SIAM J. OPTIM.
Vol. 11, No. 3, pp. 818–836
© 2001 Society for Industrial and Applied Mathematics

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

CONDITION-MEASURE BOUNDS ON THE BEHAVIOR OF THE CENTRAL TRAJECTORY OF A SEMIDEFINITE PROGRAM*

MANUEL A. NUNEZ† AND ROBERT M. FREUND‡

Abstract. We present bounds on various quantities of interest regarding the central trajectory of a semidefinite program, where the bounds are functions of Renegar's condition number $\mathcal{C}(d)$ and other naturally occurring quantities such as the dimensions n and m. The condition number $\mathcal{C}(d)$ is defined in terms of the data instance $d = (A, b, C)$ for a semidefinite program; it is the inverse of a relative measure of the distance of the data instance to the set of ill-posed data instances, that is, data instances for which arbitrary perturbations would make the corresponding semidefinite program either feasible or infeasible. We provide upper and lower bounds on the solutions along the central trajectory, and upper bounds on changes in solutions and objective function values along the central trajectory when the data instance is perturbed and/or when the path parameter defining the central trajectory is changed. Based on these bounds, we prove that the solutions along the central trajectory grow at most linearly and at a rate proportional to the inverse of the distance to ill-posedness, and grow at least linearly and at a rate proportional to the inverse of $\mathcal{C}(d)^2$, as the trajectory approaches an optimal solution to the semidefinite program. Furthermore, the change in solutions and in objective function values along the central trajectory is at most linear in the size of the changes in the data. All such bounds involve polynomial functions of $\mathcal{C}(d)$, the size of the data, the distance to ill-posedness of the data, and the dimensions n and m of the semidefinite program.

Key words. semidefinite programming, perturbation of convex programs, central trajectory, interior-point methods, ill-posed problems, condition numbers

AMS subject classifications. 90C22, 90C31, 90C51

PII. S105262349936063X

1. Introduction.

We study various properties of the central trajectory of a semidefinite program $P(d) : \min\{C \bullet X : AX = b, X \succeq 0\}$. Here X and C are symmetric matrices; $C \bullet X$ denotes the trace inner product; A is a linear operator that maps symmetric matrices into $\Re^m$; $b \in \Re^m$; $X \succeq 0$ denotes that X is a symmetric positive semidefinite matrix; and the data for $P(d)$ is the array $d = (A, b, C)$. The central trajectory of $P(d)$ is the solution to the logarithmic barrier problem $P_\mu(d) : \min\{C \bullet X - \mu \ln \det(X) : AX = b, X \succeq 0\}$ as the trajectory parameter μ ranges over the interval $(0, \infty)$. Semidefinite programming (SDP) has been the focus of an enormous amount of research in the past decade and has proven to be a unifying model for many convex programming problems amenable to efficient solution by interior-point methods; see [1, 14, 27], and [2], among others. Our primary concern lies in bounding a variety of measures of the behavior of the central trajectory of $P(d)$ in terms of the condition number $\mathcal{C}(d)$ for $P(d)$ originally developed by Renegar.

By the condition number $\mathcal{C}(d)$ of the data $d = (A, b, C)$, we mean a scale-invariant positive measure depending on a given feasible data instance $d = (A, b, C)$ and with the following property: the condition number approaches infinity as the data approaches the set of data instances for which the problem $P(d)$, or its dual, becomes infeasible. In particular, we say that a data instance is *ill-posed* whenever its

*Received by the editors August 26, 1999; accepted for publication (in revised form) August 25, 2000; published electronically February 2, 2001.
http://www.siam.org/journals/siopt/11-3/36063.html

†Argyros School of Business and Economics, Chapman University, One University Drive, Orange, CA 92866 (mnunez@chapman.edu).

‡Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139-4307 (rfreund@mit.edu).

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

corresponding condition number is unbounded, that is, whenever the data instance is on the boundary of the set of primal-dual feasible data instances. This notion of conditioning (formally presented in subsection 2.3) was originally developed by Renegar in [17] within a more general convex programming context and has proven to be a key concept in the understanding of the continuous complexity of convex optimization methods (see, for instance, [4, 5, 6, 7, 8, 9, 16, 17, 18, 19, 20, 28, 29] among others). In this paper, we show the relevance of using this measure of conditioning in the analysis of the central trajectory of a semidefinite program of the form $P(d)$.

More specifically, in section 3 we present a variety of results that bound certain behavioral measures of the central trajectory of $P(d)$ in terms of the condition number $\mathcal{C}(d)$. In Theorem 3.1, we present upper bounds on the norms of solutions along the central trajectory. These bounds show that the solutions along the central trajectory grow at most linearly in the trajectory parameter μ and at a rate proportional to the inverse of the distance to ill-posedness of d. In Theorem 3.2, we present lower bounds on the values of the eigenvalues of solutions along the central trajectory. These bounds show that the eigenvalues of solutions along the central trajectory grow at least linearly in the trajectory parameter μ and at a rate proportional to $\mathcal{C}(d)^{-2}$.

In Theorem 3.3, we present bounds on changes in solutions along the central trajectory under simultaneous changes (perturbations) in the data d as well as changes in the trajectory parameter μ. These bounds are linear in the size of the data perturbation, quadratic in the inverse of the trajectory parameter, and are polynomial functions of the condition number and the dimensions m and n. Finally, in Theorem 3.4 we present similar bounds on the change in the optimal objective function values of the barrier problem along the central trajectory, under data and trajectory parameter perturbations. These bounds also are linear in the size of the data perturbation and in the size of the change in the trajectory parameter.

The use of continuous complexity theory in convex optimization, especially the theory developed by Renegar in [17, 18, 19, 20], has added significant insight into what makes certain convex optimization problems better or worse behaved (in terms of the deformation of problem characteristics under data perturbations) and consequently what makes certain convex optimization problems easier or harder to solve. We believe that the results presented in this paper contribute to this understanding by providing behavioral bounds on relevant aspects of the central trajectory of a semidefinite program.

The main results presented in this paper can be viewed as extensions of related results for the linear programming (LP) case presented in [15]. While some of the extensions contained herein are rather straightforward generalizations of analogous results for the LP case, other extensions have proven to be mathematically challenging to us and have necessitated (in their proofs) the development of further properties of matrices arising in the analysis of SDP; see Propositions 5.1 and 5.3, for example. One reason why we have found the extension from LP to SDP to be mathematically challenging has to do with the linear algebra of certain linear operators that arise in the study of the central trajectory. In the case of LP, we have $X\bar{X} = \bar{X}X$ whenever X and $\bar{X}$ are diagonal matrices. Matrix products like this appear when dealing with solutions x and $\bar{x}$ on the central trajectory of a data instance and its perturbation, respectively, thus streamlining the proofs of results in the LP case. When dealing with analogous solutions in the case of SDP, we no longer have the same commutative property of the matrix product, and so it is necessary to develop more complicated linear operators in the analysis of the central trajectory. Another difficulty in the extension from the

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

LP case to the SDP case is the lack of closedness of certain projections of the cone of positive semidefinite symmetric matrices. This lack of closedness prevents the use of "nice" LP properties such as strict complementarity of solutions.

Literature review. The study of perturbation theory and continuous complexity for convex programs in terms of the distance to ill-posedness and condition number of a given data instance was introduced in [17] by Renegar, who studied perturbations in a very general setting of the problem $(RLP) : \sup\{c^*x : Ax \leq b, x \geq 0, x \in \mathcal{X}\}$, where $\mathcal{X}$ and $\mathcal{Y}$ denote real normed vector spaces, $A : \mathcal{X} \to \mathcal{Y}$ is a continuous linear operator, $c^* : \mathcal{X} \to \Re$ is a continuous linear functional, and the inequalities $Ax \leq b$ and $x \geq 0$ are induced by any closed convex cones (linear or nonlinear) containing the origin in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Previous to the paper of Renegar, many papers were written on perturbations of linear programs and systems of linear inequalities, but not in terms of the distance to ill-posedness (see, for instance, [12, 22, 23, 24, 25]).

Even though there is now a vast literature on SDP, there are only a few papers that study SDP in terms of some notion of a condition measure. Renegar [17] presents a bound on solutions to RLP, a bound on the change in optimal solutions when only the right-hand side vector b is perturbed, and a bound on changes in optimal objective function values when the whole data instance is perturbed. All of these bounds depend on the distance to ill-posedness of the given data instance. Because of their generality, these results also apply to the SDP case studied in this paper. Later, in [19] and [20] Renegar presented upper and lower bounds on the inverse of the Hessian matrix resulting from the application of Newton's method to the optimality conditions of RLP along the central trajectory. Again, these bounds depend on the distance to ill-posedness of the data instance, and they apply to the SDP case. These bounds are important because they can be used to study the continuous complexity of interior-point methods for solving semidefinite programs (see [19]) as well as the use of the conjugate gradient method in the solution of semidefinite programs (see [20]).

Nayakkankuppam and Overton in [13] study the conditioning of SDP in terms of a condition measure that depends on the inverse of a certain Jacobian matrix. This Jacobian matrix arises when applying Newton's method to find a root of a semidefinite system of equations equivalent to the system of equations that arise from the Karush–Kuhn–Tucker optimality conditions for $P(d)$. In particular, under the assumption that both $P(d)$ and its dual have unique optimal solutions, they present a bound on the change in the optimal solution to $P(d)$ and $P(d + \Delta d)$, where Δd is a data perturbation, in terms of their condition number. This bound is linear in the norm of Δd. Their analysis pertains to the study of the optimal solution of $P(d)$, but is not readily applicable to the central trajectory of a semidefinite program.

Sturm and Zhang [26] study the sensitivity of the central trajectory of a semidefinite program in terms of changes in the right-hand side of the constraints $AX = b$ in $P(d)$. Given a data instance $d = (A, b, C)$ of a semidefinite program, they consider data perturbations of the form $d + \Delta d = (A, b + \Delta b, C)$. Using this kind of perturbation, and under a primal and dual Slater condition as well as a strict complementarity condition, they show several properties of the derivatives of central trajectory solutions with respect to the right-hand side vector. The results presented herein differ from these results in that we use data perturbations of the form $d+\Delta d = (A+\Delta A, b+\Delta b, C+\Delta C)$, and we express our results in terms of the distance to ill-posedness of the data. As a result, our results are not as strong in terms of the size of bounds, but our results are more general, as they do not rely on any particular assumptions.

