SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1105

September 1994

**Condition Numbers,
the Barrier Method
and the Conjugate
Gradient Method**

by

James Renegar¹

¹Research was supported by the National Science Foundation grant no. CCR-9103285.

1 Introduction

1.1

We consider the (theoretical) efficiency of the conjugate gradient method in approximately solving the linear equations which arise during application of a particular interior point method. The interior point method (approximately) solves systems of constraints of the form

$$\begin{aligned} Ax &= b \\ x &\geq 0, \end{aligned} \tag{1}$$

where “ $x \geq 0$ ” is interpreted quite generally. The interior point method we consider is simply the barrier method applied to a carefully chosen reformulation of the constraints (1).

Actually, our main results are nice relations between the condition numbers of the systems of linear equations which arise during application of the barrier method and a particular quantity which is naturally interpreted as the condition number of the constraints (1). These relations have ramifications well beyond our analysis of the conjugate gradient method. Readers interested in condition numbers, but not interested in our conjugate gradient method analysis, should read Sections 1.4-1.8, 1.10 and 1.12 (through Corollary 1.7), ignoring sentences which refer to the conjugate gradient method.

This paper continues in the conceptual vein developed in the (thoroughly revised)² introductory section of [6].

We begin by recalling a few important, widely known, theoretical features of the conjugate gradient method (cgm); some of our results will be seen as analogues to these.

1.2

Assume that X is a Hilbert space, $A : X \rightarrow X$ is a strictly positive definite³ linear operator, and $b \in X$. Define the inner product $\langle \cdot, \cdot \rangle_A$ and norm $\| \cdot \|_A$ by

$$\langle u, v \rangle_A := \langle u, Av \rangle, \quad \|u\|_A := \langle u, u \rangle_A^{1/2},$$

where $\langle \cdot, \cdot \rangle$ is the (original) inner product on X .

Let x^* denote the solution of the system $Ax = b$. Provided with an initial iterate x_0 , cgm can be applied to solve the system; cgm produces the sequence

²The first version of [6] was circulated (but not widely so) at the end of Summer, 1993. The paper, especially the introduction, was thoroughly revised in Summer, 1994 and now provides much more motivation than the original version.

³ A is self-adjoint and $\langle x, Ax \rangle > 0$ for all non-zero $x \in X$.

of iterates $x_1, x_2 \dots$, where x_i is the optimal solution for the following problem:

$$\begin{aligned} \min \quad & \|x^* - x\|_A \\ \text{s.t.} \quad & x - x_0 \in \text{Span}\{Ae_0, A^2e_0, \dots, A^ie_0\}; \end{aligned} \quad (2)$$

here, $e_0 := x^* - x_0$. In the literature, one often finds this optimality property possessed by cgm expressed somewhat differently, as we now recall.

Clearly, the vectors in $\text{Span}\{Ae_0, A^2e_0, \dots, A^ie_0\}$ are precisely the vectors of the form $p(A)e_0$ where p ranges over polynomials of degree at most i which satisfy $p(0) = 0$. Since

$$x^* - x = e_0 - (x - x_0),$$

one thus sees that the optimal value of (2) (i.e., $\|x^* - x_i\|_A$) equals the optimal value of the following problem:

$$\begin{aligned} \min \quad & \|p(A)e_0\|_A \\ \text{s.t.} \quad & p \in \mathcal{P}_i, \end{aligned} \quad (3)$$

where $\mathcal{P}_i$ is the set consisting of those polynomials p of degree at most i which satisfy $p(0) = 1$. (It follows, for example, that if A has at most k distinct values in its spectrum, $\text{Spec}(A)$, then $x_i = x^*$ whenever $i \geq k$.)

As is well known, the fact that $\|x^* - x_i\|_A$ equals the optimal value of (3) allows one to deduce, via Chebyshev polynomials, a bound on $\|x^* - x_i\|_A$ in terms of the relative condition number of $\sqrt{A}$,

$$\text{relcond}(\sqrt{A}) := \|\sqrt{A}\| \|\sqrt{A}^{-1}\| (= \sqrt{\text{relcond}(A)}).$$

Precisely, one can prove that

$$\|x^* - x_i\|_A \leq 2 \left(\frac{\text{relcond}(\sqrt{A}) - 1}{\text{relcond}(\sqrt{A}) + 1} \right)^i \|x^* - x_0\|_A; \quad (4)$$

c.f., [1, Appendix B and (5.30)]. From this it is easily deduced that if $0 < \epsilon \leq 1$ then cgm produces x_i satisfying

$$\|x^* - x_i\|_A \leq \epsilon \|x^* - x_0\|_A$$

within

$$O \left(\text{relcond}(\sqrt{A}) \log_2 \frac{2}{\epsilon} \right)$$

iterations, a fact we express in short-hand as

$$i = \Omega \left(\text{relcond}(\sqrt{A}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \|x^* - x_i\|_A \leq \epsilon \|x^* - x_0\|_A. \quad (5)$$

Although (4) does not in general imply

$$\|x^* - x_i\|_A \leq \|x^* - x_0\|_A \text{ for all } i, \quad (6)$$

the relation (6) is indeed valid (virtually by definition of cgm).

1.3

If one wishes to solve a least squares problem

$$\min \|Ax - b\|$$

where $A : X \rightarrow Y$ and X, Y are Hilbert spaces, one can apply cgm to solving the system

$$A^T A x = A^T b$$

if $0 \notin \text{Spec}(A^T A)$. Letting $x_1, x_2, \dots$ denote the resulting sequence of iterates, as an immediate corollary to (5) one obtains

$$i = \Omega \left(\text{relcond}(\sqrt{A^T A}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \|x^* - x_i\|_{A^T A} \leq \epsilon \|x^* - x_0\|_{A^T A}$$

where x^* is the least squares solution. To understand the relevance of the norm $\| \cdot \|_{A^T A}$ in this context, note that

$$\|x^* - x_i\|_{A^T A}^2 = \|Ax_i - b\|^2 - \min\{\|Ax - b\|^2; x \in X\}.$$

In a similar vein, if $0 \notin \text{Spec}(AA^T)$ then one can use cgm to solve $Ax = b$ where $A : X \rightarrow Y$. Given initial $y_0 \in Y$, one applies cgm to the system

$$AA^T y = b$$

to obtain a sequence $y_1, y_2, \dots$. As an immediate corollary to (5) one obtains

$$i = \Omega \left(\text{relcond}(\sqrt{AA^T}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \|x^* - A^T y_i\| \leq \epsilon \|x^* - A^T y_0\| \quad (7)$$

where x^* is the minimal norm solution for the system $Ax = b$. This is the application of cgm that we will refer to later.

For future reference we also record the inequality analogous to (6), namely

$$\|x^* - A^T y_i\| \leq \|x^* - A^T y_0\| \text{ for all } i. \quad (8)$$

1.4

We now begin considering (very general) systems of constrained linear equations to be solved via the barrier method.

Consider a system of constraints

$$\begin{aligned} Ax &= b \\ x &\geq 0, \end{aligned} \quad (9)$$

where

- $A : X \rightarrow Y$ is a continuous linear operator,
- X and Y are real Hilbert spaces and
- “ $x \geq 0$ ” means $x \in C$ where $C \subseteq X$ is a fixed, closed, convex cone with vertex at the origin.

Such very general definitions of constraints date at least to Duffin[2].

By specializing the cone C , one obtains customary forms of constraints appearing in the linear programming literature. For example, if one lets $X = \mathbb{R}^n$ and wishes to impose non-negativity constraints on the first J variables then one lets

$$C := \{x \in \mathbb{R}^n; x_j \geq 0 \text{ for all } j = 1, \dots, J\}. \quad (10)$$

We consider perturbations in A and b but we always consider the cone C as fixed (once it has been chosen); we will impose only minor restrictions on the choice of C .

1.5

Much of our recent research (c.f., [6]) examines the efficiency of algorithms for solving general constraints like (9), where the efficiency is measured in terms of a parameter which is a type of condition number; the parameter is the reciprocal of the minimal relative amount by which the system $Ax = b$, $x \geq 0$ can be perturbed to obtain a system which is inconsistent. Formally, letting (A, b) denote the system $Ax = b$, $x \geq 0$, and defining

$$\|(A, b)\| := \max\{\|A\|, \|b\|\},$$

(where $\|A\|$ is the operator norm induced by the norms on X and Y), we define

$$\text{relcond}(A, b) := \frac{\|(A, b)\|}{\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}}$$

where “ $(A + \Delta A, b + \Delta b)$ is inconsistent” means that the system

$$\begin{aligned} (A + \Delta A)x &= b + \Delta b \\ x &\geq 0 \end{aligned}$$

is inconsistent. As is discussed at great length in [6, §1], $\text{relcond}(A, b)$ is a naturally defined condition number for systems of constraints.

It might be useful for the reader to think of the constraints as being scaled so that $\|A\| \approx \|b\|$.

It is not difficult to verify that in the case of general systems of linear equations $Ax = b$ (i.e., in the case that $C = X$) where $\|b\| \leq \|A\|$, one has

$$\text{relcond}(A, b) = \text{relcond}(\sqrt{AA^T}), \quad (11)$$

precisely the principal parameter in (7); moreover, in the case of “square” systems of linear equations (i.e., when, in addition, $X = Y$), the values in (11) equal $\|A\| \|A^{-1}\|$, the traditional condition number.

1.6

We make minor assumptions regarding the cone C so that now-standard barrier method theory is readily applicable.

We assume that the interior C° of the cone C is dense in C . Moreover, in the terminology of [5] and [6, §1], we assume C° to be the domain of a functional $\phi \in \mathcal{F}'(K)$ where $K \geq 0$; in the terminology of Nesterov and Nemirovskii[3], “ $\phi \in \mathcal{F}'(K)$ ” essentially means that ϕ is a self-concordant barrier functional whose barrier parameter is of magnitude $O(K^2)$.

For example, if $X := \mathbb{R}^n$ and C is the cone (10) then one may let

$$\phi_x := -12 \sum_{j=1}^J \ln x_j \text{ and } K = \sqrt{12J}. \quad (12)$$

The factor “12” is not crucial.⁴ What is important is the “order” of K , i.e., $K = O(\sqrt{J})$.

We write “ $x > 0$ ” to mean $x \in C^\circ$.

Let $\tilde{x}$ denote the point which minimizes the functional

$$x \mapsto \phi_x - 12 \ln(1 - \|x\|^2),$$

a functional whose domain is

$$\{x; x > 0 \text{ and } \|x\| < 1\}, \quad (13)$$

i.e., $\tilde{x}$ is the “analytic center” of the region (13). For simplicity, we assume that $\tilde{x}$ is known; this assumption can be relaxed at the price of increased technicalities.⁵

⁴The factor “12” is simply chosen so that the theory described in [5] and [6] is immediately applicable and hence we can avoid proving standard details concerning the barrier method; all of the details we need are undoubtedly contained somewhere in the fundamental work of Nesterov and Nemirovskii[3] and elsewhere. The theory in [5] and [6] involves a natural normalization constant (namely, “1”) which leads to the rather unnatural factor “12” above. Instead, one can incorporate the unnatural constant “12” into the theory, thus eliminating the need for the factor “12” above; but since the theory applies beyond the natural log function, it seems preferable to have the natural constant “1” in the theory. However, all of this discussion is moot, for in the present paper we hide all constants in big-O’s; consequently, the normalization is actually irrelevant—all of our results are in fact true if the above factor “12” is replaced by “1.” (Likewise with *all* of the “12’s” one encounters in this paper.) Not wanting to go through the labor of making this claim rigorous, we leave the factor “12” so that the theory described in [5] and [6] is indeed immediately applicable.

