A VARIANT OF THE VAVASIS–YE LAYERED-STEP INTERIOR-POINT ALGORITHM FOR LINEAR PROGRAMMING[*]

RENATO D. C. MONTEIRO[†] AND TAKASHI TSUCHIYA[‡]

Abstract. In this paper we present a variant of Vavasis and Ye's layered-step path-following primal-dual interior-point algorithm for linear programming. Our algorithm is a predictor–corrector-type algorithm which uses from time to time the layered least squares (LLS) direction in place of the affine scaling (AS) direction. It has the same iteration-complexity bound of Vavasis and Ye's algorithm, namely $\mathcal{O}(n^{3.5} \log(\bar{\chi}_A + n))$, where n is the number of nonnegative variables and $\bar{\chi}_A$ is a certain condition number associated with the constraint matrix A. Vavasis and Ye's algorithm requires explicit knowledge of $\bar{\chi}_A$ (which is very hard to compute or even estimate) in order to compute the layers for the LLS direction. In contrast, our algorithm uses the AS direction at the current iterate to determine the layers for the LLS direction, and hence does not require the knowledge of $\bar{\chi}_A$. A variant with similar properties and with the same complexity has been developed by Megiddo, Mizuno, and Tsuchiya [*Math. Programming*, 82 (1998), pp. 339–355]. However, their algorithm needs to compute n LLS directions on every iteration, while ours computes at most one LLS direction on any given iteration.

Key words. interior-point algorithms, primal-dual algorithms, path-following, central path, layered steps, condition number, polynomial complexity, predictor-corrector, affine scaling, strongly polynomial, linear programming

AMS subject classifications. 65K05, 68Q25, 90C05, 90C51, 90C60

PII. S1052623401388926

1. Introduction. We consider the linear programming (LP) problem

$$
\begin{aligned}
\text{minimize}_x \quad & c^T x \\
\text{subject to} \quad & Ax = b, \quad x \geq 0,
\end{aligned}
\tag{1}
$$

and its associated dual problem

$$
\begin{aligned}
\text{maximize}_{(y,s)} \quad & b^T y \\
\text{subject to} \quad & A^T y + s = c, \quad s \geq 0,
\end{aligned}
\tag{2}
$$

where $A \in \Re^{m \times n}$, $c \in \Re^n$, and $b \in \Re^m$ are given, and the vectors $x, s \in \Re^n$, and $y \in \Re^m$ are the unknown variables. This paper proposes a primal-dual layered-step predictor-corrector interior-point algorithm that is a variant of the Vavasis–Ye layered-step interior-point algorithm proposed in [26, 27].

Karmarkar in his seminal paper [5] proposed the first polynomially convergent interior-point method with an $\mathcal{O}(nL)$ iteration-complexity bound, where L is the size

of the LP instance (1). The first path-following interior-point algorithm was proposed by Renegar in his breakthrough paper [16]. Renegar's method closely follows the primal central path and exhibits an $\mathcal{O}(\sqrt{n}L)$ iteration-complexity bound. The first path-following algorithm which simultaneously generates iterates in both the primal and dual spaces has been proposed by Kojima, Mizuno, and Yoshise [6] and Tanabe [18], based on ideas suggested by Megiddo [9]. In contrast to Renegar's algorithm, Kojima, Mizuno, and Yoshise's algorithm has an $\mathcal{O}(nL)$ iteration-complexity bound. A primal-dual path-following with an $\mathcal{O}(\sqrt{n}L)$ iteration-complexity bound was subsequently obtained by Kojima, Mizuno, and Yoshise [7] and Monteiro and Adler [13, 14] independently. Following these developments, many other primal-dual interior-point algorithms for linear programming have been proposed.