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2. Notation, definitions, and preliminaries.

2.1. Space of symmetric matrices. Given two matrices U and V in $\Re^{n\times n}$, we define the inner product of U and V as $U \bullet V := \mathit{trace}(U^T V)$, where $\mathit{trace}(W) := \sum_{j=1}^{n} W_{jj}$ for all $W \in \Re^{n\times n}$. Given a matrix $U \in \Re^{n\times n}$, we denote by $\sigma(U) = (\sigma_1, \ldots, \sigma_n)^T$ the vector in $\Re^n$ whose components are the ordered singular values of U; that is, each σ_j is a singular value of U, and $0 \le \sigma_1 \le \cdots \le \sigma_n$. Furthermore, we denote by $\sigma_j(U)$ the jth singular value of U chosen according to the increasing order in $\sigma(U)$. In particular, $\sigma_1(U)$ and $\sigma_n(U)$ are the smallest and the largest singular values of U, respectively. We use the following norms in the space $\Re^{n\times n}$:

$$\|U\|_1 := \sum_{j=1}^{n} \sigma_j(U), \tag{1}$$

$$\|U\|_2 := \left(\sum_{j=1}^{n} \sigma_j(U)^2\right)^{1/2} = (U \bullet U)^{1/2}, \tag{2}$$

$$\|U\|_\infty := \max_{1\le j\le n} \sigma_j(U) = \sigma_n(U) \tag{3}$$

for all matrices $U \in \Re^{n\times n}$. The norm (1) is known as the *Ky Fan n*-norm or *trace* norm (see [3]); the norm (2) is known as the *Hilbert–Schmidt* norm or *Frobenius* norm and is induced by the inner product $\bullet$ defined above; (3) is the operator norm induced by the Euclidean norm on $\Re^n$. Notice that all these norms are *unitarily invariant* in that $\|U\| = \|PUQ\|$ for all unitary matrices P and Q in $\Re^{n\times n}$. We also have the following proposition that summarizes a few properties of these norms.

PROPOSITION 2.1. *For all $U, V \in \Re^{n\times n}$ we have*

(i) *Hölder's inequalities* (*see* [3])

$$|U \bullet V| \le \|U\|_\infty \|V\|_1, \tag{4}$$

$$|U \bullet V| \le \|U\|_2 \|V\|_2. \tag{5}$$

(ii) $\|UV\|_2 \le \|U\|_2 \|V\|_2$.
(iii) $\|U\|_\infty \le \|U\|_2 \le \sqrt{n}\|U\|_\infty$.
(iv) $\frac{1}{\sqrt{n}}\|U\|_1 \le \|U\|_2 \le \|U\|_1$.

From now on, whenever we use a Euclidean norm over any space, we will omit subscripts. Hence, $\|U\| := \|U\|_2$ for all U in $\Re^{n\times n}$.

Let $\mathcal{S}_n$ denote the subspace of $\Re^{n\times n}$ consisting of symmetric matrices. Given a matrix $U \in \mathcal{S}_n$, let $U \succeq 0$ denote that U is a positive semidefinite matrix, and let $U \succ 0$ denote that U is a positive definite matrix. We denote by $\mathcal{S}_n^+$ the set of positive semidefinite matrices in $\mathcal{S}_n$, that is, $\mathcal{S}_n^+ = \{U \in \mathcal{S}_n : U \succeq 0\}$. Observe that $\mathcal{S}_n^+$ is a closed convex pointed cone in $\mathcal{S}_n$ with nonempty interior given by $\{U \in \mathcal{S}_n : U \succ 0\}$. Furthermore, notice that the polar $(\mathcal{S}_n^+)^*$ of the cone $\mathcal{S}_n^+$ is the cone $\mathcal{S}_n^+$ itself. When $U \in \mathcal{S}_n$, we denote by $\lambda(U) := (\lambda_1, \ldots, \lambda_n)^T$ the vector in $\Re^n$ whose components are the real eigenvalues of U ordered as $0 \le |\lambda_1| \le \cdots \le |\lambda_n|$. Moreover, we denote by $\lambda_j(U)$ the jth eigenvalue of U chosen according to the order in $\lambda(U)$. In particular, notice that $\sigma_j(U) = |\lambda_j(U)|$ whenever $U \in \mathcal{S}_n$.

Given matrices $A_1, \ldots, A_m \in \mathcal{S}_n$, we define the linear operator $A = (A_1, \ldots, A_m)$ from $\mathcal{S}_n$ to $\Re^m$ as follows:

$$AX := (A_1 \bullet X, \ldots, A_m \bullet X)^T \tag{6}$$

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

for all $X \in \mathcal{S}_n$. We denote by $\mathcal{L}_{m,n}$ the space of linear operators from $\mathcal{S}_n$ to $\Re^m$ of the form (6). Given a linear operator $A = (A_1, \ldots, A_m) \in \mathcal{L}_{m,n}$, we define the *rank* of A as the dimension of the subspace generated by the matrices $A_1, \ldots, A_m$, that is, $\text{rank}(A) := \dim(\langle A_1, \ldots, A_m \rangle)$. We say that A has full-rank whenever $\text{rank}(A) = \min\{m, n(n-1)/2\}$. Throughout the remainder of this paper we will assume that $m \leq n(n-1)/2$, so that A has full-rank if and only if $\text{rank}(A) = m$. The corresponding adjoint transformation $A^T : \Re^m \mapsto \mathcal{S}_n$, associated with A, is given by

$$A^T[y] = \sum_{i=1}^{m} y_i A_i$$

for all $y \in \Re^m$. Furthermore, we endow the space $\mathcal{L}_{m,n}$ with the operator norm $\|A\| := \max\{\|AX\| : X \in \mathcal{S}_n, \|X\| \leq 1\}$ for all operators $A \in \mathcal{L}_{m,n}$. Finally, if we define the norm of the adjoint operator as $\|A^T\| := \max\{\|A^T[y]\| : y \in \Re^m, \|y\| \leq 1\}$, then it follows that $\|A^T\| = \|A\|$.

2.2. Data instance space. Consider the vector space $\mathcal{D}$ defined as $\mathcal{D} := \{d = (A, b, C) : A \in \mathcal{L}_{m,n}, b \in \Re^m, C \in \mathcal{S}_n\}$. We regard $\mathcal{D}$ as the space of data instances associated with the following pair of dual semidefinite programs:

$$\begin{aligned} P(d) &: \min \{C \bullet X : AX = b, X \succeq 0\}, \\ D(d) &: \max \left\{b^T y : A^T[y] + S = C, S \succeq 0\right\}, \end{aligned}$$

where $d = (A, b, C) \in \mathcal{D}$. To study the central trajectory of a data instance in $\mathcal{D}$, we use the functional $p(\cdot)$ defined as $p(U) = -\ln \det U$ for all $U \succ 0$. Notice that, as proven in [14], $p(\cdot)$ is a strictly convex n-normal barrier for the cone $\mathcal{S}_n^+$. Given a data instance $d = (A, b, C) \in \mathcal{D}$ and a fixed scalar $\mu > 0$, we study the following parametric family of dual logarithmic barrier problems associated with $P(d)$ and $D(d)$:

$$\begin{aligned} P_\mu(d) &: \min \{C \bullet X + \mu p(X) : AX = b, X \succ 0\}, \\ D_\mu(d) &: \max \left\{b^T y - \mu p(S) : A^T[y] + S = C, S \succ 0\right\}. \end{aligned}$$

Let $X(\mu)$ and $(y(\mu), S(\mu))$ denote the optimal solutions of $P_\mu(d)$ and $D_\mu(d)$, respectively (when they exist). Then the primal central trajectory is the set $\{X(\mu) : \mu > 0\}$ and is a smooth mapping from $(0, \infty)$ to S_n^+ [10, 27]. Similarly, the dual central trajectory is the set $\{(y(\mu), S(\mu)) : \mu > 0\}$ and is a smooth mapping from $(0, \infty)$ to $\Re^m \times S_n^+$.

We provide the data instance space $\mathcal{D}$ with the norm

$$\|d\| := \max\{\|A\|, \|b\|, \|C\|\} \tag{7}$$

for all data instances $d = (A, b, C) \in \mathcal{D}$. Using this norm, we denote by $B(d, \delta)$ the open ball $\{d + \Delta d \in \mathcal{D} : \|\Delta d\| < \delta\}$ in $\mathcal{D}$ centered at d and with radius $\delta > 0$ for all $d \in \mathcal{D}$.

2.3. Distance to ill-posedness. We are interested in studying data instances for which both programs $P(\cdot)$ and $D(\cdot)$ are feasible. Consequently, consider the following subset of the data set $\mathcal{D}$:

$$\mathcal{F} := \left\{(A, b, C) \in \mathcal{D} : b \in A(\mathcal{S}_n^+) \text{ and } C \in A^T[\Re^m] + \mathcal{S}_n^+\right\},$$

that is, the elements in $\mathcal{F}$ correspond to those data instances d in $\mathcal{D}$ for which $P(d)$ and $D(d)$ are feasible. The complement of $\mathcal{F}$, denoted by $\mathcal{F}^C$, is the set of data

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

instances $d = (A, b, C)$ for which $P(d)$ or $D(d)$ is infeasible. The boundary of $\mathcal{F}$, denoted by $\partial\mathcal{F}$, is called the set of *ill-posed* data instances. This is because arbitrarily small changes in a data instance $d \in \partial\mathcal{F}$ can yield data instances in $\mathcal{F}$ as well as data instances in $\mathcal{F}^C$.

