An Accelerated Interior Point Method Whose Running Time Depends Only on A

Stephen A. Vavasis* Yinyu Ye †

October 14, 1993

Abstract

We propose a "layered-step" interior point (LIP) algorithm for linear programming. This algorithm follows the central path, either with short steps or with a new type of step called a "layered least squares" (LLS) step. The algorithm returns the exact global minimum after a finite number of steps—in particular, after $O(n^{3.5}c(A))$ iterations, where $c(A)$ is a function of the coefficient matrix. The LLS steps can be thought of as accelerating a path-following interior point method whenever near-degeneracies occur. One consequence of the new method is a new characterization of the central path: we show that it composed of at most n^2 alternating straight and curved

*Department of Computer Science, Upson Hall, Cornell University, Ithaca, NY 14853. Email: vavasis@cs.cornell.edu. This work is supported in part by the National Science Foundation, the Air Force Office of Scientific Research, and the Office of Naval Research, through NSF grant DMS-8920550. Also supported in part by an NSF Presidential Young Investigator award with matching funds received from AT&T and Xerox Corp. Part of this work was done while the author was visiting Sandia National Laboratories, supported by the U.S. Department of Energy under contract DE-AC04-76DP00789.

†Department of Management Sciences, The University of Iowa, Iowa City, IA 52242. E-mail: bbuyin@vaxa.weeg.uiowa.edu. This author is supported in part by NSF Grant DDM-9207347. Part of this work was done while the author was on a sabbatical leave from the University of Iowa and visiting the Cornell Theory Center, Cornell University, Ithaca, NY 14853, supported in part by the Cornell Center for Applied Mathematics and by the Advanced Computing Research Institute, a unit of the Cornell Theory Center, which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute.

segments. If the LIP algorithm is applied to integer data, we get as another corollary a new proof of a well-known theorem by Tardos that linear programming can be solved in strongly polynomial time provided that A contains small-integer entries.

1 Interior point methods

In this report we consider solving a linear programming problem (LP) in the following form:

$$\begin{array}{rl} \text{minimize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & A^T\mathbf{y} \geq \mathbf{c}. \end{array} \tag{1}$$

Here, A is an $m \times n$ matrix assumed to have rank m, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^n$ are given vectors, and $\mathbf{y} \in \mathbb{R}^m$ is the unknown vector. This format of LP is commonly known as "dual format."

We propose a path-following interior point method for this problem. In the taxonomy of interior point methods, our method would be called a "dual" path following method, but it has some features of a "primal-dual" method. In particular, our algorithm produces both a primal and dual optimal solution.

Interior point methods were originally introduced by Karmarkar [8], and the first path-following method is due to Renegar [19]. Traditional path-following methods take small steps along the *central path* until they are "sufficiently" close to the optimum. Once sufficiently close, a "rounding" procedure such as Khachiyan's [9] or least-squares computation such as Ye's [31] is used to obtain a global minimum.

In our new method, we interleave small steps with longer *layered least squares* (LLS) steps to follow the central path. Thus, our method is always at least as efficient as existing path-following interior point methods. The last LLS step moves directly to the global optimum. Thus, our algorithm, which we call "layered-step interior point" (LIP), terminates in a finite number of steps. Furthermore, the total number of iterations depends only on A: the running time is $O(n^{3.5}c(A))$ iterations, where $c(A)$ is defined by (63) below. This is in contrast to all previous interior methods, whose complexity depends on $\mathbf{b}$ and $\mathbf{c}$ as well as A. This is important because there are many classes of problems (see, e.g., Section 10) in which A is "well-behaved" but $\mathbf{b}$ and $\mathbf{c}$ are arbitrary vectors.

In order to provide intuition on how the LIP algorithm works, and why our complexity bounds are stronger, we consider the linear programming problem presented in Figure 1. The problem solved in this figure is:

$$\begin{array}{rl} \text{minimize} & 2y_1 + 5y_2 \\ \text{subject to} & 0 \le y_1 \le 1, \\ & 0 \le y_2 \le 1, \\ & y_1 + 2y_2 \ge \epsilon, \end{array}$$

where $\epsilon = 0.1$. The dashed line indicates the boundaries of the feasible region defined by these constraints. The solid line is the *central path*, which is defined in Section 2. It connects the point A, known as the *analytic center*, to the point D, the optimum solution. Note that the optimum in this example is at a vertex of the feasible region—it is always the case that the optimum is achieved at a vertex, and for a nondegenerate problem (such as this one) the optimum is unique. The asterisks show the iterates of a conventional path-following interior point method. Such a method generates iterates approximately on the central path with spacing between the iterates decreasing geometrically. Once the iterates are sufficiently close to the optimum D, where "sufficiently close" means that the current iterate is closer to D than to any other vertex of the feasible region, then the exact optimum may be computed.

Now, consider what happens as we let ϵ get smaller in this problem. This creates a "near degeneracy" near $(0,0)$, and the diagonal segment in Figure 1 gets shorter and shorter. This means that the conventional method must take more and more iterates in order to distinguish the optimal vertex $(\epsilon, 0)$ from the nearby vertex $(0, \epsilon/2)$. In fact, it can be proved that a standard method applied to this problem would require $O(|\log \epsilon|)$ iterates. Note that ϵ is a component of right-hand side vector $\mathbf{c}$; this explains why the existing methods have complexity depending on $\mathbf{c}$. The running time of all existing methods is written as $O(\sqrt{n}L)$ (or sometimes $O(nL)$) iterations, where L is the number of total bits required to write $A, \mathbf{b}, \mathbf{c}$. Thus, for this example, L depends logarithmically on ϵ.

In contrast, our method would jump from point B directly to point C without intervening iterations. This accomplished by solving a weighted least-squares problem involving the three constraints $y_1 \ge 0, y_2 \ge 0, y_1 + 2y_2 \ge \epsilon$, which are "near active" at optimum. In a more complex example, there would several "layers" involved in the least-squares computation. When

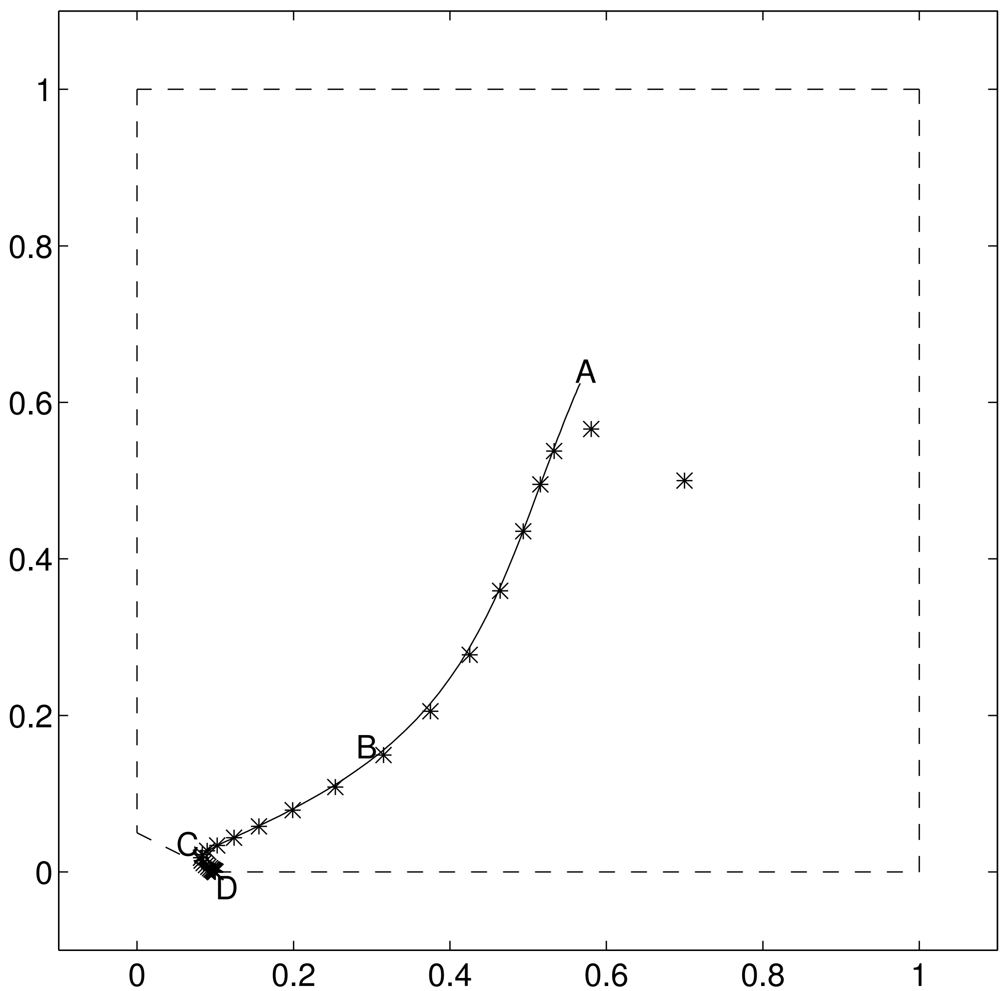

Figure 1: Example of an interior point method.

point C is reached, the LIP method takes small interior point steps, the number of which depends only on the constraint matrix

$$A^T = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 2 \end{pmatrix}.$$

After these small steps, the LIP method jumps again directly to optimal point D.

An observation concerning this figure is that the central path appears to consist of a curved segment from A to B, an approximately straight segment from B to C, a curved segment near C, and another approximately straight segment from near C to D. Indeed, a consequence of the LIP method is a new characterization of the central path as being composed of at most n^2 curved segments alternating with approximately straight segments. For a problem with no near-degeneracies, there is only one curved segment followed by an approximately straight segment leading to the optimum.

Ours is not the first complexity bound for linear programming that depends only on A; Tardos [23] earlier proposed such a method. Tardos' method, however, "probably should be considered a purely theoretical contribution" [23, p. 251] because it requires the solution of n complete LP's. In contrast, our method is a fairly standard kind of interior point method accompanied by an acceleration step; accordingly, we believe that it is quite practical.

Another difference is that we regard our method as primarily a real-number method. For example, our method is the first polynomial-time linear programming algorithm that also has a complexity bound depending only on A for finding the global minimum in the real-number model of computation. In contrast, Tardos uses the assumption of integer data in a fairly central way because an important tool in [23] is the operation of rounding down to the nearest integer.

The remainder of this paper is organized as follows. In Section 2 we define the central path and characterize points that are "approximately" centered. In Section 3 we describe quantities $\chi_A, \bar{\chi}_A$ that were discovered independently by Stewart [22] and Todd [24], and state some of their properties. Our complexity bounds depend on these parameters. In Section 4 we describe the layered least squares step and the LIP method. In Section 5 we prove the first main theorem concerning the LIP method, namely, that every point on the line segment defined by the LLS step is approximately centered. In Section 6 we define *crossover events,* a combinatorial event concerning the central path that forms the basis of our complexity analysis. We prove the second main theorem that each LLS step forces at least one crossover event. In Section 7 we assemble the pieces in order to obtain a complexity result for solving LP problems. In Section 8 we explain a method for initializing the LIP method, and analyze its complexity. This complexity is no worse than the complexity of the LIP method itself. In Section 9 we specialize to integer data and provide a new proof of Tardos' theorem. Finally, in Section 10 we apply our bounds to the special case of minimum-cost flow to explore how our interior-point method compares to other strongly polynomial time algorithms for this problem.

2 The central path

It is common to rewrite (1) by introducing *slack variables* as follows:

$$\begin{array}{rl} \text{minimize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & A^T\mathbf{y} - \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} \geq \mathbf{0}. \end{array} \tag{2}$$

We will assume this format for the remainder of the paper. Assume that a strictly interior point, that is, a point with $\mathbf{s} > \mathbf{0}$, exists. Let us introduce a *log-barrier term* into the objective function, yielding the new optimization problem:

$$\begin{array}{rl} \text{minimize} & \mathbf{b}^T\mathbf{y} - \mu\sum_{i=1}^n \ln s_i \\ \text{subject to} & A^T\mathbf{y} - \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} > \mathbf{0}. \end{array} \tag{3}$$

Here, μ is a positive parameter. This problem is strictly convex. This means that if we assume the set of feasible points for (2) is bounded and nonempty, there is a unique minimizer for (3). This point $(\mathbf{s}(\mu), \mathbf{y}(\mu))$, the minimizer, is called the *central path point* for parameter μ, for problem (2). Note that as $\mu \to \infty$, the objective function $\mathbf{b}^T\mathbf{y}$ no longer matters and $(\mathbf{s}(\mu), \mathbf{y}(\mu))$ tends to the *analytic center* of the feasible region. As $\mu \to 0$, $(\mathbf{s}(\mu), \mathbf{y}(\mu))$ tends to an optimum since (3) tends to (2) in this case. The definition of the central path is credited to Bayer and Lagarias [2], Megiddo [13], and Sonnevend [21], and it was partially analyzed by McLinden [12] earlier. For the relationship between the central path and the classical log-barrier method, see Wright [30]. For interesting properties of the central path, see the review article by Gonzaga [7].

If we introduce Lagrange multipliers into (3), we find that $(\mathbf{s}(\mu), \mathbf{y}(\mu))$ exactly satisfy the equations:

$$\begin{aligned} \mu A S^{-1}\mathbf{e} &= \mathbf{b}, \\ A^T\mathbf{y} - \mathbf{s} &= \mathbf{c}. \end{aligned}$$

Here and for the rest of the paper, S denotes diag($\mathbf{s}$) and $\mathbf{e}$ denotes the vector of all 1's. These nonlinear equations generally have no closed form solution nor any finite algorithm. Accordingly, an interior point method generates points that are *approximately centered.*

In order to further develop the theory, we consider the *primal* problem associated with (2):

$$\begin{array}{ll} \text{maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array} \tag{4}$$

This problem is written in "standard form." It can be shown that the central path for (4) is essentially identical to the central path for (2) if we make the identification $\mathbf{x} = \mu S^{-1}\mathbf{e}$.

When one is interested in approximate centering, however, several criteria for measuring proximity to the central path are used, depending on whether primal or dual information is available. For the remainder of the paper, all norms are 2-norms.

The first is the dual measure: Given $\mathbf{y}$ that is strictly feasible for the dual, define

$$\delta(\mathbf{y}, \mu) = \|S^{-1}A^T(\mathbf{b}/\mu - AS^{-1}\mathbf{e})\|.$$

Note that δ depends on $\mathbf{s}$ also; given a point $\mathbf{y}$; we let $\mathbf{s} = A^T\mathbf{y} - \mathbf{c}$ to apply this formula. This is the measure that we will use; it requires only dual information. Observe that for an exactly centered point, $\delta(\mathbf{y}, \mu) = 0$.

The second is the primal-dual measure: Given an $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{b}/\mu$ and given $\mathbf{y}$ that is strictly feasible for the dual, define

$$\eta(\mathbf{x}, \mathbf{y}, \mu) = \|S\mathbf{x} - \mathbf{e}\|.$$

Note that $\mu\mathbf{x}$ is a primal point with $A(\mu\mathbf{x}) = \mathbf{b}$.

The third is: Given $\mathbf{y}$ that is strictly feasible for the dual, define

$$\gamma(\mathbf{y}, \mathbf{v}, \mu) = \|\mathbf{v}\|, \quad \text{where } \mathbf{v} \text{ satisfies} \quad \mathbf{b}/\mu - AS^{-1}\mathbf{e} = AS^{-1}\mathbf{v}.$$

In the following lemma, which is proved by Gonzaga [7] and also in Goffin et al. [5], it is shown these three measures are essentially equivalent. Moreover, starting from an approximate central path point, Newton's method can be applied and converges to the central path point at a quadratic rate.

Lemma 1 *Given* $\mathbf{y}$ *that is strictly feasible for the dual and* $\mu > 0$*, we have*

1. *For all* $\mathbf{x}$ *satisfying* $A\mathbf{x} = \mathbf{b}/\mu$*, and for all* $\mathbf{v}$ *satisfying* $\mathbf{b}/\mu - AS^{-1}\mathbf{e} = AS^{-1}\mathbf{v}$*,*

$$\delta(\mathbf{y}, \mu) \leq \eta(\mathbf{x}, \mathbf{y}, \mu) \quad \textit{and} \quad \delta(\mathbf{y}, \mu) \leq \gamma(\mathbf{y}, \mathbf{v}, \mu).$$

2. *There exists an* $\mathbf{x}$ *such that* $A\mathbf{x} = \mathbf{b}/\mu$ *and*

$$\eta(\mathbf{x}, \mathbf{y}, \mu) \leq \delta(\mathbf{y}, \mu).$$

3. *There exist a* $\mathbf{v}$ *such that* $AS^{-1}\mathbf{v} = \mathbf{b}/\mu - AS^{-1}\mathbf{e}$ *and*

$$\|\mathbf{v}\| \leq \delta(\mathbf{y}, \mu).$$

4. *Suppose* $\delta(\mathbf{y}, \mu) < 1$*, and let*

$$\mathbf{r} = (AS^{-2}A^T)^{-1}(\mathbf{b}/\mu - AS^{-1}\mathbf{e})$$

and $\bar{\mathbf{y}} = \mathbf{y} - \mathbf{r}$ *(the Newton step). Then* $\mathbf{y}$ *is strictly feasible for the dual and*

$$\delta(\bar{\mathbf{y}}, \mu) \leq \delta(\mathbf{y}, \mu)^2.$$

The main point of the proof is that solving the least squares problem

$$\text{minimize} \quad \|S\mathbf{x} - \mathbf{e}\|, \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b}/\mu \tag{5}$$

whose solution is

$$\mathbf{x}(\mathbf{y}) = S^{-2}A^T(AS^{-2}A^T)^{-1}(\mathbf{b}/\mu) + S^{-1}P_{AS^{-1}}\mathbf{e}, \tag{6}$$

where P_{AD} is the nullspace projection matrix

$$P_{AD} = I - DA^T(AD^2A^T)^{-1}AD$$

simultaneously minimizes all the quantities in the lemma for a fixed $\mathbf{y}$.

Suppose for some $\bar{\mathbf{y}}$ that $\delta(\bar{\mathbf{y}}, \mu) = \delta < 1$, and let the slacks be $\bar{\mathbf{s}}$. It follows from quadratic convergence that for all i,

$$\prod_{p=0}^{\infty}(1 - \delta^{(2^p)})\bar{s}_i \leq s_i(\mu) \leq \prod_{p=0}^{\infty}(1 + \delta^{(2^p)})\bar{s}_i. \tag{7}$$

Here, $\mathbf{s}(\mu)$ is the slack of the central path point. Multiplying out the right-hand product gives exactly the geometric series

$$1 + \delta + \delta^2 + \cdots.$$

The left-hand product is bounded below by the geometric series

$$1 - \delta - \delta^2 - \cdots$$

although this bound is not exact. Thus, we have the following corollary which is proved in various places, including [7].

Corollary 1 *Let $\bar{\mathbf{y}}$ be an approximately centered point with $\delta(\bar{\mathbf{y}}, \mu) = \delta$ and $\delta < 1$. Let $\bar{\mathbf{s}}$ be the slacks for $\bar{\mathbf{y}}$, and let $(\mathbf{y}(\mu), \mathbf{s}(\mu))$ be the central path point for μ. Then for all $i = 1, \ldots, n$,*

$$(1 - \epsilon)\bar{s}_i \leq s_i(\mu) \leq (1 + \epsilon)\bar{s}_i$$

where $\epsilon = \delta/(1 - \delta)$.

3 The parameters χ_A and $\bar{\chi}_A$

Let A be an $m \times n$ matrix of rank m. Let us define $\mathcal{D}$ to be the set of all positive definite $n \times n$ diagonal matrices. Let us define

$$\bar{\chi}_A = \sup\{\|A^T(ADA^T)^{-1}AD\| : D \in \mathcal{D}\}.$$

Stewart [22] and Todd [24] independently proved that this quantity is always finite. Equivalently, one can show that

$$\chi_A = \sup\{\|(ADA^T)^{-1}AD\| : D \in \mathcal{D}\}.$$

is also finite. These bounds apply for any induced matrix norm, but for this paper, we confine ourselves to 2-norms.

An equivalent way to define these parameters, which is perhaps more relevant for this paper, is in terms of weighted least-squares:

$$\chi_A = \sup\left\{\frac{\|\mathbf{y}\|}{\|\mathbf{c}\|} : \mathbf{y} \text{ minimizes } \|D(A^T\mathbf{y}-\mathbf{c})\| \text{ for some } \mathbf{c} \in \mathbb{R}^n, D \in \mathcal{D}\right\} \tag{8}$$

and

$$\bar{\chi}_A = \sup\left\{\frac{\|A^T\mathbf{y}\|}{\|\mathbf{c}\|} : \mathbf{y} \text{ minimizes } \|D(A^T\mathbf{y}-\mathbf{c})\| \text{ for some } \mathbf{c} \in \mathbb{R}^n, D \in \mathcal{D}\right\} \tag{9}$$

Our complexity bounds will depend on these two parameters. Note that $\bar{\chi}_A$ is unchanged when A is scaled by a constant, whereas χ_A scales inversely. Since our complexity bounds are independent of constant scaling, we naturally can expect the bounds to depend on $\bar{\chi}_A$ and on the product $\chi_A\|A\|$.