⁵For example, if one only wishes only to assume that *some* point

$$\tilde{x} \in \{x \in C^\circ; \|x\| < 1\}$$

For example, if $X := \mathbb{R}^n$ is equipped with the Euclidean inner product, C is the cone (10) and ϕ is defined by (12) then $\check{x}$ is the vector with coordinates

$$\check{x}_j := \begin{cases} \sqrt{J/(J^2 + 2)} & \text{if } j \leq J \\ 0 & \text{if } j > J. \end{cases}$$

1.7

We now reformulate the constraints $Ax = b$, $x \geq 0$ so that they might be approximately solved by the barrier method. Our reformulation depends on two additional variables, s and t . Roughly, s is a “homogenization variable” used to transform the (relative interior of the) possibly unbounded feasible region for the constraints into a bounded set, and t is a “homotopy parameter.” Thus, the barrier method iterates will lie in the vector space

$$\mathcal{H} := X \times \mathbb{R} \times \mathbb{R}$$

which is a real Hilbert space (since X is a real Hilbert space) if we define an inner product by

$$\langle (x_1, s_1, t_1), (x_2, s_2, t_2) \rangle := \langle x_1, x_2 \rangle + s_1 s_2 + t_1 t_2.$$

We apply the barrier method to approximately solve the following optimization problem:

$$\begin{aligned} \inf \quad & t \\ \text{s.t.} \quad & \bar{A}(x, s, t) = 0 \\ & x > 0 \\ & \|x\| < 1 \\ & 0 < s < 1 \\ & 0 < t < 2 \end{aligned} \tag{14}$$

where $\bar{A} : \mathcal{H} \rightarrow Y$ is the continuous linear operator defined by

$$\bar{A}(x, s, t) := Ax - sb - t(A\check{x} - \tfrac{1}{2}b). \tag{15}$$

To understand the relevance of the optimization problem (14), observe that if $(x, s, 0)$ is the limit of a sequence of points which are feasible for the problem, and if $s > 0$, then $\frac{1}{s}x$ satisfies the original constraints to be solved, i.e., satisfies $Ax = b$, $x \geq 0$.

Let $\text{val}_{\inf}$ denote the optimal value of (14) and let $\text{val}_{\sup}$ denote the optimal value of the problem obtained by replacing “inf” in the objective of (14) with “sup.” Observe that the point

$$w_0 := (\check{x}, \tfrac{1}{2}, 1)$$

is known then, in the terminology of [6, §1], one can develop our results in terms of the quantity $\text{sym}(\check{x}, C(1))$.

satisfies the constraints of (14); hence,

$$1 \leq \text{val}_{\text{sup}} \leq 2.$$

Moreover, observe that the feasible region for (14) is closed under scaling by positive constants less than unity (i.e., if (x, s, t) satisfies the constraints and $0 < r < 1$ then (rx, rs, rt) satisfies the constraints); consequently,

$$\text{val}_{\text{inf}} = 0.$$

In applying the barrier method to approximately solve (14), we rely on the functional

$$f_{x,s,t} := \phi_x - 12 [\ln(1 - \|x\|^2) + \ln(s) + \ln(1 - s) + \ln(t) + \ln(2 - t)], \quad (16)$$

whose domain is

$$D_f := \{(x, s, t); x > 0, \|x\| < 1, 0 < s < 1 \text{ and } 0 < t < 2\}.$$

In the terminology of [5] and [6] we have that $f \in \mathcal{F}(\sqrt{K^2 + 72})$, that is, $f \in \mathcal{F}(\bar{K})$ where

$$\bar{K} = K + O(1);$$

in the terminology of Nesterov and Nemirovskii[3], “ $f \in \mathcal{F}(\bar{K})$ ” essentially means that f is a non-degenerate self-concordant barrier functional whose barrier parameter is of magnitude $O(\bar{K}^2)$.

Defining $c := (\vec{0}, 0, 1) \in \mathcal{H}$, we find it convenient to rewrite the optimization problem (14) in the following general format:

$$\begin{aligned} \inf \quad & \langle c, w \rangle \\ \text{s.t.} \quad & \bar{A}w = 0 \\ & w \in D_f. \end{aligned} \quad (17)$$

1.8

Due to our assumptions concerning the point $\check{x}$, it is not difficult to show that $w_0 := (\check{x}, \frac{1}{2}, 1)$ minimizes f . Consequently, as arguments which are well-known to the ipm community show, if $\mu_1 = \mu$ is a constant satisfying

$$0 < \mu = O\left(\frac{1}{\text{val}_{\text{sup}} - \text{val}_{\text{inf}}}\right) \quad (18)$$

then w_0 is “close” to the unique optimal solution $z_{\mu_1} = z_\mu$ of the following linearly constrained optimization problem:

$$\begin{aligned} \min_w \quad & \mu \langle c, w \rangle + f_w \\ \text{s.t.} \quad & \bar{A}w = 0. \end{aligned} \quad (19)$$

Being “close” implies that one iteration of Newton’s method, initiated at w_0 , will yield a point $w_1 \in D_f$ which is “closer” than w_0 to z_{μ_1} . The barrier method computes w_1 and then replaces the constant μ_1 by a larger value μ_2 . One iteration of Newton’s method, initiated at w_1 , then yields w_2 . And so on, with an increasing sequence of constants μ_i . If all of the constants μ_i are chosen to satisfy

$$\frac{\mu_{i+1}}{\mu_i} - 1 = O\left(\frac{1}{\bar{K}}\right) \quad (20)$$

then the barrier method “stays on track,” i.e., w_i will be “close” to $z_{\mu_{i+1}}$ for all i . Moreover, if all of the constants μ_i are chosen to satisfy

$$\frac{\mu_{i+1}}{\mu_i} - 1 = \Theta\left(\frac{1}{\bar{K}}\right), \quad (21)$$

then for each ϵ satisfying $0 < \epsilon \leq 1$,

$$i = \Omega\left(\bar{K} \log_2(\bar{K} + 1 + \frac{1}{\epsilon})\right) \Rightarrow \frac{\langle c, w_i \rangle - \text{val}_{\inf}}{\text{val}_{\sup} - \text{val}_{\inf}} < \epsilon. \quad (22)$$

Letting $w_i := (x_i, s_i, t_i)$ and recalling that in our application, $\text{val}_{\sup} \geq 1$ and $\text{val}_{\inf} = 0$, (22) can be rewritten as

$$i = \Omega\left((K + 1) \log_2(K + 1 + \frac{1}{\epsilon})\right) \Rightarrow t_i < \epsilon. \quad (23)$$

“Close” is defined in terms of a particular norm. Precisely, for each $w \in D_f$, one defines an inner product on $\mathcal{H}$ by

$$\langle u, v \rangle_w := \langle u, H_w v \rangle$$

where H_w is the Hessian of f at w . In turn, this inner product gives rise to the norm

$$\|v\|_w := \langle v, v \rangle_w^{1/2}.$$

To say that “ w is ‘close’ to z_μ ” is to say that $\|z_\mu - w\|_w$ is less than a particular universal positive constant, say, κ . The particular constant κ is not particularly important and, in fact, differs from paper to paper; that the constant is universal (i.e., problem instance independent) is important. (Moreover, it is also important that the theory allows one to replace the universal constant κ defining “close” with any smaller positive constant *if* one alters the other universal constants occurring in the theory appropriately; e.g., if one decreases κ then one must also decrease the universal constants hidden in the “big-O” notation of (18), (20) and (21), and increase the universal constant hidden in (22).)

1.9

As is well-known, to ensure that the barrier method “stays on track,” it suffices for the Newton steps to be accurately approximated rather than computed exactly. More precisely, letting n_{w_i} denote the exact *Newton step* at w_i for the linearly constrained optimization problem (19) when $\mu = \mu_{i+1}$, it suffices to compute an approximation $n_{w_i} + \Delta n_{w_i}$ where $\|\Delta n_{w_i}\|_{w_i}$ is “small”; one then lets the next iterate be

$$w_{i+1} := w_i + n_{w_i} + \Delta n_{w_i}.$$

If the iterate w_i is “close” to z_{μ_i} and if $\|\Delta n_{w_i}\|_{w_i}$ is indeed “small” then the iterate w_{i+1} will be “close” to $z_{\mu_{i+1}}$ (assuming (20)).

To say that “ $\|\Delta n_{w_i}\|_{w_i}$ is ‘small’ ” is to say that $\|\Delta n_{w_i}\|_{w_i}$ is less than a particular universal positive constant, say, γ . The determination of an appropriate universal constant γ is intertwined with the determination of an appropriate universal constant κ (defining “close”) and the determination of the appropriate universal constant hidden in (20). Although somewhat involved, such analysis is now commonplace; c.f., [3], [5].

Now, for general problems of the form (17), when one only accurately approximates the Newton steps in computing the sequence $\{w_i\}$, the implication (22) need not be valid; it can happen that w is “close” to z_μ and $\langle c, z_\mu \rangle \approx \text{val}_{\inf}$ but $\langle c, w \rangle$ is nowhere near $\text{val}_{\inf}$ if $\bar{A}w \neq 0$. However, if the optimal value $\text{val}_{\inf}$ of a general problem of the form (17) satisfies

$$\text{val}_{\inf} = \inf\{\langle c, w \rangle; w \in D_f\}, \quad (24)$$

then indeed the implication (22) is valid even when one only accurately approximates the Newton steps rather than computing them exactly; this is not difficult to prove, c.f., [4].

Recalling that $\text{val}_{\inf} = 0$ for the problem (14) in which we are interested, and noting the constraint $t > 0$ in defining the domain D_f for that problem, we see that for us the identity (24) is indeed fulfilled; hence, (22) (equivalently, (23)) is valid even when the Newton steps are accurately approximated rather than computed exactly.

1.10

Since the theory only requires that the Newton steps be accurately approximated, it is natural to investigate the efficiency of cgm in computing accurate approximations.

As is explained in §2, if $w \in D_f$ then the Newton step n_w for the linearly constrained optimization problem (19) is

$$n_w := -H_w^{-1}(\mu c + g_w + \bar{A}^T y)$$

where y solves

$$\bar{A}H_w^{-1}\bar{A}^T y = \bar{A}(w - H_w^{-1}(\mu c + g_w)); \quad (25)$$

here, g_w denotes the gradient of f at w and H_w denotes the Hessian. The operator H_w^{-1} is easily computed for the cones C of common interest, e.g., (10); in fact, H_w , and hence H_w^{-1} , is typically a diagonal matrix. Thus, assuming H_w^{-1} to be readily available, the main work in computing the Newton step lies in solving the linear equations (25); it is to these equations that one applies cgm.

Of course when one applies cgm to the linear equations (25) one does not form the “matrix product” $\bar{A}H_w^{-1}\bar{A}^T$ but instead limits oneself to “matrix-vector multiplications” so as to keep the number of arithmetic operations low and so as to take full advantage of sparsity in $\bar{A}$ and H_w^{-1} ; note that any sparsity in A and b is reflected in $\bar{A}$ (by choice of $\bar{A}$).

Our discussion of cgm indicates that the efficiency of cgm in approximately solving (25) is largely governed by the condition number

$$\text{relcond}(\sqrt{\bar{A}H_w^{-1}\bar{A}^T}). \quad (26)$$

In order to measure the efficiency of cgm in terms that are relevant to the original constraints $Ax = b$, $x \geq 0$, it is thus natural to attempt to relate the condition number (26) to the condition number $\text{relcond}(A, b)$. The following theorem and proposition yield a corollary which provides such a relation.

Recall that the *central trajectory* of the optimization problem (14) is the path

$$\text{CentTraj} := \{z_\mu; \mu \geq 0\}.$$

For $w \in D_f$, define

$$\text{dist}(w, \text{CentTraj}) := \inf\{\|z - w\|_w; z \in \text{CentTraj}\}.$$

As with all of our results, the universal constants hidden in the big-O notation of the following theorem are independent of X , Y , C , A and b , i.e, they are truly universal.

Theorem 1.1 *Assume $\text{relcond}(A, b) < \infty$. If $w \in D_f$ satisfies*

$$\text{dist}(w, \text{CentTraj}) = O(1)$$

then

$$\|\sqrt{(\bar{A}H_w^{-1}\bar{A}^T)^{-1}}\| \leq \frac{O(K^2 + 1)}{\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}}.$$

Proof: See §5. □

Proposition 1.2 *If $w \in D_f$ then*

$$\|\sqrt{\bar{A}H_w^{-1}\bar{A}^T}\| = O(\|(A, b)\|).$$

Proof: See §3. □

Corollary 1.3 *Assume $\text{relcond}(A, b) < \infty$. If $w \in D_f$ satisfies*

$$\text{dist}(w, \text{CentTraj}) = O(1)$$

then

$$\text{relcond}(\sqrt{\bar{A}H_w^{-1}\bar{A}^T}) \leq O(K^2 + 1) \text{relcond}(A, b).$$

Proof: Immediate from Theorem 1.1 and Proposition 1.2. □

In the author's opinion, Corollary 1.3 is one of the two most interesting results in this paper, both of which have ramifications beyond the analysis of cgm. Our discussion of cgm serves primarily as a showcase for the two results.