An outstanding open problem in optimization is whether there exists a strongly polynomial algorithm for linear programming, that is, one whose complexity is bounded by a polynomial of m and n only. A major effort in this direction is due to Tardos [19], who developed a polynomial-time algorithm whose complexity is bounded by a polynomial of m, n, and L_A, where L_A denotes the size of A. Such an algorithm gives a strongly polynomial method for the important class of LP problems where the entries of A are either 1, -1, or 0, e.g., LP formulations of network flow problems. Tardos's algorithm consists of solving a sequence of "low-sized" LP problems by a standard polynomially convergent LP method and using their solutions to obtain the solution of the original LP problem.

The development of a method which works entirely in the context of the original LP problem and whose complexity is also bounded by a polynomial of m, n, and L_A is due to Vavasis and Ye [26]. Their method is a primal-dual path-following interior-point algorithm similar to the ones mentioned above except that it uses from time to time a crucial step, namely the layered least squares (LLS) direction. They showed that their method has an $\mathcal{O}(n^{3.5}(\log \bar{\chi}_A + \log n))$ iteration-complexity bound, where $\bar{\chi}_A$ is a condition number associated with A having the property that $\log \bar{\chi}_A = \mathcal{O}(L_A)$. The number $\bar{\chi}_A$ was first introduced implicitly by Dikin and Zorkalcev [1] in the study of primal affine scaling algorithms and was later studied by several researchers including Vanderbei and Lagarias [25], Todd [20], and Stewart [17]. Properties of $\bar{\chi}_A$ are studied in [3, 23, 24].

The complexity analysis of Vavasis and Ye's algorithm is based on the notion of a crossover event, a combinatorial event concerning the central path. Intuitively, a crossover event occurs between two variables when one of them is larger than the other at a point in the central path and then becomes smaller asymptotically as the optimal solution set is approached. Vavasis and Ye showed that there can be at most $n(n-1)/2$ crossover events and that a distinct crossover event occurs every $\mathcal{O}(n^{1.5}(\log \bar{\chi}_A + \log n))$ iterations, from which they deduced the overall $\mathcal{O}(n^{3.5}(\log \bar{\chi}_A + \log n))$ iteration-complexity bound. In [12], an LP instance is given where the number of crossover events is $\Theta(n^2)$.

One disadvantage of Vavasis and Ye's method is that it requires the explicit knowledge of $\bar{\chi}_A$ in order to determine a partition of the variables into layers used in the computation of the LLS step. This difficulty was remedied in a variant proposed by Megiddo, Mizuno, and Tsuchiya [10] which does not require the explicit knowledge of the number $\bar{\chi}_A$. They observed that at most n types of partitions arise as $\bar{\chi}_A$ varies from 1 to ∞ and that one of these can be used to compute the LLS step. Based on this idea, they developed a variant which computes the LLS steps for all these partitions and picks the one that yields the greatest duality gap reduction at

the current iteration. Moreover, using the argument that once the first LLS step is computed the other ones can be cheaply computed by performing rank-one updates, they show that the overall complexity of their algorithm is exactly the same as Vavasis and Ye's algorithm.

In this paper, we propose another variant of Vavasis and Ye's algorithm which has the same complexity as theirs and computes only one LLS step per iteration without any explicit knowledge of $\bar{\chi}_A$. Our algorithm is a predictor–corrector-type algorithm like the one described in [11] except that at the predictor stage it takes a step along either the primal-dual affine scaling (AS) step or the LLS step. More specifically, first the AS direction is computed and a test involving this direction is performed to determine whether the LLS step is needed. If the LLS direction is not needed, a step along the AS direction is taken as usual. Otherwise, the AS direction is used to determine a partition of the variables into layers, and the LLS step with respect to this partition is computed. The algorithm then takes a step along the direction (either the AS or the LLS) which yields the largest duality gap reduction.

It is worth noting that our algorithm computes the LLS step only when a step along the AS direction has the potential to yield a significant duality gap decrease. In such a case, the LLS direction seems to be even better suited and it is used whenever the current iteration permits it. Another advantage of the LLS step is that it possesses the ability to determine an exact primal-dual optimal solution, and hence imply finite termination of the algorithm.

