A Primal-Dual Interior Point Method Whose Running Time Depends Only on the Constraint Matrix*

Stephen A. Vavasis† Yinyu Ye‡

August 31, 1995

Abstract

We propose a primal-dual "layered-step" interior point (LIP) algorithm for linear programming with data given by real numbers. 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 an exact optimum after a finite number of steps—in particular, after $O(n^{3.5}c(A))$ iterations, where $c(A)$ is a function of the

*This paper represents a simplification of an earlier manuscript "An accelerated interior point method whose running time depends only on A" by the same authors.

†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: yyye@col.biz.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.

coefficient matrix. The LLS steps can be thought of as accelerating a classical path-following interior point method. 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 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 pair of linear programming (LP) problems in primal-dual form:

$$\begin{array}{rl} \text{Primal: maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \tag{1}$$

and

$$\begin{array}{rl} \text{Dual: minimize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & A^T\mathbf{y} \geq \mathbf{c}. \end{array} \tag{2}$$

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 $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ are unknown vectors. We propose a path-following interior point method for these problems that simultaneously produces a primal and dual solution.

Interior point methods were originally introduced by Karmarkar [12], and the first path-following method is due to Renegar [27]. The first primal-dual path following algorithm is due to Kojima et al. [15]. See Nesterov and Nemirovsky [23] for a high-level and very general description of interior point methods. Traditional path-following methods take small steps along the *central path* until they are "sufficiently" close to an optimum. Once sufficiently close, a "rounding" procedure such as Khachiyan's [14] or least-squares computation such as Ye's [43] is used to obtain an exact optimum.

In our new method, we interleave small steps with longer *layered least-squares* (LLS) steps to follow the central path. The last LLS step moves directly to an exact 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 (67) 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 12) in which A is "well-behaved" but $\mathbf{b}$ and $\mathbf{c}$ are arbitrary vectors. (Because our algorithm takes small path-following steps as well as LLS steps, our algorithm also attains the complexity bound of $O(n^{1/2}L)$ iterations, where L is described below. This is the best previously known complexity bound for interior point methods.)

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 the dual-form problem:

$$\begin{array}{rl} \text{minimize} & 2y_1 + 5y_2 \\ \text{subject to} & y_1 \geq 0, \\ & y_1 \leq 1, \\ & y_2 \geq 0, \\ & y_2 \leq 1, \\ & y_1 + 2y_2 \geq \epsilon, \end{array} \tag{3}$$

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. Figure 2 provides a close-up view of the region near $(0, 0)$ from this problem.

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 iterations in order to distinguish the optimal vertex $(\epsilon, 0)$ from

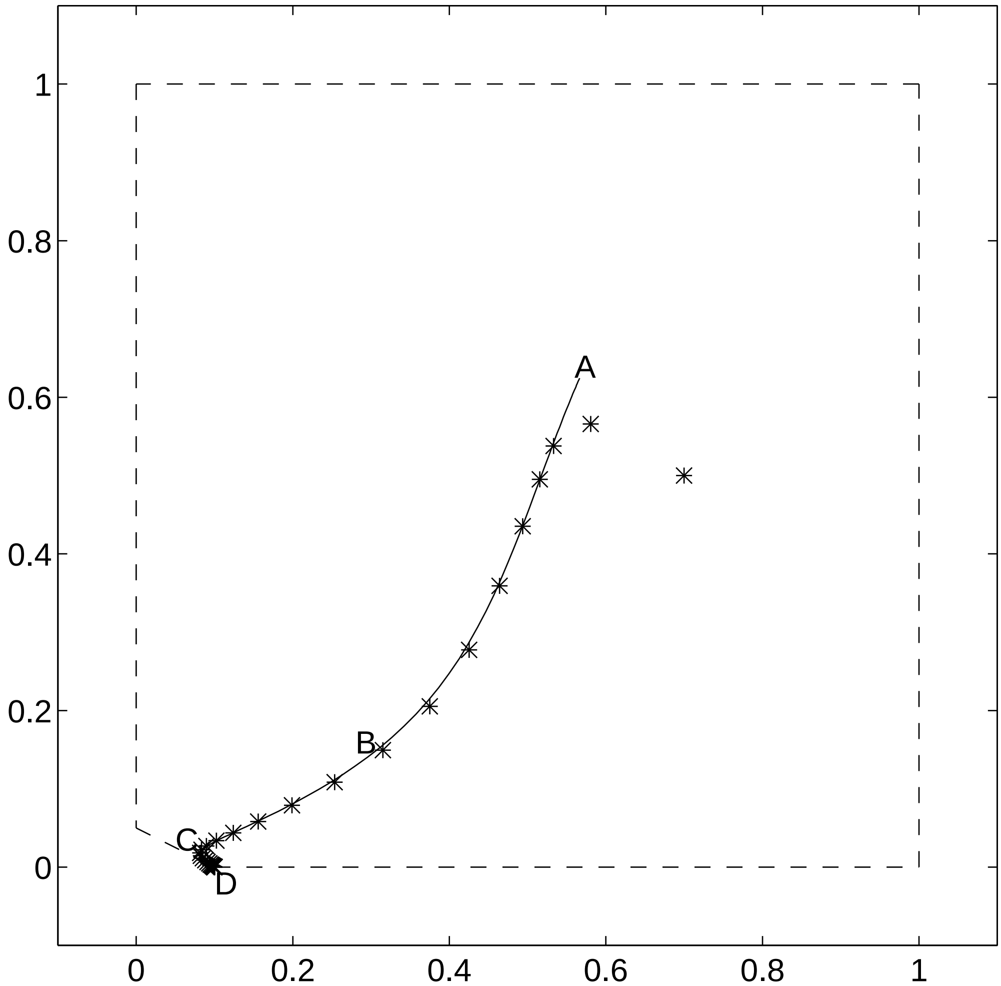

Figure 1: Example of an interior point method. A closeup of the lower left corner of the feasible region is shown in Figure 2.

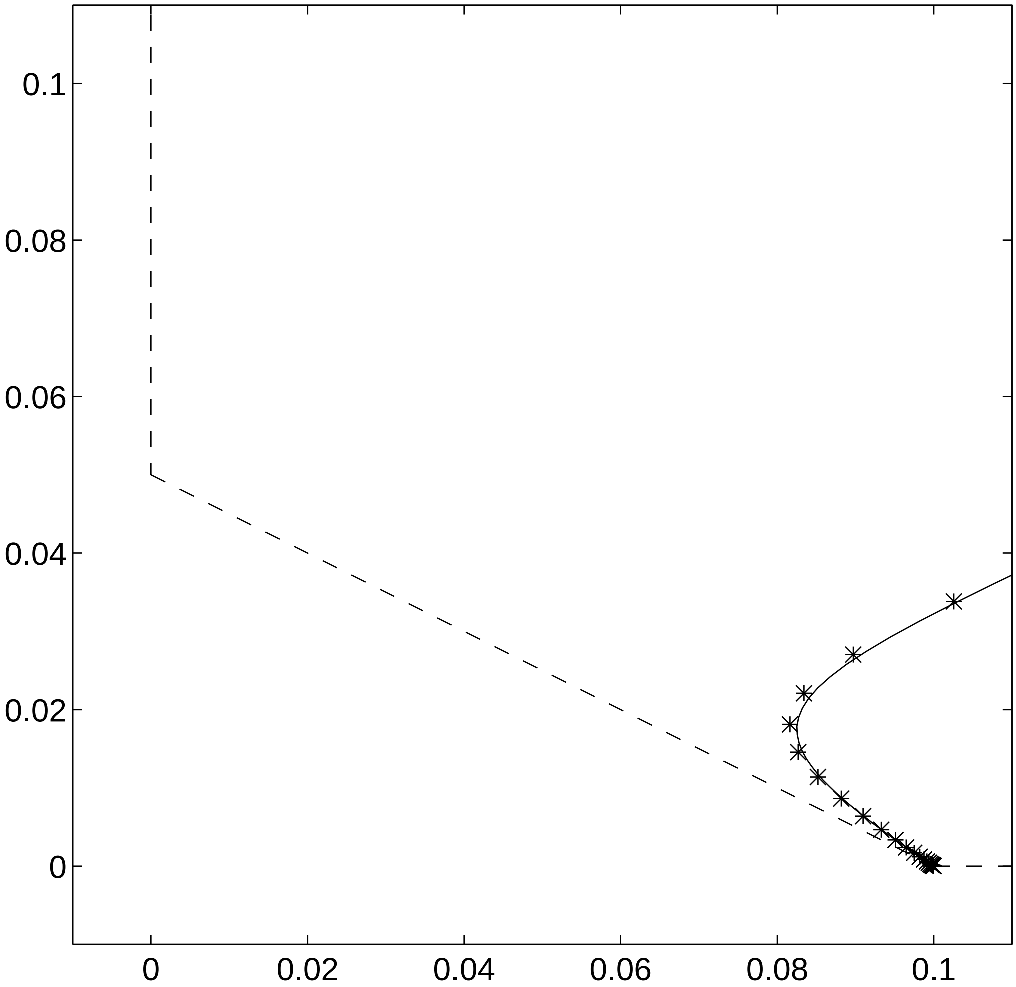

Figure 2: A close-up of the lower-left corner of Figure 1.

the nearby vertex $(0, \epsilon/2)$. In fact, it can be proved that a standard method applied to this problem would require $O(|\log \epsilon|)$ iterations. 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 is accomplished by solving a weighted least-squares problem involving the three constraints $y_1 \geq 0, y_2 \geq 0, y_1+2y_2 \geq \epsilon$, which are "near active" at optimum. In a more complex example, there would several "layers" involved in the least-squares computation. When 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-degeneracy, 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 [33] earlier proposed such a method. Tardos' method, however, "probably should be considered a purely theoretical contribution" [33, 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 major difference is that we regard our method as primarily a real-number method. Our method is the first polynomial-time linear programming algorithm that also has a complexity bound depending only on A in the real-number model of computation for finding an optimum. In contrast, Tardos uses the assumption of integer data in a fairly central way: an important tool in [33] is the operation of rounding down to the nearest integer. It is not clear how to generalize the rounding operation to noninteger data.

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 [32], Todd [34], and others, and state some of their properties. Our complexity bounds depend on $\bar{\chi}_A$. In Section 4 we describe the layered least-squares step. Section 5 explains one main-loop iteration of the LIP method, which uses the LLS computation. In Section 6 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 7 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 main-loop iteration forces at least one crossover event. In Section 8 we deduce a corollary concerning straight and curved segments in the central path. In Section 9 we assemble the pieces in order to obtain a complexity result for solving LP problems. In Section 10 we explain a technique for initializing the LIP method, and establish its complexity. In Section 11 we specialize to integer data and provide a new proof of Tardos' theorem. Finally, in Section 12 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 (2) by introducing *slack variables* as follows:

$$\begin{array}{rl} \text{minimize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & A^T\mathbf{y} - \mathbf{s} = \mathbf{c}, \\ & \mathbf{s} \geq \mathbf{0}. \end{array} \tag{4}$$

We will assume this format for the dual for the remainder of the paper. Assume that a strictly interior point pair, that is, a pair $(\mathbf{x}, \mathbf{s}) > \mathbf{0}$, exists. The *optimality conditions* for the linear programming problem may be written as follows:

$$\begin{aligned} A\mathbf{x} &= \mathbf{b}, \\ A^T\mathbf{y} - \mathbf{s} &= \mathbf{c}, \\ SX\mathbf{e} &= \mathbf{0}, \\ \mathbf{x} \geq \mathbf{0}, &\quad \mathbf{s} \geq \mathbf{0} \end{aligned} \tag{5}$$

where X denotes $\mathrm{diag}(\mathbf{x})$ and S denotes $\mathrm{diag}(\mathbf{s})$, and $\mathbf{e}$ denotes the vector of all 1's. To define a primal-dual interior point method, we introduce a parameter $\mu > 0$ in the last equation of (5) to obtain the following equations:

$$\begin{aligned} A\mathbf{x} &= \mathbf{b}, \\ A^T\mathbf{y} - \mathbf{s} &= \mathbf{c}, \\ SX\mathbf{e} &= \mu\mathbf{e}, \\ \mathbf{x} > \mathbf{0}, &\quad \mathbf{s} > \mathbf{0}. \end{aligned} \tag{6}$$

These equations always have a unique solution for any $\mu > 0$ provided that the primal and dual problem both have interior feasible points. The solution to these equations, written $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$, is called the *central path point* for μ, and the aggregate of all solutions, as μ ranges from 0 to ∞, is the *central path* of the linear programming problem. Note that as $\mu \to 0^+$, (6) approaches (5) and $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ approaches an optimal solution. The definition of the central path is credited to Bayer and Lagarias [2, 3, 4], Megiddo [18], and Sonnevend [30], and it was partially analyzed by McLinden [17] earlier. For the relationship between the central path and the classical log-barrier method, see Wright [42]. For interesting properties of the central path, see the excellent review article by Gonzaga [10].

In a generic path-following interior point method, one solves (6) approximately (an "approximately centered" point) for a large initial value of μ. Then each iteration decreases μ towards zero, and for each new value of μ, the approximate solution $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ to (6) is updated. The LIP algorithm also fits into this framework; the number of times that μ is decreased is finite, and on the final step μ is driven to zero.

An exactly central point is never computed because there is no finite algorithm to solve the nonlinear equations (6). Therefore, one defines *approximate centering.* Many definitions are possible, but we use the following proximity measure. Let $\mu > 0$ be fixed, and let $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ be an interior point feasible for the both the primal and dual. Then we define the proximity measure

$$\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) = \|SX\mathbf{e}/\mu - \mathbf{e}\|$$

(e.g., Kojima et al. [15] and Monteiro and Adler [22]). The norm used in this definition, and all norms in the remainder of the paper, are 2-norms unless otherwise noted. One sees that for an exactly centered point $(\mathbf{x}, \mathbf{y}, \mathbf{s})$, we have $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) = 0$ because $SX\mathbf{e} = \mu\mathbf{e}$. Our algorithm maintains proximity bounded by a fixed constant less than 1 (see Section 5).

We have the following useful lemma which is proved in Mizuno et al. [20] (also see references therein).

Lemma 1 *Let $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ be a given feasible primal and dual interior solution, and let $\mu > 0$ be fixed. Suppose $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) = \eta < 1$. Then the step of Newton's method for solving* (6) *is defined by*

$$\begin{aligned} A\delta\mathbf{x} &= \mathbf{0}, \\ A^T\delta\mathbf{y} &= \delta\mathbf{s}, \\ X\delta\mathbf{s} + S\delta\mathbf{x} &= \mu\mathbf{e} - SX\mathbf{e} \end{aligned} \tag{7}$$

and returns a point $\bar{\mathbf{x}} = \mathbf{x} + \delta\mathbf{x}$, $\bar{\mathbf{y}} = \mathbf{y} + \delta\mathbf{y}$ and $\bar{\mathbf{s}} = \mathbf{s} + \delta\mathbf{s}$ such that

$$\eta(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, \mu) \leq \frac{\sqrt{2}(\eta)^2}{4(1-\eta)}.$$

We also have the following corollary, which appears in Gonzaga [10].

Corollary 1 *Suppose $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) = \eta < 1$, and let $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ be the center path point for parameter μ. Then for each i,*

$$\left(1 - \frac{\eta}{1-\eta}\right) s_i \leq s_i(\mu) \leq \left(1 + \frac{\eta}{1-\eta}\right) s_i$$

and

$$\left(1 - \frac{\eta}{1-\eta}\right) x_i \leq x_i(\mu) \leq \left(1 + \frac{\eta}{1-\eta}\right) x_i.$$

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 [32] and Todd [34] 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. Other authors besides Stewart and Todd have independently proved the finiteness of $\bar{\chi}_A$ or χ_A in the previous literature: see Forsgren [8], whose bibliography lists several other proofs, the earliest being in work of Dikin.