It is worth noting that Corollary 1.3 cannot be established by bounding the condition numbers of, say, $\sqrt{\bar{A}\bar{A}^T}$ and $\sqrt{H_w^{-1}}$ separately because

$$\lim_{\mu \rightarrow \infty} \text{relcond}(\sqrt{H_w^{-1}}) = \infty.$$

One of the main tools in proving Theorem 1.1 is, not surprisingly, the Hahn-Banach Theorem.

1.11

It is not immediately evident how Corollary 1.3 provides an upper bound on the number of cgm iterations which are sufficient to compute an appropriately accurate approximation $\tilde{y}$ to the exact solution of the linear equations (25); that is, a vector $\tilde{y}$ such that if

$$\Delta n_w := -H_w^{-1}(\mu c + g_w + \bar{A}^T \tilde{y}) - n_w$$

then $\|\Delta n_w\|_w < \gamma$ (i.e., $\|\Delta n_w\|_w$ is “small”) and hence

$$-H_w^{-1}(\mu c + g_w + \bar{A}^T \tilde{y})$$

is an appropriately accurate approximation to the Newton step n_w . In §2 we derive an upper bound on the number of cgm iterations which are indeed sufficient, letting the initial cgm iterate be $y_0 := 0$. Assuming $w \in D_f$ satisfies $\|z_\mu - w\|_w = O(1)$ (i.e., w is “close” to z_μ), we use Corollary 1.3 to show that

$$O((K^2 + 1) \text{relcond}(A, b) \log_2(K + 2)) \quad (27)$$

cgm iterations suffice.

The factor “ $(K^2 + 1) \log_2(K + 2)$ ” appearing in the iteration bound (27) certainly limits the interest of that bound. In particular, if Y is finite-dimensional, i.e., $Y = \mathbb{R}^m$, then the bound is of no interest if $K^2 \geq m$ (or $\text{relcond}(A, b) \geq m$) because the linear equations (25) will involve only m variables and hence will be solved exactly by cgm within m iterations. The bound is only of interest if K^2 is small relative to m . For example, the bound is of interest if $X := \mathbb{R}^n$ is equipped with the Euclidean inner product, C is the cone (10) where $J \ll m$, and ϕ is defined by (12); then $K^2 = 12J \ll m$.

1.12

Of course the efficient approximation of the Newton steps is only a means-to-an-end, the end-goal being the computation of an accurate approximation to a solution of the constraints $Ax = b$, $x \geq 0$. One would like to know something about the overall amount of work which is sufficient to achieve the end-goal, where the overall amount of work is measured in terms of $\text{relcond}(A, b)$. Towards providing a bound on the overall amount of work, we describe an approach to computing an accurate approximation to a solution of the constraints $Ax = b$, $x \geq 0$ when given a barrier method iterate $(x, s, t) \in D_f$.

Let $f' : X \times \mathbb{R} \rightarrow \mathbb{R}$ denote the functional

$$f'_{x,s} := \phi_x - 12 [\ln(1 - \|x\|^2) + \ln(s) + \ln(1 - s)],$$

with domain

$$D_{f'} := \{(x, s); x > 0, \|x\| < 1, \text{ and } 0 < s < 1\}.$$

Note that f' is identical to f except that in the definition of f' there is no reference to the “homotopy parameter” t . If $(x, s, t) \in D_f$ then $(x, s) \in D_{f'}$.

We let H'_w denote the Hessian of f' at $w = (x, s) \in D_{f'}$.

In the terminology of [5] and [6] we have that $f' \in \mathcal{F}(\sqrt{K^2 + 48})$, that is, $f' \in \mathcal{F}(K')$ where

$$K' = K + O(1);$$

in the terminology of Nesterov and Nemirovskii[3], “ $f' \in \mathcal{F}(K')$ ” essentially means that f' is a non-degenerate self-concordant barrier functional whose barrier parameter is of magnitude $O((K')^2)$.

Let $\hat{A} : X \times \mathbb{R} \rightarrow Y$ denote the linear operator defined by

$$\hat{A}(x, s) := Ax - sb.$$

Given $w = (x, s) \in D_{f'}$ and assuming $A\hat{A}^T$ is invertible, one approach to solving the linear equations $Ax = b$ is to apply cgm to the linear equations (in the variable y)

$$\hat{A}(H'_w)^{-1}\hat{A}^T y = -\hat{A}w,$$

thus obtaining a convergent sequence $y_0 := 0, y_1, \dots$. Defining

$$w_i = (x_i, s_i) := w + (H'_w)^{-1}\hat{A}y_i, \quad (28)$$

the sequence $\{w_i\}$ then converges to a point $w_\infty = (x_\infty, s_\infty)$ satisfying $\hat{A}w_\infty = 0$. Under “certain conditions” one will have $s_i, s_\infty > 0$ and hence the sequence $\{\frac{1}{s_i}x_i\}$ will then converge to $\frac{1}{s_\infty}x_\infty$ which will be a solution of $Ax = b$. We now describe such “certain conditions” which are relevant to solving $Ax = b, x \geq 0$.

If $\text{relcond}(A, b) < \infty$ then the constraints $Ax = b, x > 0$ are consistent; c.f., [6, Corollary 3.7]. Consequently, if $\text{relcond}(A, b) < \infty$ then the constraints

$$\begin{aligned} \hat{A}w &= 0 \\ w &\in D_{f'} \end{aligned} \quad (29)$$

are consistent. Since $D_{f'}$ is bounded and $f' \in \mathcal{F}(K')$, it follows that the convex optimization problem

$$\begin{aligned} \inf \quad & f'_w \\ \text{s.t.} \quad & \hat{A}w = 0 \\ & w \in D_{f'} \end{aligned}$$

has an optimal solution z' , i.e., the constraints (29) have an “analytic center.”

Now, in the same manner that the Hessian H_w of f at $w = (x, s, t) \in D_f$ induces the inner product $\langle \cdot, \cdot \rangle_w$ and norm $\|\cdot\|_w$ on $X \times \mathbb{R} \times \mathbb{R}$, the Hessian H'_w of f' at $w = (x, s) \in D_{f'}$ induces an inner product $\langle \cdot, \cdot \rangle'_w$ and norm $\|\cdot\|'_w$ on $X \times \mathbb{R}$.

Theorem 1.4 *Assume $\text{relcond}(A, b) < \infty$. If $w \in D_{f'}$ satisfies*

$$\|z' - w\|'_w = O(1)$$

then

$$\|(\sqrt{(\hat{A}(H'_w)^{-1}\hat{A}^T)^{-1}})\| \leq \frac{O(K^2 + 1)}{\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}}. \quad (30)$$

Proof: See §4. □

Theorem 1.5 Assume $\text{relcond}(A, b) < \infty$. If $w \in D_{f'}$ and $z \in X \times \mathbb{R}$ satisfy

$$\hat{A}z = 0 \text{ and } \|z - w\|'_w = O(1)$$

then

$$\|\sqrt{(\hat{A}(H'_w)^{-1}\hat{A}^T)^{-1}}\| \geq \frac{\Omega(1)}{\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}}. \quad (31)$$

Proof: See §4. □

Proposition 1.6 If $w \in D_{f'}$ then

$$\|\sqrt{\hat{A}(H'_w)^{-1}\hat{A}^T}\| = O(\|(A, b)\|).$$

Proof: See §3. □

Corollary 1.7 If $\text{relcond}(A, b) < \infty$ and $w \in D_{f'}$ satisfies

$$\|z' - w\|'_w = O(1)$$

then

$$\text{relcond}(\sqrt{\hat{A}(H'_w)^{-1}\hat{A}^T}) = O(K^2 + 1) \text{relcond}(A, b),$$

and

$$\begin{aligned} & \Omega(1) \text{relcond}(A, b) \\ & \leq \|(A, b)\| \|\sqrt{\hat{A}(H'_w)^{-1}\hat{A}^T}\| \\ & \leq O(K^2 + 1) \text{relcond}(A, b). \end{aligned}$$

Proof: Immediate from Theorems 1.4, 1.5 and Proposition 1.6. □

The author considers Corollary 1.7 to be the other of the two most interesting results in this paper.

Corollary 1.7 is used to prove the following theorem.

Theorem 1.8 Assume $\text{relcond}(A, b) < \infty$ and $w = (x, s) \in D_{f'}$ satisfies $\|z' - w\|'_w = O(1)$. Then the sequence $\{w_i\}$ defined by (28) and the limit point w_∞ satisfy $w_i, w_\infty \in D_{f'}$; in particular, $\frac{1}{s_i}x_i, \frac{1}{s_\infty}x_\infty > 0$ and $A(\frac{1}{s_\infty}x_\infty) = b$.

Moreover, for all ϵ satisfying $0 < \epsilon \leq 1$,

$$i = \Omega((K^2 + 1) \text{relcond}(A, b) \log_2 \frac{K+2}{\epsilon}) \Rightarrow \frac{\|\frac{1}{s_\infty}x_\infty - \frac{1}{s_i}x_i\|}{\max\{\|\frac{1}{s_\infty}x_\infty\|, 1\}} \leq \epsilon.$$

Also,

$$\begin{aligned} \|\frac{1}{s_\infty}x_\infty\| &= O(K^2 + 1)(1 + \inf\{\|x\|; Ax = b \text{ and } x \geq 0\}) \\ &\leq O(K^2 + 1) \text{relcond}(A, b), \end{aligned}$$

i.e., $\frac{1}{s_\infty}x_\infty$ is a “not-too-far-from-minimal” norm solution for the constraints $Ax = b, x \geq 0$.

Proof: See §6. □

We would like to be in position to apply Theorem 1.8 to (x, s) where (x, s, t) is a point computed by the barrier method. To this end, we have the following theorem.

Theorem 1.9 Assume $\text{relcond}(A, b) < \infty$. If $w = (x, s, t) \in D_f$ satisfies

$$\text{dist}(w, \text{CentTraj}) = O(1) \text{ and } t \leq \frac{O(1)}{(K^2 + 1) \text{relcond}(A, b)}$$

then

$$\|z' - (x, s)\|'_{x,s} = O(\text{dist}(w, \text{CentTraj}) + t(K^2 + 1) \text{relcond}(A, b)).$$

Proof: See §7. □

Theorems 1.8, 1.9 and (23) readily yield bounds on the number of barrier method iterations sufficient to obtain (x, s, t) for which (x, s) yields a sequence $\{\frac{1}{s_i}x_i\}$ converging to a solution $\frac{1}{s_\infty}x_\infty$ of $Ax = b, x > 0$; precisely,

$$O((K + 1) \log_2(K + 1 + \text{relcond}(A, b)))$$

barrier method iterations suffice. Hence, if the Newton step at each barrier method iteration is approximated using cgm then (27) and Theorem 1.8 imply that the total number of cgm iterations (including those used in constructing the sequence $\{\frac{1}{s_i}x_i\}$) which are sufficient to obtain a point $\frac{1}{s_i}x_i$ which accurately approximates a solution of $Ax = b, x > 0$ is roughly

$$O((K^3 + 1) \text{relcond}(A, b) \log_2(K + 2) \log_2(K + 1 + \text{relcond}(A, b))). \quad (32)$$

It is instructive to consider the iteration bound (32) in the case that $C = X$, that is, in the case of linear equations $Ax = b$, assuming $\|b\| \leq \|A\|$. (Of course one would never use the barrier method to solve linear equations.) An appropriate functional ϕ is then the functional whose value is identically zero; for this functional, $K = 0$. Hence, the bound (32) becomes

$$O(\text{relcond}(A, b) \log(1 + \text{relcond}(A, b))),$$

which, upon substitution of (11), gives

$$O\left(\text{relcond}(\sqrt{AA^T}) \log(1 + \text{relcond}(\sqrt{AA^T}))\right).$$

Up to a logarithmic factor, this is consistent with the well-known bound (7); the logarithmic factor is precisely the number of barrier method iterations.