The organization of the paper is as follows. Section 2 consists of five subsections. In subsection 2.1, we review the notion of the primal-dual central path and its associated two-norm neighborhoods. Subsection 2.2 introduces the notion of the condition number $\bar{\chi}_A$ of a matrix A and describes the properties of $\bar{\chi}_A$ that will be useful in our analysis. Subsection 2.3 reviews the AS step and the corrector (or centrality) step which are the basic ingredients of several well-known interior-point algorithms. Subsection 2.4 describes the LLS step. Subsection 2.5 describes our algorithm in detail and states the main convergence result of this paper. Section 3, which consists of three subsections, introduces some basic tools which are used in our convergence analysis. Subsection 3.1 discusses the notion of crossover events. Subsection 3.2 states an approximation result that provides an estimation of the closeness between the AS direction and the LLS direction. Subsection 3.3 reviews from a different perspective an important result from Vavasis and Ye [26], which basically provides sufficient conditions for the occurrence of crossover events. Section 4 is dedicated to the proof of the main result stated in subsection 2.5. Section 5 gives some concluding remarks. Finally, the appendix gives the proof of the approximation result between the AS and the LLS directions stated in subsection 3.2.

The following notation is used throughout our paper. We denote the vector of all ones by e. Its dimension is always clear from the context. The symbols $\Re^n$, $\Re^n_+$, and $\Re^n_{++}$ denote the n-dimensional Euclidean space, the nonnegative orthant of $\Re^n$, and the positive orthant of $\Re^n$, respectively. The set of all $m \times n$ matrices with real entries is denoted by $\Re^{m \times n}$. If J is a finite index set, then $|J|$ denotes its cardinality, that is, the number of elements of J. For $J \subseteq \{1, \ldots, n\}$ and $w \in \Re^n$, we let w_J denote the subvector $[w_i]_{i \in J}$; moreover, if E is an $m \times n$ matrix, then E_J denotes the $m \times |J|$ submatrix of E corresponding to J. For a vector $w \in \Re^n$, we let $\max(w)$ and $\min(w)$ denote the largest and the smallest component of w, respectively, $\mathrm{Diag}(w)$ denote the diagonal matrix whose ith diagonal element is w_i for $i = 1, \ldots, n$, and w^{-1} denote the vector $[\mathrm{Diag}(w)]^{-1}e$ whenever it is well-defined. For two vectors $u, v \in \Re^n$, uv

denotes their Hadamard product, i.e., the vector in $\Re^n$ whose ith component is $u_i v_i$. The Euclidean norm, the 1-norm, and the ∞-norm are denoted by $\|\cdot\|$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$, respectively. For a matrix E, $\text{Im}(E)$ denotes the subspace generated by the columns of E and $\text{Ker}(E)$ denotes the subspace orthogonal to the rows of E. The superscript ^{T} denotes transpose.

2. Problem and primal-dual predictor-corrector interior-point algorithms. In this section we describe the proposed feasible interior-point primal-dual predictor-corrector algorithm for solving the pair of LP problems (1) and (2). We also present the main convergence result which establishes a polynomial iteration-complexity bound for the algorithm that depends only on the constraint matrix A.

This section is divided into five subsections. In subsection 2.1, we describe the primal-dual central path and its associated two-norm neighborhoods. In subsection 2.2, we describe the notion of the condition number of a matrix and describe the properties of the condition number that will be useful in our analysis. In subsection 2.3, we review the AS step and the corrector (or centrality) step which are the basic ingredients of several well-known interior-point algorithms. We also derive a lower bound on the stepsize along the AS step. In subsection 2.4, we describe an alternative step, namely the LLS step, which is sometimes used in place of the AS direction by our algorithm. In subsection 2.5, we describe our algorithm in detail and state the main convergence result of this paper.

2.1. The problem, the central path, and its associated neighborhoods. In this subsection we introduce the pair of dual linear programs and the assumptions used in our development. We also describe the associated primal-dual central path and its corresponding two-norm neighborhoods.