For a data instance $d \in \mathcal{D}$, the *distance to ill-posedness* is defined as

$$\rho(d) := \inf\{\|\Delta d\| : d + \Delta d \in \partial\mathcal{F}\}$$

(see [17, 21, 18]), and so $\rho(d)$ is the distance of the data instance d to the set of ill-posed instances $\partial\mathcal{F}$. The *condition number* $\mathcal{C}(d)$ of the data instance d is defined as

$$\mathcal{C}(d) := \begin{cases} \frac{\|d\|}{\rho(d)} & \text{if } \rho(d) > 0, \\ \infty & \text{if } \rho(d) = 0. \end{cases}$$

The condition number $\mathcal{C}(d)$ can be viewed as a scale-invariant reciprocal of $\rho(d)$, as it is elementary to demonstrate that $\mathcal{C}(d) = \mathcal{C}(\alpha d)$ for any positive scalar α. Moreover, for $d = (A, b, C) \notin \partial\mathcal{F}$, let $\Delta d = (-A, -b, -C)$. Observe that $d + \Delta d = (0, 0, 0) \in \partial\mathcal{F}$ and, since $\partial\mathcal{F}$ is a closed set, we have $\|d\| = \|\Delta d\| \geq \rho(d) > 0$ so that $\mathcal{C}(d) \geq 1$. The value of $\mathcal{C}(d)$ is a measure of the relative conditioning of the data instance d.

As proven in [24], the interior of $\mathcal{F}$, denoted $\mathrm{Int}(\mathcal{F})$, is characterized as follows:

(8)
$$\mathrm{Int}(\mathcal{F}) = \left\{(A, b, C) \in \mathcal{D} : b \in A(\mathrm{Int}(\mathcal{S}_n^+)), C \in A^T[\Re^m] + \mathrm{Int}(\mathcal{S}_n^+), A \text{ has full-rank}\right\}.$$

In particular, notice that data instances in $\mathrm{Int}(\mathcal{F})$ correspond to data instances for which both $P_\mu(\cdot)$ and $D_\mu(\cdot)$ are feasible (for any $\mu > 0$). Also, observe that $d = (A, b, C) \in \mathcal{F}$ and $\rho(d) > 0$ if and only if $d \in \mathrm{Int}(\mathcal{F})$, and so, if and only if the characterization given in (8) holds for d. We will use this characterization of the interior of $\mathcal{F}$ throughout the remainder of this paper.

We will also make use of the following elementary sufficient certificates of infeasibilty.

PROPOSITION 2.2. *Let $d = (A, b, C) \in \mathcal{D}$.*

1. *If there exists $y \in \Re^m$ satisfying $A^T[y] \prec 0$ and $b^T y \geq 0$, then $P_\mu(d)$ is infeasible.*
2. *If there exists $X \in \mathcal{S}_n$ satisfying $AX = 0, X \succ 0$, and $C \bullet X \leq 0$, then $D_\mu(d)$ is infeasible.*

3. Statement of main results. For a given data instance $d \in \mathrm{Int}(\mathcal{F})$ and a scalar $\mu > 0$, we denote by $X(d, \mu)$ the unique optimal solution to $P_\mu(d)$ and by $(y(d, \mu), S(d, \mu))$ the unique optimal solution to $D_\mu(d)$. Furthermore, we use the following function of d and μ as a condition measure for the programs $P_\mu(d)$ and $D_\mu(d)$:

$$\mathcal{K}(d, \mu) := \mathcal{C}(d)^2 + \frac{\mu n}{\rho(d)}. \tag{9}$$

As with the condition number $\mathcal{C}(d)$, it is not difficult to show that $\mathcal{K}(d, \mu) \geq 1$ and $\mathcal{K}(d, \mu)$ is scale-invariant in the sense that $\mathcal{K}(\lambda d, \lambda\mu) = \mathcal{K}(d, \mu)$ for all $\lambda > 0$. The reason why we call $\mathcal{K}(d, \mu)$ a condition measure will become apparent from the theorems stated in this section.

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Our first theorem concerns upper bounds on the optimal solutions to $P_\mu(d)$ and $D_\mu(d)$, respectively. The bounds are given in terms of the condition measure $\mathcal{K}(d,\mu)$ and the size of the data $\|d\|$. In particular, the theorem shows that the norm of the optimal primal solution along the central trajectory grows at most linearly in the barrier parameter μ and at a rate no larger than $n/\rho(d)$. The proof of this theorem is presented in section 4.

THEOREM 3.1. *Let $d \in \text{Int}(\mathcal{F})$ and μ be a positive scalar. Then*

$$\|X(d,\mu)\| \le \mathcal{K}(d,\mu), \tag{10}$$

$$\|y(d,\mu)\| \le \mathcal{K}(d,\mu), \tag{11}$$

$$\|S(d,\mu)\| \le 2\|d\|\mathcal{K}(d,\mu), \tag{12}$$

where $\mathcal{K}(d,\mu)$ is the condition measure defined in (9).

As the proof of Theorem 3.1 will show, there is a tighter bound on $\|X(d,\mu)\|$, namely, $\|X(d,\mu)\| \le \mathcal{M}(d,\mu)$, where

$$\mathcal{M}(d,\mu) := \begin{cases} \mathcal{C}(d) & \text{if } C \bullet X(d,\mu) \le 0, \\ \max\left\{\mathcal{C}(d), \frac{\mu n}{\rho(d)}\right\} & \text{if } 0 < C \bullet X(d,\mu) \le \mu n, \\ \mathcal{C}(d)^2 + \frac{\mu n}{\rho(d)} & \text{if } \mu n < C \bullet X(d,\mu), \end{cases} \tag{13}$$

whenever $d \in \text{Int}(\mathcal{F})$ and $\mu > 0$. Notice that because of the uniqueness of the optimal solution to $P_\mu(d)$ for $\mu > 0$, the condition measure $\mathcal{M}(d,\mu)$ is well defined. Also, observe that $\mathcal{M}(d,\mu)$ can always be bounded from above by $\mathcal{K}(d,\mu)$.

It is not difficult to create data instances for which the condition measure $\mathcal{M}(d,\mu)$ is a tight bound on $\|X(d,\mu)\|$. Even though the condition measure $\mathcal{M}(d,\mu)$ provides a tighter bound on $\|X(d,\mu)\|$ than $\mathcal{K}(d,\mu)$, we will use the condition measure $\mathcal{K}(d,\mu)$ for the remainder of this paper. This is because $\mathcal{K}(d,\mu)$ conveys the same general asymptotic behavior as $\mathcal{M}(d,\mu)$ and also because using $\mathcal{K}(d,\mu)$ simplifies most of the expressions in the theorems to follow. Similar remarks apply to the bounds on $\|y(d,\mu)\|$ and $\|S(d,\mu)\|$.

In particular, when $C = 0$, that is, when we are solving a semidefinite analytic center program, we obtain the following corollary.

COROLLARY 3.1. *Let $d = (A,b,C) \in \text{Int}(\mathcal{F})$ be such that $C = 0$ and μ be a positive scalar. Then*

$$\|X(d,\mu)\| \le \mathcal{C}(d).$$

The following theorem presents lower bounds on the eigenvalues of solutions along the primal and dual central trajectories. In particular, the lower bound on the eigenvalues of solutions along the primal central trajectory implies that the convergence of $X(d,\mu)$ to an optimal solution to $P(d)$, as μ goes to zero, is at least asymptotically linear in μ and at a rate of $1/(2\|d\|\mathcal{C}(d)^2)$.

THEOREM 3.2. *Let $d \in \text{Int}(\mathcal{F})$ and μ be a positive scalar. Then for all $j = 1,\ldots,n$,*

$$\lambda_j(X(d,\mu)) \ge \frac{\mu}{2\|d\|\mathcal{K}(d,\mu)},$$

$$\lambda_j(S(d,\mu)) \ge \frac{\mu}{\mathcal{K}(d,\mu)}.$$

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The proof of Theorem 3.2 is presented in section 4.

The next theorem concerns bounds on changes in optimal solutions to $P_\mu(d)$ and $D_\mu(d)$ as the data instance d and the parameter μ are perturbed. In particular, we present these bounds in terms of an asymptotically polynomial expression of the condition number $\mathcal{C}(d)$, the condition measure $\mathcal{K}(d,\mu)$, the size of the data $\|d\|$, the scalar μ, and the dimensions m and n. It is also important to notice the linear dependence of the bound on the size of the data perturbation $\|\Delta d\|$ and the parameter perturbation $|\Delta\mu|$.

THEOREM 3.3. *Let $d = (A, b, C)$ be a data instance in $\mathrm{Int}(\mathcal{F})$, μ be a positive scalar, $\Delta d = (\Delta A, \Delta b, \Delta C) \in \mathcal{D}$ be a data perturbation such that $\|\Delta d\| \le \rho(d)/3$, and $\Delta\mu$ be a scalar such that $|\Delta\mu| \le \mu/3$. Then,*

$$\begin{aligned}\|X(d+\Delta d,\mu+\Delta\mu)-X(d,\mu)\| \le \|\Delta d\|\ &\frac{640n\sqrt{m}\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5(\mu+\|d\|)}{\mu^2}\\ &+|\Delta\mu|\ \frac{6n\|d\|\mathcal{K}(d,\mu)^2}{\mu^2},\end{aligned}\tag{14}$$

$$\begin{aligned}\|y(d+\Delta d,\mu+\Delta\mu)-y(d,\mu)\| \le \|\Delta d\|\ &\frac{640\sqrt{m}\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5(\mu+\|d\|)}{\mu^2}\\ &+|\Delta\mu|\ \frac{32\sqrt{m}\|d\|\mathcal{C}(d)^2\mathcal{K}(d,\mu)^2}{\mu^2},\end{aligned}\tag{15}$$

$$\begin{aligned}\|S(d+\Delta d,\mu+\Delta\mu)-S(d,\mu)\| \le \|\Delta d\|\ &\frac{640\sqrt{m}\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5(\mu+\|d\|)^2}{\mu^2}\\ &+|\Delta\mu|\ \frac{32\sqrt{m}\|d\|^2\mathcal{C}(d)^2\mathcal{K}(d,\mu)^2}{\mu^2}.\end{aligned}\tag{16}$$

The proof of Theorem 3.3 is presented in section 5.

Finally, we present a theorem concerning changes in optimal objective function values of the program $P_\mu(d)$ as the data instance d and the parameter μ are perturbed. We denote by $z(d,\mu)$ the optimal objective function value of the program $P_\mu(d)$, namely, $z(d,\mu) := C \bullet X(d,\mu) + \mu p(X(d,\mu))$, where $X(d,\mu)$ is the optimal solution of $P_\mu(d)$.