A further result concerning these parameters is due to Stewart [22] and O'Leary [16]. Let

$$S = \{\mathbf{s} \in \mathbb{R}^n : \|\mathbf{s}\| = 1 \text{ and } \mathbf{s} = A^T\mathbf{w} \text{ for some } \mathbf{w} \in \mathbb{R}^m\}.$$

Let

$$X = \{\mathbf{x} \in \mathbb{R}^n : AD\mathbf{x} = \mathbf{0} \text{ for some } D \in \mathcal{D}\}.$$

Define

$$\rho_0 = \inf\{\|\mathbf{x}-\mathbf{s}\| : \mathbf{x} \in X, \mathbf{s} \in S\}.$$

Their result is that $1/\bar{\chi}_A = \rho_0$. Thus, we could use the reciprocal of this infimum as an alternative definition of $\bar{\chi}_A$. Indeed, it is more general because it applies to matrices A that are rank deficient.

Further theory concerning these parameters has been developed, for example, by Vavasis [27, 28]. The first paper shows that $\chi_A, \bar{\chi}_A$ act like condition numbers for numerically stable solution of weighted least-squares problems involving A. The second paper estimates variants of these parameters for finite-element problems in terms of the geometry of the finite-element triangulation.

Here is an example of a useful result stated explicitly in [27] but also implicit in [16] and [22].

Lemma 2 *Let B be an $m \times m$ submatrix of A that is nonsingular. Then*

$$\|B^{-1}\| \le \chi_A.$$

PROOF. Let $\mathbf{y}$ be arbitrary, and consider solving $B^T\mathbf{x} = \mathbf{y}$. If we could bound $\|\mathbf{x}\|$ in terms of $\|\mathbf{y}\|$, then we would have a bound on $\|B^{-T}\|$. Let D_ϵ be an $n \times n$ diagonal matrix with 1's in the positions corresponding to columns selected for B, and a small quantity $\epsilon > 0$ in other positions. Then we have $B^T\mathbf{x} = \mathbf{y}$, so $BB^T\mathbf{x} = B\mathbf{y}$, so $\mathbf{x} = (BB^T)^{-1}B\mathbf{y}$. Let $\mathbf{y}$ be extended to an n-vector $\bar{\mathbf{y}}$ by filling in zeros in positions not corresponding to B; then we have $BB^T\mathbf{x} = AD_\epsilon\bar{\mathbf{y}}$. Now, let $\mathbf{x}(\epsilon)$ be the solution to $AD_\epsilon A^T\mathbf{x}(\epsilon) = AD_\epsilon\bar{\mathbf{y}}$, i.e., $\mathbf{x}(\epsilon) = (AD_\epsilon A^T)^{-1}AD_\epsilon\bar{\mathbf{y}}$, so $\|\mathbf{x}(\epsilon)\| \le \chi_A\|\bar{\mathbf{y}}\|$, i.e.,

$$\|\mathbf{x}(\epsilon)\| \le \chi_A\|\mathbf{y}\|.$$

Since $AD_\epsilon A^T$ tends to BB^T as $\epsilon \to 0$ and the limit matrix BB^T is invertible, then $(AD_\epsilon A^T)^{-1}$ tends to $(BB^T)^{-1}$, i.e., $\mathbf{x}(\epsilon)$ tends to $\mathbf{x}$. This shows that $\|\mathbf{x}\| \le \chi_A\|\mathbf{y}\|$. This gives an upper bound on $\|B^{-T}\|$; since the matrix 2-norm is preserved under transpose, we get the same bound for $\|B^{-1}\|$. ∎

This paper provides some additional results concerning these parameters such as Lemma 3. One major gap in our understanding of these parameters is the question of computing them; we do not know of efficient algorithms for computing either parameter. See Section 11 for more comments on this matter.

A final note is that in some earlier work such as [22] and [27] it was assumed that A was an $m \times n$ matrix of rank n, and the definitions of $\chi_A, \bar{\chi}_A$ in those papers involved the transpose matrix rather than the definition used above. Thus, the notation is slightly different.

4 Layered least squares

In this section we describe one main-loop iteration of the LIP algorithm. Assume at the beginning of the iteration that we have a current approximate central path point $(\mathbf{y}, \mu)$ with $\delta(\mathbf{y}, \mu) \leq \delta$. Recall that $\delta(\cdot, \cdot)$ was defined as proximity to the central path in Section 2. We assume throughout that $\delta = 1/4$. Partition the constraints into p *layers* $J_1, \ldots, J_p$ by using $\mathbf{s} = A^T\mathbf{y} - \mathbf{c}$ as follows. Find a permutation π that sorts the slacks into ascending order:

$$s_{\pi(1)} \leq s_{\pi(2)} \leq \cdots \leq s_{\pi(n)}.$$

Let $g > 1$ be a "gap size" defined in terms of A by equation (59) below. Find the leftmost ratio-gap of size greater than g in the sorted slacks, i.e., find the smallest i such that $s_{\pi(i+1)}/s_{\pi(i)} > g$. Then let $J_1 = \{\pi(1), \ldots, \pi(i)\}$. Now, put $\pi(i+1), \pi(i+2), \ldots$ in J_2, until another ratio-gap greater than g is encountered, and so on. Thus, the constraints indexed by J_k for any k are within a factor of g^n of each other, and are separated by more than a factor of g from constraints in J_{k+1}.

To formalize this, define

$$\theta_k = \max_{i \in J_k} s_i$$

and

$$\phi_k = \min_{i \in J_k} s_i.$$

The construction above ensures that for each k,

$$\theta_k < \phi_{k+1}/g.$$

and that

$$\phi_k \leq \theta_k \leq g^n \phi_k.$$

We continue using ϕ_k, θ_k with these definitions throughout the paper.

Define $\mathbf{x}$ as the solution to weighted least-squares problem (5). Note that if $\mathbf{y}$ were exactly centered, then $\mathbf{x}$ would be identically $S^{-1}\mathbf{e}$. Let $X = \operatorname{diag}(\mathbf{x})$, and let

$$\Delta = X^{-1/2}S^{1/2}. \tag{10}$$

Again, note that if $\mathbf{y}$ were exactly centered, we would have $\Delta = S$.

Consider the matrix XS; if $\mathbf{y}$ were exactly centered, then $XS = I$. Since $\mathbf{y}$ is approximately centered, we find from Lemma 1 each diagonal entry of XS is between $1-\delta$ and $1+\delta$. Thus, we have the two inequalities

$$\|X^{1/2}S^{1/2}\| \leq \sqrt{1+\delta}$$

and

$$\|X^{-1/2}S^{-1/2}\| \leq 1/\sqrt{1-\delta},$$

which will be used frequently during the upcoming analysis.

We introduce a new kind of step, the LLS step, to generate a direction $\mathbf{r}$ for the interior point method. Let $A_1, \ldots, A_p$ be the partitioning of the columns of A according to $J_1, \ldots, J_p$. Similarly, let $\Delta_1, \ldots, \Delta_p$ be diagonal matrices that are the submatrices of Δ indexed by $J_1, \ldots, J_p$. Various other vectors and matrices are partitioned in the same way.

The *layered least squares* (LLS) step from point $\mathbf{y}$ is as follows. Define $L_0 = \mathbb{R}^m$. Then for $k = 1, \ldots, p$ compute affine subspace

$$L_k := \{\text{minimizers } \mathbf{r} \text{ of } \|\Delta_k^{-1}(A_k^T\mathbf{r} - \mathbf{s}_k)\| \text{ subject to } \mathbf{r} \in L_{k-1}\}, \tag{11}$$

so that $L_0 \supset L_1 \supset \cdots \supset L_p$. Finally, let $\mathbf{r}$ be the unique element in L_p (unique because A is assumed to be full rank). Note that, for example, if A_1^T spans $\mathbb{R}^m$, then L_1 has a unique element, and thus $L_1 = \cdots = L_p = \{\mathbf{r}\}$.

Thus, the data items necessary for specifying an LLS step are (a) a partitioning $J_1, \ldots, J_p$, (b) positive definite diagonal weight matrix Δ, (c) coefficient matrix A, and (d) right-hand side vector $\mathbf{s}$. The solution is a linear function of $\mathbf{s}$. Thus, by linearity, we can restate the computation of $\mathbf{r}$ as solving an LLS problem for $\bar{\mathbf{y}}$ where on the kth step we minimize $\|\Delta_k^{-1}(A_k^T\bar{\mathbf{y}} - \mathbf{c}_k)\|$, and then taking $\mathbf{r} = \mathbf{y} - \bar{\mathbf{y}}$. This alternate viewpoint is useful for understanding some of the examples below.

The layered least squares step may be thought of as weighted least-squares with the weights in J_1 "infinitely higher" than the weights of J_2, and so on. Indeed, this idea is formalized by Lemma 3 below.

The vector $\mathbf{r}$ can be compared to the Newton step associated with advancing an ordinary path-following method. For instance, if all the constraints were in J_1, then $\mathbf{r}$ would be precisely the direction of the Newton-step for a primal-dual interior point method (e.g., Kojima et al. [10], and Monteiro and Adler [15]).

We now update $\bar{\mathbf{y}} = \mathbf{y} - (1-\alpha)\mathbf{r}$, and $\bar{\mu} = \mu\alpha$. Here, α is a scalar parameter between 0 and 1. The idea is to choose α as close to zero as possible (or perhaps equal to zero, in the case of the final LLS step) as long as the point $\bar{\mathbf{y}}$ remains approximately centered. In Section 5 we give a precise formula for such an α, which involves solving some additional least-squares problems. An inexact line-search is also possible; see (61) below.

After the LLS step we improve the centering of $\bar{\mathbf{y}}$ with three Newton iterations, holding the central path parameter fixed at $\mu\alpha$. Following the recentering, we perform ordinary path-following steps to decrease the central path parameter from $\mu\alpha$ to $\mu\alpha/C(A)$ where $C(A)$ is defined by (60) below.

We now state the LIP method.

Algorithm LIP
Input: $A, \mathbf{b}, \mathbf{c}$ and $(\mathbf{y}, \mu)$ satisfying $\delta(\mathbf{y}, \mu) \leq \delta$.
Output: Optimal points $\mathbf{y}^*$ for (2) and $\mathbf{x}^*$ for the primal (4).

Repeat:

1. Compute LLS solution $\mathbf{r}$ as above.
2. Compute $\alpha \in [0, 1]$, either by the method of Section 5 or by line-search.
3. Update $\bar{\mathbf{y}} := \mathbf{y} - (1-\alpha)\mathbf{r}$, $\bar{\mu} := \mu\alpha$.
4. If $\alpha = 0$, then set $\mathbf{y}^* := \bar{\mathbf{y}}$, compute $\mathbf{x}^*$, and HALT.
5. Recenter $\bar{\mathbf{y}}$ with three Newton steps, yielding $\mathbf{y}'$.
6. From $(\mathbf{y}', \mu\alpha)$, compute approximately centered point $(\mathbf{y}'', \mu\alpha/C(A))$, using existing path-following techniques.
7. Update $\mathbf{y} := \mathbf{y}''$, $\mu := \mu\alpha/C(A)$, and continue.

It is not *a priori* clear that this algorithm has finite termination, let alone polynomial running time. Most of the remainder of the paper will be devoted to establishing various properties of Algorithm LIP. Note that the algorithm assumes that an approximately centered point is part of the input. This in turn implies that the feasible region for (2) is nonempty and contains a strictly interior point. In Section 8 we explain an initialization procedure that makes these assumptions valid. The output $\mathbf{y}^*$ of Algorithm LIP is a dual optimal solution, but below we explain how to use information from the final LLS step to also compute a primal optimal solution $\mathbf{x}^*$.

The above algorithm uses existing path-following techniques applied to the dual problem (2) to reduce the central path parameter from $\mu\alpha$ to

$\mu\alpha/C(A)$. For the analysis of such a technique, see, for example, [19] or [7]. Usual path-following techniques decrease the value of central path parameter by a factor of $1-1/(\mathrm{const}\sqrt{n})$ on each iteration using a Newton step. Thus, $O(n^{1/2}\log C(A))$ traditional interior point steps are used for step 6 above.

We now show that the LLS step is in fact the limiting case of an ordinary weighted least-squares problem.

Lemma 3 *Consider the general LLS problem with input data given by a partition $J_1,\ldots,J_p$, positive definite diagonal weights Δ, $m\times n$ coefficient matrix A of rank m, and right-hand side $\mathbf{c}$. Let $D(\epsilon)$ denote the $n\times n$ diagonal matrix depending on $\epsilon>0$ whose entries are as follows. For $k=1,\ldots,p$, for indices $i\in J_k$, the (i,i) entry is ϵ^k. Let $\mathbf{y}^*$ be the LLS solution, and let $\mathbf{y}(\epsilon)$ be the solution to*

$$\textit{minimize } \|D(\epsilon)\Delta^{-1}(A^T\mathbf{y}-\mathbf{c})\|.$$

Then

$$\lim_{\epsilon\to 0^+}\mathbf{y}(\epsilon)=\mathbf{y}^*.$$

Proof. Let us assume $\epsilon\le 1/2$. If we write out the Lagrange multiplier conditions defining the LLS step, they are:

$$\begin{aligned}
A_1\Delta_1^{-2}A_1^T\mathbf{y}^*-A_1\Delta_1^{-2}\mathbf{c}_1 &= \mathbf{0},\\
A_2\Delta_2^{-2}A_2^T\mathbf{y}^*-A_2\Delta_2^{-2}\mathbf{c}_2 &= A_1\boldsymbol{\lambda}_{2,1},\\
&\;\vdots\\
A_p\Delta_p^{-2}A_p^T\mathbf{y}^*-A_p\Delta_p^{-2}\mathbf{c}_p &= A_1\boldsymbol{\lambda}_{p,1}+\cdots+A_{p-1}\boldsymbol{\lambda}_{p,p-1}.
\end{aligned}\tag{12}$$

We claim these conditions uniquely define $\mathbf{y}^*$, assuming that A is full rank. To see this, suppose $\mathbf{y},\mathbf{y}^*$ are two different solutions. The first equation implies that $\mathbf{y}-\mathbf{y}^*$ lies in the nullspace of A_1^T. The second equation implies that that $A_2\Delta_2^{-2}A_2^T(\mathbf{y}-\mathbf{y}^*)$ lies in the image of A_1. Taking the inner product with $(\mathbf{y}-\mathbf{y}^*)$ shows that in fact $\mathbf{y}-\mathbf{y}^*$ lies in the nullspace of A_2^T. Proceeding in this manner by induction, we conclude that $\mathbf{y}-\mathbf{y}^*$ lies in the nullspace of A_k^T for all k, i.e., $\mathbf{y}-\mathbf{y}^*=\mathbf{0}$.

Therefore, there must exist pseudoinverse matrix P that maps the data vector $A\Delta^{-2}\mathbf{c}$ to the unique solution $\mathbf{y}^*$ of (12). Note that the Lagrange multipliers are not unique.

Write $\mathbf{y}$ in place of $\mathbf{y}(\epsilon)$. Note that $\mathbf{y}$, being the solution to a weighted least-squares problem, satisfies

$$\|\mathbf{y}\| \leq \chi_A \|\mathbf{c}\|$$

independently of ϵ by (8). Now, consider the optimality condition defining $\mathbf{y}$:

$$\epsilon^2 A_1 \Delta_1^{-2} A_1^T \mathbf{y} + \cdots + \epsilon^{2p} A_p \Delta_p^{-2} A_p^T \mathbf{y} - \epsilon^2 A_1 \Delta_1^{-2} \mathbf{c}_1 - \cdots - \epsilon^{2p} A_p \Delta_p^{-2} \mathbf{c}_p = \mathbf{0}. \quad (13)$$

Let C be the maximum of $\|A_k \Delta_k^{-2} A_k^T\| + \|A_k \Delta_k^{-2}\|$ taken over $k = 1, \ldots, p$. If we separate the terms involving A_1, we obtain:

$$\begin{aligned} \|\epsilon^2 (A_1 \Delta_1^{-2} A_1^T \mathbf{y} - A_1 \Delta_1^{-2} \mathbf{c}_1)\| &\leq (\epsilon^4 + \epsilon^6 + \cdots + \epsilon^{2p})(C \chi_A \|\mathbf{c}\|) \\ &\leq \epsilon^4 C' \end{aligned}$$

where C' denotes $2C\chi_A\|\mathbf{c}\|$. Thus,

$$\|A_1 \Delta_1^{-2} A_1^T \mathbf{y} - A_1 \Delta_1^{-2} \mathbf{c}_1\| \leq \epsilon^2 C'.$$

Similarly, from (13) we can obtain the inequality

$$\|A_2 \Delta_2^{-2} A_2^T \mathbf{y} - A_2 \Delta_2^{-2} \mathbf{c}_2 + A_1 \Delta_1^{-2} A_1^T \mathbf{y}/\epsilon^2 - A_1 \Delta_1^{-2} \mathbf{c}_1/\epsilon^2\| \leq \epsilon^2 C'$$

which we can write as

$$\|A_2 \Delta_2^{-2} A_2^T \mathbf{y} - A_2 \Delta_2^{-2} \mathbf{c}_2 - A_1 \mathbf{u}\| \leq \epsilon^2 C'.$$

Proceeding in this manner, we see that $\mathbf{y}$, accompanied by appropriate choices for the Lagrange multipliers, is an $\epsilon^2 C'$-approximate solution for (12). Since these conditions have a pseudoinverse P, we conclude that

$$\|\mathbf{y} - \mathbf{y}^*\| \leq \|P\| \epsilon^2 C'.$$

This proves the lemma. ∎

Note that as a corollary, we can conclude from (8) and (9) that the solution $\mathbf{y}$ to the LLS problem is bounded in norm by

$$\|\mathbf{y}\| \leq \chi_A \|\mathbf{c}\| \quad (14)$$

and

$$\|A^T \mathbf{y}\| \leq \bar{\chi}_A \|\mathbf{c}\| \quad (15)$$

independently of the choice of weights and partitions.

5 On the centering of the LLS step

In this section we propose a specific rule for choosing $\bar{\alpha} \in [0,1]$ in the LIP algorithm, and we prove the first main theorem concerning the method. This theorem states that for all choices of α between $\bar{\alpha}$ and 1, the point $\bar{\mathbf{y}} = \mathbf{y} - (1-\alpha)\mathbf{r}$ is approximately centered for central path parameter $\mu\alpha$.

First, we define $\bar{\alpha}$. Define an auxiliary sequence of vector $\mathbf{r}^{(k)}$, $k = 1, \ldots, p$ as follows. Vector $\mathbf{r}^{(k)}$ is also defined as the solution to a LLS problem with the same coefficients used for $\mathbf{r}$ but different right-hand sides. Specifically, the LLS data for $\mathbf{r}^{(k)}$ is the same partitions $J_1, \ldots, J_p$, the same weight matrix Δ, the same coefficient matrix A, but right-hand side vector $(\mathbf{s}_1^T, \ldots, \mathbf{s}_{k-1}^T, \mathbf{0}, \ldots, \mathbf{0})^T$, i.e., zeros filled in for positions indexed by $J_k \cup \cdots \cup J_p$. For example, note that $\mathbf{r}^{(1)} = \mathbf{0}$. Intuitively, $\mathbf{r}^{(k)}$ "agrees" with $\mathbf{r}$ in affine spaces up to L_{k-1} in (11), but from k on, we choose $\mathbf{r}^{(k)}$ so as to cause minimal weighted change on constraints $A_k^T\mathbf{y} - \mathbf{c}_k, A_{k+1}^T\mathbf{y} - \mathbf{c}_{k+1}, \ldots, A_p^T\mathbf{y} - \mathbf{c}_p$.

Define a sequence of nonorthogonal projection operators $P_1, \ldots, P_p$ as follows. For matrix A, find a subset of columns of A_1 that span the range of A_1 but are linearly independent. Next, find a subset of columns B_2 of A_2 such that $[B_1, B_2]$ are linearly independent and span the range of $[A_1, A_2]$.

Repeat this process until we obtain an $n \times n$ nonsingular matrix $[B_1, \ldots, B_p]$. Denote this matrix as B. Now, for $k = 1, \ldots, p$, define $P_k = [0, B_k, 0]B^{-1}$, where the sizes of the zero blocks are chosen so that B_k in $[0, B_k, 0]$ lies in the same position as B_k in B. (If B_k has no columns, take $P_k = 0$.)

These projection operators have the following properties:

1. They are truly projections because $P_k P_k = P_k$.
2. They are complementary in the sense that $P_k P_l = 0$ if $k \neq l$.
3. The image of P_k lies in the span of A_k.
4. The sum $P_1 + \cdots + P_p$ is the identity operator.

The other important property of these projections is stated in Lemma 5 below.

Define δ_k, for $k = 1, \ldots, p$, as follows:

$$\delta_k = \|\Delta_k^{-1} A_k^T (\mathbf{r} - \mathbf{r}^{(k)})\|. \tag{16}$$

Define

$$\epsilon_k = \|\Delta_k^{-1}(A_k^T\mathbf{r} - \mathbf{s}_k)\|. \tag{17}$$

Intuitively, δ_k measures how much improvement we can make decreasing $A_k^T\mathbf{y} - \mathbf{c}_k$ on the kth LLS step, and ϵ_k is the residual from the kth LLS step defining $\mathbf{r}$.

We define α_k, for $k = 1, \ldots, p$, as follows. If

$$\delta_k \geq \frac{1}{30\sqrt{n}} \tag{18}$$

then we take

$$\alpha_k = \min(1, 15\epsilon_k\sqrt{n}). \tag{19}$$

Else we take

$$\alpha_k = \min\left(1, \frac{30n^{1.5}\theta_k\chi_A \cdot \|P_k\mathbf{b}\|}{\mu}\right). \tag{20}$$

Let us say that an index k is of **Type I** if (18) holds and α_k is defined by (19), else k is of **Type II**.