Given $A \in \Re^{m \times n}$, $c \in \Re^n$, and $b \in \Re^m$, consider the pairs of linear programs (1) and (2), where $x \in \Re^n$ and $(y, s) \in \Re^m \times \Re^n$ are their respective variables. The set of strictly feasible solutions for these problems are

$$\mathcal{P}^{++} \equiv \{x \in \Re^n : Ax = b,\ x > 0\},$$
$$\mathcal{D}^{++} \equiv \{(y, s) \in \Re^{m \times n} : A^T y + s = c,\ s > 0\},$$

respectively. Throughout the paper we make the following assumptions on the pair of problems (1) and (2).

A.1 $\mathcal{P}^{++}$ and $\mathcal{D}^{++}$ are nonempty.

A.2 The rows of A are linearly independent.

Under the above assumptions, it is well known that for any $\nu > 0$ the system

$$(3) \qquad\qquad\qquad xs = \nu e,$$
$$(4) \qquad\qquad\qquad Ax = b,\quad x > 0,$$
$$(5) \qquad\qquad A^T y + s = c,\quad s > 0,$$

has a unique solution (x, y, s), which we denote by $(x(\nu), y(\nu), s(\nu))$. The central path is the set consisting of all these solutions as ν varies in $(0, \infty)$. As ν converges to zero, the path $(x(\nu), y(\nu), s(\nu))$ converges to a primal-dual optimal solution (x^*, y^*, s^*) for problems (1) and (2). Given a point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$, its duality gap and its normalized duality gap are defined as $x^T s$ and $\mu = \mu(x, s) \equiv x^T s / n$, respectively, and the point $(x(\mu), y(\mu), s(\mu))$ is said to be the central point associated with w. Note that $(x(\mu), y(\mu), s(\mu))$ also has normalized duality gap μ. We define the proximity

measure of a point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ with respect to the central path by

$$\eta(w) \equiv \|xs/\mu - e\|.$$

Clearly, $\eta(w) = 0$ if and only if $w = (x(\mu), y(\mu), s(\mu))$ or, equivalently, w coincides with its associated central point. The two-norm neighborhood of the central path with opening $\beta > 0$ is defined as

$$\mathcal{N}(\beta) \equiv \{w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++} : \eta(w) \leq \beta\}.$$

Finally, for any point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ we define

$$(6) \qquad \delta(w) \equiv s^{1/2} x^{-1/2} \in \Re^n.$$

The following proposition provides important estimates which are used throughout our analysis.

PROPOSITION 2.1. *Let $w = (x, y, s) \in \mathcal{N}(\beta)$ for some $\beta \in (0, 1)$ be given and define $\delta \equiv \delta(w)$. Let $w(\mu) = (x(\mu), y(\mu), s(\mu))$ be the central point associated with w. Then*

$$(7) \qquad \frac{1-\beta}{1+\beta} s \leq s(\mu) \leq \frac{1}{1-\beta} s, \qquad \frac{1-\beta}{1+\beta} x \leq x(\mu) \leq \frac{1}{1-\beta} x,$$

$$(8) \qquad \frac{1-\beta}{(1+\beta)^{1/2}} \delta \leq \frac{s(\mu)}{\sqrt{\mu}} \leq \frac{(1+\beta)^{1/2}}{1-\beta} \delta,$$

$$(9) \qquad \frac{(1-\beta)^2}{(1+\beta)} \frac{\delta_i}{\delta_j} \leq \frac{s_i(\mu)}{s_j(\mu)} \leq \frac{(1+\beta)}{(1-\beta)^2} \frac{\delta_i}{\delta_j} \quad \forall i, j \in \{1, \ldots, n\}.$$

Proof. The second and fourth inequalities in (7) follow from Lemma 2.4(ii) of Gonzaga [2]. Using these two inequalities together with $xs \leq (1+\beta)\mu e$ and $x(\mu)s(\mu) = \mu e$, we obtain the other two inequalities in (7). Using the definition of $\delta = \delta(w)$ in (6) together with the relations $xs \leq (1 + \beta)\mu e$ and $x(\mu)s(\mu) = \mu e$, we easily see that the first and second inequalities of (8) follow from the fourth and second inequalities of (7), respectively. Finally, (9) immediately follows from (8). $\square$

2.2. Condition number. In this subsection we define a certain condition number associated with the constraint matrix A and state the properties of $\bar{\chi}_A$ which will play an important role in our analysis.