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

$$\chi_A = \sup\left\{\frac{\|\mathbf{y}\|}{\|\mathbf{c}\|} : \mathbf{y} \text{ minimizes } \|D^{1/2}(A^T\mathbf{y} - \mathbf{c})\| \text{ for some } \mathbf{c} \in \mathbb{R}^n, D \in \mathcal{D}\right\} \tag{8}$$

and

$$\bar{\chi}_A = \sup\left\{\frac{\|A^T\mathbf{y}\|}{\|\mathbf{c}\|} : \mathbf{y} \text{ minimizes } \|D^{1/2}(A^T\mathbf{y} - \mathbf{c})\| \text{ for some } \mathbf{c} \in \mathbb{R}^n, D \in \mathcal{D}\right\}. \tag{9}$$

In the remainder of the paper, we use $\bar{\chi}_A$ only. Observe that this parameter is invariant when A is scaled by a constant. More generally, it is invariant if A is replaced by RA for any $m \times m$ nonsingular matrix R. This means that our complexity bound actually depends on $\mathcal{N}(A)$, the nullspace of A, rather than on A itself. The following is a useful property of $\bar{\chi}_A$.

Lemma 2 *Let B index an m-element subset of $\{1, \ldots, n\}$ such that A_B (the columns of A indexed by B) is nonsingular. Then $\|A_B^{-1}A\| \leq \bar{\chi}_A$.*

PROOF. Let D_ϵ be the matrix with 1's in positions corresponding to B and $\epsilon > 0$ (very small) in the remaining positions. Consider the matrix

$U_\epsilon = D_\epsilon A^T(AD_\epsilon A^T)^{-1}A$. Observe that the $\|U_\epsilon\|$ is bounded above by $\bar{\chi}_A$ since U_ϵ is the transpose of the matrix whose supremum is taken in the definition of $\bar{\chi}_A$, and transpose preserves matrix 2-norms. In the limit as $\epsilon \to 0^+$, we have that $AD_\epsilon A^T$ tends to $A_B A_B^T$. Since A_B is invertible, $(AD_\epsilon A^T)^{-1}$ tends to $(A_B A_B^T)^{-1}$. Also, $D_\epsilon A^T$ tends to a matrix with a copy of A_B^T, and the remaining rows filled in by zeros. Thus, numbering these rows first for simpler notation, we have

$$U_\epsilon \to \begin{bmatrix} A_B^T \\ 0 \end{bmatrix} A_B^{-T} A_B^{-1} A = \begin{bmatrix} A_B^{-1}A \\ 0 \end{bmatrix}.$$

∎

Further theory concerning these parameters has been developed, for example, by Vavasis [38, 39]. The result of [39] is that $\chi_A, \bar{\chi}_A$ act like condition numbers for numerically stable solution of weighted least-squares problems involving A. In [38], variants of these parameters are estimated for finite-element problems in terms of the geometry of the finite-element triangulation.

The definitions above for $\chi_A, \bar{\chi}_A$ are useful for analyzing the Newton step of an interior point method.

Lemma 3 *(a) Let $\delta\mathbf{x}^*$ be the solution to the weighted least-squares problem*

$$\min \|D^{1/2}(\delta\mathbf{x} + \mathbf{x}_0)\| \text{ subject to } A\delta\mathbf{x} = \mathbf{0} \text{ or } \delta\mathbf{x} \in \mathcal{N}(A),$$

where $D \in \mathcal{D}$ and $\mathbf{x}_0 \in \mathbb{R}^n$ are given, and $\mathcal{N}(A)$ denotes the null space of A. Then

$$\|\delta\mathbf{x}^* + \mathbf{x}_0\| \le \bar{\chi}_A \|\mathbf{x}_0\| \tag{10}$$

and

$$\|\delta\mathbf{x}^*\| \le (\bar{\chi}_A + 1)\|\mathbf{x}_0\|. \tag{11}$$

(b) Let $(\delta\mathbf{y}^*, \delta\mathbf{s}^*)$ be the solution to the weighted least-squares problem

$$\min \|D^{1/2}(\delta\mathbf{s} + \mathbf{s}_0)\| \text{ subject to } \delta\mathbf{s} = A^T\delta\mathbf{y} \text{ or } \delta\mathbf{s} \in \mathcal{R}(A^T),$$

where $D \in \mathcal{D}$ and $\mathbf{s}_0 \in \mathbb{R}^n$ are given, and $\mathcal{R}(A^T)$ denotes the range space of A^T. Then

$$\|\delta\mathbf{s}^* + \mathbf{s}_0\| \le (\bar{\chi}_A + 1)\|\mathbf{s}_0\| \tag{12}$$

and

$$\|\delta\mathbf{s}^*\| \le \bar{\chi}_A \|\mathbf{s}_0\|. \tag{13}$$

PROOF. We prove (a) first. By writing out the Lagrange multiplier conditions on the weighted least-squares problem, we obtain the following formula for $\delta\mathbf{x}^*$:

$$\delta\mathbf{x}^* + \mathbf{x}_0 = D^{-1}A^T(AD^{-1}A^T)^{-1}A\mathbf{x}_0$$

so

$$\|\delta\mathbf{x}^* + \mathbf{x}_0\| \leq \|D^{-1}A^T(AD^{-1}A^T)^{-1}A\| \cdot \|\mathbf{x}_0\|$$

and

$$\|\delta\mathbf{x}^*\| \leq (\|D^{-1}A^T(AD^{-1}A^T)^{-1}A\| + 1) \cdot \|\mathbf{x}_0\|.$$

Observe that the matrix $D^{-1}A^T(AD^{-1}A^T)^{-1}A$ is the transpose of the matrix whose supremum we take to define $\bar{\chi}_A$. Since transposition preserves matrix 2-norms, we conclude the proof of (a).

Part (b) is proved with a similar argument. Note that the least-squares solution is

$$\delta\mathbf{s}^* = -A^T(ADA^T)^{-1}AD\mathbf{s}_0.$$

∎

This paper provides some additional results concerning $\bar{\chi}_A$ such as Lemma 4, Lemma 5, and Lemma 6. One major gap in our understanding of χ_A and $\bar{\chi}_A$ is the question of computing them; we do not know of efficient algorithms for computing either parameter. See Section 13 for more comments on this matter.

A final note is that in some earlier work such as [32] and [39] 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

Our interior-point method uses standard path-following primal-dual interior point steps, plus an occasional use of a new type of step, the *layered least-squares* (LLS) step. In this section we define the primal and dual layered least-squares computations. In the next section we explain how to incorporate them into an interior point method.

The data specified for a dual LLS computation is an n-vector $\mathbf{s}$, an $m \times n$ matrix A of rank n, an $n \times n$ positive definite diagonal matrix Δ, and a partition $J_1, \ldots, J_p$ of the index set $\{1, \ldots, n\}$. 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 $J_1, \ldots, J_p$. Vectors and $\mathbf{s}$ and $\mathbf{x}$ are partitioned in the same way.

The dual LLS step computes a vector $\delta\mathbf{s}^*$ defined as follows. Let $L_0^{\mathrm{D}} = \mathcal{R}(A^T)$. Then for $k = 1, \ldots, p$ let

$$L_k^{\mathrm{D}} := \left\{ \text{minimizers } \delta\mathbf{s} \text{ of } \|[\Delta^{-1}(\delta\mathbf{s} + \mathbf{s})]_k\| \text{ subject to } \delta\mathbf{s} \in L_{k-1}^{\mathrm{D}} \right\} \tag{14}$$

so that $L_0^{\mathrm{D}} \supset L_1^{\mathrm{D}} \supset \cdots \supset L_p^{\mathrm{D}}$, where each L_k^{D} is an affine subset of $\mathbb{R}^n$. Here, the notation $[\mathbf{v}]_k$ denotes the subvector of $\mathbf{v}$ indexed by J_k. Finally, let $\delta\mathbf{s}^*$ be the unique element in L_p^{D}. (It will follow from the next paragraph that this vector is unique.)

We claim that the LLS step $\delta\mathbf{s}^*$ depends linearly on the input vector $\mathbf{s}$. To see this, let us write out the Lagrange multiplier conditions defining the dual LLS step. They are:

$$\begin{aligned} A_1\Delta_1^{-2}\delta\mathbf{s}_1^* + A_1\Delta_1^{-2}\mathbf{s}_1 &= \mathbf{0}, \\ A_2\Delta_2^{-2}\delta\mathbf{s}_2^* + A_2\Delta_2^{-2}\mathbf{s}_2 &= A_1\boldsymbol{\lambda}_{2,1}, \\ &\vdots \\ A_p\Delta_p^{-2}\delta\mathbf{s}_p^* + A_p\Delta_p^{-2}\mathbf{s}_p &= A_1\boldsymbol{\lambda}_{p,1} + \cdots + A_{p-1}\boldsymbol{\lambda}_{p,p-1}. \\ \delta\mathbf{s}^* &= A^T\delta\mathbf{y}. \end{aligned} \tag{15}$$

We see that if $(\delta\mathbf{s}, \mathbf{s})$ and $(\delta\mathbf{s}', \mathbf{s}')$ are two solutions for (15), then so is $(\alpha\delta\mathbf{s} + \beta\delta\mathbf{s}', \alpha\mathbf{s} + \beta\mathbf{s}')$ for any choice of α, β (and for some choice of Lagrange multipliers). Thus, there is a linear relation between $\delta\mathbf{s}$ and $\mathbf{s}$, i.e., the set of all possible $(\delta\mathbf{s}^*, \mathbf{s})$ satisfying (15) is the solution space to some system of equations of the form $Q\delta\mathbf{s}^* = P\mathbf{s}$.

In fact, we claim this relation is actually a mapping from $\mathbf{s}$ to $\delta\mathbf{s}$ (i.e., Q at the end of the last paragraph may be taken to be the identity matrix). If $\mathbf{s}_1 = \mathbf{0}$, then we must have $\delta\mathbf{s}_1^* = \mathbf{0}$, since $\Delta_1^{-1}\delta\mathbf{s}_1^* \in \mathcal{N}(A_1\Delta_1^{-1})$ from the first equation of (15) and also $\Delta_1^{-1}\delta\mathbf{s}_1^* \in \mathcal{R}((A_1\Delta_1^{-1})^T)$. If in addition $\mathbf{s}_2 = \mathbf{0}$, then we have $(-\boldsymbol{\lambda}_{2,1}; \Delta_2^{-1}\delta\mathbf{s}_2^*) \in \mathcal{N}((A_1, A_2\Delta_2^{-1}))$ from the second equation of (15) and also $(\delta\mathbf{s}_1^*; \Delta_2^{-1}\delta\mathbf{s}_2^*) \in \mathcal{R}((A_1, A_2\Delta_2^{-1})^T)$. Thus, these two vectors are orthogonal, which implies $\delta\mathbf{s}_2^* = \mathbf{0}$ since $\delta\mathbf{s}_1^* = \mathbf{0}$. Proceeding in this manner by induction, we conclude that $\mathbf{s} = \mathbf{0}$ maps to $\delta\mathbf{s}^* = \mathbf{0}$. Thus, there is a matrix P (which depends on A, Δ, and $J_1, \ldots, J_p$) mapping $\mathbf{s}$ to $\delta\mathbf{s}$ and defining the LLS step.

The LLS step for the primal is similar, except we work with the nullspace of A, Δ instead of Δ^{-1}, the opposite order of the partitions and we call the data vector $\mathbf{x}$ instead of $\mathbf{s}$. We define the primal layered least-squares step for coefficient matrix A, partition J, and weight matrix Δ as follows. We construct a sequence of nested subspaces $L_0^{\mathrm{P}} \subset L_1^{\mathrm{P}} \subset \cdots \subset L_p^{\mathrm{P}}$. We define $L_p^{\mathrm{P}} = \mathcal{N}(A)$. Then, for each $k = p, p-1 \ldots, 2, 1$, we define the subspace

$$L_{k-1}^{\mathrm{P}} := \{\text{minimizers } \delta\mathbf{x} \text{ of } \|[\Delta(\delta\mathbf{x} + \mathbf{x})]_k\| \text{ subject to } \delta\mathbf{x} \in L_k^{\mathrm{P}}\},$$

Let $\delta\mathbf{x}^*$ be the minimizer in L_0^{P}.

The LLS 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. We now formalize this idea in the next two lemmas by showing that the two LLS steps are in fact the limiting cases of ordinary weighted least-squares problems. In the remainder of this section, we apply these two lemmas to derive norm bounds on the LLS solutions. These bounds are used for analysis in Section 7, but we present them at this early point so that the reader may gain more intuition about properties of the LLS computation.

Lemma 4 *Consider the dual LLS problem with input data given by an arbitrary partition $J_1, \ldots, J_p$, positive definite diagonal weights Δ, $m \times n$ coefficient matrix A of rank m, and an input data vector $\mathbf{s}$. Let Ξ_J denote the $n \times n$ diagonal matrix whose entries are as follows. For $k = 1, \ldots, p$, for indices $i \in J_k$, the (i,i) entry of Ξ_J is 2^k. Let $\delta\mathbf{s}^*$ be the dual LLS solution, and let $\delta\mathbf{s}(t)$ be the solution to the ordinary least-squares problem*

$$\textit{minimize } \|\Delta^{-1}\Xi_J^{-t}(\delta\mathbf{s} + \mathbf{s})\| \textit{ subject to } \delta\mathbf{s} = A^T\delta\mathbf{y} \textit{ or } \delta\mathbf{s} \in \mathcal{R}(A^T).$$

Then

$$\lim_{t\to\infty} \delta\mathbf{s}(t) = \delta\mathbf{s}^*.$$

Proof. Fix a particular t and write $\delta\mathbf{s}$ in place of $\delta\mathbf{s}(t)$. Note that $\delta\mathbf{s}$, being the solution to a weighted least-squares problem, satisfies

$$\|\delta\mathbf{s}\| \leq \bar{\chi}_A \|\mathbf{s}\|$$

independently of t by (13). Now, consider the optimality condition defining $\delta\mathbf{s}$:

$$2^{-2t}A_1\Delta_1^{-2}(\delta\mathbf{s}_1+\mathbf{s}_1)+\cdots+2^{-2pt}A_p\Delta_p^{-2}(\delta\mathbf{s}_p+\mathbf{s}_p)=\mathbf{0}. \qquad (16)$$

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

$$\begin{aligned}\|2^{-2t}A_1\Delta_1^{-2}(\delta\mathbf{s}_1+\mathbf{s}_1)\| &\le (2^{-4t}+2^{-6t}+\cdots+2^{-2pt})(C(\bar\chi_A+1)\|\mathbf{s}\|)\\ &\le 2^{-4t}C'\end{aligned}$$

where C' denotes $2C(\bar\chi_A+1)\|\mathbf{s}\|$. Thus, we can write

$$A_1\Delta_1^{-2}(\delta\mathbf{s}_1+\mathbf{s}_1)-\mathbf{u}_1=\mathbf{0},$$

where $\|\mathbf{u}_1\|\le 2^{-2t}C'$. Similarly, from (16) we can obtain the inequality

$$\|A_2\Delta_2^{-2}(\delta\mathbf{s}_2+\mathbf{s}_2)+A_1\Delta_1^{-2}2^{2t}(\delta\mathbf{s}_1+\mathbf{s}_1)\|\le 2^{-2t}C'$$

which we can write as

$$A_2\Delta_2^{-2}(\delta\mathbf{s}_2+\mathbf{s}_2)+A_1\boldsymbol{\lambda}-\mathbf{u}_2=\mathbf{0}.$$

where $\|\mathbf{u}_2\|\le 2^{-2t}C'$. Proceeding in this manner, we see that $(\delta\mathbf{s},\mathbf{s})$, accompanied by appropriate choices for the Lagrange multipliers, is a $2^{-2t}C'$-approximate solution for (15).

