $\lambda_{min}(X)$ with $\|u\| = 1$. Then, by assumption on the "$-\infty$-norm" neighborhood,

$$\sigma\mu \le u^T X^{1/2} S X^{1/2} u = \lambda_{min}(X) u^T S u \le \lambda_{min}(X) \|S\|_2.$$

Since $\|X^{-1}\|_2 = 1/\lambda_{min}(X)$, this implies

$$\|S\|_2 \ge \sigma\mu \|X^{-1}\|_2.$$

Inserting this in the left hands side of

$$\epsilon \|X\|_2 \|S\|_2 \le \epsilon \|X\|_F \|S\|_F \le X \bullet S = n\mu.$$

and dividing by $\epsilon\sigma\mu$ it follows that $\text{cond}(X) \le n/(\epsilon\sigma)$ from which the claim follows. Linear programs being a special case of semidefinite programs, this bound is asymptotically tight as well. □

The generalization of this result to mappings $\mathcal{A}$ that do not have orthonormal rows follows as in Corollary 1.

Acknowledgment

The author likes to thank Felix Lieder for comments that helped to improve the presentation of this paper.

References

[1] J.D. Lawson and Y. Lim: The Geometric Mean, Matrices, Metrics, and More, *The American Mathematical Monthly* 108(9) (2001) pp. 797–812.

[2] S. Mizuno, Infeasible-interior-point algorithms, in: T. Terlaky, ed. "Interior Point Methods of Mathematical Programming" *Applied Optimization* 5 Springer (1996) pp. 159–187

[3] Y.E. Nesterov and M.J. Todd, Primal-dual interior-point methods for self-scaled cones, *SIAM Journal on Optimization* 8(2) (1998) pp. 324–364.

[4] S.J. Wright, *Primal-Dual Interior-Point Methods* SIAM, Philadelphia (1997).