If Y is finite-dimensional, i.e., $Y = \mathbb{R}^m$, then since cgm solves each of the systems of linear equations within m iterations, the total number of cgm iterations can be bounded by

$$O((K + 1)m \log_2(K + 1 + \text{relcond}(A, b))).$$

If $(K^2 + 1)\text{relcond}(A, b) \ll m$ then (32) provides a better bound.

2 The Newton Step

2.1

In this section we explain why, assuming $w \in D_f$, the Newton step n_w for the linearly constrained optimization problem

$$\begin{aligned} \min_w \quad & \mu \langle c, w \rangle + f_w \\ \text{s.t.} \quad & \bar{A}w = 0 \end{aligned} \quad (33)$$

satisfies

$$n_w = -H_w^{-1}(\mu c + g_w + \bar{A}^T y) \quad (34)$$

where y solves

$$\bar{A}H_w^{-1}\bar{A}^T y = \bar{A}(w - H_w^{-1}(\mu c + g_w)). \quad (35)$$

Our explanation follows a rather unconventional route, the advantage of this route being that once one understands it, one also understands how cgm fits into the customary analysis of the barrier method. In particular, one understands how to proceed to bound the number of cgm iterations which are sufficient to obtain an appropriately accurate approximation $\tilde{y}$ to the exact solution of the linear equations (35); that is, a vector $\tilde{y}$ such that if

$$\Delta n_w := -H_w^{-1}(\mu c + g_w + \bar{A}^T \tilde{y}) - n_w$$

then $\|\Delta n_w\|_w < \gamma$ (i.e., $\|\Delta n_w\|_w$ is “small”) and hence

$$-H_w^{-1}(\mu c + g_w + \bar{A}^T \tilde{y})$$

is an appropriately accurate approximation to the Newton step n_w .

Our explanation is somewhat unconventional in that it focuses on the inner product $\langle \cdot, \cdot \rangle_w$ rather than the original inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{H}$. Since from a theoretical perspective, the natural way to measure the proximity of w to z_μ is in terms of the norm $\|\cdot\|_w$ which arises from the inner product $\langle \cdot, \cdot \rangle_w$, it is not surprising that an explanation which focuses on the inner product $\langle \cdot, \cdot \rangle_w$ provides deeper insight than an explanation which focuses on the inner product $\langle \cdot, \cdot \rangle$.

2.2

In terms of the inner product $\langle \cdot, \cdot \rangle_w$, the linearly constrained optimization problem (33) becomes

$$\begin{aligned} \min_v \quad & \mu \langle \tilde{c}, v \rangle_w + f_v \\ \text{s.t.} \quad & \bar{A}v = 0, \end{aligned} \quad (36)$$

where $\tilde{c} := H_w^{-1}c$.

Now, H_v denotes the Hessian of f at $v \in D_f$ with respect to the original inner product $\langle \cdot, \cdot \rangle$. However, the Hessian of f at v with respect to the inner product $\langle \cdot, \cdot \rangle_w$ is

$$\tilde{H}_v := H_w^{-1} H_v;$$

in particular, $\tilde{H}_w = I$, the identity operator. Similarly, the gradient of f at v with respect to the inner product $\langle \cdot, \cdot \rangle_w$ is

$$\tilde{g}_v := H_w^{-1} g_v;$$

in particular, $\tilde{g}_w := H_w^{-1} g_w$.

Just as the choice of inner product affects gradients and Hessians, it also affects adjoint operators. In particular, whereas the adjoint of $\bar{A}$ is denoted $\bar{A}^T$ when $\mathcal{H}$ is equipped with the original inner product $\langle \cdot, \cdot \rangle$, if $\mathcal{H}$ is equipped with the inner product $\langle \cdot, \cdot \rangle_w$ then the adjoint is

$$\bar{A}^{\tilde{T}} := H_w^{-1} \bar{A}^T.$$

(We never vary the inner product on Y .)

Finally, of course the particular inner product affects what is meant by (orthogonal) projection.

Throughout the remainder of §2, whenever we speak of a gradient, a Hessian, an adjoint, or (orthogonal) projection, we mean with respect to the inner product $\langle \cdot, \cdot \rangle_w$. Similarly, whenever a norm on $\mathcal{H}$ is referred to, it is the norm $\| \cdot \|_w$.

2.3

Let $\tilde{P}$ denote the operator projecting $\mathcal{H}$ onto the nullspace of $\bar{A}$ and let $\tilde{P}_\perp$ denote the operator projecting $\mathcal{H}$ onto the orthogonal complement of the nullspace, that is, onto the range space of $\bar{A}^{\tilde{T}}$.

Since the gradient of the objective functional of (36) at w is $\mu\tilde{c} + \tilde{g}_w$ and since the Hessian is $\tilde{H}_w = I$, the Newton model functional at w for the objective functional of (36) is

$$v \mapsto (\mu\langle \tilde{c}, w \rangle_w + f_w) + \langle \mu\tilde{c} + \tilde{g}_w, v - w \rangle_w + \frac{1}{2} \|v - w\|_w^2.$$

Consequently, when the objective functional in (36) is replaced by the Newton model function, the optimal solution of the resulting linearly constrained optimization problem is

$$\begin{aligned} w + n_w &:= \tilde{P}(w - \mu\tilde{c} - \tilde{g}_w) \\ &= w - \mu\tilde{c} - \tilde{g}_w - \tilde{P}_\perp(w - \mu\tilde{c} - \tilde{g}_w). \end{aligned}$$

Hence,

$$n_w = -\mu\tilde{c} - \tilde{g}_w - w^* \tag{37}$$

where

$$w^* := \tilde{P}_\perp(w - \mu\tilde{c} - \tilde{g}_w). \quad (38)$$

Observe that w^* is the minimal norm solution to the follow system of linear equations in the variable v :

$$\bar{A}v = \bar{A}(w - \mu\tilde{c} - \tilde{g}_w). \quad (39)$$

Consequently,

$$w^* = \bar{A}^{\tilde{T}}y \quad (40)$$

where y solves

$$\bar{A}\bar{A}^{\tilde{T}}y = \bar{A}(w - \mu\tilde{c} - \tilde{g}_w). \quad (41)$$

The linear equations (41) are exactly the same as the linear equations (35) and the identity (37) is exactly the same as the identity (34), i.e., we have now derived the linear equations underlying Newton's method. Moreover, we are in position to understand a bit about the behavior of cgm when applied to the linear equations (35) (or, equivalently, the linear equations (41)) in order to approximate n_w .

2.4

Recall that for the barrier method theory, one wants to approximate n_w to within error γ (as measured in the norm $\|\cdot\|_w$) where γ is a certain universal positive constant. To obtain such an approximation, (37) shows that it suffices to approximate w^* to within error γ and then subtract the resulting approximation from $-\mu\tilde{c} - \tilde{g}_w$.

As we observed, w^* is the minimal norm solution to the linear equations (39). Our discussion of cgm in §1.3 showed that the minimal norm solution for a system of linear equations can be approximated using cgm. Translating that discussion to the present context, to solve (39) one applies cgm to solve (41), thus obtaining a sequence $\{y_i\}$; the sequence $\{\bar{A}^{\tilde{T}}y_i\}$ then converges to the minimal norm solution. In other words, when one thinks in terms of the inner product $\langle \cdot, \cdot \rangle_w$, our application of cgm is exactly as prescribed in §1.3. Moreover, in terms of the present context, (7) is rewritten as

$$i = \Omega \left(\text{relcond}(\sqrt{\bar{A}\bar{A}^{\tilde{T}}}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \|w^* - \bar{A}^{\tilde{T}}y_i\|_w < \epsilon \|w^* - \bar{A}^{\tilde{T}}y_0\|_w. \quad (42)$$

If we take our initial iterate as $y_0 := 0$, (42) yields

$$i = \Omega \left(\text{relcond}(\sqrt{\bar{A}\bar{A}^{\tilde{T}}}) \log_2 \left(2 + \frac{\|w^*\|_w}{\gamma} \right) \right) \Rightarrow \|w^* - \bar{A}^{\tilde{T}}y_i\|_w < \gamma. \quad (43)$$

To prove the claimed upper bound (27) on the number of cgm iterations which are sufficient to compute an appropriately accurate approximation to the Newton step during an iteration of the barrier method, (43) shows it suffices to prove that if

$$\|z_\mu - w\|_w = O(1)$$

then

$$\text{relcond}(\sqrt{\bar{A}\bar{A}^T}) = O((K^2 + 1) \text{relcond}(A, b)), \quad (44)$$

and

$$\|w^*\|_w = O(K^2 + 1). \quad (45)$$

The relation (44) is precisely the content of Corollary 1.3. The relation (45) is established via now-standard arguments; the proof of the following proposition provides an outline of those arguments.

Proposition 2.1 *If $\|z_\mu - w\|_w = O(1)$ and w^* is the minimal norm solution of the linear equations (39) then*

$$\|w^*\|_w = O(K^2 + 1).$$

Proof: We have

$$\begin{aligned} \|w^*\|_w &= \|\tilde{P}_\perp(w - \mu\tilde{c} - \tilde{g}_w)\|_w \\ &\leq \|\tilde{P}_\perp w\|_w + \mu\|\tilde{c}\|_w + \|\tilde{g}_w\|_w \\ &\leq \|z_\mu - w\|_w + \mu\|\tilde{c}\|_w + O(\bar{K}) \end{aligned}$$

and thus, assuming $\|z_\mu - w\|_w = O(1)$, it suffices to show that

$$\mu\|\tilde{c}\|_w = O(K^2 + 1). \quad (46)$$

It follows from the definitions of f (i.e., (16)), $\tilde{c} := H_w^{-1}c$ and $\tilde{c} := (\vec{0}, 0, 1)$ that if $w = (x, s, t)$ then

$$\|\tilde{c}\|_w = \frac{1}{\sqrt{12}} \left(\frac{1}{t^2} + \frac{1}{(2-t^2)} \right)^{-1/2}. \quad (47)$$

Also, since

$$\forall v, \|v\|_w = O(1) \Rightarrow w + v \in D_f$$

and

$$(\bar{x}, \bar{s}, \bar{t}) \in D_f \Rightarrow \bar{t} > 0$$

we have that if $z_\mu = (x_\mu, s_\mu, t_\mu)$ then

$$\|z_\mu - w\|_w = O(1) \Rightarrow t_\mu = \Theta(t). \quad (48)$$

The relations (47) and (48) imply that to prove (46) it suffices to show

$$t_\mu = O(\frac{1}{\mu} \bar{K}^2),$$

that is,

$$\langle c, z_\mu \rangle = O(\frac{1}{\mu} \bar{K}^2). \quad (49)$$

However, for general problems of the form

$$\begin{aligned} \inf_v \quad & \langle c, v \rangle \\ \text{s.t.} \quad & \bar{A}v = \bar{b} \\ & v \in D_f \end{aligned} \quad (50)$$

where $f \in \mathcal{F}(\bar{K})$, one has that

$$\langle c, z_\mu \rangle - \inf_{u \in D_f} \langle c, u \rangle = O(\frac{1}{\mu} \bar{K}^2)$$

(c.f., [5]). Since in our setting we have that

$$\inf_{u \in D_f} \langle c, u \rangle = 0,$$

the relation (49) is indeed valid, completing the proof. $\square$

We now have our cgm iteration bound (27).