Recall that we showed that there is a matrix P such that the relation $\delta\mathbf{s}^*=P\mathbf{s}$ yields all solutions for (15). Since (15) depends only on $A\Delta^{-2}\mathbf{s}$, there is some other matrix P' such that the relation $\delta\mathbf{s}^*=P'(A\Delta^{-2}\mathbf{s})$ defines all solutions. On the other hand, the argument in the previous paragraph shows that the relation $\delta\mathbf{s}(t)=P'(A\Delta^{-2}\mathbf{s}+O(2^{-2t}))$ holds for all t. This proves the lemma. ∎

We have a similar lemma for the primal LLS step. We omit the proof, which is very similar to the preceding proof.

Lemma 5 *Consider a primal LLS problem with input data given by an arbitrary partition $J_1,\ldots,J_p$, positive definite diagonal weights Δ, $m\times n$ coefficient matrix A of rank m, and an input data vector $\mathbf{x}$. Let Ξ_J denote the $n\times n$*

diagonal matrix defined in Lemma 4. Let $\delta\mathbf{x}^*$ *be the primal LLS solution, and let* $\delta\mathbf{x}(t)$ *be the solution to*

$$\textit{minimize } \|\Delta\Xi_J^t(\delta\mathbf{x}+\mathbf{x})\| \textit{ subject to } A\delta\mathbf{x}=\mathbf{0} \textit{ or } \delta\mathbf{x}\in\mathcal{N}(A).$$

Then

$$\lim_{t\to\infty}\delta\mathbf{x}(t)=\delta\mathbf{x}^*.$$

For the rest of the paper we use the notation $PLLS(\mathbf{x})$ and $DLLS(\mathbf{s})$ to denote the primal and dual LLS solutions $\delta\mathbf{x}^*$ and $\delta\mathbf{s}^*$ for data vectors $\mathbf{x}$ and $\mathbf{s}$ respectively. Note that these solutions also depend on A, Δ and the partition $J_1,\ldots,J_p$ (in addition to the data vector). Note that in the dual case, the choice of $\delta\mathbf{y}^*$ is also uniquely determined (i.e., there is a unique solution to $A^T\delta\mathbf{y}^*=\delta\mathbf{s}^*$) because A is assumed to have full rank.

As a corollary to the preceding lemmas, we can conclude from (10), (11), (12), and (13) that the LLS minimizers satisfy

$$\|\delta\mathbf{x}^*+\mathbf{x}\|\le\bar{\chi}_A\|\mathbf{x}\|, \tag{17}$$

$$\|\delta\mathbf{x}^*\|\le(\bar{\chi}_A+1)\|\mathbf{x}\|, \tag{18}$$

$$\|\delta\mathbf{s}^*+\mathbf{s}\|\le(\bar{\chi}_A+1)\|\mathbf{s}\|, \tag{19}$$

and

$$\|\delta\mathbf{s}^*\|\le\bar{\chi}_A\|\mathbf{s}\|. \tag{20}$$

In addition, we have the following properties of the LLS step based on our early discussions:

1. Vector $\delta\mathbf{x}^*$ ($\delta\mathbf{s}^*$) is uniquely determined and linear in $\mathbf{x}$ ($\mathbf{s}$). In particular, $\mathbf{0}$ maps to $\mathbf{0}$.

2. If $\mathbf{x}_p=\cdots=\mathbf{x}_k=\mathbf{0}$ ($\mathbf{s}_1=\cdots=\mathbf{s}_k=\mathbf{0}$), then $\delta\mathbf{x}_p^*=\cdots=\delta\mathbf{x}_k^*=\mathbf{0}$ ($\delta\mathbf{s}_1^*=\cdots=\delta\mathbf{s}_k^*=\mathbf{0}$). Note the reverse order between the primal and dual steps.

3. If $\mathbf{x}\in\mathcal{N}(A)$ ($\mathbf{s}\in\mathcal{R}(A^T)$), then $\delta\mathbf{x}^*=-\mathbf{x}$ ($\delta\mathbf{s}^*=-\mathbf{s}$). This is because $-\mathbf{x}\in L_p^{\mathrm{P}}$ ($-\mathbf{s}\in L_0^{\mathrm{D}}$) and leaves a residual norm of 0 at every layer of the LLS computation—the minimum possible residual.

We end this section by presenting a lemma which is an enhancement of (17) to (20).

Lemma 6 *(a) Consider the minimizer $\delta\mathbf{x}_0^*$ of the general primal LLS problem $PLLS(\mathbf{x}_0)$ for any given $\mathbf{x}_0$. Let E be the $n\times n$ diagonal matrix with 1's in positions corresponding to $J_k\cup\cdots\cup J_p$ and 0's elsewhere. Let $\|\mathbf{u}\|_E$ denote the seminorm $(\mathbf{u}^T E\mathbf{u})^{1/2}$. Then*

$$\|\delta\mathbf{x}_0^*+\mathbf{x}_0\|_E\le\bar{\chi}_A\cdot\|\mathbf{x}_0\|_E$$

and

$$\|\delta\mathbf{x}_0^*\|_E\le(\bar{\chi}_A+1)\cdot\|\mathbf{x}_0\|_E.$$

Furthermore, for an $\mathbf{x}$ such that $\mathbf{x}_0-\mathbf{x}\in\mathcal{N}(A)$ and such that $\delta\mathbf{x}^*$ is the minimizer of $PLLS(\mathbf{x})$,

$$\|\delta\mathbf{x}^*+\mathbf{x}\|_E\le\bar{\chi}_A\cdot\|\mathbf{x}_0\|_E.$$

(b) Consider the minimizer $\delta\mathbf{s}_0^*$ of the general dual LLS problem $DLLS(\mathbf{s}_0)$ for any given $\mathbf{s}_0$. 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 $(\mathbf{u}^T D\mathbf{u})^{1/2}$. Then

$$\|\delta\mathbf{s}_0^*+\mathbf{s}_0\|_D\le(\bar{\chi}_A+1)\cdot\|\mathbf{s}_0\|_D$$

and

$$\|\delta\mathbf{s}_0^*\|_D\le\bar{\chi}_A\cdot\|\mathbf{s}_0\|_D.$$

Furthermore, for an $\mathbf{s}$ such that $\mathbf{s}_0-\mathbf{s}\in\mathcal{R}(A^T)$ and such that $\delta\mathbf{s}^*$ is the minimizer of $DLLS(\mathbf{s})$,

$$\|\delta\mathbf{s}^*+\mathbf{s}\|_D\le(\bar{\chi}_A+1)\cdot\|\mathbf{s}_0\|_D.$$

PROOF. Let $\bar{\mathbf{x}}$ be defined to have 0's in positions $J_1\cup\cdots\cup J_{k-1}$ and to agree with $\mathbf{x}_0$ in positions $J_k\cup\cdots\cup J_p$. Then consider the primal LLS minimizer $\delta\bar{\mathbf{x}}^*$ for $PLLS(\bar{\mathbf{x}})$. From (17) and (18) we have

$$\|\delta\bar{\mathbf{x}}^*+\bar{\mathbf{x}}\|_E\le\|\delta\bar{\mathbf{x}}^*+\bar{\mathbf{x}}\|\le\bar{\chi}_A\cdot\|\bar{\mathbf{x}}\|=\bar{\chi}_A\cdot\|\mathbf{x}_0\|_E$$

and

$$\|\delta\bar{\mathbf{x}}^*\|_E \le \|\delta\bar{\mathbf{x}}^*\| \le (\bar{\chi}_A + 1)\cdot\|\bar{\mathbf{x}}\| = (\bar{\chi}_A + 1)\cdot\|\mathbf{x}_0\|_E.$$

Next, we claim that $\delta\mathbf{x}_0^*$ and $\delta\bar{\mathbf{x}}^*$ agree in positions $J_k \cup \cdots \cup J_p$. Consider $\delta\mathbf{x}_0^* - \delta\bar{\mathbf{x}}^*$ which is the primal LLS minimizer of $PLLS(\mathbf{x}_0 - \bar{\mathbf{x}})$. Since $(\mathbf{x}_0 - \bar{\mathbf{x}})$ has all zeros in positions $J_k \cup \cdots \cup J_p$, the minimizer $\delta\mathbf{x}_0^* - \delta\bar{\mathbf{x}}^*$ must also have all zeros in these positions as demonstrated above. Thus, since $\delta\mathbf{x}_0^*$ and $\delta\bar{\mathbf{x}}^*$ agree in positions $J_k \cup \cdots \cup J_p$, $\|\delta\mathbf{x}_0^*\|_E = \|\delta\bar{\mathbf{x}}^*\|_E$ and $\|\delta\mathbf{x}_0^* + \mathbf{x}_0\|_E = \|\delta\bar{\mathbf{x}}^* + \bar{\mathbf{x}}\|_E$. Therefore, the first two inequalities in (a) follow from the two preceding inequalities. For the third inequality in (a), since $A(\mathbf{x} - \mathbf{x}_0) = \mathbf{0}$, we must have

$$\delta\mathbf{x}^* - \delta\mathbf{x}_0^* = -(\mathbf{x} - \mathbf{x}_0)$$

again as demonstrated above. Thus, we have

$$\mathbf{x} + \delta\mathbf{x}^* = \mathbf{x}_0 + \delta\mathbf{x}_0^*$$

which implies the third inequality in (a).

Using (19) and (20) in place of (17) and (18) yields an analogous proof for (b). ∎

5 One step of the algorithm

We now describe one iteration of the main loop of our layered interior point method. Assume at the beginning of the iteration that we have a current approximate central path point $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ with $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \le \eta_0$, where we define $\eta_0 = 0.2$. Recall that $\eta(\cdot,\cdot,\cdot,\cdot)$ was defined as proximity to the central path in Section 2. (An initialization procedure that finds an approximately central point for the first main-loop iteration is described in Section 10.) Define S and X as in Section 2. Partition the index set $\{1, \ldots, n\}$ into p *layers* $J_1, \ldots, J_p$ by using $\mathbf{s}$ and $\mathbf{x}$ as follows. Let

$$\delta_i = \sqrt{\mu s_i / x_i}.$$

Let

$$\Delta = \mathrm{diag}(\boldsymbol{\delta}) = \mu^{1/2} S^{1/2} X^{-1/2}. \tag{21}$$

Note that if $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ is perfectly centered, then $SX\mathbf{e} = \mu\mathbf{e}$, i.e., $\boldsymbol{\delta} = \mathbf{s} = \mu X^{-1}\mathbf{e}$. Now find a permutation π that sorts these quantities into increasing order:

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

For the next three sections let $g > 1$ be a "gap size" parameter, which will be defined in terms of A by equation (66) below. Find the leftmost ratio-gap of size greater than g in the sorted slacks, i.e., find the smallest i such that $\delta_{\pi(i+1)}/\delta_{\pi(i)} > g$. 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 values of δ_i for constraints indexed by J_k for any k are within a factor of $g^{|J_k|} \leq g^n$ of each other, and are separated by more than a factor of g from constraints in J_{k+1}.

To formalize this, define

$$\theta_k = \max_{i \in J_k} \delta_i \tag{22}$$

and

$$\phi_k = \min_{i \in J_k} \delta_i. \tag{23}$$

The construction above ensures that for each k,

$$\theta_k < \phi_{k+1}/g. \tag{24}$$

and

$$\phi_k \leq \theta_k \leq g^n \phi_k. \tag{25}$$

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

The idea of partitioning the slacks into layers based on their relative sizes has been proposed by Kaliski and Ye [11] and Tone [35], who both propose a decomposition into two layers. The interest of these authors is in improving the running-time of computing one iteration of an interior point method, rather than in obtaining new bounds on the number of iterations. Much earlier, Wright [41] also proposed a two-layer approach in the more general setting of barrier methods for constrained nonlinear optimization.

By the assumption of approximate centrality, each diagonal entry of SX is between $\mu(1-\eta_0)$ and $\mu(1+\eta_0)$. Thus, we have the two inequalities

$$\mu^{1/2}\sqrt{1-\eta_0} \leq \|S^{1/2}X^{1/2}\| \leq \mu^{1/2}\sqrt{1+\eta_0} \tag{26}$$

and

$$\mu^{-1/2}(1+\eta_0)^{-1/2} \leq \|S^{-1/2}X^{-1/2}\| \leq \mu^{-1/2}(1-\eta_0)^{-1/2}, \tag{27}$$

which will be used frequently during the upcoming analysis.

The main-loop iteration begins with a current iterate $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ that is feasible and approximately centered with $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \leq \eta_0$. Let $\delta\mathbf{x}^*$ be the minimizer of $PLLS(\mathbf{x})$ and $(\delta\mathbf{y}^*, \delta\mathbf{s}^*)$ the minimizer of $DLLS(\mathbf{s})$ with partition $J_1, \ldots, J_p$ and weights Δ defined in the preceding paragraphs. We also let

$$\mathbf{x}^* = \mathbf{x} + \delta\mathbf{x}^*, \quad \mathbf{y}^* = \mathbf{y} + \delta\mathbf{y}^*, \quad \text{and} \quad \mathbf{s}^* = \mathbf{s} + \delta\mathbf{s}^*.$$

If Case III of our algorithm (defined below) holds, then the we will compute a point somewhere in the segment connecting $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ and $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$.

For each layer in the LLS step, we define ϵ_k^{P} to be the scaled primal residual, that is,

$$\epsilon_k^{\mathrm{P}} = \|\Delta_k(\delta\mathbf{x}_k^* + \mathbf{x}_k)\|_\infty/\mu = \|\Delta_k\mathbf{x}_k^*\|_\infty/\mu. \tag{28}$$

Notice the use of the infinity-norm in this definition, and the three upcoming definitions. We define the dual residual as well:

$$\epsilon_k^{\mathrm{D}} = \|\Delta_k^{-1}(\delta\mathbf{s}_k^* + \mathbf{s}_k)\|_\infty = \|\Delta_k^{-1}\mathbf{s}_k^*\|_\infty. \tag{29}$$

Define γ_k^{P} and γ_k^{D} for $k = 1, \ldots, p$, as follows:

$$\gamma_k^{\mathrm{P}} = \|\Delta_k\delta\mathbf{x}_k^*\|_\infty/\mu = \|\Delta_k(\mathbf{x}_k^* - \mathbf{x}_k)\|_\infty/\mu, \tag{30}$$

$$\gamma_k^{\mathrm{D}} = \|\Delta_k^{-1}\delta\mathbf{s}_k^*\|_\infty = \|\Delta_k^{-1}(\mathbf{s}_k^* - \mathbf{s}_k)\|_\infty. \tag{31}$$

There are three possibilities for one main-loop iteration of our algorithm, depending on the following cases.

Case I. There is a layer k such that

$$\gamma_k^{\mathrm{P}} \geq \frac{1}{4\sqrt{n}} \text{ and } \gamma_k^{\mathrm{D}} \geq \frac{1}{4\sqrt{n}}. \tag{32}$$

In this case, for one main-loop iteration we take $n_1(A)$ ordinary path-following interior points steps. The formula for $n_1(A)$ is given by (63) below.