THEOREM 3.4. *Let $d = (A, b, C)$ be a data instance in $\mathrm{Int}(\mathcal{F})$, μ be a positive scalar, $\Delta d = (\Delta A, \Delta b, \Delta C) \in \mathcal{D}$ be a data perturbation such that $\|\Delta d\| \le \rho(d)/3$, and $\Delta\mu$ be a scalar such that $|\Delta\mu| \le \mu/3$. Then*

$$\begin{aligned}|z(d+\Delta d,\mu+\Delta\mu)-z(d,\mu)| \le \|\Delta d\|\ &9\mathcal{K}(d,\mu)^2\\ &+|\Delta\mu|\ n\left(\ln 16+|\ln\mu|+|\ln\|d\||+\ln\mathcal{K}(d,\mu)\right).\end{aligned}$$

Notice that it follows from Theorem 3.4 that changes in objective function values of $P_\mu(d)$ are at most linear in the size of the perturbation of the data instance d and the parameter μ. As with Theorem 3.3, the bound is polynomial in terms of the condition measure $\mathcal{K}(d,\mu)$ and the size of the data instance d. Also observe that if $\Delta d = 0$, and we let $\Delta\mu$ go to zero, from the analytic properties of the central trajectory [10, 27] we obtain the following bound on the derivative of $z(\cdot)$ with respect to μ:

$$\left|\frac{\partial z(d,\mu)}{\partial\mu}\right| \le n\left(\ln 16+|\ln\mu|+|\ln\|d\||+\ln\mathcal{K}(d,\mu)\right).$$

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We remark that it is not known to us if the bounds in Theorem 3.1, 3.2, 3.3, or 3.4 are tight (even up to a constant) for some data instances, but we suspect that they are not. However, our concern herein is not the exploration of the best possible bounds but rather the demonstration of bounds that are some polynomial function of appropriate natural behavior measures of a semidefinite program.

The remaining two sections of this paper are devoted to proving the four theorems stated in this section.

4. Proof of bounds on optimal solutions. This section presents the proofs of the results on lower and upper bounds on sizes of optimal solutions along the central trajectory for the pair of dual logarithmic barrier problems $P_\mu(d)$ and $D_\mu(d)$. We start by proving Theorem 3.1. Our proof is an immediate generalization to the semidefinite case of the proof of Theorem 3.1 in [15] for the case of a linear program.

Proof of Theorem 3.1. Let $\hat{X} := X(d,\mu)$ be the optimal solution to $P_\mu(d)$ and $(\hat{y},\hat{S}) := (y(d,\mu), S(d,\mu))$ be the optimal solution to the corresponding dual problem $D_\mu(d)$. Notice that the optimality conditions of $P_\mu(d)$ and $D_\mu(d)$ imply that $C \bullet \hat{X} = b^T\hat{y} + \mu n$.

Observe that since $\hat{S} = C - A^T[\hat{y}]$, then $\|\hat{S}\| \le \|C\| + \|A^T\|\|\hat{y}\|$. Since $\|A^T\| = \|A\|$, we have that $\|\hat{S}\| \le \|d\|(1+\|\hat{y}\|)$, and using the fact that $\mathcal{K}(d,\mu) \ge 1$ the bound (12) on $\|\hat{S}\|$ is a consequence of the bound (11) on $\|\hat{y}\|$. It therefore is sufficient to prove the bounds on $\|\hat{X}\|$ and on $\|\hat{y}\|$. Furthermore, the bound on $\|\hat{y}\|$ is trivial if $\hat{y} = 0$; therefore from now on we assume that $\hat{y} \ne 0$. Also, let $\bar{X} = \hat{X}/\|\hat{X}\|$ and $\bar{y} = \hat{y}/\|\hat{y}\|$. Clearly, $\bar{X} \bullet \hat{X} = \|\hat{X}\|$, $\|\bar{X}\| = 1$, $\bar{y}^T\hat{y} = \|\hat{y}\|$, and $\|\bar{y}\| = 1$.

The rest of the proof proceeds by examining three cases:

(i) $C \bullet \hat{X} \le 0$,
(ii) $0 < C \bullet \hat{X} \le \mu n$, and
(iii) $\mu n < C \bullet \hat{X}$.

In case (i), let $\Delta A_i := -b_i\bar{X}/\|\hat{X}\|$ for $i = 1,\dots,m$. Then, by letting the operator $\Delta A := (\Delta A_1,\dots,\Delta A_m)$ and $\Delta d := (\Delta A, 0, 0) \in \mathcal{D}$, we have $(A+\Delta A)\hat{X} = 0$, $\hat{X} \succ 0$, and $C \bullet \hat{X} \le 0$. It then follows from Proposition 2.2 that $D_\mu(d+\Delta d)$ is infeasible, and so $\rho(d) \le \|\Delta d\| = \|\Delta A\| = \|b\|/\|\hat{X}\| \le \|d\|/\|\hat{X}\|$. Therefore, $\|\hat{X}\| \le \|d\|/\rho(d) = \mathcal{C}(d) \le \mathcal{K}(d,\mu)$. This proves (10) in this case.

Consider the following notation: $\theta := b^T\hat{y}$, $\Delta b := -\theta\bar{y}/\|\hat{y}\|$, $\Delta A_i := -\bar{y}_i C/\|\hat{y}\|$ for $i = 1,\dots,m$, $\Delta A := (\Delta A_1,\dots,\Delta A_m)$, and $\Delta d := (\Delta A, \Delta b, 0) \in \mathcal{D}$. Observe that $(b+\Delta b)^T\hat{y} = 0$ and $(A+\Delta A)^T[\hat{y}] \prec 0$, so from Proposition 2.2 we conclude that $P_\mu(d+\Delta d)$ is infeasible. Therefore, $\rho(d) \le \|\Delta d\| = \max\{\|C\|, |\theta|\}/\|\hat{y}\|$. Hence, $\|\hat{y}\| \le \max\{\mathcal{C}(d), |\theta|/\rho(d)\}$. Furthermore, $|\theta| = |b^T\hat{y}| = |C\bullet\hat{X} - \mu n| \le \|\hat{X}\|\|C\| + \mu n \le \mathcal{C}(d)\|d\| + \mu n$. Therefore, using the fact that $\mathcal{C}(d) \ge 1$ for any d, we have (11).

In case (ii), let $\Delta d := (\Delta A, 0, \Delta C) \in \mathcal{D}$, where $\Delta A_i := -b_i\bar{X}/\|\hat{X}\|$ for $i = 1,\dots,m$ and $\Delta C := -\mu n\bar{X}/\|\hat{X}\|$. Observe that $(A+\Delta A)\hat{X} = 0$ and $(C+\Delta C)\bullet\hat{X} \le 0$. Hence, from Proposition 2.2 $D_\mu(d+\Delta d)$ is infeasible, and so we conclude that $\rho(d) \le \|\Delta d\| = \max\{\|\Delta A\|, \|\Delta C\|\} = \max\{\|b\|, \mu n\}/\|\hat{X}\| \le \max\{\|d\|, \mu n\}/\|\hat{X}\|$. Therefore, $\|\hat{X}\| \le \max\{\mathcal{C}(d), \mu n/\rho(d)\} \le \mathcal{K}(d,\mu)$. This proves (10) for this case.

Now let $\Delta d := (\Delta A, \Delta b, 0)$, where $\Delta A_i := -\bar{y}_i C/\|\hat{y}\|$ for $i = 1,\dots,m$ and $\Delta b := \mu n\bar{y}/\|\hat{y}\|$. Observe that $(b+\Delta b)^T\hat{y} = b^T\hat{y} + \mu n = C\bullet\hat{X} > 0$ and $(A+\Delta A)^T[\hat{y}] \prec 0$. As before, we have from Proposition 2.2 that $P_\mu(d+\Delta d)$ is infeasible, and so we conclude that $\rho(d) \le \|\Delta d\| = \max\{\|\Delta A\|, \|\Delta b\|\} = \max\{\|C\|, \mu n\}/\|\hat{y}\| \le \max\{\|d\|, \mu n\}/\|\hat{y}\|$. Therefore, we obtain $\|\hat{y}\| \le \max\{\mathcal{C}(d), \mu n/\rho(d)\} \le \mathcal{K}(d,\mu)$.

In case (iii), we first consider the bound on $\|\hat{y}\|$. Let $\Delta d := (\Delta A, 0, 0) \in \mathcal{D}$, where $\Delta A_i := -\bar{y}_i C/\|\hat{y}\|$ for $i = 1,\dots,m$. Since $(A+\Delta A)^T[\hat{y}] \prec 0$ and $b^T\hat{y} = C\bullet\hat{X} - \mu n > 0$,

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

it follows from Proposition 2.2 that $P_\mu(d+\Delta d)$ is infeasible, and so $\rho(d) \le \|\Delta d\| = \|C\|/\|\hat{y}\|$. Therefore, $\|\hat{y}\| \le \mathcal{C}(d) \le \mathcal{K}(d,\mu)$.

Finally, let $\Delta A_i := -b_i\bar{X}/\|\hat{X}\|$ for $i = 1,\dots,m$, and $\Delta C := -\theta\bar{X}/\|\hat{X}\|$, where $\theta := C \bullet \hat{X}$. Observe that $(A+\Delta A)\hat{X} = 0$ and $(C+\Delta C)\bullet\hat{X} = 0$. Thus, from Proposition 2.2 we conclude that $D_\mu(d+\Delta d)$ is infeasible, and so $\rho(d) \le \|\Delta d\| = \max\{\|\Delta A\|, \|\Delta C\|\} = \max\{\|b\|,\theta\}/\|\hat{X}\|$ so that $\|\hat{X}\| \le \max\{\mathcal{C}(d), \theta/\rho(d)\}$. Furthermore, $\theta = C\bullet\hat{X} = b^T\hat{y} + \mu n \le \|b\|\|\hat{y}\| + \mu n \le \|d\|\mathcal{C}(d) + \mu n$. Therefore, $\|\hat{X}\| \le \mathcal{K}(d,\mu)$. ∎

The following corollary presents upper bounds on optimal solutions to $P_{\mu+\Delta\mu}(d+\Delta d)$ and $D_{\mu+\Delta\mu}(d+\Delta d)$, where Δd is a data instance in $\mathcal{D}$ representing a small perturbation of the data instance d, and $\Delta\mu$ is a scalar.