Now we define scalar $\beta_{k,l}$ for all choices of (k,l) such that k is Type I, l is Type II, and $k > l$. The scalar $\beta_{k,l}$ is defined to be

$$\beta_{k,l} = \min(1, 30\bar{\chi}_A\theta_l n^2/\phi_k). \tag{21}$$

Now we choose $\bar{\alpha}$ to be the maximum among all the α_k's and the $\beta_{k,l}$'s (which is always between 0 and 1). We assume for now that the parameter value α used in Algorithm LIP is precisely this $\bar{\alpha}$.

The upcoming analysis of the properties of $\bar{\alpha}$ is rather involved, so we start by providing four examples of the LLS step to help the reader to gain intuition about this parameter. We discuss further properties these examples in Section 6.

Examples.

1. The first example is the case when there is no near-degeneracy in the problem. Consider

$$\begin{array}{ll} \text{minimize} & y_1 + 2y_2 \\ \text{subject to} & 0 \leq y_1 \leq 1, \\ & 0 \leq y_2 \leq 1. \end{array}$$

Suppose we are at an approximately centered point such as $\mathbf{y} = (2\eta, \eta)$ where $\eta > 0$ is fairly small. Then $J_1 = \{y_1 \geq 0, y_2 \geq 0\}$ and $J_2 = \{y_1 \leq 1, y_2 \leq 1\}$. In this case, J_1 is spanning, so the LLS step is completely determined by J_1. Furthermore, the constraints in J_1 are linearly independent, so the least-squares computation associated with J_1 is solved as an equation. In particular, the value $\mathbf{r}$ computed is identically equal to $\mathbf{y}$. Note that $\alpha_1 = 0$ because the least-squares problem has no residual. Also, $\alpha_2 = 0$ since $P_2 = 0$. Finally, there is no $\beta_{k,l}$ defined since there is no k of Type I, l of Type I, such that $k > l$. Thus, $\bar{\alpha} = 0$. Note that $\mathbf{y} - \mathbf{r}$ is the exact optimum solution of the LP.

2. The next example is the problem illustrated in Figure 1. Suppose $\epsilon > 0$ in the figure is very small, and suppose $\mathbf{y}$ is a point like point B. Then the near-active constraints will be $J_1 = \{y_1 \geq 0, y_2 \geq 0, y_1 + y_2 \geq \epsilon\}$. The second layer will be $J_2 = \{y_1 \leq 1, y_2 \leq 1\}$. Note that A_1 spanning, so the step $\mathbf{r}$ is determined entirely by a weighted least-squares problem involving J_1. Thus, the vector $-\mathbf{r}$, an offset from $\mathbf{y}$, connects $\mathbf{y}$ to a point that is $O(\epsilon)$ away from $(0, 0)$ (possibly infeasible). Furthermore, J_1 is Type I and J_2 is Type II. The parameters are $\alpha_1 = c \cdot \epsilon$ for some constant (the least-squares residual) and $\alpha_2 = 0$ (note that $P_2 = 0$ in this case). There is no $\beta_{k,l}$ for this example. Thus, $\bar{\alpha} = O(\epsilon)$. Thus, the LLS step takes us to a point that is $O(\epsilon)$ away from the origin, that is, a point like C in the figure.

3. For the third example, consider the problem

$$\begin{array}{rl} \text{minimize} & y_1 + \epsilon y_2 \\ \text{subject to} & 0 \leq y_1 \leq 1, \\ & 0 \leq y_2 \leq 1. \end{array}$$

Assume ϵ is very close to zero, but may have either sign. Suppose we are at an approximately centered point like $(\eta, 0.5)$ where $\eta > 0$ is small but much larger than $|\epsilon|$. In this case, $J_1 = \{y_1 \geq 0\}$ and $J_2 = \{y_2 \geq 0, y_2 \leq 1, y_1 \leq 1\}$. Here, J_1 is Type I and J_2 is Type II. The least-squares computation associated with J_1 has a exact solution, so that $r_1 = y_1 = \eta$, i.e., $\mathbf{y} - \mathbf{r}$ exactly solves the constraint in J_1. The second component of $\mathbf{r}$ is left undetermined by J_1. The least-squares computation associated with J_2 causes a small perturbation, so $\mathbf{r} = (\eta, \theta)$, where θ is on the order of ϵ. Note that the first least-squares computation was exact, so $\alpha_1 = 0$. For J_2, which is Type II, we observe that $\alpha_2 = O(\|P_2\mathbf{b}\|)$. For this simple example, P_2 is projection onto the second coordinate, so we see that $\alpha_2 = O(|\epsilon|)$. Thus, $\bar{\alpha} = O(|\epsilon|)$, and we follow $\mathbf{r}$ so that the LLS step carries us to a point approximately $(O(|\epsilon|), 0.5)$. Again, we prove in this section that such a point is approximately centered.

4. For the last example, consider the problem

$$\begin{array}{rl} \text{minimize} & y_1 \\ \text{subject to} & 0 \le y_2 \le \epsilon, \\ & y_1 \ge y_2. \end{array}$$

Here, $\epsilon > 0$ is very small. Suppose we are at an approximately centered point such as $(100, \epsilon/2)$. In this case, $J_1 = \{y_2 \ge 0, y_2 \le \epsilon\}$, and $J_2 = \{y_1 \ge y_2\}$. Observe that J_1 is of Type II, and the least squares-computation associated with J_1 forces r_2 to be a number η very close to zero—much smaller than ϵ. The least-squares computation associated with J_2 determines r_1 and has no residual. Thus, $\mathbf{r}$ is something like $(-100 + \eta - \epsilon/2, \eta)$, so that $\mathbf{y} - \mathbf{r}$ satisfies $y_1 \ge y_2$ as an equation. Note that $\alpha_1 = 0$ because $\mathbf{b}$ is in the nullspace of P_1. Note also that $\alpha_2 = 0$ because the least-squares computation has no residual. However, in this case $\beta_{2,1}$ is proportional to $\epsilon/100$, so $\bar{\alpha}$ is proportional to $\epsilon/100$. After updating $\mathbf{y}$, we are at a point close to $(O(\epsilon), \epsilon/2)$.

Before stating the main theorem, we have a few preliminary lemmas.

Lemma 4 *Suppose* $\mathbf{v} = A\mathbf{x}$ *for two arbitrary vectors* $\mathbf{v}$ *and* $\mathbf{x}$*. Let* B *be a subset of* m *linearly independent columns of* A*. Let* $\mathbf{u}$ *be the (unique) solution to* $B\mathbf{u} = \mathbf{v}$*. Then*

$$\|\mathbf{u}\| \le \bar{\chi}_A \|\mathbf{x}\|.$$

PROOF. Since $B\mathbf{u} = \mathbf{v} = A\mathbf{x}$,

$$\mathbf{u} = B^{-1}A\mathbf{x} = B^T(BB^T)^{-1}A\mathbf{x}$$

By filling in zeros, $B^T(BB^T)^{-1}A\mathbf{x}$ can be extended as

$$DA^T(ADA^T)^{-1}A\mathbf{x}$$

where D is a diagonal matrix with $d_{jj}=1$ if column j is in B and $d_{jj}=\epsilon>0$ otherwise. Thus, we have

$$\|\mathbf{u}\| = \lim_{\epsilon\to 0^+} \|DA^T(ADA^T)^{-1}A\mathbf{x}\| \le \bar{\chi}_A\|\mathbf{x}\|.$$

∎

Lemma 5 *Suppose that $\mathbf{q}=P_lA'\mathbf{x}$ for vectors $\mathbf{q},\mathbf{x}$, where A' is a subset of columns of A. Then we can also find a vector $\mathbf{w}$ such that $\mathbf{q}=A_l\mathbf{w}$ and*

$$\|\mathbf{w}\| \le \bar{\chi}_A\|\mathbf{x}\|.$$

PROOF. It is sufficient to prove the case that $A'=A$, since we can always pad $\mathbf{x}$ with zeros to make $A'=A$. By definition of P_l, $\mathbf{q}=[0,B_l,0]B^{-1}A\mathbf{x}$. This is the same as writing $\mathbf{q}=B_l\mathbf{y}$, where $\mathbf{y}$ is a subvector of $\hat{\mathbf{y}}=B^{-1}A\mathbf{x}$. That is, $B\hat{\mathbf{y}}=A\mathbf{x}$. Recall that B is a subset of columns of A that is a basis. Thus, by Lemma 4, we can see that $\|\hat{\mathbf{y}}\|\le\bar{\chi}_A\|\mathbf{x}\|$. Then $\|\mathbf{y}\|\le\|\hat{\mathbf{y}}\|$.

Thus, we have shown that if $\mathbf{q}=P_lA\mathbf{x}$, then $\mathbf{q}=B_l\mathbf{y}$ with $\|\mathbf{y}\|\le\bar{\chi}_A\|\mathbf{x}\|$. We can also claim that $\mathbf{q}=A_l\mathbf{w}$ with the same bound on $\mathbf{w}$, since B_l is a subset of the columns of A_l. ∎

Lemma 6 *For each k,*

$$\|\Delta_k^{-1}A_k^T\mathbf{r}^{(k)}\| \le \sqrt{1+\delta}\cdot\sqrt{n}\bar{\chi}_A/g.$$

PROOF. Since $\mathbf{r}^{(k)}$ is an LLS solution, then by (15),

$$\|A^T\mathbf{r}^{(k)}\| \le \bar{\chi}_A\|\mathbf{u}\|,$$

where $\mathbf{u}$ is a vector that agrees with $\mathbf{s}$ in positions indexed by $J_1\cup\cdots\cup J_{k-1}$ and has zeros in the remaining positions, By definition of $\mathbf{u}$, we can see that $\|\mathbf{u}\|\le\sqrt{n}\theta_{k-1}$, so

$$\|A^T\mathbf{r}^{(k)}\| \le \sqrt{n}\bar{\chi}_A\theta_{k-1}.$$

Thus,

$$\begin{aligned}
\|\Delta_k^{-1} A_k^T \mathbf{r}^{(k)}\| &\le \|\Delta_k^{-1}\| \cdot \|A_k^T \mathbf{r}^{(k)}\| \\
&= \|X_k^{1/2} S_k^{-1/2}\| \cdot \|A_k^T \mathbf{r}^{(k)}\| \\
&\le \|X_k^{1/2} S_k^{1/2}\| \cdot \|S_k^{-1}\| \cdot \|A_k^T \mathbf{r}^{(k)}\| \\
&\le \sqrt{1+\delta} \cdot (1/\phi_k) \sqrt{n} \bar{\chi}_A \theta_{k-1} \\
&\le \sqrt{1+\delta} \cdot \sqrt{n} \bar{\chi}_A / g.
\end{aligned}$$

∎

Now for the first main result of the paper.

Theorem 1 *Let* $\alpha \in [\bar{\alpha}, 1]$ *be chosen arbitrarily. In the case that* $\bar{\alpha} = 0$*, assume also that* $\alpha > 0$*. Then* $\bar{\mathbf{y}}$*, defined to be* $\mathbf{y} - (1-\alpha)\mathbf{r}$*, is a strictly feasible point. Furthermore,*

$$\delta(\bar{\mathbf{y}}, \mu\alpha) \le 2\delta + (1+\delta)/5.$$

PROOF. Note that we are assuming $\delta \le 0.25$, that is, $\delta(\mathbf{y}, \mu) \le 0.25$. Thus, the bound on the right-hand side is at most 0.75. We denote the slack vector $A^T\bar{\mathbf{y}} - \mathbf{c}$ by $\bar{\mathbf{s}}$. Note that

$$\bar{\mathbf{s}} = A^T(\mathbf{y} - (1-\alpha)\mathbf{r}) - \mathbf{c} = \mathbf{s} - (1-\alpha)A^T\mathbf{r}. \tag{22}$$

Observe that if $\alpha_k = 1$ in (19) or (20), or $\beta_{k,l} = 1$ in (21), then $\bar{\alpha} = 1$, so $\bar{\mathbf{y}} = \mathbf{y}$ and the claim is trivial. Thus assume that $\alpha < 1$ for this proof, and that none of the minima in (19), (20) or (21) is achieved at 1.

Before proving the theorem, we develop several inequalities. Let I and II denote the partition of $\{1, \ldots, p\}$ into Type I and Type II layers. For $k \in I$, since $\|\Delta_k^{-1}(A_k^T\mathbf{r} - \mathbf{s}_k)\| = \epsilon_k$, we have

$$\begin{aligned}
\|S_k^{-1} A_k^T \mathbf{r} - \mathbf{e}\| &= \|(X_k S_k)^{-1/2} \Delta_k^{-1}(A_k^T\mathbf{r} - \mathbf{s}_k)\| \\
&\le \|(X_k S_k)^{-1/2}\| \cdot \|\Delta_k^{-1}(A_k^T\mathbf{r} - \mathbf{s}_k)\| \\
&\le \epsilon_k / \sqrt{1-\delta}.
\end{aligned}$$

This, with (22), means that

$$\begin{aligned}
\|S_k^{-1}\bar{\mathbf{s}}_k - \alpha\mathbf{e}\| &= \|\mathbf{e} - (1-\alpha)S_k^{-1}A_k^T\mathbf{r} - \alpha\mathbf{e}\| \\
&= (1-\alpha)\|\mathbf{e} - S_k^{-1}A_k^T\mathbf{r}\| \\
&\le (1-\alpha)\epsilon_k/\sqrt{1-\delta} \\
&\le \epsilon_k/\sqrt{1-\delta}.
\end{aligned}$$

This implies

$$\left\|\frac{1}{\alpha}S_k^{-1}\bar{\mathbf{s}}_k - \mathbf{e}\right\| \le \frac{\epsilon_k}{\alpha\sqrt{1-\delta}} \le \frac{1}{15\sqrt{n}\sqrt{1-\delta}}. \tag{23}$$

Note that we used the fact that $(\epsilon_k/\alpha) \le 1/(15\sqrt{n})$, which follows from the fact that $\alpha \ge \alpha_k = 15\sqrt{n}\epsilon_k$, taken from (19). Therefore, for $i \in J_k$ and $k \in I$

$$\frac{\mathbf{a}_i^T\bar{\mathbf{y}} - c_i}{\alpha(\mathbf{a}_i^T\mathbf{y} - c_i)} = \frac{\bar{s}_i}{\alpha s_i} \le 1 + \frac{1}{15\sqrt{n}\sqrt{1-\delta}} \le 2, \tag{24}$$

where we require

$$\frac{1}{15\sqrt{n}\sqrt{1-\delta}} < 1. \tag{25}$$

Obviously, $\delta = 1/4$ meets the requirement.

For $k \in II$, observe from Lemma 6 and (16) that

$$\|\Delta_k^{-1}A_k^T\mathbf{r}\| \le \|\Delta_k^{-1}A_k^T\mathbf{r} - \Delta_k^{-1}A_k^T\mathbf{r}^{(k)}\| + \|\Delta_k^{-1}A_k^T\mathbf{r}^{(k)}\| \le \delta_k + \sqrt{1+\delta}\cdot\sqrt{n}(\bar{\chi}_A/g).$$

Since k is a Type-II layer, we have an upper bound on δ_k given by the opposite of (18). Similarly, we have the same upper bound on $\sqrt{1+\delta}\cdot\sqrt{n}(\bar{\chi}_A/g)$ if we assume that

$$g \ge 30\sqrt{1+\delta}\cdot n\bar{\chi}_A. \tag{26}$$

When we define g in (59) below, we will take (26) and other similar inequalities into account. Assuming (26) and the opposite of (18), we conclude that

$$\|\Delta_k^{-1}A_k^T\mathbf{r}\| \le \frac{1}{15\sqrt{n}}. \tag{27}$$

This means

$$\|S_k^{-1}\bar{\mathbf{s}}_k - \mathbf{e}\| = \|\mathbf{e} - (1-\alpha)S_k^{-1}A_k^T\mathbf{r} - \mathbf{e}\|$$

$$
\begin{aligned}
&= \|(1-\alpha)S_k^{-1}A_k^T\mathbf{r}\| \\
&= \|(1-\alpha)(X_kS_k)^{-1/2}\Delta_k^{-1}A_k^T\mathbf{r}\| \\
&\le (1-\alpha)\|(X_kS_k)^{-1/2}\| \cdot \|\Delta_k^{-1}A_k^T\mathbf{r}\| \\
&\le \frac{(1-\alpha)}{15\sqrt{n}\sqrt{1-\delta}} \\
&\le \frac{1}{15\sqrt{n}\sqrt{1-\delta}}.
\end{aligned}
$$

Thus, for $i \in J_k$ and $k \in II$

$$
\begin{aligned}
\left|\frac{\mathbf{a}_i^T\bar{\mathbf{y}} - c_i}{\mathbf{a}_i^T\mathbf{y} - c_i} - 1\right| &= \left|\frac{\bar{s}_i}{s_i} - 1\right| \\
&\le \frac{1}{15\sqrt{n}\sqrt{1-\delta}} \qquad (28) \\
&< 1. \qquad (29)
\end{aligned}
$$

Again, we used requirement (25) to derive the last line. Note that (23) and (29) show that $\bar{\mathbf{y}}$ is strictly feasible because they imply $\bar{\mathbf{s}} > \mathbf{0}$.

Observe that we have proved the following statement: for constraints in Type-I layers, along the segment connecting $\mathbf{y}$ to $\mathbf{y} - (1-\alpha)\mathbf{r}$, the slacks decrease roughly proportionate to α. For Type-II layers, the slacks stay roughly constant. These observations are crucial for most of the remainder of the paper.

As we discussed earlier, to prove the theorem, it is sufficient to prove that the central path residual

$$
\mathbf{q} = \mathbf{b}/\mu - \alpha A\bar{S}^{-1}\mathbf{e}
$$

can be written as $\alpha A\bar{S}^{-1}\mathbf{v}$ where $\|\mathbf{v}\| \le \beta$, where we can take $\beta = 3/4$. This establishes the approximate centering property.

Note that $\mathbf{b}/\mu = A\mathbf{x}$, where $\mathbf{x}$ is given by (6), so we can write $\mathbf{q}$ as

$$
\begin{aligned}
\mathbf{q} &= \sum_{k=1}^{p}(A_k\mathbf{x}_k - \alpha A_k\bar{S}_k^{-1}\mathbf{e}) \\
&= \sum_{k\in I}(A_k\mathbf{x}_k - \alpha A_k\bar{S}_k^{-1}\mathbf{e}) + \sum_{k\in II}(A_k\mathbf{x}_k - \alpha A_k\bar{S}_k^{-1}\mathbf{e}) \\
&= \sum_{k\in I}(A_k\mathbf{x}_k - \alpha A_k\bar{S}_k^{-1}\mathbf{e}) + \sum_{k\in II}(\alpha A_k\mathbf{x}_k - \alpha A_k\bar{S}_k^{-1}\mathbf{e}) + \sum_{k\in II}(1-\alpha)A_k\mathbf{x}_k \\
&= \mathbf{q}_I + \mathbf{q}_{II} + \mathbf{q}'_{II},
\end{aligned}
$$

where

$$\mathbf{q}_I = \sum_{k\in I}(A_k\mathbf{x}_k - \alpha A_k\bar{S}_k^{-1}\mathbf{e}),$$

$$\mathbf{q}_{II} = \sum_{k\in II}(\alpha A_k\mathbf{x}_k - \alpha A_k\bar{S}_k^{-1}\mathbf{e}),$$

and

$$\mathbf{q}'_{II} = \sum_{k\in II}(1-\alpha)A_k\mathbf{x}_k.$$

We will prove that

$$\mathbf{q}_I = \alpha A\bar{S}^{-1}\mathbf{v}_I, \quad \mathbf{q}_{II} = \alpha A\bar{S}^{-1}\mathbf{v}_{II}, \quad \text{and} \quad \mathbf{q}'_{II} = \alpha A\bar{S}^{-1}\mathbf{v}'_{II} \tag{30}$$

where

$$\|\mathbf{v}_I\| \le \delta + (1+\delta)/15, \quad \|\mathbf{v}_{II}\| \le \delta + (1+\delta)/15, \quad \text{and} \quad \|\mathbf{v}'_{II}\| \le (1+\delta)/15.$$

This proves the theorem by taking $\mathbf{v} = \mathbf{v}_I + \mathbf{v}_{II} + \mathbf{v}'_{II}$, with

$$\|\mathbf{v}\| \le 2\delta + (1+\delta)/5.$$

If $\delta = 1/4$, then $\|\mathbf{v}\| \le 3/4$.