Case II. There is a layer k such that

$$\gamma_k^{\mathrm{P}} < \frac{1}{4\sqrt{n}} \text{ and } \gamma_k^{\mathrm{D}} < \frac{1}{4\sqrt{n}}. \tag{33}$$

In this case we take $n_2(A)$ ordinary path-following interior point steps. The formula for $n_2(A)$ is given by (64) below.

Case III. For every layer k, either

$$\gamma_k^{\mathrm{P}} \geq \frac{1}{4\sqrt{n}} \text{ and } \gamma_k^{\mathrm{D}} < \frac{1}{4\sqrt{n}} \tag{34}$$

holds, or

$$\gamma_k^{\mathrm{P}} < \frac{1}{4\sqrt{n}} \text{ and } \gamma_k^{\mathrm{D}} \geq \frac{1}{4\sqrt{n}} \tag{35}$$

holds. In this case, we define α_k for each layer as follows. If (34) holds, we take

$$\alpha_k = \min(1, 8\epsilon_k^{\mathrm{P}}\sqrt{n}) \tag{36}$$

else if (35) holds, we take

$$\alpha_k = \min(1, 8\epsilon_k^{\mathrm{D}}\sqrt{n}). \tag{37}$$

Then we define

$$\bar{\alpha} = \max\{\alpha_1, \ldots, \alpha_p\}. \tag{38}$$

Now we take a step defined by the primal and dual LLS directions; we compute a new feasible iterate

$$\mathbf{x}^+ = \mathbf{x} + (1-\bar{\alpha})\delta\mathbf{x}^* = \bar{\alpha}\mathbf{x} + (1-\bar{\alpha})\mathbf{x}^*,$$

$$\mathbf{y}^+ = \mathbf{y} + (1-\bar{\alpha})\delta\mathbf{y}^* = \bar{\alpha}\mathbf{y} + (1-\bar{\alpha})\mathbf{y}^*,$$

and

$$\mathbf{s}^+ = \mathbf{s} + (1-\bar{\alpha})\delta\mathbf{s}^* = \bar{\alpha}\mathbf{s} + (1-\bar{\alpha})\mathbf{s}^* = A^T\mathbf{y}^+ - \mathbf{c}.$$

Note that if (34) holds for k and $8\epsilon_k^{\mathrm{P}}\sqrt{n} \geq 1$, then the minimizer in (36) is 1, and thus $\bar{\alpha} = 1$ also. Thus, in this case $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+) = (\mathbf{x}, \mathbf{y}, \mathbf{s})$.

On the other hand, if $\bar{\alpha} = 0$, this is the termination of our algorithm: $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ is an optimal solution pair for the primal and dual problems as proved in Section 6.

If the termination criterion is not satisfied, we set $\mu^+ = \mu\bar{\alpha}$. In Theorem 1 below we claim that $\eta(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+, \mu^+) < 0.65$. Then we take two Newton steps with this fixed μ^+ to restore the proximity to 0.2 (by Lemma 1). Note that if $\bar{\alpha} = 1$, then no restoration is needed.

Note that if $p = 1$, then this approach is precisely the so-called predictor and corrector algorithm developed by Mizuno et al. [20]. This is because, in the $p = 1$ case, the LLS steps $\delta\mathbf{s}^*$ and $\delta\mathbf{x}^*$ are exactly the affine-scaling directions computed by the standard method. More specifically, in this case $\delta\mathbf{s}^*$ is the minimizer of $\|\Delta^{-1}(\delta\mathbf{s} + \mathbf{s})\|$ (an ordinary least-squares problem) subject to $\delta\mathbf{s} \in \mathcal{R}(A^T)$, and similarly for $\delta\mathbf{x}^*$.

After these two Newton steps to restore approximate centering, we take $n_3(A)$ ordinary interior point steps, where $n_3(A)$ is given by (65) below. This concludes the description of a main-loop iteration for Case III.

To summarize, a main-loop iteration of Algorithm LIP reduces μ using one of the three cases in this section. Main loop iterations continue until the termination criterion stated in Case III holds.

Example. We return to the example (3), with the hypothesis that ϵ is very close to zero. This example is illustrated by Figure 1. Assume that μ is chosen so that $|\epsilon| \ll \mu \ll 1$. Two distinct layers of constraints are observed: those with "small" slacks, namely, $y_1 \geq 0$, $y_2 \geq 0$ and $y_1 + 2y_2 \geq \epsilon$, and the remaining two constraints $y_1 \leq 1$, $y_2 \leq 1$. In other words, $J_1 = \{1, 3, 5\}$ and $J_2 = \{2, 4\}$.

The central path point for the dual is approximately $(y_1(\mu), y_2(\mu)) \approx (0.791\mu, 0.283\mu)$ with $\mathbf{s}(\mu) = A^T\mathbf{y}(\mu) - \mathbf{c} \approx (0.791\mu, 1, 0.283\mu, 1, 1.358\mu)$ and $\mathbf{x}(\mu) = \mu S(\mu)^{-1}\mathbf{e} \approx (1.264, \mu, 3.527, \mu, 0.736)$.

Now, the dual LLS step can be derived as follows. We want to find $\delta\mathbf{y}^*$ so that $\|\Delta_1^{-1}(A_1^T(\mathbf{y} + \delta\mathbf{y}^*) - \mathbf{c}_1)\|$ is minimized (because $\delta\mathbf{s}_1^* = A_1^T\delta\mathbf{y}^*$ and $A_1^T\mathbf{y} - \mathbf{c}_1 = \mathbf{s}_1$ by definition). Since A_1^T has rank 2, this first layer completely determines $\delta\mathbf{y}^*$ and therefore $\delta\mathbf{s}^*$ as well. The solution to the dual LLS problem is $\delta\mathbf{y}^* = -\mathbf{y}^* + O(\epsilon)$ since $\|\mathbf{c}_1\| = O(\epsilon)$. Thus, $\delta\mathbf{s}^* = (-s_1 + O(\epsilon), O(\mu), -s_3 + O(\epsilon), O(\mu), -s_5 + O(\mu))$. As for the primal LLS step, we have $\delta\mathbf{x}_2^* = (-\mu, -\mu)$, and then $\delta\mathbf{x}_1^* = (O(\mu), O(\mu), O(\mu))$.

Now, we have $\gamma_1^{\mathrm{D}} = O(1)$ and $\gamma_2^{\mathrm{D}} = O(\mu)$. Also, $\gamma_1^{\mathrm{P}} = O(\mu)$ and $\gamma_2^{\mathrm{P}} = O(1)$. Therefore, we are in Case III.

Next, we have $\epsilon_1^{\mathrm{D}} = O(\epsilon/\mu)$ and $\epsilon_2^{\mathrm{D}} = O(1)$. Also, $\epsilon_1^{\mathrm{P}} = O(1)$ and $\epsilon_2^{\mathrm{P}} = 0$. This means that $\alpha_1 = O(\epsilon/\mu)$ and $\alpha_2 = 0$, and hence $\bar{\alpha} = O(\epsilon/\mu)$. Therefore, we can decrease the central path parameter by a factor of ϵ/μ, i.e., down to $O(\epsilon)$. Thus, as explained in Section 1, no matter how small ϵ is, we can always follow the central path with a straight-line segment until we are in a neighborhood of the near-degeneracy. Also, observe that if $\epsilon = 0$ then $\bar{\alpha} = 0$ and we step exactly to $(0, 0)$ in the dual (optimal), and we step to a primal optimal solution (the primal is degenerate in this case).

6 On the centering of the LLS step

This section is devoted to establishing the result that $(\mathbf{x}^+, \mathbf{y}^+, \mathbf{s}^+)$ is approximately centered in Case III above.

Theorem 1 *Let $\alpha \in [\bar{\alpha}, 1]$ be chosen arbitrarily, where $\bar{\alpha}$ is defined by* (38) *and Case III in the last section holds. In the case that $\bar{\alpha} = 0$, assume further that $\alpha > 0$. Then $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$, defined by $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) = \alpha(\mathbf{x}, \mathbf{y}, \mathbf{s}) + (1-\alpha)(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$, is a strictly feasible point. Furthermore,*

$$\eta(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, \mu\alpha) \leq \eta_0 + 3\sqrt{1+\eta_0}/8 + 1/32.$$

Remark: If $\eta_0 = 0.2$, then the right-hand side of the preceding inequality is at most 0.65.

PROOF. First, note that if $\bar{\alpha} = 1$ then the claim is trivial. Therefore, assume that none of the minima in (36) or (37) is attained at 1. Thus, we have for each k either $\alpha_k = 8\epsilon_k^{\mathrm{P}}\sqrt{n} < 1$ or $\alpha_k = 8\epsilon_k^{\mathrm{D}}\sqrt{n} < 1$. Recall that

$$\eta(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, \mu\alpha) = \|\bar{S}\bar{X}\mathbf{e}/(\mu\alpha) - \mathbf{e}\|.$$

Let us focus on the subvector of $\bar{S}\bar{X}\mathbf{e} - (\mu\alpha)\mathbf{e}$ indexed by layer k. For this layer, either (34) holds or (35) holds. We first treat the case that (34) holds for k. Then, we have

$$\bar{S}_k\bar{X}_k\mathbf{e} - \mu\alpha\mathbf{e} = (\alpha S_k + (1-\alpha)S_k^*)(\alpha X_k + (1-\alpha)X_k^*)\mathbf{e} - \mu\alpha\mathbf{e}$$

or

$$\bar{S}_k\bar{X}_k\mathbf{e} - \mu\alpha\mathbf{e} = \mathbf{t}_{k,1} + \mathbf{t}_{k,2} + \mathbf{t}_{k,3} + \mathbf{t}_{k,4}$$

where

$$\begin{aligned}
\mathbf{t}_{k,1} &= \mu\alpha(S_kX_k\mathbf{e}/\mu - \mathbf{e}),\\
\mathbf{t}_{k,2} &= (1-\alpha)S_kX_k^*\mathbf{e},\\
\mathbf{t}_{k,3} &= \alpha(1-\alpha)X_k(\mathbf{s}_k^* - \mathbf{s}_k),\\
\mathbf{t}_{k,4} &= (1-\alpha)^2(S_k^* - S_k)\mathbf{x}_k^*.
\end{aligned}$$

Let us first analyze $\mathbf{t}_{k,2}$:

$$\begin{aligned}\|\mathbf{t}_{k,2}\|_\infty &= (1-\alpha)\cdot\|S_k\mathbf{x}_k^*\|_\infty \\ &\le \|S_k\mathbf{x}_k^*\|_\infty \\ &\le \|S_k^{1/2}X_k^{1/2}\mu^{-1/2}\|\cdot\|\Delta_k\mathbf{x}_k^*\|_\infty \\ &\le \sqrt{1+\eta_0}\cdot\epsilon_k^{\mathrm{P}}\mu \\ &\le \mu\alpha\sqrt{1+\eta_0}/(8\sqrt{n}).\end{aligned}$$

Next we analyze $\mathbf{t}_{k,3}$:

$$\begin{aligned}\|\mathbf{t}_{k,3}\|_\infty &= \alpha(1-\alpha)\cdot\|X_k(\mathbf{s}_k^*-\mathbf{s}_k)\|_\infty, \\ &\le \alpha\|\mu^{1/2}S_k^{1/2}X_k^{1/2}\Delta_k^{-1}(\mathbf{s}_k^*-\mathbf{s}_k)\|_\infty \\ &\le \alpha\mu^{1/2}\|S_k^{1/2}X_k^{1/2}\|\cdot\|\Delta_k^{-1}(\mathbf{s}_k^*-\mathbf{s}_k)\|_\infty \\ &\le \alpha\mu\sqrt{1+\eta_0}\cdot\gamma_k^{\mathrm{D}} \\ &\le \mu\alpha\sqrt{1+\eta_0}/(4\sqrt{n}).\end{aligned}$$

Then we turn to $\mathbf{t}_{k,4}$:

$$\begin{aligned}\|\mathbf{t}_{k,4}\|_\infty &= (1-\alpha)^2\cdot\|(S_k^*-S_k)\mathbf{x}_k^*\|_\infty \\ &\le \|\Delta_k^{-1}(S_k^*-S_k)\|\cdot\|\Delta_k\mathbf{x}_k^*\|_\infty \\ &= \gamma_k^{\mathrm{D}}\cdot\mu\epsilon_k^{\mathrm{P}} \\ &\le (1/(4\sqrt{n}))\cdot\mu\epsilon_k^{\mathrm{P}} \\ &\le \mu\alpha/(32n).\end{aligned}$$

To derive these bounds we used (26), (28), (31), (34), and the fact (partly by assumption) that

$$\alpha \ge \bar{\alpha} \ge \alpha_k = 8\epsilon_k^{\mathrm{P}}\sqrt{n}.$$

On the other hand, if (35) holds for layer k, again we have

$$\bar{S}_k\bar{X}_k\mathbf{e} - \mu\alpha\mathbf{e} = \mathbf{t}_{k,1}+\mathbf{t}_{k,2}+\mathbf{t}_{k,3}+\mathbf{t}_{k,4}$$

where

$$\begin{aligned}\mathbf{t}_{k,1} &= \mu\alpha(S_kX_k\mathbf{e}/\mu-\mathbf{e}), \\ \mathbf{t}_{k,2} &= (1-\alpha)X_kS_k^*\mathbf{e}, \\ \mathbf{t}_{k,3} &= \alpha(1-\alpha)S_k(\mathbf{x}_k^*-\mathbf{x}_k), \\ \mathbf{t}_{k,4} &= (1-\alpha)^2(X_k^*-X_k)\mathbf{s}_k^*.\end{aligned}$$

We can bound the infinity norms of $\mathbf{t}_{k,2}, \mathbf{t}_{k,3}, \mathbf{t}_{k,4}$ as before.

Now let

$$\begin{aligned}
\mathbf{t}_1 &= (\mathbf{t}_{1,1}; \mathbf{t}_{2,1}; \ldots; \mathbf{t}_{p,1}), \\
\mathbf{t}_2 &= (\mathbf{t}_{1,2}; \mathbf{t}_{2,2}; \ldots; \mathbf{t}_{p,2}), \\
\mathbf{t}_3 &= (\mathbf{t}_{1,3}; \mathbf{t}_{2,3}; \ldots; \mathbf{t}_{p,3}), \\
\mathbf{t}_4 &= (\mathbf{t}_{1,4}; \mathbf{t}_{2,4}; \ldots; \mathbf{t}_{p,4}).
\end{aligned}$$

Then, we have

$$\bar{S}\bar{X}\mathbf{e} - (\mu\alpha)\mathbf{e} = \mathbf{t}_1 + \mathbf{t}_2 + \mathbf{t}_3 + \mathbf{t}_4$$

where

$$\mathbf{t}_1 = \mu\alpha(SX\mathbf{e}/\mu - \mathbf{e}), \quad \text{hence} \quad \|\mathbf{t}_1\| \leq \mu\alpha\|SX\mathbf{e}/\mu - \mathbf{e}\| \leq \mu\alpha\eta_0;$$

$$\|\mathbf{t}_2\|_\infty \leq \mu\alpha\sqrt{1+\eta_0}/(8\sqrt{n}), \quad \text{hence} \quad \|\mathbf{t}_2\| \leq \mu\alpha\sqrt{1+\eta_0}/8;$$

$$\|\mathbf{t}_3\|_\infty \leq \mu\alpha\sqrt{1+\eta_0}/(4\sqrt{n}), \quad \text{hence} \quad \|\mathbf{t}_3\| \leq \mu\alpha\sqrt{1+\eta_0}/4;$$

and

$$\|\mathbf{t}_4\|_\infty \leq \mu\alpha/(32n), \quad \text{hence} \quad \|\mathbf{t}_4\| \leq \mu\alpha/32.$$

Therefore,

$$\|\bar{S}\bar{X}\mathbf{e} - (\mu\alpha)\mathbf{e}\| \leq \|\mathbf{t}_1\| + \|\mathbf{t}_2\| + \|\mathbf{t}_3\| + \|\mathbf{t}_4\| \leq \mu\alpha(\eta_0 + 3\sqrt{1+\eta_0}/8 + 1/32)$$

which proves the approximate centering claimed by the theorem.