Let $\mathcal{D}$ denote the set of all positive definite $n \times n$ diagonal matrices and define

$$\bar{\chi}_A \equiv \sup\{\|A^T(A\tilde{D}A^T)^{-1}A\tilde{D}\| : \tilde{D} \in \mathcal{D}\}$$
$$= \sup\left\{\frac{\|A^T y\|}{\|c\|} : y = \underset{\tilde{y} \in \Re^n}{\operatorname{argmin}} \|\tilde{D}^{1/2}(A^T\tilde{y} - c)\| \text{ for some } 0 \neq c \in \Re^n \text{ and } \tilde{D} \in \mathcal{D}\right\}.$$
(10)

The parameter $\bar{\chi}_A$ plays a fundamental role in the complexity analysis of algorithms for linear programming and least squares problems (see [26] and references therein). Its finiteness has been established first by Dikin and Zorkalcev [1]. Other authors have also given alternative derivations of the finiteness of $\bar{\chi}_A$ (see, for example, Stewart [17], Todd [20], and Vanderbei and Lagarias [25]).

We summarize in the next proposition a few important facts about the parameter $\bar{\chi}_A$.

PROPOSITION 2.2. *Let $A \in \Re^{m \times n}$ with full row rank be given. Then the following statements hold:*

(a) $\bar{\chi}_{GA} = \bar{\chi}_A$ for any nonsingular matrix $G \in \Re^{m \times m}$.

(b) $\bar{\chi}_A = \max\{\|G^{-1}A\| : G \in \mathcal{G}\}$, where $\mathcal{G}$ denote the set of all $m \times m$ nonsingular submatrices of A.

(c) If the entries of A are all integers, then $\bar{\chi}_A$ is bounded by $2^{\mathcal{O}(L_A)}$, where L_A is the input bit length of A.

(d) $\bar{\chi}_A = \bar{\chi}_F$ for any $F \in \Re^{(n-m) \times n}$ such that $\mathrm{Ker}(A) = \mathrm{Im}(F^T)$.

(e) If the $m \times m$ identity matrix is a submatrix of A and $\tilde{A}$ is an $r \times n$ submatrix of A, then $\|\tilde{G}^{-1}\tilde{A}\| \le \bar{\chi}_A$ for every $r \times r$ nonsingular submatrix $\tilde{G}$ of $\tilde{A}$.

Proof. Statement (a) readily follows from the definition (10). The inequality $\bar{\chi}_A \ge \max\{\|G^{-1}A\| : G \in \mathcal{G}\}$ is established in Lemma 3 of [26], while the proof of the reverse inequality is given in [20] (see also Theorem 1 of [21]). Hence, (b) holds. The proof of (c) can be found in Lemma 24 of [26]. A proof of (d) can be found in [3].

We now consider (e). Using the assumption that the $m \times m$ identity matrix is a submatrix of A, we easily see that A has an $m \times m$ nonsingular submatrix G which, after some symmetric permutation of its rows and columns, can be put into the form

$$\begin{bmatrix} \tilde{G} & 0 \\ \tilde{E} & I \end{bmatrix}$$

for some matrix $\tilde{E} \in \Re^{(n-r) \times r}$. Since the inverse of the above matrix is

$$\begin{bmatrix} \tilde{G}^{-1} & 0 \\ -\tilde{E}\tilde{G}^{-1} & I \end{bmatrix},$$

we easily see that $\tilde{G}^{-1}\tilde{A}$ is a submatrix of $G^{-1}A$. Hence, $\|\tilde{G}^{-1}\tilde{A}\| \le \|G^{-1}A\| \le \bar{\chi}_A$, where the last inequality is due to (b). □

We now state a Hoffman-type result for a system of linear equalities that will be used in the proof of an approximation result given in the appendix.