3 Proofs of Propositions 1.2 and 1.6

Proof of Proposition 1.2: First recall that if $w \in D_f$ then

$$\forall v, \quad \|v - w\|_w = O(1) \Rightarrow v \in D_f$$

and hence, by definition of the norm $\|\cdot\|_w$,

$$\forall u \quad \|u\| = O(1) \Rightarrow w + \sqrt{H_w^{-1}}u \in D_f.$$

It follows that

$$\|\sqrt{H_w^{-1}}\| = O(\text{diam}(D_f))$$

where

$$\text{diam}(D_f) := \sup\{\|u - v\|; u, v \in D_f\}.$$

Since for the particular domain D_f of interest to us,

$$D_f \subseteq \{(x, s, t); \|x\| < 1, 0 < s < 1 \text{ and } 0 < t < 2\},$$

we thus have

$$\|\sqrt{H_w^{-1}}\| = O(1)$$

and hence

$$\|H_w^{-1}\| = O(1).$$

Consequently,

$$\begin{aligned} \|\sqrt{\bar{A}H_w^{-1}\bar{A}^T}\| &= \sqrt{\|\bar{A}H_w^{-1}\bar{A}^T\|} \\ &\leq \sqrt{\|\bar{A}\| \|H_w^{-1}\| \|\bar{A}^T\|} \\ &= O(\sqrt{\|\bar{A}\| \|\bar{A}^T\|}) \\ &\leq O(\|\bar{A}\|) \\ &\leq O(\|(A, b)\|), \end{aligned}$$

the final equality being a simple consequence of the definition of $\bar{A}$ (and the bound $\|\tilde{x}\| < 1$). $\square$

Proof of Proposition 1.6: Entirely analogous to the proof of Proposition 1.2 and thus left to the reader. $\square$

4 Proofs of Theorems 1.4 and 1.5

4.1

In proving Theorems 1.4 and 1.5 we rely on the inner product $\langle \cdot, \cdot \rangle'_w$ for which the adjoint of $\hat{A}$ is

$$\hat{A}^{T'} := (H'_w)^{-1} \hat{A}^T.$$

A system $Ax = b$, $x \geq 0$ and vector $\bar{b} \in Y$ give rise to the following constraints in the variables $(x, s) \in X \times \mathbb{R}$:

$$\begin{aligned} Ax &= sb + \bar{b} \\ x &> 0 \\ s &> 0 \\ \|x\| + s &< 1. \end{aligned} \tag{51}$$

The proofs of Theorems 1.4 and 1.5 depend heavily on the following theorem regarding the different systems (51) obtained by varying $\bar{b}$.

Theorem 4.1

$$\begin{aligned} &\inf\{ \|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent} \} \\ &= \inf\{ \|\bar{b}\|; \bar{b} \in Y \text{ and the constraints (51) are inconsistent} \}. \end{aligned}$$

Proof: This is a special case of [6, Theorem 3.5]; the main ingredient in the proof is the Hahn-Banach Theorem. $\square$

4.2

Proof of Theorem 1.5: We wish to prove that if $w \in D_{f'}$ and z satisfy

$$\hat{A}z = 0 \text{ and } \|z - w\|'_w = O(1) \tag{52}$$

then

$$\|\sqrt{(\hat{A}\hat{A}^{T'})^{-1}}\| \geq \frac{\Omega(1)}{\inf\{ \|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent} \}}.$$

Recalling the well-known (and easily proven) Hilbert-space relation (which is based solely on the fact that $\hat{A}^{T'}$ is the adjoint of $\hat{A}$)

$$\frac{1}{\|\sqrt{(\hat{A}\hat{A}^{T'})^{-1}}\|} = \inf\{ \|\bar{b}\|; \|v\|'_w < 1 \Rightarrow \hat{A}v \neq \bar{b} \},$$

we see it suffices to show that if w and z satisfy (52) then

$$\begin{aligned} & \inf\{\|\bar{b}\|; \|v\|'_w < 1 \Rightarrow \hat{A}v \neq \bar{b}\} \\ & \leq O(1) \inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}. \end{aligned} \quad (53)$$

In proving the inequality (53) we may assume that $w = z$. For if

$$\|z - w\|'_w = O(1)$$

then $z \in D_{f'}$ and for all v we have

$$\|v\|'_z = O(\|v\|'_w);$$

thus

$$\begin{aligned} & \inf\{\|\bar{b}\|; \|v\|'_w < 1 \Rightarrow \hat{A}v \neq \bar{b}\} \\ & = O(\inf\{\|\bar{b}\|; \|v\|'_z < 1 \Rightarrow \hat{A}v \neq \bar{b}\}). \end{aligned}$$

Consequently, if the inequality (53) is valid for $w = z$ then it is valid for all $w \in D_{f'}$ satisfying $\|z - w\|'_w = O(1)$ (but perhaps with a larger universal constant hidden in the big-O).

Since $\hat{A}z = 0$, when $w = z$ the quantity on the left of the inequality (53) equals

$$\inf\{\|\bar{b}\|; \|u - z\|'_z < 1 \Rightarrow \hat{A}u \neq \bar{b}\}. \quad (54)$$

Since

$$\|u - z\|'_z = O(1) \Rightarrow u \in D_{f'},$$

the quantity (54) is bounded above by

$$O(1) \inf\{\|\bar{b}\|; u \in D_{f'} \Rightarrow \hat{A}u \neq \bar{b}\}.$$

Consequently, to establish the inequality (53) when $w = z$ it suffices to show that

$$\begin{aligned} & \inf\{\|\bar{b}\|; u \in D_{f'} \Rightarrow \hat{A}u \neq \bar{b}\} \\ & \leq O(1) \inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}. \end{aligned} \quad (55)$$

For $\bar{b} \in Y$, consider the following constraints in the variables (x, s) :

$$\begin{aligned} Ax &= sb + \bar{b} \\ x &> 0 \\ s &> 0 \\ \|x\| + s &< 2. \end{aligned} \quad (56)$$

Note that each element of $D_{f'}$ satisfies the constraints $x > 0$, $s > 0$ and $\|x\| + s < 2$. Hence,

$$\inf\{\|\bar{b}\|; u \in D_{f'} \Rightarrow \hat{A}u \neq \bar{b}\}$$

$$\leq \inf\{\|\bar{b}\|; \bar{b} \in Y \text{ and the constraints (56) are inconsistent}\}. \quad (57)$$

Observe that the quantity on the right of the inequality (57) is exactly twice the quantity

$$\inf\{\|\bar{b}\|; \bar{b} \in Y \text{ and the constraints (51) are inconsistent}\}.$$

Hence, (57) and Theorem 4.1 give the desired inequality (55). $\square$

4.3

The proof of Theorem 1.4 depends on the following proposition.

Proposition 4.2 *Let $z' = (x', s')$. If $\bar{b} \in Y$ satisfies*

$$\forall v, \quad \|v\|'_{z'} < 1 \Rightarrow \hat{A}v \neq \bar{b}$$

and if $r \in \mathbb{R}$ satisfies $r = \Omega(K^2 + 1)$ then either

$$\hat{A}w \neq r\bar{b} \text{ for all } w \in D_{f'},$$

or

$$\hat{A}w \neq r\bar{b} \text{ for all } w \in D_{f'}.$$

Proof: See Subsection 4.4. $\square$

Proof of Theorem 1.4: We wish to prove that if $\|z' - w\|'_w = O(1)$ then

$$\|\sqrt{(\hat{A}\hat{A}^{T'})^{-1}}\| \leq \frac{O(K^2 + 1)}{\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}}.$$

Recalling the well-known (and easily proven) Hilbert-space relation (which is based solely on the fact that $\hat{A}^{T'}$ is the adjoint of $\hat{A}$)

$$\frac{1}{\|\sqrt{(\hat{A}\hat{A}^{T'})^{-1}}\|} = \inf\{\|\bar{b}\|; \|v\|'_w < 1 \Rightarrow \hat{A}v \neq \bar{b}\},$$

we see it suffices to show that if $\|z' - w\|'_w = O(1)$ then

$$\begin{aligned} & \inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\} \\ & \leq O(K^2 + 1) \inf\{\|\bar{b}\|; \|v\|'_w < 1 \Rightarrow \hat{A}v \neq \bar{b}\}. \end{aligned} \quad (58)$$

In proving the inequality (58) we may assume that $w = z'$. For if $\|z' - w\|'_w = O(1)$ then for all v we have $\|v\|'_w = O(\|v\|'_{z'})$ and hence

$$\inf\{\|\bar{b}\|; \|v\|'_{z'} < 1 \Rightarrow \hat{A}v \neq \bar{b}\}$$

$$= O(\inf\{\|\bar{b}\|; \|v\|'_w < 1 \Rightarrow \hat{A}v \neq \bar{b}\}).$$

Consequently, if the inequality (58) is valid for $w = z'$ then it is valid for all w satisfying $\|z' - w\|'_w = O(1)$ (but perhaps with a larger universal constant hidden in the big-O).

Proposition 4.2 implies that there exists $\hat{b} \in Y$ such that

$$\hat{A}w \neq \hat{b} \text{ for all } w \in D_{f'} \quad (59)$$

and

$$\|\hat{b}\| \leq O(K^2 + 1) \inf\{\|\bar{b}\|; \|v\|'_{z'} < 1 \Rightarrow \hat{A}v \neq \bar{b}\} \quad (60)$$

The definitions of $\hat{A}$ and $D_{f'}$, along with (59), imply that the following system of constraints in the variables $(x, s) \in X \times \mathbb{R}$ is inconsistent:

$$\begin{aligned} Ax &= sb + \hat{b} \\ x &> 0 \\ \|x\| &< 1 \\ 0 &< s < 1. \end{aligned}$$

Consequently, Theorem 4.1 gives

$$\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\} \leq \|\hat{b}\|. \quad (61)$$

Together, (60) and (61) yield the inequality (58) when $w = z'$, completing the proof. $\square$

4.4

Proof of Proposition 4.2: Letting g'_z , denote the gradient of f' at z' , consider the functional

$$f''_w := f'_w - 12 \ln(\langle g'_{z'}, w - z' \rangle + 12),$$

whose domain is the set

$$D_{f''} := \{w \in D_{f'}; \langle g'_{z'}, w - z' \rangle > -12\}. \quad (62)$$

Recalling that $f' \in \mathcal{F}(K')$ (where $K' = K + O(1)$), in the terminology of [5] and [6] we have $f'' \in \mathcal{F}(\sqrt{(K')^2 + 12})$; in the terminology of Nesterov and Nemirovskii[3], f'' is essentially a non-degenerate self-concordant barrier functional whose barrier parameter is of magnitude $O(K^2 + 1)$.

We let g''_w , H''_w , $\langle \cdot, \cdot \rangle''_w$ and $\|\cdot\|''_w$ denote, respectively, the gradient, the Hessian, the inner product and the norm at w with respect to the functional f'' .

We claim that to prove the proposition it suffices to prove the following statement (which in the sequel we refer to simply as “the Statement”): *If $\bar{b} \in Y$ satisfies*

$$\forall v, \|v\|''_{z'} < 1 \Rightarrow \hat{A}v \neq \bar{b}$$

and if $r \in \mathbb{R}$ satisfies $r = \Omega(K^2 + 1)$ then

$$\hat{A}w \neq r\bar{b} \text{ for all } w \in D_{f''}.$$

To understand why it suffices to prove the Statement, assume that $\bar{b} \in Y$ satisfies the assumption of the proposition, that is,

$$\forall v, \|v\|_{z'}' < 1 \Rightarrow \hat{A}v \neq \bar{b}.$$

Since $\|v\|_{z'}' \leq \|v\|_{z'}''$ for all v , we then have

$$\forall v, \|v\|_{z'}'' < 1 \Rightarrow \hat{A}v \neq \bar{b}.$$

Hence, if $r = \Omega(K^2 + 1)$ then according to the Statement,

$$\hat{A}w \neq r\bar{b} \text{ for all } w \in D_{f''}. \quad (63)$$

Now, since z' minimizes f' subject to the constraints $\hat{A}x = 0$, we have that $g_{z'}'$ lies in the range space of $\hat{A}^T$. Hence, for any $r \in \mathbb{R}$, either

$$\forall w, \hat{A}w = r\bar{b} \Rightarrow \langle g_{z'}', w - z' \rangle \geq 0 \quad (64)$$

or

$$\forall w, \hat{A}w = -r\bar{b} \Rightarrow \langle g_{z'}', w - z' \rangle \geq 0. \quad (65)$$

If (64) is valid then (62) and (63) imply

$$\hat{A}w \neq r\bar{b} \text{ for all } w \in D_{f'}.$$

Similarly, if (65) is valid then (62) and (63) imply

$$\hat{A}w \neq -r\bar{b} \text{ for all } w \in D_{f'}.$$

Consequently, the proposition does indeed follow from the Statement.