Now, we argue about the feasibility of $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$. Observe that the proximity measure $\|\bar{S}\bar{X}\mathbf{e}/(\mu\alpha) - \mathbf{e}\|$ is a continuous function of α as α ranges over $(\bar{\alpha}, 1]$. (Recall from our definitions that $\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}$ depend on α). We have just proved that the proximity measure is bounded by 0.65 for all α in this range, and in particular, the proximity is strictly less than 1. This means no $\bar{x}_i$ or $\bar{s}_i$ for α in this range can be equal to 0. Again, by continuity, this implies that $\bar{\mathbf{s}} > \mathbf{0}$ and $\bar{\mathbf{x}} > \mathbf{0}$ for all $\alpha \in (\bar{\alpha}, 1]$. ∎

Recall that the *analytic center* of a polytope $\{\mathbf{y} : A^T\mathbf{y} \geq \mathbf{c}\}$ (in dual form) is defined to be a point $\mathbf{y}^*$ such that $AS^{-1}\mathbf{e} = \mathbf{0}$, where as usual $\mathbf{s} = A^T\mathbf{y}^* - \mathbf{c}$ and $S = \mathrm{diag}(\mathbf{s})$. If the polytope is not full dimensional (which is always the case for the following 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. [21])

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

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

$$A_0\mathbf{x}_0 + A\mathbf{x} = \mathbf{0}, \quad \text{for some} \quad \mathbf{x}_0, \mathbf{x} > \mathbf{0} \tag{40}$$

where $\mathbf{s} = A^T\mathbf{y} - \mathbf{c}$, $S = \mathrm{diag}(\mathbf{s})$, and $\|S\mathbf{x} - \mathbf{e}\| \leq 0.65$.

Corollary 2 *If $\bar{\alpha} = 0$ in Case III of Algorithm LIP, then $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ is a primal-dual optimal pair. Furthermore, $\mathbf{y}^*$ is an approximate analytic center of the dual optimal face and $\mathbf{x}^*$ is an approximate analytic center of the primal optimal face.*

Proof. If $\bar{\alpha} = 0$, then the LLS step remains strictly interior for every choice of $\alpha > 0$ as proved by Theorem 1. Therefore, the full LLS step to $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ is feasible. Furthermore, since $\alpha_k = 0$ for each k, either $\epsilon_k^{\mathrm{P}} = 0$ or $\epsilon_k^{\mathrm{D}} = 0$ for every layer, depending on whether (34) or (35) is satisfied. Thus, for every layer, either $\mathbf{x}_k^* = \mathbf{0}$ or $\mathbf{s}_k^* = \mathbf{0}$. Therefore, $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ satisfies the complementary slackness conditions, so $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ is optimal.

In fact, more strongly, we can claim that these vectors are strictly complementary. Consider a layer k in which (34) holds. Since $\gamma_k^{\mathrm{D}} < 1/(4\sqrt{n})$, by (31), $|(s_i - s_i^*)/\delta_i| < 1/(4\sqrt{n})$ for each $i \in J_k$, that is, $\left|\sqrt{s_i x_i/\mu} - \sqrt{s_i^* x_i/\mu}\right| < 1/(4\sqrt{n})$. Notice that the first term $\sqrt{s_i x_i/\mu}$ is close to 1 by (26), and therefore the second term must be positive. This shows $\mathbf{s}_k^* > \mathbf{0}$ for layers in which (34) holds, and a similar argument applies to the complementary layers. Therefore $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ determines the so-called complementary partition of $\{1, \ldots, n\}$ into N, B according to whether $x_i^* = 0$ or $s_i^* = 0$.

Next we argue that the $(\mathbf{y}^*, \mathbf{s}^*)$ is approximately centered on the dual optimal face. A similar argument may be applied to the primal face. First, since $\mathbf{x}^*$ is feasible, we have $A\mathbf{x}^* = \mathbf{b}$, which can be written $A_B\mathbf{x}_B^* = \mathbf{b}$ (where subscript of B indicates to take the columns of A indexed by B).

The result of Theorem 1 is that for all $\alpha > 0$,

$$\|[\alpha S + (1-\alpha)S^*][\alpha X + (1-\alpha)X^*]\mathbf{e} - \mu\alpha\mathbf{e}\| \leq 0.65\mu\alpha\mathbf{e},$$

where $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ is the feasible point from which the last LLS step was computed. Let us restrict the preceding inequality to entries indexed by N. Using the fact that $X_N^* = 0$, and dividing through by α, we obtain

$$\|[\alpha S_N + (1-\alpha)S_N^*]X_N\mathbf{e} - \mu\mathbf{e}\| \leq 0.65\mu\mathbf{e}.$$

Taking the limit $\alpha \to 0$ yields

$$\|S_N^* X_N \mathbf{e}/\mu - \mathbf{e}\| \leq 0.65\mathbf{e}.$$

Let $\mathbf{x}_0 = \mathbf{x}_B/\mu - \mathbf{x}_B^*/\mu$. Then, we have

$$\begin{aligned} A_B\mathbf{x}_0 + A_N\mathbf{x}_N/\mu &= A_B(\mathbf{x}_B/\mu - \mathbf{x}_B^*/\mu) + A_N\mathbf{x}_N/\mu \\ &= (A_B\mathbf{x}_B + A_N\mathbf{x}_N)/\mu - A_B\mathbf{x}_B^*/\mu \\ &= \mathbf{b}/\mu - \mathbf{b}/\mu = \mathbf{0}. \end{aligned}$$

This proves the approximate centering of $(\mathbf{y}^*, \mathbf{s}^*)$ in the dual optimal face. A similar argument applies to the primal. ∎

7 Crossover events

The theorem and corollary in the last section prove 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 $O(n^2)$. The key concept for developing a complexity theory is a crossover event.

We start with a simple lemma:

Lemma 7 *Consider the current approximately-centered iterate $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ in Algorithm LIP. Suppose $i, j \in \{1, \ldots, n\}$ are two indices such that $i \in J_1 \cup \cdots \cup J_k$ and $j \in J_k \cup \cdots \cup J_p$ for some $k \in \{1, \ldots, p\}$. Then*

$$s_i(\mu) \leq 3g^n s_j(\mu). \tag{41}$$

PROOF. The hypothesis means that either i is in a lower layer than j or in the same layer. In the first case we conclude that $\delta_i < \delta_j$. In the second case, $\delta_i \leq g^n \delta_j$. Thus, the latter inequality applies to both cases, i.e.,

$$\sqrt{s_i/x_i} \leq g^n \sqrt{s_j/x_j}$$

which implies by (26) and (27) that

$$s_i \leq \frac{\sqrt{1+\eta_0}}{\sqrt{1-\eta_0}} \cdot g^n s_j.$$

Now we apply Corollary 1 with the choice $\eta_0 = 0.2$ to arrive at (41). ∎

We now define a crossover event:

Definition. Given an LP problem in primal-dual form, and given a current approximately centered iterate $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$, we say that the 4-tuple (μ', μ, i, j) defines a **crossover event** for $\mu' > 0$, and for $i, j \in \{1, \ldots, n\}$ if, for some k,

$$i \in J_1 \cup \cdots \cup J_k \quad \text{and} \quad j \in J_k \cup \cdots \cup J_p \tag{42}$$

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

$$s_i(\mu'') \geq 5 g^n s_j(\mu''). \tag{43}$$

Note that one must have $i \neq j$ for (43) to hold, since $g > 1$. Notice also that $\mu' < \mu$ since (43) cannot hold for $\mu'' = \mu$ (because this would contradict Lemma 7).

In simple terms, (μ', μ, i, j) specifies a crossover event provided that $s_j(\mu)$ is on the same order as, or is much larger than $s_i(\mu)$, but $s_j(\mu') \ll s_i(\mu')$, and furthermore, the relation $s_j(\mu'') \ll s_i(\mu'')$ remains in force for all $\mu'' \in (0, \mu']$.

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 8 *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 primal-dual LP. 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 $\mu_1' < \mu_1 \leq \mu_2' < \mu_2$. Suppose $i_1 = i_2$ and $j_1 = j_2$ both hold. Then (43) for the second event takes the form $s_{i_2}(\mu'') \geq 5g^n s_{j_2}(\mu'')$. Substituting equal indices and taking the special case of $\mu'' = \mu_1$ yields $s_{i_1}(\mu_1) \geq 5g^n s_{j_1}(\mu_1)$. But this contradicts (42) for the first crossover event: Lemma 7 combined with (42) implies that $s_{i_1}(\mu_1) \leq 3g^n s_{j_1}(\mu_1)$.

Similarly, we could not have $i_1 = j_2$ and $j_1 = i_2$ either, because then (43) would be contradictory for any μ'' chosen less than $\min(\mu_1', \mu_2')$ (i.e., (43) 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 $n(n-1)/2$ main-loop iterations of Algorithm LIP before termination.

Before proving this theorem, we require some preliminary lemmas. The purpose of these lemmas may be described as follows. At a given main-loop iteration with parameter μ, we have information $\epsilon_1^{\mathrm{P}}, \ldots, \epsilon_p^{\mathrm{P}}$, $\gamma_1^{\mathrm{P}}, \ldots, \gamma_p^{\mathrm{P}}$ and the corresponding dual information for the central path at parameter value μ. We would like to use these residuals to deduce bounds for the remaining part of the central path, that is, the part of the central path parameterized by $\mu' \in (0, \mu]$. Information about the rest of the central path is necessary in order to establish inequality (43). Lemma 9 establishes a basic monotonicity property that is crucial for making inferences about the remainder of the central path. Then Lemma 10 and Lemma 12 actually give us bounds about the remainder of the central path in terms of the current residual parameters. Lemma 11 is a preliminary result to Lemma 12.

Lemma 9 *Let $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ and $(\mathbf{x}', \mathbf{y}', \mathbf{s}', \mu')$ be two points on the central path such that $0 < \mu' \leq \mu$. Then for any i,*

$$s_i' \leq n s_i \quad \text{and} \quad x_i' \leq n x_i.$$

Proof. Assume $n > 1$ since the $n = 1$ case may be trivially analyzed. Since the points are exactly central, $x_i s_i = \mu$ and $x'_i s'_i = \mu'$ for all i. Note that

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

since $\mathbf{x} - \mathbf{x}'$ is in $\mathcal{N}(A)$, and $\mathbf{s} - \mathbf{s}'$ is in $\mathcal{R}(A^T)$. 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

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

i.e.,

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

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

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

Thus, for any i,

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

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,

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

which proves the lemma. ∎

This lemma leads to another lemma:

Lemma 10 *Given a point* $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ *in Algorithm LIP, let* ϵ_k^{P} *and* ϵ_k^{D} *be defined by* (28) *and* (29)*. Then there is an index* $i \in J_1 \cup \cdots \cup J_k$ *such that for all* $\mu' \in (0, \mu]$,

$$s_i(\mu') \geq \frac{\epsilon_k^{\mathrm{D}} \phi_k}{(\bar{\chi}_A + 1) n^{1.5}}. \tag{44}$$

There is also an index $j \in J_k \cup \cdots \cup J_p$ *such that for all* $\mu' \in (0, \mu]$,

$$x_j(\mu') \geq \frac{\mu \epsilon_k^{\mathrm{P}}}{\bar{\chi}_A \theta_k n^{1.5}}. \tag{45}$$

Proof. We start with (44). Let D and the seminorm $(\mathbf{u}^T D \mathbf{u})^{1/2}$ be as in (b) of Lemma 6. Recall that D denotes the $n \times n$ diagonal matrix with 1's in positions corresponding to $J_1 \cup \cdots \cup J_k$ and zeros elsewhere.

Let $(\mathbf{x}', \mathbf{y}', \mathbf{s}')$ be any feasible vector for the original LP problem and let $(\delta \mathbf{x}^*, \delta \mathbf{y}^*, \delta \mathbf{s}^*)$ be the minimizers of $PLLS(\mathbf{x})$ and $DLLS(\mathbf{s})$, respectively. Then, since $\mathbf{s}' - \mathbf{s} \in \mathcal{R}(A^T)$, from the third inequality in (b) of Lemma 6

$$\begin{aligned} \|\mathbf{s}'\|_D &\geq \|\delta \mathbf{s}^* + \mathbf{s}\|_D / (\bar{\chi}_A + 1) \\ &= \|\mathbf{s}^*\|_D / (\bar{\chi}_A + 1). \end{aligned}$$

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

$$\|\mathbf{s}(\mu')\|_D \geq \|\mathbf{s}^*\|_D / (\bar{\chi}_A + 1).$$

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

$$\|\Delta_k^{-1}\| = \max\{1/\delta_i : i \in J_k\} = 1/\phi_k.$$

Now we have:

$$\begin{aligned} \|\mathbf{s}^*\|_D &\geq \|\mathbf{s}_k^*\| \\ &\geq \|\Delta_k^{-1} \mathbf{s}_k^*\|_\infty / \|\Delta_k^{-1}\| \\ &= \epsilon_k^{\mathrm{D}} \phi_k. \end{aligned}$$

Combining, we have shown that

$$\|\mathbf{s}(\mu')\|_D \geq \epsilon_k^{\mathrm{D}} \phi_k / (\bar{\chi}_A + 1).$$

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

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

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

$$s_i(\mu'') < \frac{\epsilon_k^{\mathrm{D}} \phi_k}{(\bar{\chi}_A + 1) \cdot n^{1.5}} \tag{47}$$

holds, then for all $\mu' \in (0, \mu'']$, (46) could never hold because of Lemma 9. Therefore, there has to be at least one fixed i such that (47) never holds for any $\mu'' \in (0, \mu]$, i.e., (44) always holds for all $\mu' \in (0, \mu]$ for this i.

We have a similar proof for the primal case. Let E and $\| \cdot \|_E$ be as in (a) of Lemma 6. Recall E denote the $n \times n$ diagonal matrix with 1's in positions corresponding to $J_k \cup \cdots \cup J_p$ and zeros elsewhere.

We have

$$\|\mathbf{x}(\mu')\|_E \geq \|\mathbf{x}^*\|_E / \bar{\chi}_A$$

by the lemma, and

$$\begin{aligned} \|\mathbf{x}^*\|_E &\geq \|\mathbf{x}_k^*\| \\ &\geq \|\Delta_k \mathbf{x}_k^*\|_\infty / \|\Delta_k\| \\ &\geq \mu \epsilon_k^{\mathrm{P}} / \theta_k. \end{aligned}$$

Thus,

$$\|\mathbf{x}(\mu')\|_E \geq \frac{\mu \epsilon_k^{\mathrm{P}}}{\bar{\chi}_A \theta_k}.$$

Now we argue as in the first part of this proof, again using Lemma 9, to establish (45). ∎

It should be noted that the preceding lemma has a curious nonconstructive aspect; the existence of a special i 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 it were possible to identify i in the lemma from the current iterate, then we could presumably delete the ith variable from the problem because we know it is not active at the solution.