First consider $\mathbf{q}_I$. Concatenate the matrices A_k for $k \in I$ into a large submatrix A_I, and index other variables similarly. Then, from (22)

$$\begin{aligned}
\mathbf{q}_I &= A_I\mathbf{x}_I - \alpha A_I\bar{S}_I^{-1}\mathbf{e} \\
&= A_I\bar{S}_I^{-1}(X_I\bar{\mathbf{s}}_I - \alpha\mathbf{e}) \\
&= A_I\bar{S}_I^{-1}(X_I(\mathbf{s}_I - (1-\alpha)A_I^T\mathbf{r}) - \alpha\mathbf{e}) \\
&= A_I\bar{S}_I^{-1}(X_I\mathbf{s}_I - (1-\alpha)X_IA_I^T\mathbf{r} - \alpha\mathbf{e}) \\
&= A_I\bar{S}_I^{-1}[\alpha(X_I\mathbf{s}_I - \mathbf{e}) + (1-\alpha)(X_I\mathbf{s}_I - X_IA_I^T\mathbf{r})] \\
&= \alpha A_I\bar{S}_I^{-1}[(X_I\mathbf{s}_I - \mathbf{e}) + \frac{1-\alpha}{\alpha}(X_I\mathbf{s}_I - X_IA_I^T\mathbf{r})].
\end{aligned}$$

Note from (5) that

$$\|X_I\mathbf{s}_I - \mathbf{e}\| \le \delta,$$

and

$$\left\|\frac{1-\alpha}{\alpha}(X_I\mathbf{s}_I - X_IA_I^T\mathbf{r})\right\| \le \frac{\|(X_I\mathbf{s}_I - X_IA_I^T\mathbf{r})\|}{\alpha}$$

$$
\begin{aligned}
&= \frac{\|(X_I S_I)^{1/2} \Delta_I^{-1} (\mathbf{s}_I - A_I^T \mathbf{r})\|}{\alpha} \\
&\le \frac{\|(X_I S_I)^{1/2}\| \cdot \|\Delta_I^{-1} (\mathbf{s}_I - A_I^T \mathbf{r})\|}{\alpha} \\
&\le \frac{\sqrt{1+\delta} \cdot \|\Delta_I^{-1} (\mathbf{s}_I - A_I^T \mathbf{r})\|}{\alpha} \\
&= \frac{\sqrt{1+\delta} \cdot \sqrt{\sum_{k \in I} \|\Delta_k^{-1} (\mathbf{s}_k - A_k^T \mathbf{r})\|^2}}{\alpha} \\
&= \frac{\sqrt{1+\delta} \cdot \sqrt{\sum_{k \in I} \epsilon_k^2}}{\alpha} \\
&\le \frac{\sqrt{1+\delta} \cdot \sqrt{n} \max_{k \in I} \epsilon_k}{\alpha} \\
&\le \frac{\sqrt{1+\delta} \cdot \sqrt{n} \max_{k \in I} \epsilon_k}{\max_{k \in I} \alpha_k} \\
&\le (\sqrt{1+\delta})/15.
\end{aligned}
$$

The last inequality due to the choice of α_k in (19). Thus, we can take

$$\mathbf{v}_I = (X_I \mathbf{s}_I - \mathbf{e}) + \frac{1-\alpha}{\alpha}(X_I \mathbf{s}_I - X_I A_I^T \mathbf{r})$$

and fill zeros up to dimension n, and have

$$\|\mathbf{v}_I\| \le \delta + (\sqrt{1+\delta})/15.$$

This concludes the first term of (30).

Second, consider $\mathbf{q}_{II}$. Define A_{II} as above. Then, from (22)

$$
\begin{aligned}
\mathbf{q}_{II} &= \alpha A_{II} (\mathbf{x}_{II} - \bar{S}_{II}^{-1} \mathbf{e}) \\
&= \alpha A_{II} \bar{S}_{II}^{-1} (X_{II} \bar{\mathbf{s}}_{II} - \mathbf{e}) \\
&= \alpha A_{II} \bar{S}_{II}^{-1} (X_{II} (\mathbf{s}_{II} - (1-\alpha) A_{II}^T r) - \mathbf{e}) \\
&= \alpha A_{II} \bar{S}_{II}^{-1} [(X_{II} \mathbf{s}_{II} - \mathbf{e}) - (1-\alpha) X_{II} A_{II}^T \mathbf{r}].
\end{aligned}
$$

Note again that

$$\|X_{II} \mathbf{s}_{II} - \mathbf{e}\| \le \delta,$$

and from relation (27)

$$\begin{aligned}\|(1-\alpha)X_{II}A_{II}^T\mathbf{r}\| &= \|(1-\alpha)(X_{II}S_{II})^{1/2}\Delta_{II}^{-1}A_{II}^T\mathbf{r}\| \\ &\le \|(X_{II}S_{II})^{1/2}\|\cdot\|\Delta_{II}^{-1}A_{II}^T\mathbf{r}\| \\ &= \|(X_{II}S_{II})^{1/2}\|\cdot\sqrt{\sum_{k\in II}\|\Delta_k^{-1}A_k^T\mathbf{r}\|^2} \\ &\le \frac{\sqrt{1+\delta}\cdot\sqrt{n}}{15\sqrt{n}} \\ &\le (\sqrt{1+\delta})/15.\end{aligned}$$

Thus, we can take

$$\mathbf{v}_{II} = (X_{II}\mathbf{s}_{II}-\mathbf{e}) - (1-\alpha)X_{II}A_{II}^T\mathbf{r}$$

and fill zeros up to dimension n, and have

$$\|\mathbf{v}_{II}\| \le \delta + (\sqrt{1+\delta})/15.$$

This concludes the second term of (30).

Third, consider

$$\begin{aligned}\mathbf{q}'_{II} &= \sum_{k\in II}(1-\alpha)A_k\mathbf{x}_k \\ &= \sum_{l=1}^{p}P_l[\sum_{k\in II}(1-\alpha)A_k\mathbf{x}_k] \\ &= \sum_{l=1}^{p}P_l[\sum_{k\ge l,k\in II}(1-\alpha)A_k\mathbf{x}_k].\end{aligned}$$

The last relation holds because $P_lA_k = 0$ when $k < l$, i.e., A_k lies in the span of $B_1,\ldots,B_k$. Let

$$\mathbf{q}_l = P_l[\sum_{k\in II}(1-\alpha)A_k\mathbf{x}_k] = P_l[\sum_{k\ge l,k\in II}(1-\alpha)A_k\mathbf{x}_k].$$

If we can prove that for $1\le l\le p$

$$\mathbf{q}_l = \alpha A_l\bar{S}_l^{-1}\mathbf{v}_l \quad \text{and} \quad \|\mathbf{v}_l\| \le (1+\delta)/(15\sqrt{n}), \tag{31}$$

then, we define

$$\mathbf{v}'_{II} = (\mathbf{v}_1^T, \mathbf{v}_2^T, \ldots, \mathbf{v}_l^T)^T, \quad \text{and} \quad \mathbf{q}'_{II} = \alpha A \bar{S}^{-1} \mathbf{v}'_{II},$$

and establish

$$\|\mathbf{v}'_{II}\| \le \sqrt{\sum_{l=1}^{p} \|\mathbf{v}_l\|^2} \le (1+\delta)/15,$$

which concludes (30).

Consider the following two cases for $\mathbf{q}_l$.

CASE 1, l is of Type I. In this case, we must have

$$\begin{aligned} \mathbf{q}_l &= P_l[\sum_{k \in II} (1-\alpha) A_k \mathbf{x}_k] \\ &= P_l[\sum_{k>l, k\in II} (1-\alpha) A_k \mathbf{x}_k]. \end{aligned}$$

Note we have

$$\begin{aligned} \|(1-\alpha)\mathbf{x}_k\| &= \|(1-\alpha) X_k S_k (S_k)^{-1} \mathbf{e}\| \\ &\le \|X_k S_k\| \cdot \|\mathbf{e}\| \cdot \|S_k^{-1}\| \\ &\le (1+\delta)\sqrt{n}/\phi_k. \end{aligned}$$

Now we apply Lemma 5 to conclude that

$$\mathbf{q}_l = \sum_{k>l, k\in II} A_l \mathbf{w}_k = A_l(\sum_{k>l, k\in II} \mathbf{w}_k), \quad \|\mathbf{w}_k\| \le \bar{\chi}_A (1+\delta)\sqrt{n}/\phi_k.$$

Next, observe that

$$\mathbf{q}_l = \alpha A_l^T \bar{S}_l^{-1} (\sum_{k>l, k\in II} \mathbf{u}_k) \quad \text{where} \quad \mathbf{u}_k = \bar{S}_l \mathbf{w}_k / \alpha$$

so

$$\begin{aligned} \|\mathbf{u}_k\| &\le \|\bar{S}_l/\alpha\| \cdot \|\mathbf{w}_k\| \\ &= \max_{i\in J_l} \frac{\bar{s}_i}{\alpha} \cdot \|\mathbf{w}_k\| \\ &\le \max_{i \in J_l} 2 s_i \cdot \|\mathbf{w}_k\| \\ &\le 2\theta_l \cdot \bar{\chi}_A (1+\delta)\sqrt{n}/\phi_k \\ &\le 2\bar{\chi}_A (1+\delta)\sqrt{n}/g. \end{aligned}$$

Note that to derive the third line we used (24) since l is of Type I. The last line was derived because $g\theta_l \leq \phi_k$ since $k > l$.

If we require that

$$g \geq 30\bar{\chi}_A n^2 \tag{32}$$

then we can claim that $\|\mathbf{u}_k\| \leq (1+\delta)/(15n^{1.5})$. Then, we let

$$\mathbf{v}_l = (\sum_{k>l, k\in II} \mathbf{u}_k),$$

and have

$$\mathbf{q}_l = \alpha A_l^T \bar{S}_l^{-1} \mathbf{v}_l, \quad \|\mathbf{v}_l\| \leq (1+\delta)/(15\sqrt{n}),$$

which gives (31) and concludes CASE 1.

CASE 2, l is of Type II. In this case, we have

$$\begin{aligned} \mathbf{q}_l &= P_l[\sum_{k\in II}(1-\alpha)A_k\mathbf{x}_k] \\ &= P_l[(1-\alpha)\mathbf{b}/\mu - \sum_{k\in I}(1-\alpha)A_k\mathbf{x}_k] \\ &= P_l[(1-\alpha)\mathbf{b}/\mu - \sum_{k>l, k\in I}(1-\alpha)A_k\mathbf{x}_k]. \end{aligned} \tag{33}$$

In the expression (33), the first term is

$$\mathbf{q}_l' = (1-\alpha)P_l\mathbf{b}/\mu \tag{34}$$

where $l \in II$. We write $(1-\alpha)P_l\mathbf{b}$ in the form $\mu\alpha A_l \bar{S}_l^{-1}\mathbf{u}_l$. In general, given the vector $P_l\mathbf{b}$, we want to write it as $A_l\mathbf{y}$ with a bound on $\|\mathbf{y}\|$. Observe that $P_l\mathbf{b} = [0, B_l, 0]B^{-1}\mathbf{b} = B_l\mathbf{y}$, where we define $\mathbf{y}$ to be a subvector of $\hat{\mathbf{y}} = B^{-1}\mathbf{b}$. Thus, we need a bound on $\mathbf{y}$ in terms of $\|P_l\mathbf{b}\|$, that is, in terms of $\|B_l\mathbf{y}\|$. Let $\mathbf{y}'$ be the extension of $\mathbf{y}$ to an m-vector with zeros filled in. Observe that

$$\begin{aligned} \|\mathbf{y}\| &= \|\mathbf{y}'\| \\ &\leq \|B^{-1}\| \cdot \|B\mathbf{y}'\| \\ &\leq \chi_A \cdot \|B\mathbf{y}'\| \\ &= \chi_A \cdot \|B_l\mathbf{y}\| \\ &= \chi_A \cdot \|P_l\mathbf{b}\|. \end{aligned} \tag{35}$$

Note that the third line was obtained from Lemma 2. Thus, we can write $(1-\alpha)P_l\mathbf{b}$ in the form $\mu\alpha A_l\bar{S}_l^{-1}\mathbf{u}_l$ such that

$$\|\mu\alpha\bar{S}_l^{-1}\mathbf{u}_l\| \le \chi_A \cdot \|P_l\mathbf{b}\|.$$

This means that

$$\|\mathbf{u}_l\| \le \frac{\|\bar{S}_l\| \cdot \chi_A\|P_l\mathbf{b}\|}{\mu\alpha} \tag{36}$$

$$\le \frac{\bar{\theta}_l\chi_A\|P_l\mathbf{b}\|}{\mu\alpha} \tag{37}$$

where $\bar{\theta}_l$ denotes the largest entry of $\bar{S}_l$. It follows from (29) that $\bar{\theta}_l \le 2\cdot\theta_l$. Proceeding from (37),

$$\|\mathbf{u}_l\| \le \frac{2\theta_l\chi_A\|P_l\mathbf{b}\|}{\mu\alpha}.$$

Since $\alpha_l \le \alpha$,

$$\|\mathbf{u}_l\| \le \frac{2\theta_l\chi_A\|P_l\mathbf{b}\|}{\mu\alpha_l}.$$

By choice of α_l in (20), this implies that

$$\|\mathbf{u}_l\| \le 1/(15n^{1.5}). \tag{38}$$

Finally, consider the second term in (33)

$$\mathbf{q}_l'' = P_l[-\sum_{k>l,k\in I}(1-\alpha)A_k\mathbf{x}_k].$$

Note we have

$$\begin{aligned}\|(1-\alpha)\mathbf{x}_k\| &= \|(1-\alpha)(X_kS_k)S_k^{-1}\mathbf{e}\| \\ &\le (1-\alpha)\|X_kS_k\|\cdot\|\mathbf{e}\|\cdot\|S_k^{-1}\| \\ &\le (1+\delta)\sqrt{n}/\phi_k.\end{aligned}$$

Now we apply Lemma 5 to conclude that

$$\mathbf{q}_l'' = \sum_{k>l,k\in I} A_l\mathbf{w}_k = A_l(\sum_{k>l,k\in I}\mathbf{w}_k), \quad \|\mathbf{w}_k\| \le \bar{\chi}_A(1+\delta)\sqrt{n}/\phi_k.$$

Next, observe that

$$\mathbf{q}_l'' = \alpha A_l^T \bar{S}_l^{-1} (\sum_{k>l, k\in I} \mathbf{u}_k) \quad \text{where} \quad \mathbf{u}_k = \bar{S}_l \mathbf{w}_k / \alpha$$

so

$$\begin{aligned} \|\mathbf{u}_k\| &= \|\bar{S}_l/\alpha\| \cdot \|\mathbf{w}_k\| \\ &\le \max_{i\in J_l} \frac{\bar{s}_i}{\alpha} \cdot \|\mathbf{w}_k\| \\ &\le \max_{i\in J_l} \frac{2s_i}{\alpha} \cdot \|\mathbf{w}_k\| \\ &\le \frac{2\theta_l \cdot \bar{\chi}_A (1+\delta)\sqrt{n}/\phi_k}{\alpha} \\ &\le (1+\delta)/(15n^{1.5}). \end{aligned}$$

Note that to derive the third line we used (29) since l is Type II, and to derive the last line we used $\alpha \ge \beta_{k,l}$ in (21), since $k > l$, and l is of Type II and k is Type I.

Thus, we have

$$\mathbf{q}_l'' = \sum_{k>l, k\in I} \alpha A_l \bar{S}_l^{-1} \mathbf{u}_k$$

where $\|\mathbf{u}_k\| \le (1+\delta)/(15n^{1.5})$, which together with (38) implies that

$$\mathbf{q}_l = \mathbf{q}_l' + \mathbf{q}_l'' = \sum_{k\in\{l\}\cup\{k>l, k\in I\}} \alpha A_l \bar{S}_l^{-1} \mathbf{u}_k, \quad \|\mathbf{u}_k\| \le (1+\delta)/(15n^{1.5})$$

or

$$\mathbf{q}_l = \alpha A_l \bar{\Delta}_l^{-1} \mathbf{v}_l, \quad \|\mathbf{v}_l\| = \left\| \sum_{k\in\{l\}\cup\{k>l, k\in I\}} \mathbf{u}_k \right\| \le (1+\delta)/(15\sqrt{n}).$$

Thus, this proves (31) for $l = 1, \ldots, p$, and establishes the third term of (30), and, therefore, proves the theorem. ∎

We can now comment further on Algorithm LIP. We have proved that after the LLS step, $\delta(\bar{\mathbf{y}}, \mu\alpha) \le 0.75$. This means that after three Newton steps in Step 5 of the algorithm, by Lemma 1, we have a point $\mathbf{y}'$ such that $\delta(\mathbf{y}', \mu\alpha) \le 0.75^8 < 0.25$, so $\mathbf{y}'$ is well-centered for continuing the algorithm.

This theorem has a second consequence as far as Algorithm LIP is concerned. In Algorithm LIP, if $\alpha = 0$ then we terminate. The following corollary shows that this termination is at an optimum point for (2), and an optimum point for the primal may be recovered by solving a least squares problem.

Not only is it an optimum, but, in the case of degeneracy (multiple optima), the point computed by Algorithm LIP is an *approximate analytic center* of the face of optimal solutions. Proving that Algorithm LIP computes an approximate analytic center in the case of degeneracy is important for our initialization procedure defined in Section 8.

Recall that the *analytic center* of a polytope $\{\mathbf{y} : A^T\mathbf{y} \geq \mathbf{c}\}$ is defined to be a point $\mathbf{y}^*$ such that $AS^{-1}\mathbf{e} = \mathbf{0}$, where $\mathbf{s} = A^T\mathbf{y}^* - \mathbf{c}$ and $S = \mathrm{diag}(\mathbf{s})$. If the polytope is not full dimensional (which is always the case for this corollary) and it is given as $\{\mathbf{y} : A_0^T\mathbf{y} = \mathbf{c}_0,\ A^T\mathbf{y} \geq \mathbf{c}\}$, then its analytic center $\mathbf{y}^*$ satisfies (see Mizuno et al. [14])

$$A_0\mathbf{x}_0 + AS^{-1}\mathbf{e} = \mathbf{0}, \quad \text{for some} \quad \mathbf{x}_0, \tag{39}$$

where $\mathbf{s} = A^T\mathbf{y}^* - \mathbf{c}$. We say that $\mathbf{y}$ is an *approximate analytic center* of a polytope $\{\mathbf{y} : A_0^T\mathbf{y} = \mathbf{c}_0,\ A^T\mathbf{y} \geq \mathbf{c}\}$ if

$$A_0\mathbf{x}_0 + AS^{-1}\mathbf{f} = \mathbf{0}, \quad \text{for some} \quad \mathbf{x}_0, \tag{40}$$

where $\mathbf{s} = A^T\mathbf{y} - \mathbf{c}$, $S = \mathrm{diag}(\mathbf{s})$, and $\|\mathbf{f} - \mathbf{e}\| \leq 0.75$.

Corollary 2 *Suppose $\bar{\alpha} = 0$, that is, $\epsilon_k = 0$ for Type I layers, $P_k\mathbf{b} = \mathbf{0}$ for Type II layers, and $\beta_{k,l} = 0$ for all k, l for which it is defined. Then $\mathbf{y}^* = \mathbf{y} - \mathbf{r}$ is an optimum solution for (2), and it is an approximate analytic center for the (dual) optimal face. Moreover, by solving a single least squares problem we generate an optimum solution $\mathbf{x}^*$ for the primal.*

Proof. Define $\bar{\mathbf{y}}(\alpha) = \mathbf{y} - (1-\alpha)\mathbf{r}$. From Theorem 1, we know that $\bar{\mathbf{y}}(\alpha)$ is strictly feasible and approximately centered for all $\alpha \in (0, 1]$. In fact, we have the following equation:

$$\mathbf{q} = \mathbf{b}/\mu - \alpha A\bar{S}(\alpha)^{-1}\mathbf{e} = \alpha A\bar{S}(\alpha)^{-1}\mathbf{v}(\alpha), \tag{41}$$

where $\|\mathbf{v}(\alpha)\| \leq 3/4$. Consider taking the limit of this equation as $\alpha \to 0^+$. Since $\bar{\mathbf{y}}(\alpha)$ is strictly feasible, the slacks satisfy $\bar{\mathbf{s}}(\alpha) \geq \mathbf{0}$ for all $\alpha \in [0, 1]$.

Partition $\{1,\ldots,n\}$ into B and N, where B is the set of indices such that $\bar{s}_i(0)=0$ and N is the set of indices such that $\bar{s}_i(0)>0$. Note that $\mathbf{s}^*$, the slacks of $\mathbf{y}^*$, are equal to $\bar{\mathbf{s}}(0)$. As $\alpha\rightarrow 0^+$, the matrix $\alpha\bar{S}_N(\alpha)$ tends to zero. Thus, in the limit, the left-hand side of (41) tends to $\mathbf{b}/\mu-\alpha A_B\bar{S}_B(\alpha)^{-1}\mathbf{e}$. But notice that $\alpha\bar{S}_B(\alpha)^{-1}$ is a constant matrix independent of α (because $\bar{\mathbf{s}}_B(\alpha)$ is linear in α), hence, if we denote $\alpha\bar{S}_B(\alpha)^{-1}$ as S_B^{-1}, we have that the limit of the left-hand side is

$$\mathbf{b}/\mu-A_BS_B^{-1}\mathbf{e}.$$

Note that from the final LLS step we can easily construct B,N and S_B because we have a closed-form expression for $\bar{\mathbf{s}}(\alpha)$.

Now consider taking limits on the right-hand side of (41). In general, we cannot claim that $\mathbf{v}(\alpha)$ converges to anything. On the other hand, because $\mathbf{v}(\alpha)$ stays bounded, a compactness argument shows that there must exist an accumulation point $\mathbf{v}^*$ as $\alpha\rightarrow 0$. Then we have the equation:

$$\mathbf{b}/\mu-A_BS_B^{-1}\mathbf{e}=A_BS_B^{-1}\mathbf{v}_B^* \text{ and } \|\mathbf{v}_B^*\|\le 3/4. \tag{42}$$

Although we cannot construct this particular $\mathbf{v}_B^*$ (its existence was proved by a compactness argument), we can certain construct some other $\mathbf{v}_B^*$ satisfying these relations by solving the least-squares problem of minimizing $\|\mathbf{v}_B\|$ subject to $\mathbf{b}/\mu-A_BS_B^{-1}\mathbf{e}=A_BS_B^{-1}\mathbf{v}_B$; clearly the minimizer has norm bounded by 3/4. This least squares problem is similar to the termination discussed in [31].