To prove the Statement, first observe that $g_{z'}'' = 0$. Hence z' minimizes the strictly convex functional f'' , i.e., z' is the “analytic center” of the domain $D_{f''}$. Thus, since $f'' \in \mathcal{F}((K')^2 + 12)$ and $K' = K + O(1)$, we have

$$w \in D_{f''} \Rightarrow \|w - z'\|_{z'}'' = O(K^2 + 1). \quad (66)$$

Assume that $\bar{b}$ satisfies the assumption of the Statement, that is,

$$\forall v, \|v\|_{z'}'' < 1 \Rightarrow \hat{A}v \neq \bar{b}. \quad (67)$$

Of course if $r \in \mathbb{R}$ then (67) implies

$$\forall v, \quad \|v\|_z'' < r \Rightarrow \hat{A}v \neq r\bar{b}. \quad (68)$$

Finally, since $\hat{A}z' = 0$, (66) and (68) imply that if $r = \Omega(K^2 + 1)$ then

$$\hat{A}w \neq r\bar{b} \text{ for all } w \in D_{f''},$$

concluding the proof of the Statement and hence the proof of the proposition.

□

5 Proof of Theorem 1.1

5.1

The proof of Theorem 1.1 is very similar to, but slightly more complicated than, the proof of Theorem 1.4 given in §4. The proof of Theorem 1.1 depends on Theorem 4.1 and the following proposition (which plays a role analogous to that of Proposition 4.2 in the proof of Theorem 1.4).

Proposition 5.1 *Let $z_\mu = (x_\mu, s_\mu, t_\mu)$. If $\bar{b} \in Y$ satisfies*

$$\forall v, \|v\|_{z_\mu} < 1 \Rightarrow \bar{A}v \neq \bar{b}$$

and if $r \in \mathbb{R}$ satisfies $r = \Omega(K^2 + 1)$ then either

$$\bar{A}w \neq r\bar{b} \text{ for all } w = (x, s, t) \in D_f \text{ satisfying } t \leq t_\mu$$

or

$$\bar{A}w \neq -r\bar{b} \text{ for all } w = (x, s, t) \in D_f \text{ satisfying } t \leq t_\mu.$$

Proof: See Subsection 5.2. □

Proof of Theorem 1.1: We rely on the inner product $\langle \cdot, \cdot \rangle_w$ for which the adjoint of $\bar{A}$ is

$$\bar{A}^{\bar{T}} := H_w^{-1} \bar{A}^T.$$

We wish to prove that if $\text{dist}(w, \text{CentTraj}) = O(1)$ then

$$\|\sqrt{(\bar{A}\bar{A}^{\bar{T}})^{-1}}\| = \frac{O(K^2 + 1)}{\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\}}.$$

Recalling the well-known (and easily proven) Hilbert-space relation (which is based solely on the fact that $\bar{A}^{\bar{T}}$ is the adjoint of $\bar{A}$)

$$\frac{1}{\|\sqrt{(\bar{A}\bar{A}^{\bar{T}})^{-1}}\|} = \inf\{\|\bar{b}\|; \|v\|_w < 1 \Rightarrow \bar{A}v \neq \bar{b}\},$$

we see it suffices to show that if $\text{dist}(w, \text{CentTraj}) = O(1)$ then

$$\begin{aligned} & \inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\} \\ & \leq O(K^2 + 1) \inf\{\|\bar{b}\|; \|v\|_w < 1 \Rightarrow \bar{A}v \neq \bar{b}\}. \end{aligned} \tag{69}$$

In proving the inequality (69) we may assume that $w = z_\mu \in \text{CentTraj}$. For if $\|z_\mu - w\|_w = O(1)$ then for all v we have $\|v\|_w = O(\|v\|_{z_\mu})$ and hence

$$\inf\{\|\bar{b}\|; \|v\|_{z_\mu} < 1 \Rightarrow \bar{A}v \neq \bar{b}\} \\ = O(\inf\{\|\bar{b}\|; \|v\|_w < 1 \Rightarrow \bar{A}v \neq \bar{b}\}).$$

Consequently, if the inequality (69) is valid for $w = z_\mu \in \text{CentTraj}$ then it is valid for all w satisfying $\text{dist}(w, \text{CentTraj}) = O(1)$ (but perhaps with a larger universal constant hidden in the big-O).

Proposition 5.1 implies that there exists $\hat{b} \in Y$ such that

$$\bar{A}w \neq \hat{b} \text{ for all } w = (x, s, t) \in D_f \text{ satisfying } t \leq t_\mu \quad (70)$$

and

$$\|\hat{b}\| \leq O(K^2 + 1) \inf\{\|\bar{b}\|; \|v\|_{z_\mu} < 1 \Rightarrow \bar{A}v \neq \bar{b}\} \quad (71)$$

The definitions of $\bar{A}$ and D_f , along with (70), imply that the following system of constraints in the variables $(x, s) \in X \times \mathbb{R}$ is inconsistent if $0 < t \leq t_\mu$:

$$\begin{aligned} Ax &= sb + t(A\tilde{x} - \tfrac{1}{2}b) + \hat{b} \\ x &> 0 \\ \|x\| &< 1 \\ 0 &< s < 1. \end{aligned}$$

Consequently, Theorem 4.1 shows that if $0 < t \leq t_\mu$ then

$$\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\} \leq \|t(A\tilde{x} - \tfrac{1}{2}b) + \hat{b}\|;$$

thus, letting $t \downarrow 0$,

$$\inf\{\|(\Delta A, \Delta b)\|; (A + \Delta A, b + \Delta b) \text{ is inconsistent}\} \leq \|\hat{b}\|. \quad (72)$$

Together, (71) and (72) yield the inequality (69) when $w = z_\mu$, completing the proof. $\square$

5.2

Proof of Proposition 5.1: Consider the functional

$$f''_w := f_w - 12 \ln(-\langle c, w - z_\mu \rangle + \tfrac{12}{\mu}),$$

whose domain is the set

$$D_{f''} := \{w = (x, s, t) \in D_f; t < t_\mu + \tfrac{12}{\mu}\}. \quad (73)$$

Recalling that $f \in \mathcal{F}(\bar{K})$ (where $\bar{K} = K + O(1)$), in the terminology of [5] and [6] we have $f'' \in \mathcal{F}(\sqrt{\bar{K}^2 + 12})$; in the terminology of Nesterov and Nemirovskii[3],

f'' is essentially a non-degenerate self-concordant barrier functional whose barrier parameter is of magnitude $O(K^2 + 1)$.

We let g''_w , H''_w , $\langle \cdot, \cdot \rangle''_w$ and $\| \cdot \|''_w$ denote, respectively, the gradient, the Hessian, the inner product and the norm at w with respect to the functional f'' .

Observe that if $w \in D_{f''}$ then $\|v\|_w \leq \|v\|''_w$ for all v . Together with (73) this implies that to prove the proposition, it suffices to prove the following statement (which in the sequel we refer to as “Statement 1”): *If $\bar{b} \in Y$ has the property that*

$$\forall v, \|v\|''_{z_\mu} < 1 \Rightarrow \bar{A}v \neq \bar{b}$$

and if $r \in \mathbb{R}$ satisfies $r = \Omega(K^2 + 1)$ then either

$$\bar{A}w \neq r\bar{b} \text{ for all } w \in D_{f''}$$

or

$$\bar{A}w \neq -r\bar{b} \text{ for all } w \in D_{f''}.$$

Note that $g''_{z_\mu} = \mu c + g_{z_\mu}$ where g_{z_μ} denotes the gradient of f at z_μ . Consequently, since z_μ is optimal for the linearly constrained convex optimization problem

$$\begin{aligned} \min \quad & \mu \langle c, w \rangle + f_w \\ \text{s.t.} \quad & \bar{A}w = 0, \end{aligned}$$

we have that g''_{z_μ} lies in the range space of $\bar{A}^T$.

Now consider the functional

$$f'''_w := f''_w - 12 \ln(\langle g''_{z_\mu}, w - z_\mu \rangle + 12),$$

whose domain is the set

$$D_{f'''} := \{w \in D_{f''}; \langle g''_{z_\mu}, w - z_\mu \rangle > -12\}. \quad (74)$$

In the terminology of [5] and [6], $f''' \in \mathcal{F}(\sqrt{K^2 + 24})$; in the terminology of Nesterov and Nemirovskii[3], f''' is essentially a non-degenerate self-concordant barrier functional whose barrier parameter is $O(K^2 + 1)$.

We let g'''_w , H'''_w , $\langle \cdot, \cdot \rangle'''_w$ and $\| \cdot \|'''_w$ denote, respectively, the gradient, the Hessian, the inner product and the norm at w with respect to the functional f''' .

We claim that to prove Statement 1 it suffices to prove the following statement (which in the sequel we refer to as “Statement 2”): *If $\bar{b} \in Y$ satisfies*

$$\forall v, \|v\|'''_{z_\mu} < 1 \Rightarrow \bar{A}v \neq \bar{b}$$

and if $r \in \mathbb{R}$ satisfies $r = \Omega(K^2 + 1)$ then

$$\bar{A}w \neq r\bar{b} \text{ for all } w \in D_{f''' }.$$

To understand why Statement 2 implies Statement 1, assume that $\bar{b} \in Y$ satisfies the assumption of Statement 1, that is,

$$\forall v, \|v\|_{z_\mu}'' < 1 \Rightarrow \bar{A}v \neq \bar{b}.$$

Since $\|v\|_{z_\mu}''' \leq \|v\|_{z_\mu}''$ for all v , we then have

$$\forall v, \|v\|_{z_\mu}''' < 1 \Rightarrow \bar{A}v \neq \bar{b}.$$

Hence, if $r = \Omega(K^2 + 1)$ then according to Statement 2,

$$\bar{A}w \neq r\bar{b} \text{ for all } w \in D_{f''' }. \quad (75)$$

Now, since g_{z_μ}'' lies in the range space of $\bar{A}^T$, given $r \in \mathbb{R}$ we have that either

$$\forall w, \bar{A}w = r\bar{b} \Rightarrow \langle g_{z_\mu}'', w - z_\mu \rangle \geq 0 \quad (76)$$

or

$$\forall w, \bar{A}w = -r\bar{b} \Rightarrow \langle g_{z_\mu}'', w - z_\mu \rangle \geq 0. \quad (77)$$

If (76) is valid then (74) and (75) imply

$$\bar{A}w \neq r\bar{b} \text{ for all } w \in D_{f''' }.$$

Similarly, if (77) is valid then (74) and (75) imply

$$\bar{A}w \neq -r\bar{b} \text{ for all } w \in D_{f''' }.$$

Consequently, Statement 1 does indeed follow from Statement 2.

To prove Statement 2, first observe that $g_{z_\mu}''' = 0$. Hence z_μ minimizes the strictly convex functional f''' , i.e., z_μ is the “analytic center” of the domain $D_{f''' }$. Thus, since $f \in \mathcal{F}(\bar{K}^2 + 24)$ and $\bar{K} = K + O(1)$, we have

$$w \in D_{f''' } \Rightarrow \|w - z_\mu\|_{z_\mu}''' = O(K^2 + 1). \quad (78)$$

Assume that $\bar{b}$ satisfies the assumption of Statement 2, that is,

$$\forall v, \|v\|_{z_\mu}''' < 1 \Rightarrow \bar{A}v \neq \bar{b}. \quad (79)$$

Of course if $r \in \mathbb{R}$ then (79) implies

$$\forall v, \|v\|_{z_\mu}''' < r \Rightarrow \bar{A}v \neq r\bar{b}. \quad (80)$$

Finally, since $\bar{A}z_\mu = 0$, (78) and (80) imply that if $r = \Omega(K^2 + 1)$ then

$$\bar{A}w \neq r\bar{b} \text{ for all } w \in D_{f''' },$$

concluding the proof of Statement 2 and hence the proof of the proposition. $\square$

6 Proof of Theorem 1.8

6.1

Recall that in the context of Theorem 1.8, we are given $w = (x, s) \in X \times \mathbb{R}$ and are considering the sequence

$$w_i = (x_i, s_i) := w + (H'_w)^{-1} \hat{A} y_i \quad (81)$$

where $y_0 := 0, y_1, \dots$, is the sequence obtained by applying cgm to the linear equations (in the variable y)

$$\hat{A}(H'_w)^{-1} \hat{A}^T y = -\hat{A}w. \quad (82)$$

As discussed in §1, if $\text{relcond}(A, b) < \infty$ then the constraints

$$\begin{aligned} \hat{A}v &= 0 \\ v &\in D_{f'} \end{aligned}$$

have an “analytic center,” that is, a solution z' of the following optimization problem:

$$\begin{aligned} \inf \quad & f'_v \\ \text{s.t.} \quad & \hat{A}v = 0 \\ & v \in D_{f'}. \end{aligned} \quad (83)$$

Our proof of Theorem 1.8 depends on the following two propositions.