We now introduce some additional notation. Let

$$\mathbf{x}^{[k]} = (\mathbf{0}; \ldots; \mathbf{0}; \mathbf{x}_k; \mathbf{0}; \ldots; \mathbf{0}).$$

Define $\delta\mathbf{x}^{*[k]}$ to be the primal LLS minimizer of $PLLS(\mathbf{x}^{[k]})$. Let

$$\mathbf{s}^{[k]} = (\mathbf{0}; \ldots; \mathbf{0}; \mathbf{s}_k; \mathbf{0}; \ldots; \mathbf{0}).$$

Define $(\delta\mathbf{y}^{*[k]}, \delta\mathbf{s}^{*[k]})$ to be the the dual LLS minimizer of $DLLS(\mathbf{s}^{[k]})$. Note that we use these vectors in our analyses only, and we do not need them in our algorithm. Recall that $\delta\mathbf{x}^*$ and $\delta\mathbf{s}^*$ are LLS minimizers for $PLLS(\mathbf{x})$ and $DLLS(\mathbf{s})$, respectively.

Lemma 11 *For each k,*

$$\|\Delta_k(\delta\mathbf{x}_k^* - \delta\mathbf{x}_k^{*[k]})\| \leq \mu(\bar{\chi}_A + 1)\sqrt{1+\eta_0} \cdot \sqrt{n}/g \tag{48}$$

and

$$\|\Delta_k^{-1}(\delta\mathbf{s}_k^* - \delta\mathbf{s}_k^{*[k]})\| \leq \bar{\chi}_A\sqrt{1+\eta_0} \cdot \sqrt{n}/g. \tag{49}$$

Proof. We start with (49). Note that $\delta\mathbf{s}^* - \delta\mathbf{s}^{*[k]}$ is the dual minimizer of $DLLS(\mathbf{s} - \mathbf{s}^{[k]})$. From the second inequality in (b) of Lemma 6 we have

$$\begin{aligned} \|\delta\mathbf{s}^* - \delta\mathbf{s}^{*[k]}\|_D &\leq \bar{\chi}_A\|\mathbf{s} - \mathbf{s}^{[k]}\|_D \\ &= \bar{\chi}_A\|(\mathbf{s}_1; \ldots; \mathbf{s}_{k-1})\|. \end{aligned} \tag{50}$$

Observe that

$$\begin{aligned} \|(\mathbf{s}_1; \ldots; \mathbf{s}_{k-1})\| &= \|(S^{1/2}X^{1/2}/\mu^{1/2})(\boldsymbol{\delta}_1; \ldots; \boldsymbol{\delta}_{k-1}; \mathbf{0}; \ldots; \mathbf{0})\| \\ &\leq \|S^{1/2}X^{1/2}/\mu^{1/2}\| \cdot \|(\boldsymbol{\delta}_1; \ldots; \boldsymbol{\delta}_{k-1})\| \\ &\leq \sqrt{1+\eta_0} \cdot \sqrt{n}\theta_{k-1}. \end{aligned} \tag{51}$$

To derive the last line we used (22) and (26). Also, we have

$$\begin{aligned} \|\Delta_k^{-1}\| &= \max\{1/\delta_i : i \in J_k\} \\ &= 1/\phi_k. \end{aligned} \tag{52}$$

Combining (50), (51) and (52) yields

$$\begin{aligned}\|\Delta_k^{-1}(\delta\mathbf{s}_k^* - \delta\mathbf{s}_k^{*[k]})\| &\le \|\Delta_k^{-1}\| \cdot \|\delta\mathbf{s}_k^* - \delta\mathbf{s}_k^{*[k]}\| \\ &\le \|\Delta_k^{-1}\| \cdot \|\delta\mathbf{s}^* - \delta\mathbf{s}^{*[k]}\|_D \\ &\le \phi_k^{-1} \cdot \bar{\chi}_A \cdot \sqrt{1+\eta_0} \cdot \sqrt{n} \cdot \theta_{k-1}.\end{aligned}$$

This proves (49), because $\theta_{k-1}/\phi_k \le 1/g$ by (24).

We prove (48) in a similar manner where we use the second inequality in (a) of Lemma 6. ∎

The next lemma again states a pair of primal/dual inequalities.

Lemma 12 *Let gap-size g satisfy* (57) *and* (62) *below.*
(a) Suppose for some k that $\gamma_k^{\mathrm{D}} \ge 1/(4\sqrt{n})$. Then there exists a $j \in J_k \cup \cdots \cup J_p$ such that for all $\mu' \in (0,\mu]$,

$$x_j(\mu') \ge \frac{\mu}{2 \cdot 8^2 \cdot \bar{\chi}_A \theta_k n^3 \sqrt{1+\eta_0}}. \tag{53}$$

(b) Suppose for some k that $\gamma_k^{\mathrm{P}} \ge 1/(4\sqrt{n})$. Then there exists an $i \in J_1 \cup \cdots \cup J_k$ such that for all $\mu' \in (0,\mu]$,

$$s_i(\mu') \ge \frac{\phi_k}{2 \cdot 8^2 \cdot (\bar{\chi}_A + 1) n^3 \sqrt{1+\eta_0}}. \tag{54}$$

Proof. Recall that $\mathbf{s}^{[k]} = (\mathbf{0};\ldots;\mathbf{0},\mathbf{s}_k;\mathbf{0};\ldots;\mathbf{0})$ and $(\delta\mathbf{y}^{*[k]}, \delta\mathbf{s}^{*[k]})$ is the the dual minimizer of $DLLS(\mathbf{s}^{[k]})$. This means that the following Lagrange multiplier condition is satisfied by $\delta\mathbf{s}^{*[k]}$ for the LLS step (15):

$$\delta\mathbf{s}_1^{*[k]} = \cdots = \delta\mathbf{s}_{k-1}^{*[k]} = \mathbf{0}$$

and

$$A_k\Delta_k^{-2}\delta\mathbf{s}_k^{*[k]} + A_k\Delta_k^{-2}\mathbf{s}_k = A_1\boldsymbol{\lambda}_1 + \cdots + A_{k-1}\boldsymbol{\lambda}_{k-1}.$$

In the above equation, we have $\Delta_k^{-2}\mathbf{s}_k = \mathbf{x}_k/\mu$ by (21), so

$$A_k\Delta_k^{-2}\delta\mathbf{s}_k^{*[k]} + A_k\mathbf{x}_k/\mu = A_1\boldsymbol{\lambda}_1 + \cdots + A_{k-1}\boldsymbol{\lambda}_{k-1}. \tag{55}$$

Let

$$\mathbf{u} = (-\boldsymbol{\lambda}_1; \ldots; -\boldsymbol{\lambda}_{k-1}; \Delta_k^{-2}\delta\mathbf{s}_k^{*[k]} + \mathbf{x}_k/\mu; \mathbf{0}; \ldots; \mathbf{0}).$$

Then, $\mathbf{u} \in \mathcal{N}(A)$ by (55). Since also $\delta\mathbf{s}^{*[k]} \in \mathcal{R}(A^T)$, we have

$$0 = \mathbf{u}^T\delta\mathbf{s}^{*[k]} = \|\Delta_k^{-1}\delta\mathbf{s}_k^{*[k]}\|^2 + \mathbf{x}_k^T\delta\mathbf{s}_k^{*[k]}/\mu$$

or

$$\mathbf{x}_k^T\delta\mathbf{s}_k^{*[k]} = -\mu\|\Delta_k^{-1}\delta\mathbf{s}_k^{*[k]}\|^2. \tag{56}$$

Here, we have used the fact that $\delta\mathbf{s}^{*[k]}$ has zeros in positions $J_1, \ldots, J_{k-1}$ and $\mathbf{u}$ has zeros in positions $J_{k+1}, \ldots, J_p$.

Note that by (31) we have

$$\begin{aligned}\gamma_k^{\mathrm{D}} &= \|\Delta_k^{-1}\delta\mathbf{s}_k^*\|_\infty \\ &\le \|\Delta_k^{-1}\delta\mathbf{s}_k^*\| \\ &\le \|\Delta_k^{-1}(\delta\mathbf{s}_k^* - \delta\mathbf{s}_k^{*[k]})\| + \|\Delta_k^{-1}\delta\mathbf{s}_k^{*[k]}\| \\ &\le \bar{\chi}_A\sqrt{1+\eta_0}\sqrt{n}/g + \|\Delta_k^{-1}\delta\mathbf{s}_k^{*[k]}\|\end{aligned}$$

using (49) in the last line. If we choose

$$g \ge 8\bar{\chi}_A\sqrt{1+\eta_0}\cdot n, \tag{57}$$

then we must have

$$\|\Delta_k^{-1}\delta\mathbf{s}_k^{*[k]}\| \ge \gamma_k^{\mathrm{D}} - 1/(8\sqrt{n}) \ge 1/(4\sqrt{n}) - 1/(8\sqrt{n}) = 1/(8\sqrt{n}).$$

This, together with (56), implies that

$$\mathbf{x}_k^T\delta\mathbf{s}_k^{*[k]} \le -\mu/(8^2 n). \tag{58}$$

Consider the product $\mathbf{x}^T\delta\mathbf{s}^{*[k]}$.

$$\begin{aligned}\mathbf{x}^T\delta\mathbf{s}^{*[k]} &= \sum_{l=1}^{p}\mathbf{x}_l^T\delta\mathbf{s}_l^{*[k]} \\ &= \sum_{l=k}^{p}\mathbf{x}_l^T\delta\mathbf{s}_l^{*[k]} \quad (\delta\mathbf{s}_1^{*[k]} = \cdots = \delta\mathbf{s}_{k-1}^{*[k]} = \mathbf{0}) \\ &= \mathbf{x}_k^T\delta\mathbf{s}_k^{*[k]} + \mathbf{x}_N^T\delta\mathbf{s}_N^{*[k]} \end{aligned} \tag{59}$$

where N denotes $J_{k+1} \cup \cdots \cup J_p$. Observe that $\delta \mathbf{s}^{*[k]}$, as the minimizer to $DLLS(\mathbf{s}^{[k]})$, has the following norm bound (20):

$$\begin{aligned}
\|\delta \mathbf{s}^{*[k]}\| &\leq \bar{\chi}_A \|\mathbf{s}^{[k]}\| \\
&= \bar{\chi}_A \|\mathbf{s}_k\| \\
&= \bar{\chi}_A \|(S_k X_k/\mu)^{1/2} \Delta_k \mathbf{e}\| \\
&\leq \bar{\chi}_A \|(S_k X_k/\mu)^{1/2}\| \cdot \|\Delta_k\| \cdot \|\mathbf{e}\| \\
&\leq \bar{\chi}_A \sqrt{1+\eta_0} \cdot \theta_k \cdot \sqrt{n}.
\end{aligned} \tag{60}$$

Then, analyzing the second term in (59) we have

$$\begin{aligned}
|\mathbf{x}_N^T \delta \mathbf{s}_N^{*[k]}| &\leq \|\delta \mathbf{s}_N^{*[k]}\| \cdot \|\mathbf{x}_N\| \\
&= \|\delta \mathbf{s}^{*[k]}\| \cdot \|(\mu^{1/2} X_N^{1/2} S_N^{1/2}) \Delta_N^{-1} \mathbf{e}\| \\
&\leq \bar{\chi}_A \sqrt{1+\eta_0} \sqrt{n} \theta_k \cdot \mu \sqrt{1+\eta_0} \cdot (1/\phi_{k+1}) \cdot \sqrt{n} \\
&\leq \mu \bar{\chi}_A (1+\eta_0) n/g.
\end{aligned}$$

If we choose g sufficiently large so that $\mu \bar{\chi}_A (1+\eta_0) n/g \leq 0.5 |\mathbf{x}_k^T \delta \mathbf{s}_k^{*[k]}|$, i.e., $\mu \bar{\chi}_A (1+\eta_0) n/g \leq 0.5 \mu/(8^2 n)$ then it can be guaranteed that

$$|\mathbf{x}_N^T \delta \mathbf{s}_N^{*[k]}| \leq \mu/(2 \cdot 8^2 n). \tag{61}$$

Thus, we must choose g to satisfy

$$g \geq 2 \cdot 8^2 (1+\eta_0) n^2 (\bar{\chi}_A + 1). \tag{62}$$

Now, let us consider a $\mu' > 0$ and $(\mathbf{x}', \mathbf{y}', \mathbf{s}')$ on the central path for μ'. Then, since $\mathbf{x} - \mathbf{x}' \in \mathcal{N}(A)$, we must have

$$(\mathbf{x} - \mathbf{x}')^T \delta \mathbf{s}^{*[k]} = 0$$

or

$$\mathbf{x}^T \delta \mathbf{s}^{*[k]} = (\mathbf{x}')^T \delta \mathbf{s}^{*[k]}.$$

Let M denote $J_k \cup \cdots \cup J_p$. The preceding equation means $\mathbf{x}_M^T \delta \mathbf{s}_M^{*[k]} = (\mathbf{x}'_M)^T \delta \mathbf{s}_M^{*[k]}$. Then from (58) and (59) we have:

$$\mu/(8^2 n) \leq -\mathbf{x}_k^T \delta \mathbf{s}_k^{*[k]}$$

$$
\begin{aligned}
&= \mathbf{x}_N^T \delta \mathbf{s}_N^{*[k]} - \mathbf{x}_M^T \delta \mathbf{s}_M^{*[k]} \\
&= \mathbf{x}_N^T \delta \mathbf{s}_N^{*[k]} - (\mathbf{x}'_M)^T \delta \mathbf{s}_M^{*[k]} \\
&\le |\mathbf{x}_N^T \delta \mathbf{s}_N^{*[k]}| - (\mathbf{x}'_M)^T \delta \mathbf{s}_M^{*[k]} \\
&\le \mu/(2 \cdot 8^2 n) - (\mathbf{x}'_M)^T \delta \mathbf{s}_M^{*[k]} \quad \text{by (61)} \\
&\le \mu/(2 \cdot 8^2 n) + \|\delta \mathbf{s}_M^{*[k]}\| \cdot \|\mathbf{x}'_M\| \\
&\le \mu/(2 \cdot 8^2 n) + \bar{\chi}_A \sqrt{1+\eta_0} \cdot \theta_k \sqrt{n} \cdot \|\mathbf{x}'_M\| \quad \text{by (60).}
\end{aligned}
$$

Thus,

$$
\|\mathbf{x}'_M\| \ge \frac{\mu}{2 \cdot 8^2 \cdot n^{1.5} \cdot \theta_k \cdot \bar{\chi}_A \sqrt{1+\eta_0}}.
$$

This in turn means that there is at least one $j \in M$ such that

$$
x'_j \ge \frac{\mu}{2 \cdot 8^2 \cdot n^2 \cdot \theta_k \cdot \bar{\chi}_A \sqrt{1+\eta_0}}.
$$

This choice of j seemingly depends on μ'. However, arguing as we did concerning (47), we can assert that we can pick a fixed j independently of μ' to satisfy (53). (Note that (53) has one more factor of n compared to the preceding inequality.)