Next, define $\mathbf{x}^*\in\mathbb{R}^n$ as follows. Let $\mathbf{x}_B^*=\mu S_B^{-1}(\mathbf{e}+\mathbf{v}_B^*)$, a strictly positive vector, and let $\mathbf{x}_N^*=\mathbf{0}$. Observe that $\mathbf{x}^*\ge\mathbf{0}$, and that $A\mathbf{x}^*=\mathbf{b}$. Thus, $\mathbf{x}^*$ is feasible for the primal problem (4). Furthermore, $(\mathbf{x}^*)^T\mathbf{s}^*=0$ because $\mathbf{x}_N^*=\mathbf{0}$ and $\mathbf{s}_B^*=\mathbf{0}$. Thus, $(\mathbf{x}^*,\mathbf{y}^*)$ are feasible points for the primal and dual respectively satisfying the complementary slackness condition. This means that both are optimal solutions. Furthermore, they are strictly complementary, so this implies that the optimal face for the dual is precisely

$$\{\mathbf{y}:A_B^T\mathbf{y}=\mathbf{c}_B, A_N^T\mathbf{y}\ge\mathbf{c}_N\}. \tag{43}$$

Now we show that $\mathbf{y}^*$ is an approximate analytic center of the optimal face. If we subtract (42) from (41) and divide by α we obtain

$$A_N\bar{S}_N(\alpha)^{-1}(\mathbf{e}+\mathbf{v}_N(\alpha))\in\mathrm{Range}(A_B).$$

By compactness, $\mathbf{v}_N(\alpha)$ must have a limit point $\mathbf{v}_N^*$ as $\alpha \to 0^+$ whose norm is bounded by 3/4. Also, $\bar{S}_N(\alpha)^{-1}$ is well-defined when $\alpha = 0$ by choice of N. Thus, for some $\mathbf{x}_0$,

$$A_N \bar{S}_N(0)^{-1}(\mathbf{e} + \mathbf{v}_N^*) = A_B \mathbf{x}_0.$$

This is precisely the condition that $\mathbf{y}^*$ is an approximate analytic center of (43). ∎

6 Crossover events

The theorem and corollary in the last section proves that Algorithm LIP is valid in the sense that it accurately tracks the central path, and terminates only when an optimum is reached. In this section, we prove that the number of main-loop iterations required by Algorithm LIP is finite, and, in particular, is bounded by n^2. The key concept for developing a complexity theory is a crossover event.

Definition. Given an instance of (2) with a strictly feasible interior point, we say that the 4-tuple (μ, μ', i, j) defines a **crossover event**, for $\mu > \mu' > 0$, and for $i, j \in \{1, \ldots, n\}$, if

$$\frac{s_i(\mu)}{s_j(\mu)} \leq 3g^n \tag{44}$$

and for all $\mu'' \in (0, \mu']$,

$$\frac{s_i(\mu'')}{s_j(\mu'')} \geq 5g^n. \tag{45}$$

Here $s_i(\mu)$ denotes the ith slack for a point on the central path of (2).

Note that one must have $i \neq j$ for (45) to hold, since $g > 1$.

Definition. We say that two crossover events (μ, μ', i, j) and $(\mu_1, \mu_1', i_1, j_1)$ are **disjoint** if $(\mu', \mu) \cap (\mu_1', \mu_1) = \emptyset$.

Lemma 7 *Let $(\mu_1, \mu_1', i_1, j_1), \ldots, (\mu_t, \mu_t', i_t, j_t)$ be a sequence of t disjoint crossover events for a particular instance of (2). Then $t \leq n(n-1)/2$.*

PROOF. We claim that a particular pair of indices (i, j) can occur at most once in this list. For example, consider the first two events, $(\mu_1, \mu'_1, i_1, j_1)$ and $(\mu_2, \mu'_2, i_2, j_2)$. Since these are disjoint, then either $\mu_1 \leq \mu'_2$ or $\mu_2 \leq \mu'_1$; assume the former. Then $\mu'_1 < \mu_1 \leq \mu'_2 < \mu_2$. But it is not possible for both $i_1 = i_2$ and $j_1 = j_2$ to hold because (45) for the second event, for μ'' set to μ_1 would contradict (44) for the first event.

Similarly, we could not have $i_1 = j_2$ and $j_1 = i_2$ either, because then (45) would be contradictory for μ'' chosen less than $\min(\mu'_1, \mu'_2)$ (i.e., (45) would have to hold, and the same relation would have to hold with i and j swapped.) ∎

The main theorem of this section is that every pass through the main loop of Algorithm LIP causes at least one crossover event to occur, until $\bar{\alpha} = 0$. This sequence of crossover events is clearly a disjoint sequence since μ decreases monotonically throughout Algorithm LIP. Therefore, there can be at most $n(n-1)/2$ iterations of Algorithm LIP before termination. We start by revisiting the examples given on p. 18.

1. In the first example, $\bar{\alpha} = 0$ so Theorem 2 doesn't apply.

2. In the second example, recall that the LLS step takes us to a point $O(\epsilon)$ away from the origin. Now observe that after a certain fixed number of ordinary interior point steps (independent of ϵ) one of the three constraints $\{y_1 \geq 0, y_2 \geq 0, y_1 + y_2 \geq \epsilon\}$ must leave J_1 because at least one of these constraints will always have residual at least $O(\epsilon)$, whereas the residuals of the other two will tend to zero. Hence there will be a crossover event involving two of the constraints in J_1.

3. For the third example, recall that after the LLS step we end up at a point like $(O(|\epsilon|), 0.5)$. Now, further steps on the central path will force one of the two constraints $\{y_2 \geq 0, y_2 \leq 1\}$ to become active, because the central path equation
$$(1, \epsilon)^T = \mu A S^{-1} \mathbf{e}$$
must be approximately satisfied, and those are the only two constraints with a nonzero component in the y_2 direction. The constraint that becomes active depends on the sign of ϵ. Either way, there will be a crossover event involving the two constraints $\{y_2 \geq 0, y_2 \leq 1\}$ that are in J_2.

4. For the last example, recall that the LLS step ends up at a point like $(O(\epsilon), \epsilon/2)$. Now, after a fixed number of interior point steps, the constraint $y_1 \geq y_2$ will become active. On the other hand, of the two constraints $0 \leq y_2 \leq \epsilon$, they cannot both stay near-active once μ gets below $O(\epsilon)$. Thus, a crossover event takes place between constraint $y_1 \geq y_2$ and constraint $y_2 \leq \epsilon$.

Before proving this theorem, we require some preliminary lemmas.

Lemma 8 *Let $(\mathbf{y}, \mu)$ and $(\mathbf{y}', \mu')$ be two points on the central path such that $0 < \mu' \leq \mu$. Let the slacks be $\mathbf{s}, \mathbf{s}'$ respectively. Then for any i,*

$$s_i' \leq n s_i.$$

PROOF. Assume $n > 1$ since the $n = 1$ case may be trivially analyzed. Let

$$\mathbf{x} = \mu S^{-1}\mathbf{e}, \quad \text{and} \quad \mathbf{x}' = \mu'(S')^{-1}\mathbf{e}.$$

(This is slightly different from the use of $\mathbf{x}$ in the last section.) Then, we have

$$A\mathbf{x} = \mathbf{b}, \quad \text{and} \quad A\mathbf{x}' = \mathbf{b}.$$

Then,

$$(\mathbf{x} - \mathbf{x}')^T(\mathbf{s} - \mathbf{s}') = 0,$$

since $\mathbf{x} - \mathbf{x}'$ is in the null space of A, and $\mathbf{s} - \mathbf{s}'$ is in the range of A^T. Thus,

$$\sum_{i=1}^{n}(x_i' s_i + s_i' x_i) = \mathbf{x}^T\mathbf{s} + (\mathbf{x}')^T\mathbf{s}' = n(\mu + \mu'),$$

which implies that

$$\sum_{i=1}^{n}\left(\frac{s_i}{s_i'}\mu' + \frac{s_i'}{s_i}\mu\right) = n(\mu + \mu'),$$

i.e.,

$$\sum_{i=1}^{n}\left(\frac{s_i}{s_i'}\frac{\mu'}{\mu} + \frac{s_i'}{s_i}\right) = n\left(1 + \frac{\mu'}{\mu}\right).$$

Note that for each $j = 1, \ldots, n$, using the arithmetic-geometric mean inequality, we derive

$$\frac{s_j}{s'_j}\frac{\mu'}{\mu} + \frac{s'_j}{s_j} \geq 2\sqrt{\mu'/\mu}.$$

Thus, for any i,

$$\begin{aligned} \frac{s_i}{s'_i}\frac{\mu'}{\mu} + \frac{s'_i}{s_i} &= n\left(1 + \frac{\mu'}{\mu}\right) - \sum_{j \neq i}\left(\frac{s_j}{s'_j}\frac{\mu'}{\mu} + \frac{s'_j}{s_j}\right) \\ &\leq n(1 + \mu'/\mu) - 2(n-1)\sqrt{\mu'/\mu} \\ &= n(1 - \sqrt{\mu'/\mu})^2 + 2\sqrt{\mu'/\mu}. \end{aligned}$$

Since $0 < \sqrt{\mu'/\mu} \leq 1$ and $n > 1$, we have

$$n(1 - \sqrt{\mu'/\mu})^2 + 2\sqrt{\mu'/\mu} \leq n.$$

Thus,

$$\frac{s_i}{s'_i}\frac{\mu'}{\mu} + \frac{s'_i}{s_i} \leq n,$$

and in particular,

$$\frac{s'_i}{s_i} \leq n$$

which proves the lemma. ∎

Lemma 9 *Consider an LLS problem with partitions $J_1, \ldots, J_p$, weights Δ, coefficient matrix A, and right-hand side $\mathbf{s}$. Let $\mathbf{r}_1$ be the LLS solution. Let D be the $n \times n$ diagonal matrix with 1's in positions corresponding to $J_1, \ldots J_k$ and 0's elsewhere. Let $\|\mathbf{u}\|_D$ denote the seminorm $(\mathbf{u}^T D\mathbf{u})^{1/2}$, and let $\mathbf{r}^*$ be the minimizer (not necessarily unique) of $\|A^T\mathbf{r} - \mathbf{s}\|_D$. Then*

$$\|A^T\mathbf{r}_1 - \mathbf{s}\|_D \leq (\bar{\chi}_A + 1) \cdot \|A^T\mathbf{r}^* - \mathbf{s}\|_D.$$

Proof. Let $\mathbf{s}_1$ be the residual for the unweighted least-squares problem, so that $\mathbf{s} - A^T\mathbf{r}^* = \mathbf{s}_1$. Let $D(\epsilon)$ be chosen as in Lemma 3, and let $D_1 = \Delta^{-1}D(\epsilon)$. Let $\mathbf{r}(\epsilon)$ be the minimizer of $\|D_1(A^T\mathbf{r} - \mathbf{s})\|$. Then $\mathbf{r}(\epsilon)$ tends to $\mathbf{r}_1$, the LLS solution, as ϵ tends to zero.

Observe that

$$D_1(A^T\mathbf{r} - \mathbf{s}) = D_1(A^T(\mathbf{r} - \mathbf{r}^*) - \mathbf{s}_1)$$

by definition of $\mathbf{s}_1$. Thus, we could equivalently define $\mathbf{r}(\epsilon)$ to minimize $\|D_1(A^T(\mathbf{r} - \mathbf{r}^*) - \mathbf{s}_1)\|$. Thus, by (9),

$$\|A^T(\mathbf{r}(\epsilon) - \mathbf{r}^*)\| \leq \bar{\chi}_A\|\mathbf{s}_1\|.$$

Since $\mathbf{r}(\epsilon) \to \mathbf{r}_1$ as $\epsilon \to 0$,

$$\|A^T(\mathbf{r}_1 - \mathbf{r}^*)\| \leq \bar{\chi}_A\|\mathbf{s}_1\|.$$

Thus,

$$\begin{aligned}\|A^T\mathbf{r}_1 - \mathbf{s}\| &= \|A^T(\mathbf{r}_1 - \mathbf{r}^*) - \mathbf{s}_1\| \\ &\leq \|A^T(\mathbf{r}_1 - \mathbf{r}^*)\| + \|\mathbf{s}_1\| \\ &\leq (\bar{\chi}_A + 1)\|\mathbf{s}_1\|.\end{aligned}$$

Substituting the definition of $\mathbf{s}_1$ proves the lemma. ∎

Lemma 10 *Given a point* $(\mathbf{y}, \mu)$ *in Algorithm LIP, let* ϵ_k *be defined by* (17), *that is,* ϵ_k *is the residual of the kth layer in the LLS step. Then, there is a constraint* $i \in J_1 \cup \cdots \cup J_k$ *such that for all* $\mu'' \in (0, \mu]$,

$$s_i(\mu'') \geq \frac{\epsilon_k\phi_k}{\sqrt{1+\delta}\cdot(\bar{\chi}_A + 1)n^{1.5}}. \tag{46}$$

Proof. Let D denote the $n \times n$ diagonal matrix with 1's in positions corresponding to $J_1 \cup \cdots \cup J_k$ and zeros elsewhere. Let $\|\mathbf{u}\|_D$ denote the seminorm $(\mathbf{u}^T D\mathbf{u})^{1/2}$.

Let $\mathbf{y}'$ be any feasible vector for (2), and let $\mathbf{s}'$ be the slacks for $\mathbf{y}'$, and let $\mathbf{r}' = \mathbf{y} - \mathbf{y}'$. Let $\mathbf{s}$ be the slack variables for $\mathbf{y}$, and let $\mathbf{r}$ be the LLS step. Then we have

$$\begin{aligned}\|\mathbf{s}'\|_D &= \|A^T\mathbf{y}' - \mathbf{c}\|_D \\ &= \|A^T(\mathbf{y}' - \mathbf{y}) + \mathbf{s}\|_D \\ &= \|A^T\mathbf{r}' - \mathbf{s}\|_D \\ &\geq \min\{\|A^T\mathbf{r}'' - \mathbf{s}\|_D : \mathbf{r}'' \in \mathbb{R}^m\} \\ &\geq \|A^T\mathbf{r} - \mathbf{s}\|_D/(\bar{\chi}_A + 1).\end{aligned}$$

To obtain the last line we applied the preceding lemma.

Thus, for any feasible point, and in particular, for $\mathbf{s}(\mu'')$, we have

$$\|\mathbf{s}(\mu'')\|_D \geq \|A^T\mathbf{r} - \mathbf{s}\|_D/(\bar{\chi}_A + 1).$$

Let us first address the right-hand side of this inequality. First, note that

$$\|\Delta_k^{-1}\| = \|(X_k S_k)^{1/2} S_k^{-1}\| \leq \frac{\sqrt{1+\delta}}{\phi_k}.$$

Now we have:

$$\begin{aligned}
\|A^T\mathbf{r} - \mathbf{s}\|_D &\geq \|A_k^T\mathbf{r} - \mathbf{s}_k\| \\
&\geq \|\Delta_k^{-1}(A_k^T\mathbf{r} - \mathbf{s}_k)\|/\|\Delta_k^{-1}\| \\
&= \epsilon_k/\|\Delta_k^{-1}\| \\
&\geq \epsilon_k\phi_k/\sqrt{1+\delta}.
\end{aligned}$$

Combining, we have shown that

$$\|\mathbf{s}(\mu'')\|_D \geq \frac{\epsilon_k\phi_k}{(\bar{\chi}_A + 1)\sqrt{1+\delta}}.$$

This means that for at least one $i \in J_1 \cup \cdots \cup J_k$,

$$s_i(\mu'') \geq \frac{\epsilon_k\phi_k}{(\bar{\chi}_A + 1)\sqrt{1+\delta} \cdot \sqrt{n}}. \tag{47}$$

For the argument in the last paragraph, the choice of i apparently depends on μ''. The claim in the lemma is that we can pick i independently of μ''. To prove this stronger claim, observe that, if, for some $\mu' \in (0, \mu]$, the inequality

$$s(\mu') < \frac{\epsilon_k\phi_k}{(\bar{\chi}_A + 1)\sqrt{1+\delta} \cdot n^{1.5}} \tag{48}$$

holds, then for all $\mu'' \in (0, \mu']$, (47) could never hold because of Lemma 8. Therefore, there has to be at least one fixed i such that (48) never holds for any $\mu' \in (0, \mu]$, i.e., (46) always holds for all $\mu'' \in (0, \mu]$ for this i. ∎

It should noted that this lemma has a curious nonconstructive property; the existence of a special $i \in J_1 \cup \cdots \cup J_k$ is demonstrated, but there is apparently no method for determining the value of i other than running the interior point method to completion to find the whole central path. Indeed, if were possible to identify i in the lemma from the current iterate, then we could presumably delete the ith constraint from the problem because we know it is not active at the solution.

We now come to the main theorem for this section:

Theorem 2 *Consider one pass through the main loop of the LIP algorithm. Assume $\bar{\alpha} > 0$ in the LLS step. Then at least one event takes place during this pass as the central path parameter drops from μ to $\mu\bar{\alpha}/C(A)$, where $C(A)$ is defined by* (60).

PROOF. Recall that $\bar{\alpha}$ is defined to be the maximum of the α_k's and $\beta_{k,l}$'s. There are therefore three cases to this proof: either $\bar{\alpha} = \alpha_k$ where k is a Type I layer, $\bar{\alpha} = \alpha_k$ where k is a Type II layer, or $\bar{\alpha} = \beta_{k,l}$. The three cases correspond to examples 2, 3, and 4 listed above, in that order.

CASE 1, $\alpha = \alpha_k$ and α_k is defined by (19).

First, we develop an inequality that applies to any layer k of Type I, not necessarily the particular k that determines $\bar{\alpha}$. This is because we will reuse the same argument for Case 3 below.

For some layer k of Type I, let $\mathbf{v} = \mathbf{r}^{(k+1)} - \mathbf{r}^{(k)}$. Recall that $\mathbf{r}^{(k)}, \mathbf{r}^{(k+1)}$ are defined by LLS problems. Therefore, $\mathbf{v}$ is also defined by a LLS problem with the same partition, weights, and coefficients but with right-hand side vector all zeros except for J_k, in which the right-hand side vector is $\mathbf{s}_k$. This means that $\mathbf{v}$ lies in the nullspace of $A_1^T, \ldots, A_{k-1}^T$, and the following Lagrange multiplier condition is satisfied by $\mathbf{v}$ for the kth LLS step:

$$A_k \Delta_k^{-2} A_k^T \mathbf{v} - A_k \mathbf{x}_k = A_1 \boldsymbol{\lambda}_1 + \cdots + A_{k-1} \boldsymbol{\lambda}_{k-1}. \tag{49}$$

Note that

$$\begin{aligned} \delta_k &= \|\Delta_k^{-1} A_k^T (\mathbf{r} - \mathbf{r}^{(k)})\| \\ &= \|\Delta_k^{-1} A_k^T \mathbf{v}\| \end{aligned}$$

because $\mathbf{v} - (\mathbf{r} - \mathbf{r}^{(k)}) = \mathbf{r}^{(k+1)} - \mathbf{r}$ lies in the nullspace of A_k^T. The assumption for this case is that δ_k is bounded below by (18).

Consider the product $\mathbf{v}^T\mathbf{b}$. Since $\mathbf{v}$ is in the nullspace of $A_1^T, \ldots, A_{k-1}^T$ and $A\mathbf{x} = \mathbf{b}/\mu$,

$$\begin{aligned} \mathbf{v}^T\mathbf{b} &= \mathbf{v}^T \sum_{l=1}^{p} \mu A_l \mathbf{x}_l \\ &= \mu \mathbf{v}^T \sum_{l=k}^{p} A_l \mathbf{x}_l. \end{aligned} \tag{50}$$

We now obtain a lower bound on this sum, first by estimating the $l = k$ term, and then by placing an upper bound on the remaining terms.

Let us focus on the kth term of this sum, which we denote as q:

$$q = \mu \mathbf{v}^T A_k \mathbf{x}_k. \tag{51}$$

Substituting the expression for $A_k\mathbf{x}_k$ from (49) into (51) yields

$$\begin{aligned} q &= \mu \mathbf{v}^T (A_k \Delta_k^{-2} A_k^T \mathbf{v} - A_1 \boldsymbol{\lambda}_1 - \cdots - A_{k-1} \boldsymbol{\lambda}_{k-1}) \\ &= \mu \mathbf{v}^T A_k \Delta_k^{-2} A_k^T \mathbf{v} \\ &= \mu \|\Delta_k^{-1} A_k^T \mathbf{v}\|^2 \\ &= \mu \delta_k^2. \end{aligned}$$

The second line follows because $\mathbf{v}$ is in the nullspace of $A_1^T, \ldots, A_{k-1}^T$. Everything else follows from definitions.