COROLLARY 4.1. *Let $d \in \mathrm{Int}(\mathcal{F})$ and $\mu > 0$. If $\|\Delta d\| \le \rho(d)/3$ and $|\Delta\mu| \le \mu/3$, then*

$$\begin{aligned}\|X(d+\Delta d,\mu+\Delta\mu)\| &\le 4\mathcal{K}(d,\mu),\\ \|y(d+\Delta d,\mu+\Delta\mu)\| &\le 4\mathcal{K}(d,\mu),\\ \|S(d+\Delta d,\mu+\Delta\mu)\| &\le 6\|d\|\mathcal{K}(d,\mu).\end{aligned}$$

Proof. The proof follows by observing that

$$\begin{aligned}\|d+\Delta d\| &\le \|d\| + \frac{\rho(d)}{3},\\ \mu+\Delta\mu &\le \frac{4\mu}{3},\\ \rho(d+\Delta d) &\ge \frac{2\rho(d)}{3}.\end{aligned}$$

From these inequalities, we have $\mathcal{C}(d+\Delta d) \le \frac{3}{2}(\|d\| + \rho(d)/3)/\rho(d) = \frac{3}{2}(\mathcal{C}(d) + 1/3) \le 2\mathcal{C}(d)$ and $\|d+\Delta d\| \le \frac{4}{3}\|d\| \le 1.5\|d\|$, since $\mathcal{C}(d) \ge 1$. Furthermore, $(\mu+\Delta\mu)n/\rho(d+\Delta d) \le 2\mu n/\rho(d)$. Therefore, $\mathcal{K}(d+\Delta d,\mu+\Delta\mu) \le 4\mathcal{K}(d,\mu)$, and the result follows. ∎

The following proof of Theorem 3.2 is a generalization of part of the proof of Theorem 3.2 in [15] for the case of a linear program.

Proof of Theorem 3.2. Because of the Karush–Kuhn–Tucker optimality conditions of the dual pair of programs $P_\mu(d)$ and $D_\mu(d)$, we have $X(d,\mu)S(d,\mu) = \mu I$. This being the case, $X(d,\mu)$ and $S(d,\mu)$ can be simultaneously diagonalized, and so there exists an orthogonal matrix U such that $X(d,\mu) = UDU^T$, where $D = \mathrm{diag}(\lambda(X(d,\mu)))$ and $S(d,\mu) = \mu UD^{-1}U^T$. Then

$$\frac{1}{\lambda_j(X(d,\mu))} \le \frac{1}{D_{11}} = \frac{\|S(d,\mu)\|_\infty}{\mu},$$

and so $\lambda_j(X(d,\mu)) \ge \frac{\mu}{\|S(d,\mu)\|_\infty}$ for $j = 1,\dots,n$, and the result for $\lambda_j(X(d,\mu))$ follows from Theorem 3.1. Similarly,

$$\frac{1}{\lambda_j(S(d,\mu))} \le \frac{D_{nn}}{\mu} = \frac{\|X(d,\mu)\|_\infty}{\mu},$$

and so $\lambda_j(S(d,\mu)) \ge \frac{\mu}{\|X(d,\mu)\|_\infty}$ for $j = 1,\dots,n$, and the result for $\lambda_j(S(d,\mu))$ again follows from Theorem 3.1. ∎

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

COROLLARY 4.2. *Let $d \in \text{Int}(\mathcal{F})$ and $\mu > 0$. If $\|\Delta d\| \leq \rho(d)/3$ and $|\Delta\mu| \leq \mu/3$, then for all $j = 1, \ldots, n$,*

$$\lambda_j\left(X(d + \Delta d, \mu + \Delta\mu)\right) \geq \frac{\mu}{16\|d\|\mathcal{K}(d,\mu)},$$
$$\lambda_j\left(S(d + \Delta d, \mu + \Delta\mu)\right) \geq \frac{\mu}{6\mathcal{K}(d,\mu)}.$$

Proof. The proof follows immediately from Theorem 3.2 by observing that $\|d + \Delta d\| \leq \frac{4}{3}\|d\|$, $\mu + \Delta\mu \geq \frac{2}{3}\mu$, and $\mathcal{K}(d + \Delta d, \mu + \Delta\mu) \leq 4\mathcal{K}(d,\mu)$. ∎

5. Proof of bounds on changes in optimal solutions. In this section we prove Theorems 3.3 and 3.4. Before presenting the proofs, we first present properties of certain linear operators that arise in our analysis, in Propositions 5.1–5.5, and Corollary 5.1.

PROPOSITION 5.1. *Given a data instance $d = (A, b, C) \in \mathcal{D}$ and matrices X and $\bar{X}$ such that $X, \bar{X} \succ 0$, let P be the linear operator from $\Re^m$ to $\Re^m$ defined as*

$$Pw := A\left(X\left(A^T[w]\right)\bar{X}\right)$$

for all $w \in \Re^m$. If A has rank m, then the following statements hold true:

1. *P corresponds to a symmetric positive definite matrix in $\Re^{m\times m}$,*
2. *$Pw = A\left(\bar{X}\left(A^T[w]\right)X\right)$ for all $w \in \Re^m$.*

Proof. By using the canonical basis for $\Re^m$ and slightly amending the notation, we have that the (i,j)-coordinate of the matrix corresponding to P is given by

$$P_{ij} = A_i \bullet (XA_j\bar{X}). \tag{17}$$

Hence, if w is such that $Pw = 0$, then for all $i = 1, \ldots, m$,

$$\sum_{j=1}^{m}\left(A_i \bullet (XA_j\bar{X})\right)w_j = 0,$$
$$\sum_{j=1}^{m}\left((X^{1/2}A_i) \bullet (X^{1/2}A_j\bar{X})\right)w_j = 0,$$
$$\sum_{j=1}^{m}\left((X^{1/2}A_i\bar{X}^{1/2}) \bullet (X^{1/2}A_j\bar{X}^{1/2})\right)w_j = 0.$$

It therefore follows that

$$w^T Pw = \sum_{i=1}^{m}\sum_{j=1}^{m} w_i\left((X^{1/2}A_i\bar{X}^{1/2}) \bullet (X^{1/2}A_j\bar{X}^{1/2})\right)w_j = 0. \tag{18}$$

This in turn can be written as

$$\|X^{1/2}(A^T[w])\bar{X}^{1/2}\|_2^2 = 0,$$

from which we obtain $A^T[w] = 0$. Using the fact that A has rank m, we have $w = 0$. Therefore the matrix corresponding to P is nonsingular.

On the other hand, notice that from (17) we have $P_{ij} = A_i \bullet (XA_j\bar{X}) = (XA_i\bar{X}) \bullet A_j = A_j \bullet (XA_i\bar{X}) = P_{ji}$ for all $1 \leq i, j \leq m$. Hence, P is a symmetric operator.

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Furthermore, if we let $\hat{A} := (X^{1/2}A_1\bar{X}^{1/2}, \ldots, X^{1/2}A_m\bar{X}^{1/2})$, we obtain from (18) $w^T P w = \|\hat{A}^T[w]\|_2^2 \geq 0$ for all $w \in \Re^m$. Hence, P is a positive semidefinite operator. Using that P is nonsingular, we conclude the first statement.

Finally, the second statement follows from well-known properties of the trace:

$$\begin{aligned} trace\left(A_i X A_j \bar{X}\right) &= trace\left(X A_j \bar{X} A_i\right) \\ &= trace\left((X A_j \bar{X} A_i)^T\right) \\ &= trace\left(A_i \bar{X} A_j X\right) \end{aligned}$$

for all $1 \leq i, j \leq m$. Therefore, $Pw = A\left(\bar{X}\left(A^T[w]\right)X\right)$ for all $w \in \Re^m$, and the result follows. ∎

PROPOSITION 5.2. *Let $d = (A, b, C) \in \mathrm{Int}(\mathcal{F})$ and P be the linear operator from $\Re^m$ to $\Re^m$ defined as*

$$Pw := A\left(A^T[w]\right)$$

for all $w \in \Re^m$. Then P is a symmetric positive definite matrix and

$$\rho(d) \leq \sqrt{\lambda_1(P)}.$$

Proof. Observe that since $d \in \mathrm{Int}(\mathcal{F})$, A has rank m, and so from Proposition 5.1 (setting $\bar{X} := X := I$), P is a symmetric and positive semidefinite matrix.

Let $\lambda := \lambda_1(P)$. There exists a vector $v \in \Re^m$ with $\|v\| = 1$ and $Pv = \lambda v$. Hence, $v^T P v = \lambda$. Let $\Delta A \in \mathcal{L}_{m,n}$ be defined as

$$\Delta A := \left(-v_1(A^T[v]), \ldots, -v_m(A^T[v])\right)^T,$$

and $\Delta b = \epsilon v$ for any $\epsilon > 0$ and small. Then, $(A + \Delta A)^T[v] = 0$ and $(b + \Delta b)^T v = b^T v + \epsilon \neq 0$ for all $\epsilon > 0$ and small. Hence, $(A + \Delta A)X = b + \Delta b$ is an inconsistent system of equations for all $\epsilon > 0$ and small. Therefore, $\rho(d) \leq \max\{\|\Delta A\|, \|\Delta b\|\} = \|\Delta A\| = \|A^T[v]\| = \sqrt{\lambda}$, thus proving this proposition. ∎

PROPOSITION 5.3. *Given a data instance $d = (A, b, C) \in \mathcal{D}$ such that A has rank m, and matrices X and $\bar{X}$ such that $X, \bar{X} \succ 0$, let Q be the linear operator from $\Re^{n\times n}$ to $\Re^{n\times n}$ defined as*

$$QV := V - \bar{X}^{1/2}\left(A^T\left[P^{-1}A\left(\bar{X}^{1/2}VX^{1/2}\right)\right]\right)X^{1/2}$$

for all $V \in \Re^{n\times n}$, where P is the matrix from Proposition 5.1. Then Q corresponds to a symmetric projection operator.