The proof of (b) is similar, except we use (18):

$$
\begin{aligned}
\|\delta \mathbf{x}^{*[k]}\| &\le (\bar{\chi}_A + 1)\|\mathbf{x}^{[k]}\| \\
&= (\bar{\chi}_A + 1)\|\mathbf{x}_k\| \\
&= (\bar{\chi}_A + 1)\|S_k^{1/2} X_k^{1/2} \mu^{1/2} \Delta_k^{-1} \mathbf{e}\| \\
&\le (\bar{\chi}_A + 1)\mu\sqrt{1+\eta_0} \cdot (1/\phi_k) \cdot \sqrt{n}
\end{aligned}
$$

to replace (60). ∎

Lemma 13 *Suppose for some k that $\gamma_k^{\mathrm{D}} \le 1/(4\sqrt{n})$. Then $\epsilon_k^{\mathrm{D}} \ge 0.5$. Similarly, if $\gamma_k^{\mathrm{P}} \le 1/(4\sqrt{n})$, then $\epsilon_k^{\mathrm{P}} \ge 0.5$.*

PROOF. We have

$$
\begin{aligned}
\epsilon_k^{\mathrm{D}} &= \|\Delta_k^{-1} \mathbf{s}_k^*\|_\infty \\
&\ge \|\Delta_k^{-1} \mathbf{s}_k\|_\infty - \|\Delta_k^{-1}(\mathbf{s}_k^* - \mathbf{s}_k)\|_\infty \\
&= \|\Delta_k^{-1} \mathbf{s}_k\|_\infty - \gamma_k^{\mathrm{D}}.
\end{aligned}
$$

Note that γ_k^{D} by assumption is bounded by $1/(4\sqrt{n})$ and

$$\|\Delta_k^{-1}\mathbf{s}_k\|_\infty = \|S_k^{1/2}X_k^{1/2}/\sqrt{\mu}\| \geq \sqrt{1-\eta_0}.$$

This proves the dual case. The primal case is similar. ∎

In the rest of this section, we use the result that a standard primal-dual path-following interior-point algorithm can reduce μ to μ' ($< \mu$) in $c_0\sqrt{n}\log(\mu/\mu')$ iterations, while maintaining proximity η_0 to the central path. The constant c_0 depends on η_0 and on the particular algorithm but has 10 as an upper bound [15] for $\eta_0 = 0.2$.

Theorem 2 *Let gap-size g satisfy* (57) *and* (62) *and consider one main loop iteration of Algorithm LIP. (If Case III holds for the iteration, assume further that $\bar{\alpha} > 0$, i.e., the algorithm does not terminate.) This iteration causes a crossover event to take place. In other words, if μ denotes the central path parameter at the beginning of the iteration and μ' at the end, then there exist two indices i, j such that (μ', μ, i, j) is a crossover event.*

PROOF. The proof breaks down into three cases, based on which case for the iteration holds.

Case I holds for the iteration. In this case, γ_k^{D} and γ_k^{P} are both greater than or equal to $1/(4\sqrt{n})$ for some layer k, and we use (54) to conclude that there exists an $i \in J_1 \cup \cdots \cup J_k$ such that

$$s_i(\mu') \geq \frac{\phi_k}{2\cdot 8^2\cdot(\bar{\chi}_A+1)n^3\sqrt{1+\eta_0}}$$

Similarly, there is an $j \in J_k \cup \cdots \cup J_p$ such that (53) holds. Because $s_j(\mu')x_j(\mu') = \mu'$ for an exactly centered point, we can rewrite (53) as

$$s_j(\mu') \leq \frac{\mu'\cdot 2\cdot 8^2\cdot\bar{\chi}_A\cdot\theta_k n^3\sqrt{1+\eta_0}}{\mu}.$$

Thus, a crossover event takes place between i and j when μ' is sufficiently small so that

$$\frac{\mu'\cdot 2\cdot 8^2\cdot\bar{\chi}_A\cdot\theta_k n^3\sqrt{1+\eta_0}}{\mu} = \frac{1}{5g^n}\cdot\frac{\phi_k}{2\cdot 8^2\cdot(\bar{\chi}_A+1)n^3\sqrt{1+\eta_0}}$$

i.e.,

$$\frac{\mu'}{\mu} = \frac{\phi_k}{20g^n \cdot 8^4 \cdot \bar{\chi}_A(\bar{\chi}_A+1)\cdot\theta_k n^6(1+\eta_0)}$$
$$\geq \frac{1}{20g^{2n} \cdot 8^4 \cdot \bar{\chi}_A(\bar{\chi}_A+1)\cdot n^6(1+\eta_0)}.$$

Here, (25) was used to obtain the second line from the first. Thus, the number of steps to achieve this reduction in μ and cause the crossover event is given by $n_1(A)$, where

$$n_1(A) = c_0\sqrt{n}(2n\log g + 2\log\bar{\chi}_A + 6\log n + \mathrm{const}). \tag{63}$$

Case II holds for the iteration. In this case, γ_k^{P} and γ_k^{D} are both bounded above by $1/(4\sqrt{n})$. Then by Lemma 13, $\epsilon_k^{\mathrm{P}} \geq 0.5$ and $\epsilon_k^{\mathrm{D}} \geq 0.5$. Now we apply Lemma 10 to conclude that there is an $i \in J_1 \cup \cdots \cup J_k$ such that for all $\mu' \in (0,\mu]$,

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

By the same lemma, applying the reciprocal of (45), we conclude that there is a $j \in J_k \cup \cdots \cup J_p$ such that

$$s_j(\mu') \leq \frac{2\mu'\bar{\chi}_A\theta_k n^{1.5}}{\mu}.$$

Thus, if we pick μ' small enough so that

$$\frac{2\mu'\bar{\chi}_A\theta_k n^{1.5}}{\mu} = \frac{1}{5g^n}\cdot\frac{\phi_k}{2g^n(\bar{\chi}_A+1)n^{1.5}}$$

i.e.

$$\frac{\mu'}{\mu} = \frac{\phi_k}{20g^n\bar{\chi}_A(\bar{\chi}_A+1)\theta_k n^3} \geq \frac{1}{20g^{2n}\bar{\chi}_A(\bar{\chi}_A+1)n^3}$$

then a crossover takes place. To reduce μ by this factor requires $n_2(A)$ small interior point steps, where

$$n_2(A) = c_0\sqrt{n}(2n\log g + 2\log\bar{\chi}_A + 3\log n + \mathrm{const}). \tag{64}$$

Case III holds for the iteration, and $\bar{\alpha} > 0$. Let k be the layer such that $\bar{\alpha} = \alpha_k$ in the definition LLS step. For now, assume that for this layer k, (34) is satisfied. The other case is similar. In this case, $\alpha_k = \min(1, 8\epsilon_k^{\mathrm{P}}\sqrt{n})$. By taking the reciprocal of (45), we arrive at a $j \in J_k \cup \cdots \cup J_p$ such that for all $\mu' \in (0, \mu]$,

$$\begin{aligned} s_j(\mu') &\le \frac{\mu' \bar{\chi}_A \theta_k n^{1.5}}{\mu \epsilon_k^{\mathrm{P}}} \\ &\le \frac{8\mu' \bar{\chi}_A \theta_k n^2}{\mu \bar{\alpha}}. \end{aligned}$$

On the other hand, by (54), there is an $i \in J_1 \cup \cdots \cup J_k$ such that

$$s_i(\mu') \ge \frac{\phi_k}{2 \cdot 8^2 \cdot (\bar{\chi}_A + 1) n^3 \sqrt{1 + \eta_0}}.$$

Thus, a crossover event takes place when

$$\frac{8\mu' \bar{\chi}_A \theta_k n^2}{\mu \bar{\alpha}} = \frac{\phi_k}{5g^n \cdot 2 \cdot 8^2 \cdot (\bar{\chi}_A + 1) n^3 \sqrt{1 + \eta_0}},$$

i.e.

$$\frac{\mu'}{\mu \bar{\alpha}} = \frac{\phi_k}{80 g^n \cdot 8^2 \bar{\chi}_A (\bar{\chi}_A + 1) n^5 \theta_k \sqrt{1 + \eta_0}} \ge \frac{1}{80 g^{2n} \cdot 8^2 \bar{\chi}_A (\bar{\chi}_A + 1) n^5 \sqrt{1 + \eta_0}}.$$

Since the LLS step decreases the central path parameter to $\mu\bar{\alpha}$, the number of small steps is the number required to reduce the parameter from $\mu\bar{\alpha}$ to μ' and cause a crossover is

$$n_3(A) = c_0 \sqrt{n} (2n \log g + 2 \log \bar{\chi}_A + 5 \log n + \mathrm{const}). \tag{65}$$

∎

Earlier we showed that there can be at most $n(n-1)/2$ crossover events. In fact, the preceding theorem shows that whenever (i, j) is a crossover event, at optimum $s_i^* > 0$ whereas $s_j^* = 0$. This means that $i \in N$ and $j \in B$, where (B, N) is the complementary partition of the set $\{1, \ldots, n\}$ defined in the proof of Corollary 2. Therefore, the total number of crossover events is bounded by $|B| \cdot |N|$. Since $|B| + |N| = n$, we have the slightly stronger upper bound of $n^2/4$ on the number of crossover events.

8 Characterizing the central path

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 approximately straight segments. By "approximately straight" we mean that there is a straight-line segment that is approximately centered over a range of central-path parameters. This is formally stated in the following corollary.

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

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

such that $k \le n^2/2$ and such that, for each $i \in \{0, \ldots, k-1\}$, either

1. *There exist vectors $\mathbf{u}_i, \bar{\mathbf{u}}_i, \mathbf{v}_i, \bar{\mathbf{v}}_i, \mathbf{w}_i, \bar{\mathbf{w}}_i$ such that for all $\mu \in (\mu_i, \mu_{i+1})$, $\eta(\mathbf{u}_i + \mu\bar{\mathbf{u}}_i, \mathbf{v}_i + \mu\bar{\mathbf{v}}_i, \mathbf{w}_i + \mu\bar{\mathbf{w}}_i, \mu) \le 0.65$, or*
2. *μ_{i+1}/μ_i is bounded above by a constant $C(A)$ depending only on A.*

PROOF. Introduce a breakpoint at the value of μ at the beginning of each iteration of Algorithm LIP. If Case III holds for some iteration, then introduce an additional breakpoint for the parameter value $\mu\bar{\alpha}$. Then for iterations in Cases I and II, and for the second part of Case III, Case 2 of the lemma holds by Theorem 2. For the LLS step, Case 1 of the lemma holds by Theorem 1. We have $k \le n^2/2$ because iterations in Cases I and II each introduce one breakpoint, and iterations in Case III introduce two breakpoints. ∎

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 many intervals $[\mu_i, \mu_{i+1}]$, 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 difficulty is that the tangent to the central path at a centered point (which is the direction computed by traditional algorithms) does not necessarily "point" in the right direction. This is most clearly seen on the last LLS step of Algorithm LIP (see, e.g., Figure 2), when the optimum is found. The tangent to the central path at $\mathbf{y}(\mu)$ is never precisely aligned with the displacement vector between $\mathbf{y}(\mu)$ and the optimum $\mathbf{y}^*$ for any positive μ. In contrast, the LLS step is precisely this displacement.

The curvature of the central path has received attention previously in the literature. It is already known that, as observed in the last paragraph, the central path becomes asymptotically straight (as a function of $\log \mu$) as the optimum is approached. This fact is the basis for results on asymptotic quadratic convergence of interior point methods such as the result of Ye et al. [44] and earlier works cited therein.

Sonnevend, Stoer and Zhao [31] have considered the total integral of the curvature in the path in order to bound complexity. Megiddo and Shub [19] have considered bends in the limiting behavior of the central path trajectory as the boundary of the feasible set is reached. We do not know of any direct connection between these results and the preceding corollary.

9 Complexity of Algorithm LIP

We can now estimate the complexity of the algorithm. There are a total of at most $n^2/4$ iterations of Algorithm LIP. The main work associated with each iteration is the inner path-following interior point small steps. (The number of LLS steps is much smaller than the number of ordinary steps). There are $O(\sqrt{n} \cdot c_1(A))$ small steps per LIP iteration, where $c_1(A)$ arises from the maximum of (63), (64) and (65), that is $c_1(A) = O(n \log g + \log \bar{\chi}_A + \log n)$. Recall that g must satisfy (57) and (62). Since (62) dominates, we use this as the definition of g:

$$g = 128(1 + \eta_0)n^2(\bar{\chi}_A + 1). \tag{66}$$

With this definition, we conclude that $c_1(A) = O(n(\log \bar{\chi}_A + \log n))$. Thus, the total number of small step iterations over all $O(n^2)$ main loop iterations is $O(n^{3.5}c(A))$ where

$$c(A) = 2c_0(\log(\bar{\chi}_A) + 2\log n + \text{const}). \tag{67}$$

Each small step requires solution of a system of linear equations, which takes $O(m^2n)$ arithmetic operations. (The LLS step also requires $O(m^2n)$ operations.)

It should be noted that this complexity bound does not depend on the value of μ given to Algorithm LIP at the start of the first main-loop iteration. (This is in contrast to other interior point methods, whose complexity depends logarithmically on the initial value of the central path parameter.) This is an important point for the initialization procedure in the next section.

10 Initialization

We now examine the question of initializing Algorithm LIP. Given $(A, \mathbf{b}, \mathbf{c})$, we consider the following artificial LP problem

$$\begin{array}{lccccccl} \text{maximize} & & & \mathbf{c}^T\mathbf{x} & - & N\mathbf{e}^T\mathbf{x}_2 & & \\ \text{subject to} & & & A\mathbf{x} & - & A\mathbf{x}_2 & = & \mathbf{b}, \\ & \mathbf{x}_1 & + & \mathbf{x} & & & = & 2N\mathbf{e}, \\ & \mathbf{x}_1 & , & \mathbf{x} & , & \mathbf{x}_2 & \geq & \mathbf{0} \end{array} \tag{68}$$

and its dual

$$\begin{array}{lccccccl} \text{minimize} & \mathbf{b}^T\mathbf{y} & + & 2N\mathbf{e}^T\mathbf{y}_1 & & & & \\ \text{subject to} & & & \mathbf{y}_1 & - & \mathbf{s}_1 & = & \mathbf{0}, \\ & A^T\mathbf{y} & + & \mathbf{y}_1 & - & \mathbf{s} & = & \mathbf{c}, \\ & -A^T\mathbf{y} & & & - & \mathbf{s}_2 & = & -N\mathbf{e}, \\ & & & & & \mathbf{s}_1, \mathbf{s}, \mathbf{s}_2 & \geq & \mathbf{0}. \end{array} \tag{69}$$

We may select an initial feasible point as

$$\mathbf{x}_1^0 = N\mathbf{e}, \quad \mathbf{x}^0 = N\mathbf{e}, \quad \mathbf{x}_2^0 = N\mathbf{e} - \mathbf{d}$$

$$\mathbf{y}^0 = \mathbf{0}, \quad \mathbf{y}_1^0 = N\mathbf{e}$$

and

$$\mathbf{s}_1^0 = N\mathbf{e}, \quad \mathbf{s}^0 = N\mathbf{e} - \mathbf{c}, \quad \mathbf{s}_2^0 = N\mathbf{e}$$

where $\mathbf{d}$ is a vector satisfying $A\mathbf{d} = \mathbf{b}$, for example, $\mathbf{d} = A^T(AA^T)^{-1}\mathbf{b}$, which is the minimum-norm vector subject to $A\mathbf{x} = \mathbf{b}$. Obviously, this selection satisfies all equations in both (68) and (69) for any $N > 0$. We now choose N sufficiently large such that all $(\mathbf{x}_1^0; \mathbf{x}^0; \mathbf{x}_2^0)$ and $(\mathbf{s}_1^0; \mathbf{s}^0; \mathbf{s}_2^0)$ are positive and they are approximately centered with parameter

$$\mu^0 = \frac{3nN^2 - N\mathbf{c}^T\mathbf{e} - N\mathbf{d}^T\mathbf{e}}{3n}.$$

Therefore, we are ready to apply Algorithm LIP to solving (68) and (69).