Proposition 6.1 Assume that $w = (x, s) \in D_{f'}$ and the linear operator

$$\hat{A}(H'_w)^{-1}\hat{A}^T$$

is invertible; hence the sequence $\{w_i\}$ given by (81) is well-defined and has limit w_∞ satisfying $\hat{A}w_\infty = 0$.

If z is any point satisfying $\hat{A}z = 0$ and $\|z - w\|'_w = O(1)$ then

$$w_i, w_\infty, z \in D_{f'}, \quad \|w_\infty - z\|'_z = O(\|z - w\|'_w)$$

and

$$i = \Omega \left(\text{relcond}(\sqrt{\hat{A}(H'_w)^{-1}\hat{A}^T}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \left\| \frac{1}{s_\infty} x_\infty - \frac{1}{s_i} x_i \right\| \leq \frac{\epsilon}{s_\infty} \|z - w\|'_w$$

for all ϵ satisfying $0 < \epsilon \leq 1$.

Proof: See Subsection 6.2. $\square$

Proposition 6.2 Assume $\text{relcond}(A, b) < \infty$ and let z' denote the optimal solution for the optimization problem (83). If $\bar{w} = (\bar{x}, \bar{s})$ is an arbitrary point satisfying $\|\bar{w} - z'\|'_{z'} = O(1)$ then

$$\left\| \frac{1}{\bar{s}} \bar{x} \right\| \leq O(K^2 + 1)(1 + \inf\{\|x\|; Ax = b \text{ and } x \geq 0\}) \quad (84)$$

$$\leq O(K^2 + 1) \text{relcond}(A, b) \quad (85)$$

and

$$\frac{1}{\bar{s}} \leq O(K^2 + 1)(\left\| \frac{1}{\bar{s}} \bar{x} \right\| + 1). \quad (86)$$

Proof: See Subsection 6.5. $\square$

Proof of Theorem 1.8: Immediate from Corollary 1.7 and Propositions 6.1 and 6.2, letting z in Proposition 6.1 be z' and letting $\bar{w}$ in Proposition 6.2 be w_∞ . $\square$

6.2

The proof of Proposition 6.1 depends on the following proposition and lemma; the lemma will also be used in the proof of Proposition 6.2.

Proposition 6.3 Assume the same conditions as in Proposition 6.1 but require no bound on $\|z - w\|'_w$. Then for all i ,

$$\|w_\infty - w_i\|'_w \leq \|z - w\|'_w \quad (87)$$

and for all ϵ satisfying $0 < \epsilon \leq 1$,

$$i = \Omega \left(\text{relcond}(\sqrt{\hat{A}(H'_w)^{-1}\hat{A}^T}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \|w_\infty - w_i\|'_w \leq \epsilon \|z - w\|'_w. \quad (88)$$

Proof: See Subsection 6.3. □

Lemma 6.4 *If $\bar{w} = (\bar{x}, \bar{s})$ and $\hat{w} = (\hat{x}, \hat{s})$ satisfy*

$$\|\bar{w} - w\|'_w = O(1) \text{ and } \|\hat{w} - w\|'_w = O(1)$$

where $w = (x, s) \in D_{f'}$, then $\bar{w}, \hat{w} \in D_{f'}$,

$$\|\frac{1}{\bar{s}}\bar{x} - \frac{1}{\hat{s}}\hat{x}\| = O(\frac{1}{\bar{s}}\|\bar{w} - \hat{w}\|'_w) \text{ and } \hat{s} = \Theta(\bar{s}).$$

Proof: See Subsection 6.4. □

Proof of Proposition 6.1: First note that since $w_0 = w$, (87) implies

$$\|w_\infty - w\|'_w \leq \|z - w\|'_w \text{ and } \|w_i - w\|'_w \leq 2\|z - w\|'_w. \quad (89)$$

Consequently, if $\|z - w\|'_w = O(1)$ then $w_i, w_\infty \in D_{f'}$.

Observe that (89) implies

$$\|w_\infty - z\|'_w = O(\|z - w\|'_w).$$

Since

$$\|z - w\|'_w = O(1) \Rightarrow \|v\|'_z = O(\|v\|'_w) \text{ for all } v,$$

it follows that if $\|z - w\|'_w = O(1)$ then

$$\|w_\infty - z\|'_z = O(\|z - w\|'_w).$$

For the remaining assertions of the proposition, simply combine Proposition 6.3 and Lemma 6.4, replacing the points $\bar{w}, \hat{w}$ in the lemma with w_∞, w_i , respectively. □

6.3

Proof of Proposition 6.3: Let w^* denote the minimal norm solution (with respect to the norm $\|\cdot\|'_w$) to the linear equations (in the variable v)

$$\hat{A}v = -\hat{A}w.$$

Note that $\|w^*\|_w \leq \|z - w\|_w$ since $\hat{A}(z - w) = -\hat{A}w$.

The adjoint of $\hat{A}$ with respect to the inner product $\langle \cdot, \cdot \rangle'_w$ is

$$\hat{A}^{T'} := (H'_w)^{-1}\hat{A}^T.$$

Consequently, the linear equations (82) can be rewritten as

$$\hat{A}\hat{A}^{T'}y = -\hat{A}w.$$

By considering the linear equations rewritten in this form we see that (8) implies for all i ,

$$\|w^* - \hat{A}^{T'} y_i\|'_w \leq \|w^*\|'_w. \quad (90)$$

Similarly, (7) gives

$$i = \Omega \left(\text{relcond}(\sqrt{\hat{A}\hat{A}^{T'}}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \|w^* - \hat{A}^{T'} y_i\|'_w < \epsilon \|w^*\|'_w. \quad (91)$$

for all ϵ satisfying $0 < \epsilon \leq 1$.

Observe that (91) implies $w_\infty = w + w^*$ and hence (90) implies

$$\|w_\infty - w_i\|'_w \leq \|w^*\|'_w$$

whereas (91) gives

$$i = \Omega \left(\text{relcond}(\sqrt{\hat{A}\hat{A}^{T'}}) \log_2 \frac{2}{\epsilon} \right) \Rightarrow \|w_\infty - w_i\|'_w < \epsilon \|w^*\|'_w.$$

Since $\|w^*\|'_w \leq \|z - w\|'_w$, the proof is complete. $\square$

6.4

Proof of Lemma 6.4: Since $f' \in \mathcal{F}(K')$ and $w \in D_{f'}$, we have that for all u ,

$$\|u - w\|'_w = O(1) \Rightarrow [u \in D_{f'} \text{ and for all } v, \|v\|'_u = \Theta(\|v\|'_w)].$$

Consequently, in proving the Lemma we may assume that $\hat{w} = w$.

Since

$$\|v\|'_w = O(1) \Rightarrow w + v \in D_{f'}$$

and

$$(x', s') \in D_{f'} \Rightarrow [\|x'\| < 1 \text{ and } 0 < s' < 1],$$

one can easily prove that for all $\tilde{v} = (\tilde{x}, \tilde{s})$,

$$\|\tilde{x}\| = O(\|\tilde{v}\|'_w) \text{ and } \tilde{s} = O(s\|\tilde{v}\|'_w).$$

Hence,

$$\|\bar{x} - x\| = O(\|\bar{w} - w\|'_w) \text{ and } |\bar{s} - s| = O(s\|\bar{w} - w\|'_w).$$

Thus, if $\|\bar{w} - w\|'_w = O(1)$ then

$$\begin{aligned} \left\| \frac{1}{\bar{s}} \bar{x} - \frac{1}{s} x \right\| &= \frac{1}{\bar{s}} \left\| \bar{x} - x + \frac{\bar{s}-s}{\bar{s}} x \right\| \\ &\leq \frac{1}{\bar{s}} \left(\|\bar{x} - x\| + \frac{\bar{s}-s}{\bar{s}} \right) \\ &= O\left(\frac{1}{\bar{s}}\|\bar{w} - w\|'_w\right), \end{aligned}$$

completing the proof. $\square$

6.5

The proof of Proposition 6.2 depends on the following lemma.

Lemma 6.5 *Assume $w = (x, s) \in D_{f'}$ and $\frac{1}{s} = O(\|\frac{1}{s}x\| + 1)$. If $\hat{w} = (\hat{x}, \hat{s})$ satisfies $\|\hat{w} - w\|_w = O(1)$ then*

$$\|\frac{1}{\hat{s}}\hat{x}\| = O(\|\frac{1}{s}x\| + 1) \quad (92)$$

and

$$\frac{1}{\hat{s}} = O(\|\frac{1}{\hat{s}}\hat{x}\| + 1). \quad (93)$$

Proof: See Subsection 6.6. $\square$

Proof of Proposition 6.2: Lemma 6.5, applied with $w = z'$ and $\hat{w} = \bar{w}$, implies that in proving the proposition we may assume $\bar{w} = z'$. Thus, letting $z' = (x', s')$, it suffices to show

$$\|\frac{1}{s'}x'\| \leq O(K^2 + 1)(1 + \inf\{\|x\|; Ax = b \text{ and } x \geq 0\}), \quad (94)$$

$$\frac{1}{s'} \leq O(K^2 + 1)(\|\frac{1}{s'}x'\| + 1) \quad (95)$$

and (since $\text{relcond}(A, b) \geq 1$),

$$\inf\{\|x\|; Ax = b \text{ and } x \geq 0\} \leq \text{relcond}(A, b). \quad (96)$$

The inequality (96) is a special case of [4, Theorem 1.1(1)]. Hence we need only prove (94) and (95).

The assumption $\text{relcond}(A, b) < \infty$ implies that the set of points which satisfy the strict constraints $Ax = b$, $x > 0$ are dense in the set of points satisfying the constraints $Ax = b$, $x \geq 0$; c.f., [6, Lemma 3.4]. Consequently, assuming that $\tilde{x}$ is an arbitrary point satisfying the strict constraints, to prove (94) it suffices to show

$$\|\frac{1}{s'}x'\| \leq O(K^2 + 1)(\|\tilde{x}\| + 1). \quad (97)$$

Observe that

$$\hat{w} = (\hat{x}, \hat{s}) := \frac{1}{2(\|\tilde{x}\| + 1)}(\tilde{x}, 1)$$

satisfies the constraints

$$\begin{aligned} \hat{A}w &= 0 \\ w &\in D_{f'}, \end{aligned} \quad (98)$$

i.e., $\hat{w}$ is an element of the region for which z' is the “analytic center.” Since $s > 0$ is among the constraints defining (98) it follows that

$$\hat{s} = O((K')^2 s').$$

Thus,

$$\|\frac{1}{s'}x'\| \leq \frac{1}{s'} = O((K')^2 \frac{1}{s}).$$

Since $\frac{1}{s} = 2(\|\tilde{x}\| + 1)$ and $K' = K + O(1)$, the proof of (97) is complete; thus, the proof of (94) is complete.

Let $z'' = (x'', s'')$ be a positive multiple of z' with the property that z'' lies in the bounded region defined by the constraints (98) but $2z''$ does not. Since z' is the “analytic center” for the region and $s > 0$ is among the constraints defining the region, we have

$$s'' = O((K')^2 s') \quad (99)$$

and hence

$$\|x''\| = O((K')^2 \|x'\|). \quad (100)$$

Since $2z''$ does not lie in the region we have

$$\|x''\| \geq \frac{1}{2} \text{ or } s'' \geq \frac{1}{2}. \quad (101)$$

Combining (99), (100) and (101) gives

$$(K')^2 \|x'\| = \Omega(1) \text{ or } (K')^2 s' = \Omega(1).$$

Since $K' = K + O(1)$ and $\|x'\| < 1$, the bound (95) follows. $\square$

6.6

Proof of Lemma 6.5: Lemma 6.4 (applied with $\bar{w} = w$) implies that if $\|\hat{w} - w\|'_w = O(1)$ then

$$\|\frac{1}{s}\hat{x}\| \leq \|\frac{1}{s}x\| + O(\frac{1}{s}).$$

Since $\frac{1}{s} = O(\|\frac{1}{s}x\| + 1)$ (by assumption), the bound (92) follows.