Proof. Let $RV := \bar{X}^{1/2}\left(A^T\left[P^{-1}A\left(\bar{X}^{1/2}VX^{1/2}\right)\right]\right)X^{1/2}$ for all $V \in \Re^{n\times n}$. Since $QV = V - RV = (I - R)V$, then Q is a symmetric projection if and only if R is a symmetric projection. It is straightforward to show that

$$RV = \sum_{i=1}^{m}\sum_{j=1}^{m} P_{ij}^{-1}\left(A_j \bullet \left(\bar{X}^{1/2}VX^{1/2}\right)\right)\left(\bar{X}^{1/2}A_iX^{1/2}\right) \tag{19}$$

for all $V \in \Re^{n\times n}$. For a fixed matrix V in $\Re^{n\times n}$, it follows from this identity that

$$W \bullet (RV) = \left(\sum_{i=1}^{m}\sum_{j=1}^{m} P_{ij}^{-1}\left(A_i \bullet \left(\bar{X}^{1/2}WX^{1/2}\right)\right)\left(\bar{X}^{1/2}A_jX^{1/2}\right)\right) \bullet V$$

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

for all W in $\Re^{n\times n}$. Hence, we have

$$R^T[W] = \sum_{i=1}^{m}\sum_{j=1}^{m} P^{-1}_{ij}\left(A_i \bullet \left(\bar{X}^{1/2} W X^{1/2}\right)\right)\left(\bar{X}^{1/2} A_j X^{1/2}\right)$$

for all W in $\Re^{n\times n}$. By noticing that P is a symmetric matrix and using (19), we obtain $R = R^T$; that is, R is a symmetric operator.

On the other hand, for a given V in $\Re^{n\times n}$, let $w := P^{-1}A\left(\bar{X}^{1/2} V X^{1/2}\right)$. Thus, using Proposition 5.1, statement 2, we obtain

$$\begin{aligned}
RRV &= \bar{X}^{1/2}\left(A^T\left[P^{-1}A\left(\bar{X}^{1/2}(RV)X^{1/2}\right)\right]\right)X^{1/2}\\
&= \bar{X}^{1/2}\left(A^T\left[P^{-1}A\left(\bar{X}^{1/2}\left(\bar{X}^{1/2}\left(A^T\left[P^{-1}A\left(\bar{X}^{1/2}VX^{1/2}\right)\right]\right)X^{1/2}\right)X^{1/2}\right)\right]\right)X^{1/2}\\
&= \bar{X}^{1/2}\left(A^T\left[P^{-1}A\left(\bar{X}\left(A^T[w]\right)X\right)\right]\right)X^{1/2}\\
&= \bar{X}^{1/2}\left(A^T\left[P^{-1}Pw\right]\right)X^{1/2}\\
&= \bar{X}^{1/2}\left(A^T[w]\right)X^{1/2}\\
&= RV,
\end{aligned}$$

where the fourth equality above follows from statement 2 of Proposition 5.1. Therefore, from [11, Theorem 1, page 73], R is a projection and the result follows. ∎

PROPOSITION 5.4. *Given a data instance $d = (A, b, C) \in \mathcal{D}$ and matrices X and $\bar{X}$ such that $X, \bar{X} \succ 0$, let P be the linear operator from $\Re^m$ to $\Re^m$ defined as*

$$Pw := A\left(X\left(A^T[w]\right)\bar{X}\right)$$

for all $w \in \Re^m$. Then if A has rank m,

$$\|P^{-1}\|_\infty \le \|X^{-1}\|_\infty \|\bar{X}^{-1}\|_\infty \|(AA^T)^{-1}\|_\infty.$$

Proof. From Proposition 5.1, it follows that P is nonsingular. Let w be any vector in $\Re^m$ normalized so that $\|w\| = 1$, and consider a spectral decomposition of X as

$$X = \sum_{k=1}^{n} \lambda_k(X) u_k u_k^T,$$

where $\{u_1, \ldots, u_n\}$ is an orthonormal basis for $\Re^n$. By using that $trace(u_k u_k^T) \ge 0$ for all $1 \le k \le n$, and $\sum_{k=1}^{n} u_k u_k^T = I$, we have

$$\begin{aligned}
w^T P w &= \sum_{i=1}^{m}\sum_{j=1}^{m} trace\left(A_i X A_j \bar{X}\right) w_i w_j\\
&= trace\left(\sum_{i=1}^{m}\sum_{j=1}^{m} A_i X A_j \bar{X} w_i w_j\right)\\
&= trace\left(\bar{X}^{1/2}\left(\sum_{i=1}^{m} A_i w_i\right) X \left(\sum_{j=1}^{m} A_j w_j\right)\bar{X}^{1/2}\right)\\
&= \sum_{k=1}^{n} \lambda_k(X)\, trace\left(\bar{X}^{1/2}\left(\sum_{i=1}^{m} A_i w_i\right) u_k u_k^T \left(\sum_{j=1}^{m} A_j w_j\right)\bar{X}^{1/2}\right)
\end{aligned}$$

$$\geq \|X^{-1}\|_\infty^{-1} trace\left(\bar{X}^{1/2}\left(\sum_{i=1}^m A_i w_i\right)\left(\sum_{j=1}^m A_j w_j\right)\bar{X}^{1/2}\right)$$
$$= \|X^{-1}\|_\infty^{-1} trace\left(\left(\sum_{i=1}^m A_i w_i\right)\bar{X}\left(\sum_{j=1}^m A_j w_j\right)\right).$$

Now, consider a spectral decomposition of $\bar{X}$ as

$$\bar{X} = \sum_{k=1}^n \lambda_k(\bar{X}) v_k v_k^T,$$

where, as before, $\{v_1, \ldots, v_n\}$ is an orthonormal basis for $\Re^n$, and so $\sum_{k=1}^n v_k v_k^T = I$. Notice that from Proposition 5.1, it follows that the operator AA^T is nonsingular. Then, we have

$$w^T P w \geq \|X^{-1}\|_\infty^{-1} \sum_{k=1}^n \lambda_k(\bar{X}) trace\left(\left(\sum_{i=1}^m A_i w_i\right) v_k v_k^T \left(\sum_{j=1}^m A_j w_j\right)\right)$$
$$\geq \|X^{-1}\|_\infty^{-1} \|\bar{X}^{-1}\|_\infty^{-1} trace\left(\left(\sum_{i=1}^m A_i w_i\right)\left(\sum_{j=1}^m A_j w_j\right)\right)$$
$$\geq \|X^{-1}\|_\infty^{-1} \|\bar{X}^{-1}\|_\infty^{-1} \|(AA^T)^{-1}\|_\infty^{-1}.$$

Notice that in the last inequality we used

$$trace\left(\left(\sum_{i=1}^m A_i w_i\right)\left(\sum_{j=1}^m A_j w_j\right)\right) = w^T \hat{P} w \geq \min_k \lambda_k(\hat{P}) = \|(AA^T)^{-1}\|_\infty^{-1},$$

where $\hat{P} = AA^T$.

Now let $\hat{w}$ be the normalized eigenvector corresponding to the smallest eigenvalue of P, i.e., $\|\hat{w}\| = 1$ and $P\hat{w} = \lambda_1(P)\hat{w}$. Then from above we have

$$\|P^{-1}\|_\infty^{-1} = \lambda_1(P) = \hat{w}^T P \hat{w} \geq \|X^{-1}\|_\infty^{-1} \|\bar{X}^{-1}\|_\infty^{-1} \|(AA^T)^{-1}\|_\infty^{-1}$$

and the result follows. ∎

COROLLARY 5.1. *Let $d = (A, b, C)$ be a data instance in $\mathrm{Int}(\mathcal{F})$, μ be a positive scalar, $\Delta d = (\Delta A, \Delta b, \Delta C) \in \mathcal{D}$ be a data perturbation such that $\|\Delta d\| \leq \rho(d)/3$, and $\Delta\mu$ be a scalar such that $|\Delta\mu| \leq \mu/3$. Then*

$$\|P^{-1}\| \leq 32\sqrt{m}\left(\frac{\mathcal{C}(d)\mathcal{K}(d,\mu)}{\mu}\right)^2,$$

where P is the linear operator from $\Re^m$ to $\Re^m$ defined as

$$Pw := A\left(X(d,\mu)\left(A^T[w]\right)X(d+\Delta d, \mu+\Delta\mu)\right)$$

for all $w \in \Re^m$, and $\mathcal{K}(d,\mu)$ is the scalar defined in (9).

Proof. Let $X = X(d,\mu)$ and $\bar{X} = X(d+\Delta d, \mu+\Delta\mu)$. From Proposition 5.4 we know that

$$\|P^{-1}\|_\infty \leq \|X^{-1}\|_\infty \|\bar{X}^{-1}\|_\infty \|(AA^T)^{-1}\|_\infty.$$

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

From Theorem 3.2 and Corollary 4.2, respectively, we have

$$\|X^{-1}\|_\infty \leq \frac{2\|d\|\mathcal{K}(d,\mu)}{\mu},$$
$$\|\bar{X}^{-1}\|_\infty \leq \frac{16\|d\|\mathcal{K}(d,\mu)}{\mu}.$$

Furthermore, from Proposition 5.2 we have

$$\|(AA^T)^{-1}\|_\infty \leq \frac{1}{\rho(d)^2}.$$

By combining these results and using Proposition 2.1, we obtain the corollary. ∎

PROPOSITION 5.5. *Let $d = (A, b, C)$ be a data instance in $\mathrm{Int}(\mathcal{F})$, μ be a positive scalar, $\Delta d = (\Delta A, \Delta b, \Delta C) \in \mathcal{D}$ be a data perturbation such that $\|\Delta d\| \leq \rho(d)/3$, and $\Delta\mu$ be a scalar such that $|\Delta\mu| \leq \mu/3$. Then,*

$$\|\Delta b - \Delta A X(d + \Delta d, \mu + \Delta\mu)\| \leq 5\|\Delta d\|\mathcal{K}(d,\mu),$$
$$\|\Delta C - \Delta A^T[y(d + \Delta d, \mu + \Delta\mu)]\| \leq 5\|\Delta d\|\mathcal{K}(d,\mu).$$

Proof. Let $\bar{X} := X(d + \Delta d, \mu + \Delta\mu)$ and $\bar{y} := y(d + \Delta d, \mu + \Delta\mu)$. From Corollary 4.1, we have

$$\begin{aligned}\|\Delta b - \Delta A\bar{X}\| &\leq \|\Delta d\|\left(1 + \|\bar{X}\|\right)\\ &\leq \|\Delta d\|\left(1 + 4\mathcal{K}(d,\mu)\right)\\ &\leq 5\|\Delta d\|\mathcal{K}(d,\mu),\end{aligned}$$

$$\begin{aligned}\|\Delta C - \Delta A^T[\bar{y}]\| &\leq \|\Delta d\|\left(1 + \|\bar{y}\|\right)\\ &\leq \|\Delta d\|\left(1 + 4\mathcal{K}(d,\mu)\right)\\ &\leq 5\|\Delta d\|\mathcal{K}(d,\mu),\end{aligned}$$

and so the proposition follows. ∎

Now we are ready to present the proof of Theorem 3.3.