If both the original (1) and (2) have a solution, then they have an optimal basis by assumption of full rank of A, that is, an index set for a basis B and a point $\mathbf{x}_B$ and $\mathbf{y}$ such that $A_B\mathbf{x}_B = \mathbf{b}$, $\mathbf{x}_B \geq 0$, $A_B^T\mathbf{y} = \mathbf{c}_B$ and $A^T\mathbf{y} \geq \mathbf{c}$. Thus,

$$\|\mathbf{x}_B\| = \|A_B^{-1}\mathbf{b}\| = \|A_B^{-1}A\mathbf{d}\| \leq \|A_B^{-1}A\| \cdot \|\mathbf{d}\| \leq \bar{\chi}_A\|\mathbf{d}\|$$

and

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

The last line follows from Lemma 2. Thus, if we choose $N > (1+\bar{\chi}_A)\|\mathbf{c}\|$ and $N > \bar{\chi}_A\|\mathbf{d}\|$, then we are guaranteed that an optimal solution generated from the artificial problem is an optimal solution for the original problem.

The only remaining issue is that we do not work with the original A in the artificial problem. We instead work with $\hat{A}$ where

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

In this regard, we have the following lemma.

Lemma 14

$$\bar{\chi}(\hat{A}) \leq 3\sqrt{2}(1+\bar{\chi}_A).$$

Proof. Let $\mathbf{c} = (\mathbf{c}_1; \mathbf{c}_2; \mathbf{c}_3)$ be arbitrary, $D = (D_1; D_2; D_3) \in \mathcal{D}$ arbitrary, and consider a solution to

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

The goal is to obtain a bound on $\|\hat{A}^T\mathbf{y}\|$ in terms of $\|\mathbf{c}\|$. Let $\mathbf{y} = (\mathbf{y}_1; \mathbf{y}_2)$. Then this equation may be written in block form:

$$\begin{pmatrix} AD_2A^T + AD_3A^T & AD_2 \\ D_2A^T & D_1 + D_2 \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} AD_2\mathbf{c}_2 - AD_3\mathbf{c}_3 \\ D_1\mathbf{c}_1 + D_2\mathbf{c}_2 \end{pmatrix}.$$

Thus, from the bottom block equation we have

$$\mathbf{y}_2 = (D_1 + D_2)^{-1}(D_1\mathbf{c}_1 + D_2\mathbf{c}_2 - D_2 A^T\mathbf{y}_1),$$

and, substituting $\mathbf{y}_2$ in the top equation,

$$A(D_{12} + D_3)A^T\mathbf{y}_1 = A(D_{12}\mathbf{c}_2 - D_{12}\mathbf{c}_1 - D_3\mathbf{c}_3).$$

where

$$D_{12} = D_1(D_1 + D_2)^{-1}D_2.$$

Let $\bar{D} = D_{12} + D_3$; then we have

$$A\bar{D}A^T\mathbf{y}_1 = A\bar{D}[\bar{D}^{-1}(D_{12}\mathbf{c}_2 - D_{12}\mathbf{c}_1 - D_3\mathbf{c}_3)].$$

Observe that because $\bar{D}$ is the sum of D_{12} and D_3, both positive, the scaling matrices $\bar{D}^{-1}D_{12}$ and $\bar{D}^{-1}\bar{D}_3$ both have positive diagonal entries bounded by 1. Thus,

$$\|\bar{D}^{-1}(D_{12}\mathbf{c}_2 - D_{12}\mathbf{c}_1 - D_3\mathbf{c}_3)\| \le \|\mathbf{c}_1\| + \|\mathbf{c}_2\| + \|\mathbf{c}_3\|,$$

so by definition

$$\|A^T\mathbf{y}_1\| \le \bar{\chi}_A\|\bar{D}^{-1}(D_{12}\mathbf{c}_2 - D_{12}\mathbf{c}_1 - D_3\mathbf{c}_3)\| \le \bar{\chi}_A(\|\mathbf{c}_1\| + \|\mathbf{c}_2\| + \|\mathbf{c}_3\|).$$

Furthermore, we have

$$\begin{aligned}\|\mathbf{y}_2\| &= \|(D_1 + D_2)^{-1}(D_1\mathbf{c}_1 + D_2\mathbf{c}_2 - D_2A^T\mathbf{y}_1)\| \\ &\le \|\mathbf{c}_1\| + \|\mathbf{c}_2\| + \|A^T\mathbf{y}_1\| \\ &\le \|\mathbf{c}_1\| + \|\mathbf{c}_2\| + \bar{\chi}_A(\|\mathbf{c}_1\| + \|\mathbf{c}_2\| + \|\mathbf{c}_3\|) \\ &\le (1 + \bar{\chi}_A)(\|\mathbf{c}_1\| + \|\mathbf{c}_2\| + \|\mathbf{c}_3\|).\end{aligned}$$

Since

$$\hat{A}^T\mathbf{y} = \begin{pmatrix} \mathbf{y}_2 \\ A^T\mathbf{y}_1 + \mathbf{y}_2 \\ -A^T\mathbf{y}_1 \end{pmatrix},$$

we have

$$\|\hat{A}^T\mathbf{y}\|^2 \le 3(\|A^T\mathbf{y}_1\|^2 + \|\mathbf{y}_2\|^2).$$

Thus

$$\begin{aligned}
\|\hat{A}^T\mathbf{y}\| &\le \sqrt{6}(1+\bar{\chi}_A)(\|\mathbf{c}_1\|+\|\mathbf{c}_2\|+\|\mathbf{c}_3\|) \\
&\le \sqrt{6}(1+\bar{\chi}_A)\sqrt{3(\|\mathbf{c}_1\|^2+\|\mathbf{c}_2\|^2+\|\mathbf{c}_3\|^2)} \\
&= \sqrt{6}(1+\bar{\chi}_A)\sqrt{3\|(\mathbf{c}_1;\mathbf{c}_2;\mathbf{c}_3)\|^2} \\
&= 3\sqrt{2}(1+\bar{\chi}_A)\|\mathbf{c}\|.
\end{aligned}$$

This proves the lemma. ∎

Of course, we have no knowledge whether the original problem has a solution or not. However, we can apply this initialization technique and the LIP algorithm to solving LP data set $(A, \mathbf{b}, \mathbf{0})$ ($(A, \mathbf{0}, \mathbf{c})$). This will generate either a primal (dual) feasible solution close to the analytic center of system $Ax = \mathbf{b}$ ($A^T\mathbf{y} \ge \mathbf{c}$) or a certificate of primal (dual) infeasibility, if N is chosen sufficiently large as described above.

11 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 [36] and Vavasis and Ye [40]):

Lemma 15 *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 $\chi_A, \bar{\chi}_A$. A similar proof is due to Ben-Tal and Teboulle [5]. 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 [37]. 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 *[33] Consider solving an LP of the form* (2) *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 [7, 37].

It can be proved that this truncation preserves approximate centering; see, for example, Chapter 3 of [37]. The running time is $O(n^{3.5}c(A))$ interior point steps, and from (67) and the lemma we see that $c(A) = O(L_A + \log n)$. ∎

12 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 (67) 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 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 [26, 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 11 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

$$\|\mathbf{y}_3\| \leq \|\hat{\mathbf{c}}\|,$$

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

and

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

Thus, we must have

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

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

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

Therefore, the overall complexity for Algorithm LIP applied to the min-cost flow problem is $O(m^{3.5}\log n)$ 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 [25]. Thus, the LIP method needs considerable improvement in order to reach this bound. On the other hand, in practice, interior point methods often perform much better than their worst-case bounds and can be specialized for network problems: see Resende and Veiga [29].

13 Conclusions

We have proposed the first known method whose running time depends only on A for linear programming with data given by real numbers. This is attained by including a step called the "LLS" step that accelerates the convergence of an interior-point method. 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 $\bar{\chi}_A$ is used—the most crucial use of this knowledge is in the formula for g given by (66), which determines the spacing between the layers. Note that we do not need to know these parameters exactly—upper bounds will suffice.

   The only known algorithm for computing these parameters is implicit in the results of Stewart [32] and O'Leary [24] and requires exponential-time. We suspect that computing them, or even getting a good upper bound, may be a hard problem. Khachiyan [13] has shown that it is NP-hard to compute or approximate χ_A, and his results may extend to $\bar{\chi}_A$.

   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 $\bar{\chi}_A = 100$ and run the entire LIP algorithm. If it fails in any way, then guess $\bar{\chi}_A = 100^2$ 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 $\log\log\bar{\chi}_A$.

   Another possibility would be to obtain a bound on $\bar{\chi}_A$ for special classes of linear programs that occur in practice. For example, we carried out this process for min-cost flow problems in Section 12. 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 [6]), 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 an exact 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 step. 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 [42].

4. Our upper bound on the total number of main-loop iterations is $n^2/4$, but we have not been able to construct any examples that required more than n main-loop iterations.

   Furthermore, another factor of n in the final complexity bound comes from the factor g^{2n} appearing (43). It is the logarithm of this factor, $n \log g$, that determines $c(A)$. But the appearance of g^{2n} in (43) rather than g alone seems to be an artifact of our analysis rather than an actual aspect of the complexity.

5. An alternative initialization procedure gaining popularity is the so-called "infeasible" interior point method (for example, see Lustig et al. [16]). It would be interesting if there were a layered version of these algorithms.

6. As mentioned in Section 3, there is a relationship between $\chi_A, \bar{\chi}_A$ and condition numbers of weighted least-squares problems as demonstrated by [39]. Thus, our new results can be thought of a link between condition number and complexity of iterative processes (also see [36] and [40]). Such links are well-known in numerical analysis; the classic example is the conjugate gradient algorithm (see [9]). Renegar [28] 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 (2).

7. 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. We remark that the result on the finiteness of $\bar{\chi}_A$ stated in Section 3 has recently been partially generalized to nondiagonal weight matrices by Forsgren [8]. 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.

Acknowledgments

The authors benefited from several discussions with J. Renegar and M. Todd on this material. After the first version of this paper, the authors improved and simplified the paper by adopting a primal-dual approach after valuable discussions with M. Todd and L. Tuncel.

The first author would like to dedicate this paper in memory of his grandparents Stavros, Valentina, Anton and Mary.

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.A. Bayer and J.C. Lagarias, The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories, Transactions of the AMS 314 (1989) 499–526.

[3] D.A. Bayer and J.C. Lagarias, The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories, Transactions of the AMS 314 (1989) 527–581.

[4] D.A. Bayer and J.C. Lagarias, The nonlinear geometry of linear programming. III. Projective Legendre transform coordinates and Hilbert geometry, Transactions of the AMS 320 (1990) 193–225.

[5] A. Ben-Tal and M. Teboulle, A geometric property of the least squares solutions of linear equations, Linear algebra and its applications 139 (1990) 165–170.

[6] T. Coleman, Large-scale numerical optimization: introduction and overview, in: Encyclopedia of computer science and technology, A. Kent and J.G. Williams, eds. (Marcel Dekker, New York, 1993).

[7] J. Edmonds, Systems of distinct representatives and linear algebra, J. Res. Nat. Bur. Standards 71B (1967) 241–245.

[8] A. Forsgren, On linear least-squares problems with diagonally dominant weight matrices, Technical Report TRITA-MAT-1995-OS2, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden.

[9] G.H. Golub and C.F.V. Loan, Matrix Computations, 2nd Edition (Johns Hopkins University Press, Baltimore, 1989).

[10] C.C. Gonzaga, Path-following methods for linear programming, SIAM Review 34 (1992) 167–224.

[11] J.A. Kaliski and Y. Ye, A short-cut potential reduction algorithm for linear programming, Management Science 39 (1993) 757–773.

[12] N. Karmarkar, A new polynomial-time algorithm for linear programming, Combinatorica 4 (1984) 373–395.

[13] L. Khachiyan, On the complexity of approximating extremal determinants in matrices (1994). Preprint.

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

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

[16] 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 (1991) 191–222.

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

[18] N. Megiddo, Pathways to the optimal set in linear programming, in: Progress in Mathematical Programming: Interior Point and Related Method, N. Megiddo, ed. (Springer, New York, 1989) pp. 131–158.

[19] N. Megiddo and M. Shub, Boundary behavior of interior point algorithms in linear programming, Mathematics of Operations Research 14 (1989) 97–146.

[20] S. Mizuno, M.J. Todd and Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming, Mathematics of Operations Research 18 (1993) 964–981.

[21] S. Mizuno, M.J. Todd and Y. Ye, A surface of analytic centers and infeasible interior-point algorithms for linear programming, Mathematics of Operations Research 20 (1995) 135–162.

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

[23] Y.E. Nesterov and A.S. Nemirovsky, Interior Point Polynomial Methods in Convex Programming: Theory and Algorithms, in: SIAM Publications (SIAM, Philadelphia, USA, 1993).

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

[25] J.B. Orlin, A faster strongly polynomial minimum cost flow algorithm, in: Proc. 20th ACM Symposium on the Theory of Computing (1988).

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

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

[28] J. Renegar, Linear Programming, Complexity Theory, and Elementary Functional Analysis (1993). Unpublished manuscript, Department of Operations Research and Industrial Engineering, Cornell University.

[29] M.G.C. Resende and G. Veiga, An efficient implementation of a network interior point method, Technical report, Mathematical Sciences Research Center, AT & T Bell Laboratories, Murray Hill, NJ 07974–0636, USA (1992). Revised March 1992. Revised version 2.0 August 1992. To appear in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science.*

[30] 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 (Springer–Verlag, New York, 1985) pp. 866–876.

[31] G. Sonnevend, J. Stoer and G. Zhao, On the complexity of following the central path of linear programs by linear extrapolation II, Mathematical Programming 52 (1991) 527–553.

[32] G.W. Stewart, On scaled projections and pseudoinverses, Linear Algebra and its Applications 112 (1989) 189–193.

[33] E. Tardos, A strongly polynomial algorithm to solve combinatorial linear programs, Operations Research 34 (1986) 250–256.

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

[35] K. Tone, An active–set strategy in an interior point method for linear programming, Mathematical Programming 59 (1993) 345–360.

[36] 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).

[37] S.A. Vavasis, Nonlinear Optimization: Complexity Issues (Oxford University Press, New York, 1991).

[38] S.A. Vavasis, Stable finite elements for problems with wild coefficients, Technical Report 93–1364, Department of Computer Science, Cornell University, Ithaca, NY (1993). To appear in *SIAM J. Numerical Analysis.*

[39] S.A. Vavasis, Stable numerical algorithms for equilibrium systems, SIAM J. Matrix Anal. Appl 15 (1994) 1108–1131.

[40] S.A. Vavasis and Y. Ye, Condition Numbers for Polyhedra with Real Number Data, Technical Report 93-1398, Department of Computer Science, Cornell University (1993).

[41] M.H. Wright, Numerical methods for nonlinearly constrained optimization, Ph.D. Thesis, Stanford University (1976).

[42] M.H. Wright, Interior methods for constrained optimization, in: Acta Numerica 1992, A. Iserles, ed. (Cambridge University Press, Cambridge, 1992) pp. 341–407.

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

[44] Y. Ye O. Güler, R.A. Tapia and Y. Zhang, A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming, Mathematical Programming 59 (1993) 151–162.