Lemma 6.4 also implies that if $\|\hat{w} - w\|'_w = O(1)$ then

$$\|\frac{1}{s}x\| \leq \|\frac{1}{s}\hat{x}\| + O(\frac{1}{s}\|\hat{w} - w\|'_w) \quad (102)$$

and

$$\frac{1}{s} = O(\frac{1}{s}). \quad (103)$$

If $\|\hat{w} - w\|'_w = O(1)$ and $\frac{1}{s} = O(\|\frac{1}{s}x\| + 1)$ then (102) implies

$$\|\frac{1}{s}x\| = O(\|\frac{1}{s}\hat{x}\| + 1)$$

and hence (again using $\frac{1}{s} = O(\|\frac{1}{s}x\| + 1)$),

$$\frac{1}{s} = O(\|\frac{1}{s}x\| + 1). \quad (104)$$

Together, (103) and (104) yield (93). $\square$

7 Proof of Theorem 1.9

7.1

In proving Theorem 1.9 it is useful to rely on the functional

$$f''_{x,s,t} := \phi_x - 12 [\ln(1 - \|x\|^2) + \ln(s) + \ln(1 - s) + \ln(1 + t) + \ln(2 - t)],$$

whose domain is

$$D_{f''} := \{(x, s, t); x > 0, \|x\| < 1, 0 < s < 1 \text{ and } -1 < t < 2\}.$$

Observe the only difference between f'' and f is that the summand “ $\ln(1 + t)$ ” in the definition of f'' is replaced by “ $\ln t$ ” in the definition of f . Similarly, the only difference between $D_{f''}$ and D_f is that the constraint “ $-1 < t$ ” in the definition of $D_{f''}$ is replaced by “ $0 < t$ ” in the definition of D_f ; thus $D_f \subseteq D_{f''}$. Also note that

$$(x, s, t) \in D_{f''} \Rightarrow (x, s) \in D_{f'}.$$

The functional f'' is an element of $\mathcal{F}(K'')$ where $K'' = \sqrt{K^2 + 72}$; recall that in the terminology of Nesterov and Nemirovskii[3], $f'' \in \mathcal{F}(K'')$ essentially means that f'' is a non-degenerate self-concordant barrier functional with barrier parameter of magnitude $O((K'')^2)$.

In the same manner that the Hessian H_w of f at $w \in D_f$ induces the inner product $\langle \cdot, \cdot \rangle_w$ and norm $\| \cdot \|_w$, the Hessian H''_w of f'' at $w \in D_{f''}$ induces an inner product $\langle \cdot, \cdot \rangle''_w$ and norm $\| \cdot \|''_w$.

Let $\text{val}''_{\inf}$ denote the optimal value of the problem (in the variable $w = (x, s, t)$),

$$\begin{aligned} & \inf \quad t \\ & \text{s.t.} \quad \bar{A}w = 0 \\ & \quad \quad w \in D_{f''} \end{aligned} \tag{105}$$

and let $\text{val}''_{\sup}$ denote the optimal value of the problem obtained by replacing “inf” in (105) with “sup.”

Our proof of Theorem 1.9 depends on the corollary to the following two propositions.

Proposition 7.1 *If $\text{relcond}(A, b) < \infty$ then*

$$\text{val}_{\inf}'' \leq \frac{-1}{2 \text{relcond}(A, b)}$$

and

$$\text{val}_{\sup}'' \geq 1.$$

Proof: The bound on $\text{val}_{\inf}''$ is a special case of [6, Proposition 3.2]. The bound on $\text{val}_{\sup}''$ follows from the fact that $(\check{x}, \frac{1}{2}, 1)$ satisfies the constraints of (105). $\square$

For a constant $\mu \in \mathbb{R}$, let z_{μ}'' denote the optimal solution of the problem

$$\begin{aligned} \inf \quad & \mu t + f_w'' \\ \text{s.t.} \quad & \bar{A}w = 0 \\ & w \in D_{f''}. \end{aligned} \tag{106}$$

(A unique optimal solution exists because $f'' \in \mathcal{F}(K'')$, $D_{f''}$ is bounded and the feasible region for (106) is non-empty (in particular, $(\check{x}, \frac{1}{2}, 1)$ is feasible); c.f., [5].)

Define

$$\text{CentTraj}_{+,-}'' := \{z_{\mu}''; \mu \in \mathbb{R}\}.$$

The subscripts “+,-” are meant to highlight the fact that all values of μ are allowed, not just non-negative values as in the definition of CentTraj in §1.

Proposition 7.2 *If*

$$z_{\mu_1}'' = (x_{\mu_1}'', s_{\mu_1}'', t_{\mu_1}'') \in \text{CentTraj}_{+,-}'' \quad \text{and} \quad z_{\mu_2}'' = (x_{\mu_2}'', s_{\mu_2}'', t_{\mu_2}'') \in \text{CentTraj}_{+,-}''$$

satisfy

$$|t_{\mu_2}'' - t_{\mu_1}''| \leq \frac{O(\min\{\text{val}_{\sup}'' - t_{\mu_2}, t_{\mu_2} - \text{val}_{\inf}''\})}{(K'')^2}$$

then

$$\|z_{\mu_2}'' - z_{\mu_1}''\|_{z_{\mu_1}''}'' \leq \frac{O((K'')^2 |t_{\mu_2}'' - t_{\mu_1}''|)}{\min\{\text{val}_{\sup}'' - t_{\mu_2}, t_{\mu_2} - \text{val}_{\inf}''\}}.$$

Proof: This follows from standard (but somewhat involved) arguments in the interior-point method literature; it is solely a consequence of f'' being an element of $\mathcal{F}(K'')$; for a detailed proof, c.f., [5]. $\square$

Recall that z' denotes the “analytic center” for the constraints

$$\begin{aligned} \hat{A}(x, s) &= 0 \\ (x, s) &\in D_{f'}. \end{aligned}$$

Corollary 7.3 Assume $\text{relcond}(A, b) < \infty$. If $z''_\mu = (x''_\mu, s''_\mu, t''_\mu) \in \text{CentTraj}''_{+,-}$ satisfies

$$|t''_\mu| \leq \frac{O(1)}{(K^2 + 1) \text{relcond}(A, b)}$$

then

$$\|(z', 0) - z''_\mu\|''_{z''_\mu} \leq O(|t''_\mu|)(K^2 + 1) \text{relcond}(A, b).$$

Proof: Follows easily from Propositions 7.1 and 7.2 because

$$(z', 0) \in \text{CentTraj}''_{+,-}$$

and $K'' = K + O(1)$. □

7.2

Proof of Theorem 1.9: Observe that if $w = (x, s, t) \in D_f$ then for all $(\bar{x}, \bar{s}, \bar{t})$,

$$\langle (\bar{x}, \bar{s}, \bar{t}), H_w(\bar{x}, \bar{s}, \bar{t}) \rangle = \langle (\bar{x}, \bar{s}, 0), H_w(\bar{x}, \bar{s}, 0) \rangle + \langle (0, 0, \bar{t}), H_w(0, 0, \bar{t}) \rangle$$

and

$$\langle (\bar{x}, \bar{s}, \bar{t}), H''_w(\bar{x}, \bar{s}, \bar{t}) \rangle = \langle (\bar{x}, \bar{s}, 0), H''_w(\bar{x}, \bar{s}, 0) \rangle + \langle (0, 0, \bar{t}), H''_w(0, 0, \bar{t}) \rangle.$$

Moreover,

$$\langle (\bar{x}, \bar{s}, 0), H_w(\bar{x}, \bar{s}, 0) \rangle = \langle (\bar{x}, \bar{s}), H'_{x,s}(\bar{x}, \bar{s}) \rangle$$

and

$$\langle (\bar{x}, \bar{s}, 0), H''_w(\bar{x}, \bar{s}, 0) \rangle = \langle (\bar{x}, \bar{s}), H'_{x,s}(\bar{x}, \bar{s}) \rangle.$$

Hence,

$$\|(\bar{x}, \bar{s}, \bar{t})\|_w^2 = (\|(\bar{x}, \bar{s})\|'_{x,s})^2 + \|(0, 0, \bar{t})\|_w^2 \quad (107)$$

and

$$(\|(\bar{x}, \bar{s}, \bar{t})\|''_w)^2 = (\|(\bar{x}, \bar{s})\|'_{x,s})^2 + (\|(0, 0, \bar{t})\|''_w)^2. \quad (108)$$

If $w = (x, s, t) \in D_f$ and $\hat{w} = (\hat{x}, \hat{s}, \hat{t})$ then (107) implies

$$\|(\hat{x}, \hat{s}) - (x, s)\|'_{x,s} \leq \|\hat{w} - w\|_w. \quad (109)$$

Consequently,

$$\|\hat{w} - w\|_w = O(1) \Rightarrow [\|v\|'_{\hat{x}, \hat{s}} = \Theta(\|v\|'_{x,s}) \text{ for all } v]. \quad (110)$$

Note that if $w = (x, s, t) \in D_f$ and $z_\mu = (x_\mu, s_\mu, t_\mu) \in \text{CentTraj}$ then (109) and (110) yield

$$\|z_\mu - w\|_w = O(1) \Rightarrow \|z' - (x, s)\|'_{x,s} = O(\|z_\mu - w\|_w + \|z' - (x_\mu, s_\mu)\|'_{x_\mu, s_\mu}).$$

Consequently, to prove the theorem it suffices to show that if

$$\|z_\mu - w\|_w = O(1) \text{ and } t \leq \frac{O(1)}{(K^2 + 1) \text{relcond}(A, b)} \quad (111)$$

then

$$\|z' - (x_\mu, s_\mu)\|'_{x_\mu, s_\mu} \leq O(t)(K^2 + 1) \text{relcond}(A, b). \quad (112)$$

Since

$$\|v\|_w = O(1) \Rightarrow w + v \in D_f$$

and since

$$(\bar{x}, \bar{s}, \bar{t}) \in D_f \Rightarrow \bar{t} > 0,$$

we have

$$\|z_\mu - w\|_w = O(1) \Rightarrow t_\mu = O(t).$$

Consequently, to prove that (111) implies (112), it suffices to show that if

$$t_\mu \leq \frac{O(1)}{(K^2 + 1) \text{relcond}(A, b)} \quad (113)$$

then

$$\|z' - (x_\mu, s_\mu)\|'_{x_\mu, s_\mu} \leq O(t_\mu)(K^2 + 1) \text{relcond}(A, b). \quad (114)$$

Since (108) implies

$$\|z' - (x_\mu, s_\mu)\|'_{x_\mu, s_\mu} \leq \|(z', 0) - z_\mu\|''_{z_\mu},$$

to prove that (113) implies (114) it suffices to show that (113) implies

$$\|(z', 0) - z_\mu\|''_{z_\mu} = O(t_\mu)(K^2 + 1) \text{relcond}(A, b). \quad (115)$$

However, since $\text{CentTraj} \subseteq \text{CentTraj}''_{+, -}$, the fact that (113) does indeed imply (115) is immediate from Corollary 7.3. $\square$

References

- [1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.
- [2] R.J. Duffin, “Infinite programs,” in: H.W. Kuhn and A.W. Tucker, eds., *Linear Inequalities and Related Systems*, Princeton University Press, Princeton, 1956, 157-170.
- [3] Yu.E. Nesterov and A.S. Nemirovskii, *Interior Point Methods in Convex Optimization: Theory and Applications*, SIAM, Philadelphia, 1993.
- [4] J. Renegar, “Some perturbation theory for linear programming,” *Mathematical Programming* 65 (1994) 73-92.
- [5] J. Renegar, “Lecture notes on the efficiency of Newton’s method for convex optimization in Hilbert spaces,” preprint, School of Operations Research and Industrial Engineering, Cornell University, 1993.
- [6] J. Renegar, “Linear programming, complexity theory and elementary functional analysis,” to appear in *Mathematical Programming*.