Proof of Theorem 3.3. To simplify the notation, let $(X, y, S) := (X(d,\mu), y(d,\mu), S(d,\mu))$, $(\bar{X}, \bar{y}, \bar{S}) := (X(d+\Delta d, \mu+\Delta\mu), y(d+\Delta d, \mu+\Delta\mu), S(d+\Delta d, \mu+\Delta\mu))$, and $\bar{\mu} := \mu + \Delta\mu$. From the Karush–Kuhn–Tucker optimality conditions associated with the programs $P_\mu(d)$ and $P_{\mu+\Delta\mu}(d + \Delta d)$, respectively, we obtain

$$\begin{array}{rclcrcl} XS &=& \mu I, & & \bar{X}\bar{S} &=& \bar{\mu} I,\\ A^T[y] + S &=& C, & (A+\Delta A)^T[\bar{y}] + \bar{S} &=& C + \Delta C,\\ AX &=& b, & (A+\Delta A)\bar{X} &=& b + \Delta b,\\ X &\succ& 0, & \bar{X} &\succ& 0.\end{array}$$

Let $\Delta E := \Delta b - \Delta A\bar{X}$ and $\Delta F := \Delta C - \Delta A^T[\bar{y}]$. Therefore,

$$\begin{aligned}\bar{X} - X &= \frac{1}{\mu\bar{\mu}}\bar{X}(\bar{\mu}S - \mu\bar{S})X\\ &= \frac{1}{\mu\bar{\mu}}\bar{X}\left(\bar{\mu}(C - A^T[y]) - \mu(C + \Delta C - (A+\Delta A)^T[\bar{y}])\right)X\\ &= \frac{1}{\mu\bar{\mu}}\bar{X}\left(\Delta\mu C - \mu(\Delta C - \Delta A^T[\bar{y}]) - A^T[\bar{\mu}y - \mu\bar{y}]\right)X\\ &= \frac{\Delta\mu}{\mu\bar{\mu}}\bar{X}CX - \frac{1}{\bar{\mu}}\bar{X}\Delta FX - \frac{1}{\mu\bar{\mu}}\bar{X}\left(A^T[\bar{\mu}y - \mu\bar{y}]\right)X. \qquad (20)\end{aligned}$$

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

On the other hand, $A(\bar{X} - X) = \Delta b - \Delta A\bar{X} = \Delta E$. Since $d \in \mathrm{Int}(\mathcal{F})$, then A has full-rank (see (8)). It follows from Proposition 5.1 that the linear operator P from $\Re^m$ to $\Re^m$ defined as $Pw := A(\bar{X}(A^T[w])X)$, for all $w \in \Re^m$, corresponds to a positive definite matrix in $\Re^{m\times m}$. By combining this result with (20), we obtain

$$\Delta E = \frac{\Delta\mu}{\mu\bar{\mu}} A\left(\bar{X}CX\right) - \frac{1}{\bar{\mu}} A\left(\bar{X}\Delta FX\right) - \frac{1}{\mu\bar{\mu}} P(\bar{\mu}y - \mu\bar{y}),$$

and so

$$P^{-1}\Delta E = \frac{\Delta\mu}{\mu\bar{\mu}} P^{-1}A\left(\bar{X}CX\right) - \frac{1}{\bar{\mu}} P^{-1}A\left(\bar{X}\Delta FX\right) - \frac{1}{\mu\bar{\mu}}(\bar{\mu}y - \mu\bar{y}).$$

Therefore, we have the following identity:

$$\frac{1}{\mu\bar{\mu}}(\bar{\mu}y - \mu\bar{y}) = \frac{\Delta\mu}{\mu\bar{\mu}} P^{-1}A\left(\bar{X}CX\right) - \frac{1}{\bar{\mu}} P^{-1}A\left(\bar{X}\Delta FX\right) - P^{-1}\Delta E. \tag{21}$$

Combining (21) and (20), we obtain

$$\begin{aligned}
\bar{X} - X &= \frac{\Delta\mu}{\mu\bar{\mu}}\bar{X}CX - \frac{1}{\bar{\mu}}\bar{X}\Delta FX \\
&\quad - \bar{X}\left(A^T\left[\frac{\Delta\mu}{\mu\bar{\mu}} P^{-1}A\left(\bar{X}CX\right) - \frac{1}{\bar{\mu}} P^{-1}A\left(\bar{X}\Delta FX\right) - P^{-1}\Delta E\right]\right)X \\
&= \frac{\Delta\mu}{\mu\bar{\mu}}\left(\bar{X}CX - \bar{X}\left(A^T\left[P^{-1}A\left(\bar{X}CX\right)\right]\right)X\right) \\
&\quad - \frac{1}{\bar{\mu}}\left(\bar{X}\Delta FX - \bar{X}\left(A^T\left[P^{-1}A\left(\bar{X}\Delta FX\right)\right]\right)X\right) \\
&\quad + \bar{X}\left(A^T\left[P^{-1}\Delta E\right]\right)X \\
&= \frac{\Delta\mu}{\mu\bar{\mu}}\bar{X}^{1/2}Q\left(\bar{X}^{1/2}CX^{1/2}\right)X^{1/2} - \frac{1}{\bar{\mu}}\bar{X}^{1/2}Q\left(\bar{X}^{1/2}\Delta FX^{1/2}\right)X^{1/2} \\
&\quad + \bar{X}\left(A^T\left[P^{-1}\Delta E\right]\right)X, \qquad (22)
\end{aligned}$$

where by Q we denote the following linear operator from $\Re^{n\times n}$ to $\Re^{n\times n}$:

$$Q(V) := V - \bar{X}^{1/2}\left(A^T\left[P^{-1}A\left(\bar{X}^{1/2}VX^{1/2}\right)\right]\right)X^{1/2}$$

for all $V \in \Re^{n\times n}$. By using Proposition 5.3, it follows that Q is a symmetric projection operator, and so $\|QV\| \le \|V\|$ for all $V \in \Re^{n\times n}$. Since $\|V^{1/2}\|^2 \le \sqrt{n}\|V\|$ for all V in $\mathcal{S}_n^+$, from (22), Theorem 3.1, Corollary 4.1, Corollary 5.1, and Proposition 5.5, it follows that

$$\begin{aligned}
\|\bar{X} - X\| &\le \frac{|\Delta\mu|}{\mu\bar{\mu}}\|\bar{X}^{1/2}\|^2\|C\|\|X^{1/2}\|^2 + \frac{1}{\bar{\mu}}\|\bar{X}^{1/2}\|^2\|\Delta F\|\|X^{1/2}\|^2 \\
&\qquad + \|\bar{X}\left(A^T[P^{-1}\Delta E]\right)X\| \\
&\le \frac{n|\Delta\mu|}{\mu\bar{\mu}}\|\bar{X}\|\|C\|\|X\| + \frac{n}{\bar{\mu}}\|\bar{X}\|\|\Delta F\|\|X\| + \|\bar{X}\|\|A^T\|\|P^{-1}\|\|\Delta E\|\|X\| \\
&\le \frac{4n|\Delta\mu|}{\mu\bar{\mu}}\|d\|\mathcal{K}(d,\mu)^2 + \frac{20n\|\Delta d\|}{\bar{\mu}}\mathcal{K}(d,\mu)^3 + \frac{640}{\mu^2}\sqrt{m}\|\Delta d\|\|d\|\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5.
\end{aligned}$$

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Therefore, by noticing that $\bar{\mu} \geq \frac{2}{3}\mu$, we obtain

$$\|\bar{X} - X\| \leq \frac{6n}{\mu^2}|\Delta\mu|\|d\|\mathcal{K}(d,\mu)^2 + \frac{640n\sqrt{m}}{\mu^2}\|\Delta d\|\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5\left(\mu + \|d\|\right),$$

and so (14) follows.

Next, we prove the bound on $\|\bar{y} - y\|$. From the identities $(A + \Delta A)^T[\bar{y}] + \bar{S} = C + \Delta C$ and $A^T[y] + S = C$, it follows that

$$\begin{aligned}
\bar{S} - S &= \Delta F - A^T[\bar{y} - y], \\
\bar{\mu}\bar{X}^{-1} - \mu X^{-1} &= \Delta F - A^T[\bar{y} - y], \\
\bar{X}^{-1}(\bar{\mu}X - \mu\bar{X})X^{-1} &= \Delta F - A^T[\bar{y} - y].
\end{aligned}$$

Hence,

$$\begin{aligned}
\bar{\mu}X - \mu\bar{X} &= \bar{X}\left(\Delta F - A^T[\bar{y} - y]\right)X \\
&= \bar{X}\Delta F X - \bar{X}\left(A^T[\bar{y} - y]\right)X.
\end{aligned}$$

By premultiplying this identity by A, we obtain

$$\Delta\mu b - \mu\Delta E = A\left(\bar{X}\Delta F X\right) - P(\bar{y} - y),$$

and so,

$$\begin{aligned}
P(\bar{y} - y) &= -\Delta\mu b + \mu\Delta E + A\left(\bar{X}\Delta F X\right), \\
\bar{y} - y &= -\Delta\mu P^{-1}b + \mu P^{-1}\Delta E + P^{-1}A\left(\bar{X}\Delta F X\right).
\end{aligned}$$

Therefore, using this identity, Theorem 3.1, Corollary 4.1, Corollary 5.1, and Proposition 5.5, we obtain

$$\begin{aligned}
\|\bar{y} - y\| &\leq |\Delta\mu|\|P^{-1}\|\|b\| + \mu\|P^{-1}\|\|\Delta E\| + \|P^{-1}\|\|A\|\|\bar{X}\|\|\Delta F\|\|X\| \\
&\leq 32\sqrt{m}|\Delta\mu|\|d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^2}{\mu^2} + 160\sqrt{m}\|\Delta d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^3}{\mu} \\
&\quad + 640\sqrt{m}\|\Delta d\|\|d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5}{\mu^2} \\
&\leq 32\sqrt{m}|\Delta\mu|\|d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^2}{\mu^2} + 640\sqrt{m}\|\Delta d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5(\mu + \|d\|)}{\mu^2},
\end{aligned}$$

and so we obtain inequality (15).