Next, let us evaluate the remaining terms in $\mathbf{v}^T\mathbf{b}$. Observe that $\mathbf{v}$, as a solution to a LLS problem, has the following norm bound:

$$\|A^T\mathbf{v}\| \le \bar{\chi}_A \theta_k \sqrt{n}.$$

This follows because the right-hand side for the LLS problem defining $\mathbf{v}$ has norm equal to $\|\mathbf{s}_k\|$, which is bounded by $\theta_k\sqrt{n}$. Let $A_N = (A_{k+1}, \ldots, A_p)$. Then, analyzing the remaining terms in (50),

$$\begin{aligned} \left\| \mu \sum_{l=k+1}^{p} \mathbf{v}^T A_l \mathbf{x}_l \right\| &= \left\| \mu \mathbf{v}^T A_N \mathbf{x}_N \right\| \\ &\le \mu \|A_N^T \mathbf{v}\| \cdot \|\mathbf{x}_N\| \\ &\le \mu \|A_N^T \mathbf{v}\| \cdot \|(X_N S_N) S_N^{-1} \mathbf{e}\| \\ &\le \mu \|A_N^T \mathbf{v}\| \cdot \|(X_N S_N)\| \cdot \|S_N^{-1}\| \cdot \|\mathbf{e}\| \end{aligned}$$

$$\begin{aligned}
&\leq \mu \bar{\chi}_A \theta_k \sqrt{n}(1+\delta)\|S_N^{-1}\| \cdot \sqrt{n} \\
&\leq \mu \bar{\chi}_A \theta_k (1+\delta) n/\phi_{k+1}. \\
&\leq \mu \bar{\chi}_A (1+\delta) n/g.
\end{aligned}$$

If we require $\mu \bar{\chi}_A(1+\delta)n/g \leq q/2$, i.e., $\mu \bar{\chi}_A(1+\delta)n/g \leq 0.5\mu\delta_k^2$ then we can be guaranteed that $\mathbf{v}^T\mathbf{b} \geq 0.5\mu\delta_k^2$. Taking into account (18), this places the following restriction on g:

$$g \geq 2 \cdot 30^2 (1+\delta) n^2 \bar{\chi}_A. \tag{52}$$

Now, we continue with the analysis of this case, using the inequality $\mathbf{v}^T\mathbf{b} \geq 0.5\mu\delta_k^2$. Let us consider a $\mu' \leq \mu$ and $\mathbf{y}'$ on the central path for μ'. We have:

$$\begin{aligned}
\mu\delta_k^2/2 &\leq \mathbf{v}^T\mathbf{b} \\
&= \mu' \sum_{l=1}^{p} \mathbf{v}^T A_l (S_l')^{-1}\mathbf{e} \\
&= \mu' \sum_{l=k}^{p} \mathbf{v}^T A_l (S_l')^{-1}\mathbf{e} \\
&= \mu' \mathbf{v}^T A_N (S_N')^{-1}\mathbf{e} \\
&\leq \mu' \|A_N^T \mathbf{v}\| \cdot \|(S_N')^{-1}\| \cdot \|\mathbf{e}\| \\
&\leq \mu' \bar{\chi}_A \theta_k n \cdot \|(S_N')^{-1}\|
\end{aligned}$$

where $A_N = (A_k, ..., A_p)$ and $S_N' = \mathrm{diag}(S_k', \ldots, S_p')$. Thus,

$$\|(S_N')^{-1}\| \geq \frac{\mu\delta_k^2/2}{\mu' \bar{\chi}_A \theta_k n}.$$

This means that there is an $j \in J_k \cup \cdots \cup J_p$ such that

$$\begin{aligned}
s_j(\mu') &\leq \frac{2\mu' \bar{\chi}_A \theta_k n}{\mu \delta_k^2} \\
&\leq \frac{2 \cdot 30^2 \mu' \bar{\chi}_A \theta_k n^2}{\mu}.
\end{aligned} \tag{53}$$

Now we focus on the particular k of Type I that satisfies the hypothesis of this case, that is, that $\bar{\alpha} = \alpha_k$. The preceding inequalities held for an arbitrary $\mu' \in (0, \mu]$; select in particular

$$\mu' = \frac{\mu\epsilon_k}{10 \cdot 30^2(\bar{\chi}_A + 1)\bar{\chi}_A n^{4.5} g^{2n}\sqrt{1+\delta}} \tag{54}$$

and fix a particular j satisfying (53). Then for all $\mu'' \leq \mu'$, combining (53) with Lemma 8,

$$\begin{aligned} s_j(\mu'') &\leq \frac{2 \cdot 30^2 \mu' \bar{\chi}_A \theta_k n^3}{\mu} \\ &= \frac{\mu\epsilon_k}{10 \cdot 30^2(\bar{\chi}_A + 1)\bar{\chi}_A n^{4.5} g^{2n}\sqrt{1+\delta}} \cdot \frac{2 \cdot 30^2 \bar{\chi}_A \theta_k n^3}{\mu} \\ &= \frac{\epsilon_k \theta_k}{5(\bar{\chi}_A + 1) n^{1.5} g^{2n}\sqrt{1+\delta}}. \end{aligned}$$

Now, consider the index i picked out by Lemma 10; divide (46) by this inequality to obtain that for all $\mu'' \leq \mu'$,

$$\begin{aligned} \frac{s_i(\mu'')}{s_j(\mu'')} &\geq \frac{\epsilon_k \phi_k}{(\bar{\chi}_A + 1) n^{1.5}\sqrt{1+\delta}} \cdot \frac{5(\bar{\chi}_A + 1) n^{1.5} g^{2n}\sqrt{1+\delta}}{\epsilon_k \theta_k} \\ &= \frac{5\phi_k g^{2n}}{\theta_k} \\ &\geq 5g^n. \end{aligned}$$

Thus, we have identified two constraints, i and j, such that (45) holds in the definition of a crossover event.

Now recall that $j \in J_k \cup \cdots \cup J_p$ and $i \in J_1 \cup \cdots \cup J_k$. If either $i \notin J_k$ or $j \notin J_k$, then $s_i < s_j/g$. If both are in J_k, then $s_i \leq g^n s_j$. Applying Corollary 1 for $\delta = 0.25$ shows that in either case, $s_i(\mu) \leq 3g^n s_j(\mu)$. Thus, (44) holds for i and j. This shows that a crossover event takes place during the main loop of Algorithm LIP, provided that at the end of the step, $\mu\alpha/C(A)$ is at least as small as μ' in (54). Recalling that $\bar{\alpha} = \alpha_k$ and $\alpha_k = \min(1, 15\epsilon_k\sqrt{n})$, i.e., $\bar{\alpha} \leq 15\epsilon_k\sqrt{n}$, we find that $\mu\bar{\alpha}/C(A) \leq \mu'$ provided

$$C(A) \geq 150 \cdot 30^2(\bar{\chi}_A + 1)\bar{\chi}_A n^5 g^{2n}\sqrt{1+\delta}.$$

CASE 2, $\alpha = \alpha_k$ and α_k is defined by (20).

First let us consider i as in (46) for an arbitrary k of Type II (not necessarily the k determining α_k). We need a lower bound on ϵ_k for layers of Type II. Recalling (27), we observe that

$$\begin{aligned}
\epsilon_k &= \|\Delta_k^{-1} A_k^T \mathbf{r} - (X_k S_k)^{1/2} \mathbf{e}\| \\
&\geq \|(X_k S_k)^{1/2} \mathbf{e}\| - \|\Delta_k^{-1} A_k^T \mathbf{r}\| \\
&\geq (1-\delta) - 1/(15\sqrt{n}) \\
&\geq 1/2.
\end{aligned}$$

Thus, from (46) there is an $i \in J_1 \cup \cdots \cup J_k$ such that for all $\mu' \leq \mu$,

$$s_i(\mu') \geq \frac{\phi_k}{2(\bar{\chi}_A + 1) n^{1.5} \sqrt{1+\delta}} \tag{55}$$

Now, consider an arbitrary $\mu' \leq \mu$. We have the following computation using Lemma 5:

$$\begin{aligned}
P_k \mathbf{b} &= P_k \cdot \mu' A (S')^{-1} \mathbf{e} \\
&= \mu' P_k A_N (S'_N)^{-1} \mathbf{e} \\
&= \mu' A_k^T \mathbf{v}
\end{aligned}$$

where $A_N = (A_k, \ldots, A_p)$, $S'_N = \mathrm{diag}(S'_k, \ldots, S'_p)$, and $\|\mathbf{v}\| \leq \bar{\chi}_A \|(S'_N)^{-1}\mathbf{e}\|$. Thus,

$$\|P_k \mathbf{b}\| \leq \mu' \bar{\chi}_A \|A_k\| \cdot \|(S'_N)^{-1} \mathbf{e}\|.$$

This means that there exists an $j \in J_k \cup \cdots \cup J_p$ such that

$$s_j(\mu') \leq \frac{\mu' \sqrt{n} \bar{\chi}_A \|A\|}{\|P_k \mathbf{b}\|}.$$

Now, fix the specific k of Type II such that $\bar{\alpha} = \alpha_k$, and select

$$\mu' = \frac{\phi_k \cdot \|P_k \mathbf{b}\|}{10 n^3 (\bar{\chi}_A + 1) \bar{\chi}_A g^n \|A\| \sqrt{1+\delta}}. \tag{56}$$

Then for all $\mu'' \leq \mu'$, combining the inequality in the last paragraph with Lemma 8, there is a j such that

$$s_j(\mu'') \leq \frac{\phi_k}{10 n^{1.5} (\bar{\chi}_A + 1) g^n \sqrt{1+\delta}}. \tag{57}$$

Dividing (55) by this inequality, yielding the result that for all $\mu'' \le \mu'$,

$$\frac{s_i(\mu'')}{s_j(\mu'')} \ge 5g^n.$$

This shows that (45) holds for a crossover event.

On the other hand, $i \in J_1 \cup \cdots \cup J_k$ and $j \in J_k \cup \cdots \cup J_p$, so initially $s_i \le s_j g^n$. Arguing as in Case I, $s_i(\mu) \le 3g^n s_j(\mu)$. Thus, an event has occurred.

Now, let us compare μ' in (56) to $\mu\bar{\alpha}/C(A)$. Observe from (20) that

$$\begin{aligned}\mu\bar{\alpha} &= \mu\alpha_k \\ &\le 30n^{1.5}\theta_k\chi_A \cdot \|P_k\mathbf{b}\|\end{aligned}$$

so

$$\begin{aligned}\mu' &= \frac{\phi_k \cdot \|P_k\mathbf{b}\|}{10n^3(\bar{\chi}_A+1)\bar{\chi}_A g^n\|A\|\sqrt{1+\delta}} \\ &\ge \frac{\phi_k\mu\bar{\alpha}}{300n^{4.5}(\bar{\chi}_A+1)\bar{\chi}_A\chi_A\|A\|g^n\theta_k\sqrt{1+\delta}} \\ &\ge \frac{\mu\bar{\alpha}}{300n^{4.5}(\bar{\chi}_A+1)^2\chi_A\|A\|g^{2n}\sqrt{1+\delta}}.\end{aligned}$$

Thus, we must pick

$$C(A) \ge 300n^{4.5}(\bar{\chi}_A+1)^2\chi_A\|A\|g^{2n}\sqrt{1+\delta}$$

to ensure that $\mu\bar{\alpha}/C(A) \le \mu'$.

CASE 3, $\alpha = \beta_{k,l}$, where k is Type I, l is Type II, and $k > l$.

From (53) in the Case 1 analysis, we conclude that for any $\mu' \le \mu$, there is a $j \in J_k \cup \cdots \cup J_p$ such that

$$s_j(\mu') \le \frac{2 \cdot 30^2\mu'\bar{\chi}_A\theta_k n^2}{\mu}.$$

On the other hand, from (55) from Case 2 applied to l, that there is an $i \in J_1 \cup \cdots \cup J_l$ such that for all $\mu' \le \mu$,

$$s_i(\mu') \ge \frac{\phi_l}{2(\bar{\chi}_A+1)n^{1.5}\sqrt{1+\delta}}.$$

Thus, if we pick

$$\mu' = \frac{\mu\phi_l}{20 \cdot 30^2 \theta_k(\bar{\chi}_A + 1)\bar{\chi}_A n^{4.5} g^n \sqrt{1+\delta}} \tag{58}$$

and apply Lemma 8, then for all $\mu'' \leq \mu'$ we obtain

$$\frac{s_i(\mu'')}{s_j(\mu'')} \geq 5g^n.$$

On the other hand, $s_i \leq s_j/g$ because $i \in J_1 \cup \cdots \cup J_l$ and $j \in J_k \cup \cdots \cup J_p$, and $k > l$. This means also that the much weaker relation $s_i(\mu) \leq 3g^n s_j(\mu)$ holds. Thus, an event has taken place.

Again, we compare μ' from (58) to $\mu\bar{\alpha}/C(A)$. Note that $\bar{\alpha} \leq 30\bar{\chi}_A\theta_l n^2/\phi_k$ by (21), so we have the following bound:

$$\begin{aligned}
\mu' &= \frac{\mu\phi_l}{20 \cdot 30^2 \theta_k(\bar{\chi}_A + 1)\bar{\chi}_A n^{4.5} g^n \sqrt{1+\delta}} \\
&\geq \frac{\mu\phi_l}{20 \cdot 30^2 \theta_k(\bar{\chi}_A + 1)\bar{\chi}_A n^{4.5} g^n \sqrt{1+\delta}} \cdot \frac{\bar{\alpha}\phi_k}{30\bar{\chi}_A\theta_l n^2} \\
&\geq \frac{\mu\bar{\alpha}\phi_k\phi_l}{20 \cdot 30^3 \theta_k\theta_l(\bar{\chi}_A + 1)\bar{\chi}_A^2 n^{6.5} g^n \sqrt{1+\delta}} \\
&\geq \frac{\mu\bar{\alpha}}{20 \cdot 30^3 (\bar{\chi}_A + 1)\bar{\chi}_A^2 n^{6.5} g^{3n} \sqrt{1+\delta}}.
\end{aligned}$$

Thus, we must pick

$$C(A) \geq 20 \cdot 30^3 (\bar{\chi}_A + 1)\bar{\chi}_A^2 n^{6.5} g^{3n} \sqrt{1+\delta}$$

for Case 3.

We now explain how to pick the two parameters g and $C(A)$ taking into account all constraints. First, g must satisfy (26), (32), and (52). Of these, (52) dominates, hence we may pick

$$g = 2 \cdot 30^2 (1+\delta) n^2 \bar{\chi}_A \tag{59}$$

or anything larger. Second $C(A)$ has three lower bounds given by the three cases of the previous proof; the third dominates except for the $\chi_A\|A\|$ factor in the second. Note that $\chi_A\|A\| \geq \bar{\chi}_A$, hence we pick

$$C(A) = 20 \cdot 30^3 (\bar{\chi}_A + 1)^2 \chi_A \|A\| n^{6.5} g^{3n} \sqrt{1+\delta}. \tag{60}$$

to satisfy all three cases. ∎

This theorem shows that Algorithm LIP has finite termination; in particular, it must terminate after $n(n-1)/2$ main-loop iterations, since each main loop iteration forces at least one crossover event to occur. These crossover events are clearly disjoint because the central path parameter decreases monotonically during the algorithm.

We now make some remarks on the choice of α in Algorithm LIP. Up to now, we have assumed that α is chosen to be equal to $\bar{\alpha}$ defined in Section 5. As an alternative, suppose we picked α as follows. First, identify α_0 so that $\mathbf{y}-(1-\alpha)\mathbf{r}$ is infeasible for any $\alpha<\alpha_0$ but feasible for $\alpha\in[\alpha_0,1]$. This α_0 can be identified with a simple ratio test. Let $\alpha_1=\max(\alpha_0,0)$. Now find an $\alpha\in[\alpha_1,1]$ such that

$$\delta(\mathbf{y}-(1-\alpha)\mathbf{r},\mu\alpha)\leq 0.75 \tag{61}$$

but

$$\delta(\mathbf{y}-(1-\alpha/2)\mathbf{r},\mu\alpha/2)>0.75 \quad \text{or} \quad \alpha/2<\alpha_1. \tag{62}$$

From this α we clearly maintain proximity to the central path by (61). On the other hand, (62) implies that $\alpha/2<\bar{\alpha}$ by Theorem 1. Thus, if we insert an additional factor of '2' in the formula for $C(A)$, then we can guarantee that Theorem 2 will also go through. Thus, all of the theorems proved so far still hold if we determine α in Algorithm LIP with some kind of a line search for a solution in $[\alpha_1,1]$ to conditions (61), (62) (and treating $\alpha=0$ as a special case). Notice that this eliminates the need for computing $P_1\mathbf{b},\ldots,P_p\mathbf{b}$ and $\mathbf{r}^{(1)},\ldots,\mathbf{r}^{(p)}$ that were used to determine $\bar{\alpha}$, presumably a substantial savings in practice.

The points t satisfying $\delta(\mathbf{y}-(1-t)\mathbf{r},\mu t)=0.75$ are roots of a certain polynomial in t, so presumably we could apply polynomial root-finding techniques to solve (61), (62). We have not worked out the complexity of this procedure.

As mentioned in the introduction, a consequence of the validity of Algorithm LIP is a new characterization of the central path as being composed of $O(n^2)$ alternating curved and straight segments.

Corollary 3 *Consider an instance of* (1) *that is bounded and has a strictly interior feasible point. Let $M>0$ be arbitrary. Then there exist a sequence of $k+1$ breakpoints*

$$0=\mu_0\leq\mu_1\leq\mu_2\leq\cdots\leq\mu_k=M$$

such that $k \leq n(n-1)/2$ and such that:

1. *For all $i \in \{0, \ldots, k-1\}$ such that i is even, there exists vectors $\mathbf{u}_i, \mathbf{v}_i$ such that $\delta(\mathbf{u}_i + \mu \mathbf{v}_i, \mu) \leq 0.75$ for all $\mu \in (\mu_i, \mu_{i+1})$. In other words, the central path is approximately a straight line segment over this range of parameters.*

2. *For all $i \in \{0, \ldots, k-1\}$ such that i is odd, μ_{i+1}/μ_i is bounded above by a constant $C(A)$ depending only on A.*

PROOF. This follows immediately from Theorem 1 and Theorem 2. Introduce breakpoints at the beginning and end of each LLS step of Algorithm LIP, and the corollary follows. ∎

Since the statement of this corollary makes no reference to any specific algorithm, it seems that the corollary ought to have a proof that is algorithm-independent. However, we do not know of a direct proof.

A natural question to ask is: if the central path is approximately straight over intervals $[\mu_i, \mu_{i+1}]$ for i even, then shouldn't a traditional path-following interior point method be able to take larger steps and achieve the same complexity as the LIP method? The problem is that the tangent to the central path at $\mathbf{y}(\mu_{i+1})$ (which is the direction followed by traditional algorithms) does not necessarily "point" in the right direction. This is most clearly seen on the last LLS step of Algorithm LIP, when the global minimum is found. The tangent to the central path at $\mathbf{y}(\mu)$ is never precisely aligned with the displacement vector between the $\mathbf{y}(\mu)$ and the global minimum $\mathbf{y}^*$ for any positive μ. In contrast, the LLS step $\mathbf{r}$ is precisely this displacement.

7 Complexity of Algorithm LIP

We can now estimate the complexity of the algorithm. There are a total of at most $n(n-1)/2$ LLS iterations. There are $O(\sqrt{n} \log C(A))$ small steps per LLS step. The small steps dominate the work, and the total number of small steps is $O(n^{2.5} \log C(A))$. If we use (60), then

$$\log C(A) = 2 \log \bar{\chi}_A + \log(\chi_A \|A\|) + \log n + 3n \log g + \text{const.}$$

The dominant terms are $3n \log g$ and $\log(\chi_A \|A\|)$. Note from (59) that

$$\log g = 2 \log n + \log \bar{\chi}_A + \text{const.}$$

Keeping the dominant terms, we see that the overall number of small steps is

$$O(n^{3.5}(\log n + \log(\bar{\chi}_A + 1)) + n^{2.5} \log(\chi_A \|A\|))$$

Let us define

$$c(A) = \log n + \log(\bar{\chi}_A + 1) + \log(\chi_A \|A\|)/n. \tag{63}$$

Then we can write the complexity as $O(n^{3.5} c(A))$ iterations, where in this context "iteration" means either a small step or an LLS step. Each iteration requires solution of a system of linear equations, which takes $O(mn^2)$ arithmetic operations.

It should be noted that this complexity bound is independent of the size of μ that is given as input data to Algorithm LIP. This is an important point for the initialization procedure in the next section.

8 Initialization

We now examine the question of initializing Algorithm LIP, that is, given $A, \mathbf{b}, \mathbf{c}$, find an approximately centered $\mathbf{y}, \mu$. The procedure used is as follows: we start with a basic subset of constraints whose *analytic center* is known in closed form. We then add the constraints one at a time, each time running Algorithm LIP to find the analytic center of a new polytope. After computing analytic centers of a sequence of $n - m$ polytopes, the last polytope is the feasible region of the problem (plus some extra, "dummy" constraints). From this point we then run Algorithm LIP. At first glance, it seems that the running time of initialization will be a factor $n - m$ larger than the running time of Algorithm LIP. With a more careful analysis, however, we are able to show that initialization has the same complexity (up to a constant factor) as Algorithm LIP. We call this algorithm "cutting-plane initialization," which is similar to an interior-point column generation algorithm (for example, see Goffin et al. [5]).

We first comment on a more standard approach to initializing interior point methods, namely, adding a dummy variable. For example, a common technique is to replace the inequality constraint in (1) with $A^T\mathbf{y} + t\mathbf{q} \geq \mathbf{c}$ where $t \geq 0$ is a new variable and $\mathbf{q}$ is a vector with all positive entries. Note that this problem always has a strictly feasible point; simply take $\mathbf{y} = \mathbf{0}$ and t sufficiently large. The difficulty with the approach, as far as our analysis is concerned, is that we do not know how to analyze $\chi_{A'}$ or $\bar{\chi}_{A'}$ in terms of A, where

$$A' = \begin{pmatrix} A \\ \mathbf{q}^T \end{pmatrix}.$$

We now describe cutting plane initialization. We generate a sequence of polytopes $F_m, F_{m+1}, \ldots, F_n$. Polytope F_k in this sequence has $k + m$ constraints, of which k are from the original problem and m are "dummies." We have a containment relationship $F_m \supset F_{m+1} \supset \cdots \supset F_n$.