Finally, to obtain the bound on $\|\bar{S} - S\|$, we proceed as follows. Notice that $\bar{S} - S = \Delta F - A^T[\bar{y} - y]$. Hence, from (15) and Proposition 5.5, we have

$$\begin{aligned}
\|\bar{S} - S\| &\leq \|\Delta F\| + \|A^T\|\|\bar{y} - y\| \\
&\leq 5\|\Delta d\|\mathcal{K}(d,\mu) + \|d\|\left(32\sqrt{m}|\Delta\mu|\|d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^2}{\mu^2}\right. \\
&\qquad\qquad \left. + 640\sqrt{m}\|\Delta d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5(\mu + \|d\|)}{\mu^2}\right) \\
&\leq 32\sqrt{m}|\Delta\mu|\|d\|^2\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^2}{\mu^2} + 640\sqrt{m}\|\Delta d\|\frac{\mathcal{C}(d)^2\mathcal{K}(d,\mu)^5(\mu + \|d\|)^2}{\mu^2},
\end{aligned}$$

which establishes (16), concluding the proof of this theorem. ∎

Finally, we present the proof of Theorem 3.4.

Proof of Theorem 3.4. To simplify the notation, let $\bar{z} := z(d+\Delta d, \mu+\Delta\mu)$ and $z = z(d,\mu)$. Consider the Lagrangian functions associated with $P_\mu(d)$ and $P_{\mu+\Delta\mu}(d+\Delta d)$, respectively:

$$L(X,y) := C \bullet X + \mu p(X) + y^T(b - AX),$$
$$\bar{L}(X,y) := (C+\Delta C) \bullet X + (\mu+\Delta\mu)p(X) + y^T(b+\Delta b - (A+\Delta A)X),$$

and define $M(X,y) := L(X,y) - \bar{L}(X,y)$. Let $\hat{X}$ and $(\hat{y},\hat{S})$ denote the optimal solutions to $P_\mu(d)$ and $D_\mu(d)$, respectively, and let $\bar{X}$ and $(\bar{y},\bar{S})$ denote the optimal solutions to $P_{\mu+\Delta\mu}(d+\Delta d)$ and $D_{\mu+\Delta\mu}(d+\Delta d)$, respectively. Hence, we have

$$\begin{aligned} z &= L(\hat{X},\hat{y}) \\ &= \max_y L(\hat{X},y) \\ &= \max_y \left\{ \bar{L}(\hat{X},y) + M(\hat{X},y) \right\} \\ &\geq \bar{L}(\hat{X},\bar{y}) + M(\hat{X},\bar{y}) \\ &\geq \min_{X\succ 0} \bar{L}(X,\bar{y}) + M(\hat{X},\bar{y}) \\ &= \bar{z} + M(\hat{X},\bar{y}). \end{aligned}$$

Thus, $z - \bar{z} \geq M(\hat{X},\bar{y})$. Similarly, we can prove that $z - \bar{z} \leq M(\bar{X},\hat{y})$. Therefore, we obtain that either $|\bar{z}-z| \leq |M(\hat{X},\bar{y})|$ or $|\bar{z}-z| \leq |M(\bar{X},\hat{y})|$. On the other hand, by using Theorem 3.1 and Corollary 4.1, we have

$$\begin{aligned} |M(\hat{X},\bar{y})| &= |\Delta C \bullet \hat{X} + \Delta\mu p(\hat{X}) + \bar{y}^T\Delta b - \bar{y}^T\Delta A\hat{X}| \\ &\leq \|\Delta C\|\|\hat{X}\| + |\Delta\mu||p(\hat{X})| + \|\bar{y}\|\|\Delta b\| + \|\bar{y}\|\|\Delta A\|\|\hat{X}\| \\ &\leq \|\Delta d\| \left(\|\hat{X}\| + \|\bar{y}\| + \|\bar{y}\|\|\hat{X}\| \right) + |\Delta\mu||p(\hat{X})| \\ &\leq 9\|\Delta d\|\mathcal{K}(d,\mu)^2 + |\Delta\mu||p(\hat{X})|. \end{aligned}$$

Similarly, it is not difficult to show that

$$|M(\bar{X},\hat{y})| \leq 9\|\Delta d\|\mathcal{K}(d,\mu)^2 + |\Delta\mu||p(\bar{X})|.$$

Therefore,

$$|\bar{z}-z| \leq 9\|\Delta d\|\mathcal{K}(d,\mu)^2 + |\Delta\mu| \max\left\{ |p(\hat{X})|, |p(\bar{X})| \right\}.$$

By using Theorems 3.1 and 3.2 and Corollaries 4.1 and 4.2, we obtain

$$\begin{aligned} -n\ln\left(\mathcal{K}(d,\mu)\right) \leq & p(\hat{X}) \leq -n\ln\left(\frac{\mu}{2\|d\|\mathcal{K}(d,\mu)}\right) \\ -n\ln\left(4\mathcal{K}(d,\mu)\right) \leq & p(\bar{X}) \leq -n\ln\left(\frac{\mu}{16\|d\|\mathcal{K}(d,\mu)}\right). \end{aligned}$$

Thus, we have the following bound:

$$\begin{aligned} \max\left\{ |p(\hat{X})|, |p(\bar{X})| \right\} &\leq n\max\left\{ \ln\left(4\mathcal{K}(d,\mu)\right), \left| \ln\left(\frac{\mu}{16\|d\|\mathcal{K}(d,\mu)}\right) \right| \right\} \\ &\leq n\left(\ln 16 + |\ln\mu| + |\ln\|d\|| + \ln\mathcal{K}(d,\mu)\right), \end{aligned}$$

and so the result follows. ∎

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/11/16 to 18.111.104.46. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

REFERENCES

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM J. Optim., 5 (1995), pp. 13–51.

[2] F. Alizadeh, J.-P. Haeberly, and M. L. Overton, *Primal-dual interior-point methods for semidefinite programming*: *Convergence rates, stability and numerical results*, SIAM J. Optim., 8 (1998), pp. 746–768.

[3] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997.

[4] M. Epelman and R. M. Freund, *Condition number complexity of an elementary algorithm for computing a reliable solution of a conic linear system*, Math. Program., 88 (2000), pp. 451–485.

[5] S. Filipowski, *On the complexity of solving feasible systems of linear inequalities specified with approximate data*, Math. Program., 71 (1995), pp. 259–288.

[6] S. Filipowski, *On the complexity of solving sparse symmetric linear programs specified with approximate data*, Math. Oper. Res., 22 (1997), pp. 769–792.

[7] S. Filipowski, *On the complexity of solving feasible linear programs specified with approximate data*, SIAM J. Optim., 9 (1999), pp. 1010–1040.

[8] R. M. Freund and J. R. Vera, *Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system*, Math. Program., 86 (1999), pp. 225–260.

[9] R. M. Freund and J. R. Vera, *Condition-based complexity of convex optimization in conic linear form via the ellipsoid algorithm*, SIAM J. Optim., 10 (1999), pp. 155–176.

[10] D. Goldfarb and K. Scheinberg, *Interior point trajectories in semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 871–886.

[11] P. R. Halmos, *Finite-Dimensional Vector Spaces*, Springer-Verlag, New York, 1993.

[12] O. L. Mangasarian, *A stable theorem of the alternative*: *An extension of the Gordan theorem*, Linear Algebra Appl., 41 (1981), pp. 209–223.

[13] M. V. Nayakkankuppam and M. L. Overton, *Conditioning of semidefinite programs*, Math. Program., 85 (1999), pp. 525–540.

[14] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.

[15] M. A. Nunez and R. M. Freund, *Condition measures and properties of the central trajectory of a linear program*, Math. Program., 83 (1998), pp. 1–28.

[16] J. Peña and J. Renegar, *Computing approximate solutions for convex conic systems of constraints*, Math. Program., 87 (2000), pp. 351–383.

[17] J. Renegar, *Some perturbation theory for linear programming*, Math. Program., 65 (1994), pp. 73–91.

[18] J. Renegar, *Incorporating condition measures into the complexity theory of linear programming*, SIAM J. Optim., 5 (1995), pp. 506–524.

[19] J. Renegar, *Linear programming, complexity theory, and elementary functional analysis*, Math. Program., 70 (1995), pp. 279–351.

[20] J. Renegar, *Condition numbers, the barrier method, and the conjugate-gradient method*, SIAM J. Optim., 6 (1996), pp. 879–912.

[21] J. Renegar, *A Mathematical View of Interior-Point Methods in Convex Optimization*, SIAM, Philadelphia, forthcoming.

[22] S. M. Robinson, *Stability theory for systems of inequalities. Part* I: *Linear systems*, SIAM J. Numer. Anal., 12 (1975), pp. 754–769.

[23] S. M. Robinson, *Stability theory for systems of inequalities. Part* II: *Differential nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[24] S. M. Robinson, *A characterization of stability in linear programming*, Oper. Res., 25 (1977), pp. 435–447.

[25] S. Scholtes, *On the Computability of the Condition Number for Certain Inconsistent Systems*, Tech. report, Department of Engineering and The Judge Institute of Management Studies, University of Cambridge, Cambridge, England, 1996.

[26] J. F. Sturm and S. Zhang, *On Sensitivity of Central Solutions in Semidefinite Programming*, Tech. report 9813/A, Econometric Institute, Erasmus University, Rotterdam, The Netherlands, 1998.

[27] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.

[28] J. R. Vera, *Ill-posedness and the Computation of Solutions to Linear Programs with Approximate Data*, Tech. report, Cornell University, Ithaca, NY, 1992.

[29] J. R. Vera, *Ill-posedness and the complexity of deciding existence of solutions to linear programs*, SIAM J. Optim., 6 (1996), pp. 549–569.