We start by describing F_m. Select a set of m linearly independent columns of A arbitrarily, which we will assume are numbered $1, \ldots, m$. Now define

$$F_m = \{\mathbf{y} : A_m^T\mathbf{y} \geq \mathbf{c}_m, A_m^T\mathbf{y} \leq N\mathbf{e}\}.$$

Here A_m is notation for $(\mathbf{a}_1, \ldots, \mathbf{a}_m)$, that is, the first m columns of A, and N is a large positive number determined below.

Recall that the *analytic center* and *approximate analytic center* were defined by (39) and (40). For polytope F_m, because of its simple structure, we can compute the analytic center in closed form. It is precisely the solution to the square linear system

$$A_m^T\mathbf{y}^* = \frac{1}{2}(\mathbf{c}_m + N\mathbf{e}).$$

(Hence, F_m is full dimensional.)

The sequence of polytopes we work with is $F_m, \ldots, F_n$, where

$$F_k = \{\mathbf{y} : A_k^T\mathbf{y} \geq \mathbf{c}_k, A_m^T\mathbf{y} \leq N\mathbf{e}\}.$$

Let $\hat{A}_k$ be the constraint matrix for this problem, that is, $\hat{A}_k = [-A_m, A_k]$ and let $\hat{\mathbf{c}}_k$ be the right-hand side, that is $\hat{\mathbf{c}}_k = (-N\mathbf{e}^T, \mathbf{c}_k^T)^T$. Thus, $F_k = \{\mathbf{y} : \hat{A}_k^T\mathbf{y} \geq \hat{\mathbf{c}}_k\}$.

Now we describe an incremental procedure: given an approximate analytic center for F_k, the procedure returns an approximate analytic center for F_{k+1}. The main idea of the procedure is that the analytic center of F_{k+1}

always lies on the central path for F_k if we use the objective function in (64) below.

More specifically, the procedure is as follows. Let $\mathbf{y}_0$ be an approximate center of F_k, let $\mathbf{s} = \hat{A}_k\mathbf{y}_0 - \hat{\mathbf{c}}_k$, let $S = \text{diag}(\mathbf{s})$, and let $\mathbf{f}$ be chosen so that $\hat{A}_k^{-1}S^{-1}\mathbf{f} = \mathbf{0}$, $\|\mathbf{f} - \mathbf{e}\| \le 0.75$. Solve the nonsingular square linear system

$$\hat{A}_k S^{-2}\hat{A}_k^T\mathbf{w} = -\mathbf{a}_{k+1}.$$

This system is nonsingular because $\hat{A}_k$ contains a basis A_m. Let $\mathbf{w}' = S^{-1}\hat{A}_k^T\mathbf{w}$ so that $\hat{A}_k S^{-1}\mathbf{w}' = -\mathbf{a}_{k+1}$, and let $\mu_0 = 100\|\mathbf{w}'\|$. Then we have

$$-\mathbf{a}_{k+1} - \mu_0\hat{A}_k S^{-1}\mathbf{e} = \mu_0\hat{A}_k S^{-1}(\mathbf{f} - \mathbf{e} + \mathbf{w}'/\mu_0)$$

and

$$\begin{aligned}\|\mathbf{f} - \mathbf{e} + \mathbf{w}'/\mu_0\| &\le \|\mathbf{f} - \mathbf{e}\| + \|\mathbf{w}'/\mu_0\| \\ &\le 0.76.\end{aligned}$$

Thus, $\mathbf{y}_0$ is approximately centered for LP problem

$$\begin{array}{ll}\text{minimize} & -\mathbf{a}_{k+1}^T\mathbf{y} \\ \text{subject to} & \hat{A}_k\mathbf{y} \ge \hat{\mathbf{c}}_k.\end{array} \tag{64}$$

with central path parameter μ_0. The centering distance can be improved from 0.76 to 0.25 with three Newton steps. Now we apply Algorithm LIP to problem (64)—note that we have an initial approximately centered starting point.

We run Algorithm LIP, not necessarily to completion, but until we can find an approximately centered point $\mathbf{y}$ with parameter μ such that

$$\mathbf{a}_{k+1}^T\mathbf{y} - c_{k+1} = \mu. \tag{65}$$

We search for such a $\mathbf{y}$ as follows. If, initially, when LIP begins, we have $\mathbf{a}_{k+1}^T\mathbf{y}_0 - c_{k+1} > \mu_0$, then we do not run LIP at all; we merely keep the same value of $\mathbf{y}_0$ and increase μ_0 so that (65) is satisfied. The computations above show that $\mathbf{y}_0$ remains approximately centered for this larger parameter value.

The other case is that $\mathbf{a}_{k+1}^T\mathbf{y}_0 - c_{k+1} < \mu_0$. Then we run Algorithm LIP until we detect two approximately centered points (either before and after an LLS steps or before and after a small steps) such that

$$\mathbf{a}_{k+1}^T\mathbf{y}_1 - c_{k+1} < \mu_1 \tag{66}$$

and

$$\mathbf{a}_{k+1}^T \mathbf{y}_2 - c_{k+1} > \mu_2 \tag{67}$$

with $\mu_1 > \mu_2$ and $\mathbf{y}_1, \mathbf{y}_2$ approximately centered for μ_1, μ_2 respectively. Now we solve the following linear interpolation problem for $\beta \in [0, 1]$:

$$\mathbf{a}_{k+1}^T (\beta \mathbf{y}_1 + (1-\beta)\mathbf{y}_2) - c_{k+1} = \mu_1 \beta + \mu_2 (1-\beta).$$

We take $\mu = \mu_1\beta + \mu_2(1-\beta)$ and $\mathbf{y} = \mathbf{y}_1\beta + \mathbf{y}_2(1-\beta)$ for (65). The fact that this choice of $\mathbf{y}$ is approximately centered follows from Theorem 1 in the case that μ_1, μ_2 are the parameter values before and after an LLS step. In this case, the centering distance is 0.75. If the transition from μ_1 to μ_2 occurs during one of the three Newton steps for recentering the LLS solution, then again the point is approximately centered with distance 0.75. Finally, if the transition from (66) to (67) occurs in a small step, then conventional interior point theory tells us that interpolating between the steps (which amounts to truncating a certain Newton step) preserves approximate centering with distance 0.25.

Once we have a solution for (65), then we have an approximate analytic center $\mathbf{y}$ for F_{k+1}. This is because the approximate centering condition of $\mathbf{y}$ for (64) means that

$$-\mathbf{a}_{k+1}/\mu - \hat{A}_k S^{-1}\mathbf{e} = \hat{A}_k S^{-1}\mathbf{v}$$

with $\|\mathbf{v}\| \leq 0.75$, i.e.,

$$-(\hat{A}_k, \mathbf{a}_{k+1}) \begin{pmatrix} S^{-1} & 0 \\ 0 & \mu^{-1} \end{pmatrix} \mathbf{e} = (\hat{A}_k, \mathbf{a}_{k+1}) \begin{pmatrix} S^{-1} & 0 \\ 0 & \mu^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ 0 \end{pmatrix}.$$

Since μ^{-1} is precisely the reciprocal of the slack $\mathbf{a}_{k+1}^T\mathbf{y} - c_{k+1}$ by (65), this latter equation means that $\mathbf{y}$ is an approximate analytic center for F_{k+1}.

We call this procedure one "phase" of the cutting plane algorithm. Thus, there are $n - m$ phases total. We have not addressed the case that (67) is never satisfied during Algorithm LIP, that is, even at the optimum point $\mathbf{y}$ for (64),

$$\mathbf{a}_{k+1}^T \mathbf{y} - c_{k+1} \leq 0.$$

If this inequality is strict, then clearly F_{k+1} is infeasible. We claim the original LP (1) is also infeasible, provided that N is chosen sufficiently large. Instead, we prove the contrapositive: suppose the original problem (1) is feasible. By assumption of full rank, there is a basic feasible solution to (1), that is, an index set for a basis B and a point $\mathbf{y}$ such that $A_B^T\mathbf{y} = \mathbf{c}_B$ and $A^T\mathbf{y} \geq \mathbf{c}$. Observe that $A_m^T\mathbf{y} = A_m^T A_B^{-T}\mathbf{c}_B$, hence

$$\begin{aligned} \|A_m^T\mathbf{y}\| &= \|A_m^T A_B^{-T}\mathbf{c}\| \\ &\leq \|A_m^T A_B^{-T}\| \cdot \|\mathbf{c}\| \\ &= \|A_B^{-1}A_m\| \cdot \|\mathbf{c}\| \\ &\leq \bar{\chi}_A\|\mathbf{c}\|. \end{aligned}$$

The last line follows from Lemma 4. Thus, we choose $N > \bar{\chi}_A\|\mathbf{c}\|$, and we are guaranteed that $A_m^T\mathbf{y} < N\mathbf{e}$, hence $\mathbf{y} \in F_{k+1}$.

Another possibility is that at the optimum point $\mathbf{y}$ for (64),

$$\mathbf{a}_{k+1}^T\mathbf{y} - c_{k+1} = 0.$$

This means that this equation must hold for every feasible point of original feasible region. Moreover, if this equation holds, and other constraints are satisfied as equalities at the optimum of (64), then these inequalities are also satisfied as equations at every feasible point of (1). These equations may be eliminated by eliminating some variables. Then the cutting-plane algorithm continues on a lower dimensional problem. Note that $\mathbf{y}$ is an approximate analytic center for the lower-dimensional problem by Corollary 2.

An alternative to eliminating variables is as follows. Suppose we run Algorithm LIP for phase k to completion, discovering that

$$\mathbf{a}_{k+1}^T\mathbf{y}^* - c_{k+1} = \mu = 0. \tag{68}$$

From Corollary 2 we know that $\mathbf{y}^*$ is approximately centered in the optimal face. Let $\hat{A}_{\text{eq}}$ be the set of constraints that

$$\hat{A}_{\text{eq}}^T\mathbf{y}^* = \hat{\mathbf{c}}_{\text{eq}}.$$

Note that LIP also generates a primal optimizer $\mathbf{x}^*$ with

$$\hat{A}_{\text{eq}}\mathbf{x}_{\text{eq}}^* = -\mathbf{a}_{k+1}, \quad \mathbf{x}_{\text{eq}}^* > \mathbf{0}.$$

Then, we must have that for every possible feasible point $\mathbf{y}$ of the original LP problem, this set of constraints, plus $\mathbf{a}_{k+1}$, must be satisfied as equalities. Let

$\hat{A}_0 = (\hat{A}_{\rm eq}, \mathbf{a}_{k+1})$ and let the rest of $\hat{A}_k$ be $\hat{A}_{k+1}$, and let the corresponding $\hat{\mathbf{c}}$ be partitioned as $\hat{\mathbf{c}}_0$ and $\hat{\mathbf{c}}_{k+1}$, respectively. Then we have a degenerate F_{k+1} given by

$$F_{k+1} = \{\mathbf{y} : \ \hat{A}_0^T\mathbf{y} = \hat{\mathbf{c}}_0, \ \hat{A}_{k+1}^T\mathbf{y} \geq \hat{\mathbf{c}}_{k+1}\}.$$

Let $\mathbf{x}_0^* = (\mathbf{x}_{\rm eq}, 1)$; then $\mathbf{x}_0^* > \mathbf{0}$ and $\hat{A}_0\mathbf{x}_0^* = \mathbf{0}$. One can verify that $\mathbf{y}^*$ is still an approximate analytic center in F_{k+1}. In fact, $\mathbf{y}^*$ satisfies

$$\hat{A}_{k+1}(S^*)^{-1}(\mathbf{e}+\mathbf{v}) \in \text{Range}(\hat{A}_0) \quad \text{for some} \quad \mathbf{v}$$

and $\|\mathbf{v}\| \leq 0.75$ and $\mathbf{s}^* = \hat{A}_{k+1}^T\mathbf{y}^* - \hat{\mathbf{c}}_{k+1}$. This is the centering condition for degenerate polyhedra. Now we continue to phase $k+1$ of initialization as we did before, starting from $\mathbf{y}^* \in F_{k+1}$. During this process, we may detect infeasibility of the original LP and stop.

Since Algorithm LIP assumes nondegenerate polyhedra, we must use a generalization of Algorithm LIP that involves one additional layer, J_0. The constraints indexed by J_0 have all their slacks identically zero throughout the algorithm. Thus, this layer has highest "priority" in the LLS step; the diagonal weight matrix for this layer in the LLS step can be arbitrary since the least-squares problem has an exact (zero-residual) solution. This layer is always of Type I and always has $\alpha_0 = \epsilon_0 = 0$. It can be proved that all of the theory developed so far for Algorithm LIP works for this generalization. Indeed, degenerate polyhedra are the limits of full-dimension polyhedra as $\mathbf{c}$ is changed, but the bounds proved for Algorithm LIP are independent of the choice of $\mathbf{c}$.

After phase n of initialization, the feasible set F_n of the original problem, if it is not empty, is represented by

$$F_n = \{\mathbf{y} : \ \hat{A}_0^T\mathbf{y} = \hat{\mathbf{c}}_0, \ \hat{A}_n^T\mathbf{y} \geq \hat{\mathbf{c}}_n\},$$

where $\hat{A}_0$ is the set of constraints active at every feasible point. Note again, we have a $\mathbf{y}^*$, which is an approximate analytic center for F_n, and an $\bar{\mathbf{x}}_0^*$ such that

$$\hat{A}_0\bar{\mathbf{x}}_0^* = \mathbf{0}, \quad \bar{\mathbf{x}}_0^* > \mathbf{0}. \tag{69}$$

If F_n is a singleton, we may stop and $\mathbf{y}^*$ is the unique optimizer. Otherwise, we move to the optimization phase solving

$$\begin{array}{ll} \text{minimize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & \hat{A}_0^T\mathbf{y} = \hat{\mathbf{c}}_0, \\ & \hat{A}_n^T\mathbf{y} \geq \hat{\mathbf{c}}_n. \end{array} \tag{70}$$

Applying Algorithm LIP, we generate $\mathbf{y}^*$ on the optimal face. Suppose in addition to $\hat{A}_0^T\mathbf{y}^* = \hat{\mathbf{c}}_0$ we have $\hat{A}_B^T\mathbf{y}^* = \hat{\mathbf{c}}_B$, then we have a primal optimizer $(\mathbf{x}_0^*, \mathbf{x}_B^*)$ with $\hat{A}_0\mathbf{x}_0^* + \hat{A}_B\mathbf{x}_B^* = \mathbf{b}$ and $\mathbf{x}_B^* > 0$. Since $\mathbf{x}_0^*$ may not be positive, $(\mathbf{x}_0^*, \mathbf{x}_B^*)$ may not be a feasible solution for the primal. However, we can use (69) and redefine

$$\mathbf{x}_0^* := \mathbf{x}_0^* + \beta\bar{\mathbf{x}}_0^* > \mathbf{0}$$

for β large enough. Then we have $(\mathbf{x}_0^*, \mathbf{x}_B^*) > \mathbf{0}$ and maintain

$$\hat{A}_0\mathbf{x}_0^* + \hat{A}_B\mathbf{x}_B^* = \mathbf{b}.$$

Thus, we generate a feasible and strictly complementary solution for the primal (4).

In general, when the initialization procedure terminates, we are left with an approximate analytic center for F_n, which is a subset of the feasible region for the original problem. We claim again that if N is at least $\bar{\chi}_A\|\mathbf{c}\|$, then the optimal value of minimizing $\mathbf{b}^T\mathbf{y}$ over F_n is the same as the optimal point of $\mathbf{b}^T\mathbf{y}$ in (1), unless the original problem is unbounded. Here, the reasoning is the same as earlier—if the original problem has a lower bound on the objective function, then there is an optimal basic solution.

In practice, there is an technique that makes the cutting-plane algorithm much more efficient. We make the observation that if the new constraint $\mathbf{a}_{k+1}^T\mathbf{y} - c_{k+1}$, evaluated at the approximate analytic center of F_k, has a sufficiently large and positive slack, then it is not necessary to use Algorithm LIP to find the approximate analytic of F_{k+1}. Instead, it suffices to use a constant number of Newton iterations starting from the old center. This technique does not seem to reduce the worst-case complexity, however.

Now we analyze the complexity of the cutting plane algorithm. First, observe that we do not work with the original A in this algorithm. We instead work with $\hat{A}_k$ for some k. In this regard, we have the following lemma.

Lemma 11 *For all $k = 1, \ldots, n$, $\chi(\hat{A}_k) \le 2\chi_A$ and $\bar{\chi}(\hat{A}_k) \le 4\bar{\chi}_A$.*

Proof. First, we address $\chi(\hat{A}_k)$. Let $\mathbf{b}$ be arbitrary, $D \in \mathcal{D}$ arbitrary, and consider a solution to

$$\hat{A}_k D\hat{A}_k^T\mathbf{y} = \hat{A}_k D\mathbf{c}.$$

The goal is to obtain a bound on $\|\mathbf{y}\|$ in terms of $\|\mathbf{c}\|$. This equation may be written in block form. Recall that $\hat{A}_k = [-A_m, A_k]$. Partition $D = \mathrm{diag}(D_1, D_2)$ and $\mathbf{c} = (\mathbf{c}_1^T, \mathbf{c}_2^T)^T$ conformally. Then

$$A_m D_1 A_m^T \mathbf{y} + A_k D_2 A_k^T \mathbf{y} = -A_m D_1 \mathbf{c}_1 + A_k D_2 \mathbf{c}_2.$$

Let $\bar{D}_1$ be an $n \times n$ matrix that agrees with D_1 in positions indexed by A_m, and has a small $\epsilon > 0$ in other diagonal positions. Similarly, let $\bar{D}_2$ agree with D_2 in positions indexed by A_k, and let it have ϵ in other diagonal entries. Finally, extend $\mathbf{c}_1, \mathbf{c}_2$ to vectors $\mathbf{c}_1', \mathbf{c}_2'$ with n entries by filling in zeros. Then we have

$$A\bar{D}_1 A^T \mathbf{y} + A\bar{D}_2 A_k^T \mathbf{y} \approx -A\bar{D}_1 \mathbf{c}_1' + A\bar{D}_2 \mathbf{c}_2'$$

where the approximation may be made arbitrarily close as ϵ gets small. Let $\bar{D} = \bar{D}_1 + \bar{D}_2$; then we have

$$A\bar{D}A^T \mathbf{y} \approx -A\bar{D}[\bar{D}^{-1}(-\bar{D}_1 \mathbf{c}_1' + \bar{D}_2 \mathbf{c}_2')].$$

Now let $\bar{\mathbf{c}} = \bar{D}^{-1}(-\bar{D}_1 \mathbf{c}_1' + \bar{D}_2 \mathbf{c}_2')$; note that this vector depends on ϵ. Observe that because $\bar{D}$ is the sum of $\bar{D}_1$ and $\bar{D}_2$, both positive, the scaling matrices $\bar{D}^{-1}\bar{D}_1$ and $\bar{D}^{-1}\bar{D}_2$ both have positive diagonal entries bounded by 1. Thus, $\|\bar{\mathbf{c}}\| \leq \|\mathbf{c}_1\| + \|\mathbf{c}_2\|$, so $\|\bar{\mathbf{c}}\| \leq 2\|\mathbf{c}\|$. Since $\mathbf{y}$ solves

$$A\bar{D}A^T \mathbf{y} \approx A\bar{D}\bar{\mathbf{c}}$$

and the matrix on the left is nonsingular as $\epsilon \to 0$ (because D has a strictly positive entry bounded independently of ϵ in positions corresponding to A_m, a basis), then by definition $\|\mathbf{y}\| \leq \chi_A \|\bar{\mathbf{c}}\| \leq 2\chi_A \|\mathbf{c}\|$. Thus, we have proved that $\chi(\hat{A}_k) \leq 2\chi_A$.

Now for the second claim in the lemma, we follow the same argument to conclude that $\|A^T \mathbf{y}\| \leq 2\bar{\chi}_A \|\mathbf{c}\|$. Since $\|\hat{A}_k^T \mathbf{y}\| = \|-A_m{}^T \mathbf{y} + A_k^T \mathbf{y}\| \leq 2\|A^T \mathbf{y}\|$, we conclude that

$$\|\hat{A}_k^T \mathbf{y}\| \leq 4\bar{\chi}_A \|\mathbf{c}\|.$$

This proves the second claim in the lemma. ∎

As noted above, during the cutting plane algorithm, it may become necessary to either eliminate variables or work with a degenerate polyhedron.

Eliminating variables is a further change to $\chi_A, \bar{\chi}_A$ that also can be bounded. We do not provide those bounds, however, because the other alternative (working with a degenerate polyhedron) does not require changes to A.

This lemma shows that the total complexity of the $n - m$ phases in the cutting-plane algorithm is no worse than $n - m$ times the complexity of the LLS method. We can do better than this, however; we can prove that the factor of $n - m$ may be omitted and that the total work of the cutting plane method is bounded by a constant multiple of the work of the LIP method.

The reason for this is as follows. Suppose a crossover event takes place during some phase of the cutting-plane method. We claim that this event is "permanent" in the sense that the same event can never recur during any other phases of the cutting-plane method, nor during the final LIP solution of (1).

Examining the proof of Theorem 2, we see that in all three cases, when a crossover takes place between two indices i and j, the slack s_i that remains large is determined by Lemma 10 and the constraint that is made small must remain small by Lemma 8. Lemma 10 holds even in the case of the cutting plane algorithm because it relies on the residual of a layer of constraints J_k in the current phase. As additional constraints are added by the cutting plane algorithm, these constraints in J_k remain in the problem, so additional constraints do not change the main idea of the lemma.

On the other hand, Lemma 8 apparently may not hold for the cutting-plane algorithm. However, we can prove a new version of the lemma. Let $\bar{n} = m + n$. Note that $\bar{n}$ denotes the number of constraints in F_n.

Lemma 12 *Let $(\mathbf{y}(\mu), \mu)$ be a point on the central path encountered in phase k of the cutting-plane initialization procedure. Let $(\mathbf{y}', \mu')$ be a central-path point either in the same phase, in which case we assume $\mu' \leq \mu$, or in a later phase, in which case μ' is arbitrary. Let the slacks be $\mathbf{s}(\mu), \mathbf{s}'$ respectively. Then for any $i \leq m + k$,*

$$s_i' \leq 80\bar{n} s_i(\mu).$$

PROOF. In the case that the two points occur in the same phase of cutting plane algorithm, the result is obvious by Lemma 8; in this case the bound holds with $m + k \leq 80\bar{n}$.

In the other case, during phase k, we do not have $\mathbf{y}(\mu)$ but rather an approximate central path point $\hat{\mathbf{y}}$ such that

$$\delta(\hat{\mathbf{y}}, \mu) \leq 0.75.$$

In this case, a variant of Corollary 1 that estimates the left-hand product of (7) more carefully shows that for all $i = 1, \ldots, m + k$ we have

$$0.05\hat{s}_i \leq s_i(\mu) \leq 4\hat{s}_i \tag{71}$$

where $\hat{\mathbf{s}}$ is the slack at $\hat{\mathbf{y}}$.

Note that $\hat{\mathbf{y}}$ satisfies the approximate central path equation

$$\hat{A}_k \hat{S}^{-1}\mathbf{f} = -\mathbf{a}_{k+1}/\mu$$

with

$$\|\mathbf{f} - \mathbf{e}\| \leq 0.75.$$

This relation can be written as

$$\hat{A}_k \hat{S}^{-1}\mathbf{f} + \mathbf{a}_{k+1}/\mu = \mathbf{0}.$$

Thus, $\hat{\mathbf{y}}$ is also an approximate analytic center for the system

$$F(\mu) = \{\mathbf{y} : \hat{A}_k^T \mathbf{y} \geq \hat{\mathbf{c}}_k, \quad \mathbf{a}_{k+1}^T \mathbf{y} \geq \mathbf{a}_{k+1}^T \hat{\mathbf{y}} - \mu\}$$

with the centering distance bounded by 0.75. Let $\mathbf{y}^F$ be the analytic center for $F(\mu)$ with slack $\mathbf{s}^F \in \mathbb{R}^{m+k+1}$. Then, for all $i = 1, \ldots, m + k$ we again have

$$0.05\hat{s}_i \leq s_i^F \leq 4\hat{s}_i. \tag{72}$$

Note also that $(S^F)^{-1}\mathbf{e}$ is in the null space of $(\hat{A}_k, \mathbf{a}_{k+1})$.

Observe that $\mathbf{a}_{k+1}^T \hat{\mathbf{y}} - \mu \leq c_{k+1}$, because we are assuming that phase k has not terminated yet. Thus,

$$F_{k+1} = \{\mathbf{y} : \hat{A}_k^T \mathbf{y} \geq \hat{\mathbf{c}}_k, \quad \mathbf{a}_{k+1}^T \mathbf{y} \geq c_{k+1}\} \subset F(\mu).$$

Now, let $\mathbf{y}'$ be any point in $F(\mu)$ (which includes any central path point in $F_{k+1}, \ldots, F_n$), and let its slack be $\mathbf{s}' \geq 0$. Then we have

$$\|(S^F)^{-1}\mathbf{s}'\| \leq \mathbf{e}^T(S^F)^{-1}\mathbf{s}' = \mathbf{e}^T(S^F)^{-1}(\mathbf{s}' - \mathbf{s}^F + \mathbf{s}^F) = \mathbf{e}^T(S^F)^{-1}\mathbf{s}^F = m+k+1,$$

which implies that

$$s_i' \le (m+k+1)s_i^F, \quad \text{for} \quad i=1,\ldots,m+k+1.$$

Thus, from (72) and (71),

$$s_i' \le (m+k+1)(4\hat{s}_i) \le (m+k+1)(80 s_i(\mu)), \quad \text{for} \quad i=1,\ldots,m+k.$$

Note that $m+k+1 \le \bar{n} = m+n$. This concludes the proof. ∎

Thus, we can reprove Lemma 10 and Theorem 2 using Lemma 12 instead of Lemma 8 to get the same global bound of $n(n-1)/2$ on the total number of crossover events. Because of the weaker constant in Lemma 12, the bound in (46) changes by a constant factor, and the definition of $C(A)$ in (60) changes by a constant factor. This means that $c(A)$ changes by a constant additive factor, and n is replace by $\bar{n}$, which is bounded by $2n$. Thus, the running time is changed by a constant factor.

9 Integer data and Tardos' theorem

In this section we specialize the LIP algorithm to the case that all of the entries of A are integers. Suppose that the total number of bits required to write A is L_A. We have the following result (similar results can be found in Tuncel [25] and Vavasis and Ye [29]):

Lemma 13 *The parameters χ_A and $\bar{\chi}_A$ are both bounded by $2^{O(L_A)}$.*

Proof. We follow Todd's proof of the finiteness of $\chi_A, \bar{\chi}_A$. Specifically, consider solving a weighted LS problem of finding $\mathbf{y}$ to minimize $\|D(A^T\mathbf{y}-\mathbf{c})\|$. The hyperplanes $\mathbf{a}_1^T\mathbf{y}=c_1, \ldots, \mathbf{a}_n^T\mathbf{y}=c_n$ partition $\mathbb{R}^m$ into an arrangement of convex polyhedral cells, some of which are bounded and some of which are infinite. The solution $\mathbf{y}$ to the weighted LS problem must always lie in one of the bounded cells. This means that it is a convex combination of vertices of the bounded cell. Therefore, the point $\mathbf{y}$ is bounded in norm by the norm of a vertex in this arrangement. A vertex $\mathbf{v}$ satisfies $A_B^T\mathbf{v}=\mathbf{c}_B$ for some basic set of constraints. Thus,

$$\chi_A \le \sup\{\|A_B^{-T}\| : B \text{ indexes a basis}\}.$$

This in turn is bounded by $2^{O(L_A)}$; see, for example, Lemma 3.1 in [26]. Similarly, $\bar{\chi}_A$ is bounded by $\|A\|\chi_A$, so we obtain a bound of the same order. ∎

This leads to a new proof of the following corollary, which is a well-known result due to Tardos.

Corollary 4 *[23] Consider solving an LP of the form* (1) *on a Turing machine. Suppose that* $A, \mathbf{b}, \mathbf{c}$ *have integer entries, and suppose that* L_A *is the number of bits to write* $m \times n$ *matrix* A*. Then there is an algorithm to solve this problem in polynomial time, and furthermore, the number of arithmetic operations is polynomial in* m, n, L_A*.*

Proof. We apply Algorithm LIP to the LP. Each LLS step and each small step is carried out in exact rational arithmetic, but after the step, we truncate $\mathbf{y}$ to $O(L)$ bits, where L is the total number of bits in $A, \mathbf{b}, \mathbf{c}$. The step involves solving linear equations, which can be done in polynomial time [4, 26].

It can be proved that this truncation preserves approximate centering; see, for example, Chapter 3 of [26]. The running time is $O(n^{3.5}c(A))$ interior point steps, and from (63) and the lemma we see that $c(A) = O(L_A + \log m)$. ∎

10 Application to flow problems

We can apply Algorithm LIP to flow problems such as max-flow or min-cost flow. Let the number of arcs and nodes in the network be m and n respectively. (For this section only, this represents a different use of "m" and "n".) For flow problems, Algorithm LIP is strongly polynomial-time because the result in the last section implies $\chi_A, \bar{\chi}_A \leq 2^{O(m)}$, and hence $c(A)$ in (63) is polynomial in m, n. In fact, an even better bound is possible for $\chi_A, \bar{\chi}_A$ in the case of minimum-cost flow. Recall that the minimum cost flow problem is to find the lowest cost flow through a network satisfying flow-balance constraints at every node of the flow, and satisfying upper and lower bound constraints on the flow on each arc. The arc cost is a linear function of the flow through the that arc, and the total cost is the sum of the arc costs.

Let us write the flow problem in primal form. Let A be the node-arc incidence matrix, let $\mathbf{l}$ denote the arc lower bounds and $\mathbf{u}$ the upper bounds. If we let $\mathbf{x}$ be the values of the flows minus the lower bounds, then we have the following problem to determine $\mathbf{x}$:

$$\begin{array}{rl} \text{minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A(\mathbf{x}+\mathbf{l}) = \mathbf{0}, \\ & \mathbf{x}+\mathbf{z} = \mathbf{u}-\mathbf{l}, \\ & \mathbf{x} \geq \mathbf{0}, \mathbf{z} \geq \mathbf{0} \end{array}$$

The vector $\mathbf{z}$ here represents the slack, that is, the residual capacities in the arcs.

Note that for every basis B of A, B^{-1} is a matrix whose every entry is either 1, -1 or 0 because A is a node-arc incidence matrix [18, sec. 13.2]. Now consider the min-cost flow constraint matrix:

$$\hat{A} = \begin{pmatrix} A & 0 \\ I & I \end{pmatrix}.$$

Note that $\hat{A}$ is an $(m+n) \times 2m$ matrix. We show that

$$\bar{\chi}(\hat{A}) \leq O(mn).$$

From the proof of Lemma 9 in the previous section, we see that it suffices to analyze basic submatrices of $\hat{A}$.

Consider any basis $\hat{B}$ of $\hat{A}$. Since $\hat{B}$ is an $(m+n) \times (m+n)$ matrix, $\hat{B}$ must contain $n+k$, $k \geq 0$, columns from the first m columns of $\hat{A}$, and $m-k$ columns among the last m columns of $\hat{A}$. Thus, $\hat{B}$ can be written as

$$\hat{B} = \begin{pmatrix} B & N & 0 \\ 0 & I_k & 0 \\ D_B & 0 & I_{m-k} \end{pmatrix},$$

where D_B is an $(m-k) \times n$ such that each row has exact one 1 and the rest are 0's, I_k is a $k \times k$ identity matrix, and B is a basis for A. Consider solving the system of linear equations with

$$\hat{B}^T\mathbf{y} = \hat{\mathbf{c}}$$

or

$$\begin{pmatrix} B^T & 0 & D_B^T \\ N^T & I_k & 0 \\ 0 & 0 & I_{m-k} \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{c}}_1 \\ \hat{\mathbf{c}}_2 \\ \hat{\mathbf{c}}_3 \end{pmatrix}.$$

To solve this system we first obtain

$$\mathbf{y}_3 = \hat{\mathbf{c}}_3,$$

and then we solve

$$B^T \mathbf{y}_1 = \hat{\mathbf{c}}_1 - D_B^T \mathbf{y}_3 = \hat{\mathbf{c}}_1 - D_B^T \hat{\mathbf{c}}_3,$$

and finally we have

$$\mathbf{y}_2 = \hat{\mathbf{c}}_2 - N^T \mathbf{y}_1.$$

Note that

$$\|\mathbf{y}_3\| \le \|\hat{\mathbf{c}}\|$$

$$\|\mathbf{y}_1\| \le O(m)\|\hat{\mathbf{c}}\|$$

and

$$\|\mathbf{y}_2\| \le O(m\sqrt{n})\|\hat{\mathbf{c}}\|.$$

Thus, we must have

$$\|\mathbf{y}\| = \|(\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{y}_3^T)^T\| \le O(m\sqrt{n}) \cdot \|\hat{\mathbf{c}}\| \le O(m\sqrt{n}) \cdot \|\mathbf{c}\|.$$

Therefore, $\chi(\hat{A}) \le O(m\sqrt{n})$ and

$$\bar{\chi}(\hat{A}) \le \sup\{\|\hat{A}^T \mathbf{y}\|/\|\mathbf{c}\|\} \le O(mn).$$

Therefore, the overall complexity for Algorithm LIP applied to the min-cost flow problem is $O(m^{3.5} \log m)$ iterations, where each iteration requires $O(m^3)$ arithmetic operations. According to Ahuja, Magnanti and Orlin [1], the best currently-known strongly polynomial algorithm for this problem is $O(m^2 \log n + mn(\log n)^2)$ and is due to Orlin [17]. Thus, the LIP method needs considerable improvement in order to reach this bound. On the other hand, in practice, interior point methods usually perform much better than their worst-case bounds, so we suspect that Algorithm LIP could be competitive in practice, especially if a good method for solving the least-squares problems is used.

11 Conclusions

We have proposed the first known interior point method whose running time depends only on A. This is attained by including a step called the "LLS" step that accelerates the convergence. There are many open questions raised by this work. Here are some of the more interesting questions.

1. The biggest barrier preventing Algorithm LIP from being fully general-purpose is the fact that we don't know how to compute or obtain good upper bounds on χ_A or $\bar{\chi}_A$. There are several places in our algorithm where explicit knowledge of these parameters is used—the most crucial use of this knowledge is in the formula for g given by (59), which determines the spacing between the layers. Note that we do not need to know these parameters exactly—good upper bounds will suffice.

   The only known algorithm for computing these parameters is implicit in the results of Stewart [22] and O'Leary [16] and requires exponential-time. We suspect that computing them, or even getting a good upper bound, may be a hard problem.

   In the absence of an algorithm to compute them, one possibility would be an LP algorithm that makes a low guess at g, and then increases the guess as more knowledge of the parameters is gained from the running of the algorithm itself. We have not developed such an algorithm.

   A very straightforward "guessing" algorithm was suggested by J. Renegar. Simply guess $\chi_A = 100/\|A\|$ and $\bar{\chi}_A = 100$ and run the entire LIP algorithm. If it fails in any way, then guess $\chi_A = 100^2/\|A\|$ and $\bar{\chi}_A = 100^2$ and try again. Repeatedly square the guesses, until Algorithm LIP works. (Note that, since Algorithm LIP produces complementary primal and dual solutions, we have a certificate concerning whether it terminates accurately.) This multiplies the running time by a factor of $\log\log(\chi_A\|A\|)$.

   Another possibility would be to obtain bounds on these parameters for special classes of linear programs that occur in practice. For example, we carried out this process for min-cost flow problems in Section 10. Other good candidates for special-case analysis would be linear programs arising in scheduling, optimization problems involving finite-element subproblems (see example 3 of Coleman [3]), and LP relaxations of combinatorial optimization problems.

2. For floating-point number computations—where one usually is interested in computing an approximate solution rather than the global optimum—perhaps there is a different strategy for picking g. For instance, perhaps g should be picked in a manner depending on machine-epsilon. This issue needs further investigation.

3. Another very important issue for practical application is the best way to compute the LLS direction $\mathbf{r}$. This computation can of course be reduced to solving linear equations, but there are many different ways to solve linear equations in interior point methods, and some are better than others—see [30].

4. Although our algorithm has primal-dual aspects, it is principally a dual algorithm. The step of the algorithm that is most "dual" is the computation of $\bar{\alpha}$, which uses no primal information. Perhaps it is not a coincidence that this is also the least elegant part of the algorithm. We suspect that there may be a fully primal-dual algorithm that uses primal and dual information symmetrically and gives a simpler formula for $\bar{\alpha}$.

5. Although our upper bound on the total number of crossover events is $O(n^2)$, we have not been able to construct any examples that have more than cn crossovers. This raises the possibility that there really are no more than $O(n)$ crossover events, but we do not know how to prove this. Furthermore, another factor of n in the final complexity bound comes from g^{3n} appearing (60). The fact that we need $C(A)$ to be proportional to $g^{O(n)}$ rather than g appears to be an artifact of our analysis rather than an actual aspect of the complexity.

6. As mentioned earlier, we do not know how to analyze the complexity of an initialization procedure that involves introduction of a dummy variable. An alternative initialization procedure gaining popularity is the so-called "infeasible" interior point method (for example, see Lustig et al. [11]). It would be interesting if there were a layered version of these algorithms.

7. As mentioned in Section 3, there is a relationship between $\chi_A, \bar{\chi}_A$ and condition numbers of weighted-least problems as demonstrated by [27]. Thus, our new results can be thought of a link between condition number and complexity of iterative processes (also see [25] and [29]). Such links are well-known in numerical analysis; the classic example is the conjugate gradient algorithm (see [6]). Renegar [20] is currently also developing a theory connecting complexity and conditioning for interior point methods. His setting, however, is more general than ours (general convex constraints in Banach spaces are allowed) but his condition measure involves $\mathbf{b}$ and $\mathbf{c}$ as well as A.

   Although $\chi_A, \bar{\chi}_A$ act as condition numbers for weighted-least squares, it is not clear to us whether they act as condition numbers (that is, measures of the distance to ill-posedness) for LP problems like (1).

8. Can the analysis be extended to convex quadratic programming? It is not clear how the Hessian matrix H of a quadratic objective function ought to enter the complexity. For the monotone linear complementarity problem: $\mathbf{s} = M\mathbf{x} + \mathbf{q}$, $\mathbf{s} \geq \mathbf{0}$, $\mathbf{x} \geq \mathbf{0}$, and $\mathbf{x}^T\mathbf{s} = 0$, we expect that the running time of LIP depends only on M.

References

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications.* Prentice-Hall, Englewood Cliffs, New Jersey, 1993.

[2] D. Bayer and J. C. Lagarias. The non-linear geometry of linear programming. AT&T Bell Laboratories preprint, 1986.

[3] T. Coleman. Large-scale numerical optimization: introduction and overview. In A. Kent and J. G. Williams, editors, *Encyclopedia of computer science and technology.* Marcel Dekker, New York, 1993.

[4] J. Edmonds. Systems of distinct representatives and linear algebra. *J. Res. Nat. Bur. Standards*, 71B:241–245, 1967.

[5] J. L. Goffin, Z.-Q. Luo, and Y. Ye. Complexity analysis of a column generation algorithm for convex or quasiconvex feasibility problems. Faculty

of Management, McGill University, Montreal, Quebec, Canada, 1993, 1993.

[6] G. H. Golub and C. F. Van Loan. *Matrix Computations, 2nd Edition.* Johns Hopkins University Press, Baltimore, 1989.

[7] C. C. Gonzaga. Path-following methods for linear programming. *SIAM Review*, 34:167–224, 1992.

[8] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[9] L. G. Khachiyan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244:1093–1086, 1979. Translated in *Soviet Math. Dokl.* 20:191–194.

[10] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial-time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

[11] I. J. Lustig, R. E. Marsten, and D. F. Shanno. Computational experience with a primal-dual interior point method for linear programming. *Linear Algebra and Its Applications*, 152:191–222, 1991.

[12] L. McLinden. An analogue of Moreau's proximation theorem, with applications to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88:101–161, 1980.

[13] N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Method*, pages 131–158. Springer, New York, 1989.

[14] S. Mizuno, M. J. Todd, and Y. Ye. A surface of analytic centers and infeasible interior-point algorithms for linear programming. Technical Report 1037, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY. To appear in *Math. Oper. Res.*

[15] R. C. Monteiro and I. Adler. Interior path following primal-dual algorithm. Part I: linear programming. *Mathematical Programming*, 44:27–41, 1989.

[16] D. P. O'Leary. On bounds for scaled projections and pseudoinverses. *Linear Algebra and its Applications*, 132:115–117, 1990.

[17] J. B. Orlin. A faster strongly polynomial minimum cost flow algorithm. In *Proc. 20th ACM Symposium on the Theory of Computing*, pages 377–387, 1988.

[18] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity.* Prentice Hall, Englewood Cliffs, New Jersey, 1982.

[19] J. Renegar. A polynomial-time algorithm based on Newton's method for linear programming. *Mathematical Programming*, 40:59–94, 1988.

[20] J. Renegar. Linear programming, complexity theory, and elementary functional analysis. Unpublished manuscript, Department of Operations Research and Industrial Engineering, Cornell University, 1993.

[21] G. Sonnevend. An analytical center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In *Lecture Notes in Control and Information Sciences 84*, pages 866–876. Springer–Verlag, New York, 1985.

[22] G. W. Stewart. On scaled projections and pseudoinverses. *Linear Algebra and its Applications*, 112:189–193, 1989.

[23] É. Tardos. A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research*, 34:250–256, 1986.

[24] M. J. Todd. A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm. *Operations Research*, 38:1006–1018, 1990.

[25] L. Tuncel. A pseudo-polynomial complexity analysis for interior-point algorithms. Technical Report CORR 93–16, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada, 1993.

[26] S. A. Vavasis. *Nonlinear Optimization: Complexity Issues.* Oxford University Press, New York, 1991.

[27] S. A. Vavasis. Stable numerical algorithms for equilibrium systems. Technical Report 92-1280, Department of Computer Science, Cornell University, Ithaca, NY, 1992. Accepted to *SIAM J. Matrix Analysis*.

[28] S. A. Vavasis. Stable finite elements for problems with wild coefficients. Technical Report 93–1364, Department of Computer Science, Cornell University, Ithaca, NY, 1993.

[29] S. A. Vavasis and Y. Ye. Condition numbers for polyhedra with real number data. manuscript, 1993.

[30] M. H. Wright. Interior methods for constrained optimization. In A. Iserles, editor, *Acta Numerica 1992*, pages 341–407. Cambridge University Press, Cambridge, 1992.

[31] Y. Ye. On the finite convergence of interior-point algorithms for linear programming. *Mathematical Programming*, 57:325–336, 1992.