LEMMA 2.3. *Let $A \in \Re^{m \times n}$ with full row rank be given, and let $(\mathcal{K}, \mathcal{L})$ be an arbitrary bipartition of the index set $\{1, \ldots, n\}$. Assume that $\bar{w} \in \Re^{|\mathcal{L}|}$ is an arbitrary vector such that the system $A_{\mathcal{K}}u = A_{\mathcal{L}}\bar{w}$ is feasible. Then this system has a feasible solution $\bar{u}$ such that $\|\bar{u}\| \le \bar{\chi}_A\|\bar{w}\|$.*

Proof. Due to Proposition 2.2(a), it is sufficient to establish the lemma for a matrix of the form GA, where G is an $m \times m$ nonsingular matrix. Hence, we may assume that A contains the $m \times m$ identity matrix. Eliminating some redundant rows from $A_{\mathcal{K}}u = A_{\mathcal{L}}\bar{w}$ and some variables from u, we obtain a nonsingular system

$$\tilde{G}\tilde{u} = \tilde{H}\bar{w},$$

where $\tilde{G}$ is a square submatrix of $A_{\mathcal{K}}$ such that $\mathrm{rank}(\tilde{G}) = \mathrm{rank}(A_{\mathcal{K}})$, $\tilde{H}$ is the corresponding submatrix of $A_{\mathcal{L}}$, and $\tilde{u}$ is a subvector of u. Clearly, the solution $\tilde{u}$ of this system satisfies $\|\tilde{u}\| \le \|\tilde{G}^{-1}\tilde{H}\| \|\bar{w}\| \le \bar{\chi}_A\|\bar{w}\|$, where the last inequality follows from Proposition 2.2(e). We can augment $\tilde{u}$ to a solution $\bar{u}$ of $A_{\mathcal{K}}u = A_{\mathcal{L}}\bar{w}$ by setting the components of u not in $\tilde{u}$ to zero. □

2.3. Predictor-corrector step and its properties. In this subsection we describe the well-known predictor-corrector (P-C) iteration which is used by several interior-point algorithms (see, for example, Mizuno, Todd, and Ye [11]). We also describe the properties of this iteration which will be used in our analysis.

The P-C iteration consists of two steps, namely the predictor (or AS) step and the corrector (or centrality) step. The search direction used by either step from a current point in $(x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ is the solution of the following linear system of equations:

$$\begin{aligned} S\Delta x + X\Delta s &= \sigma\mu e - xs, \\ A\Delta x &= 0, \\ A^T\Delta y + \Delta s &= 0, \end{aligned}$$

(11)

where $\mu = \mu(x, s)$ and $\sigma \in \Re$ is a prespecified parameter, commonly referred to as the centrality parameter. When $\sigma = 0$, we denote the solution of (11) by $(\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ and refer to it as the primal-dual AS direction at w; it is the direction used in the predictor step of the P-C iteration. When $\sigma = 1$, we denote the solution of (11) by $(\Delta x^{\mathrm{c}}, \Delta y^{\mathrm{c}}, \Delta s^{\mathrm{c}})$ and refer to it as the corrector direction at w; it is the direction used in the corrector step of the P-C iteration.

We are now ready to describe the entire P-C iteration. Suppose that a constant $\beta \in (0, 1/4]$ and a point $w = (x, y, s) \in \mathcal{N}(\beta)$ is given. The P-C iteration generates another point $(x^+, y^+, s^+) \in \mathcal{N}(\beta)$ as follows. It first moves along the direction $(\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ until it hits the boundary of the enlarged neighborhood $\mathcal{N}(2\beta)$. More specifically, it computes the point $w^{\mathrm{a}} = (x^{\mathrm{a}}, y^{\mathrm{a}}, s^{\mathrm{a}}) \equiv (x, y, s) + \alpha_{\mathrm{a}}(\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$, where

(12) $$\alpha_{\mathrm{a}} \equiv \sup\{\alpha \in [0, 1] : (x, y, s) + \alpha(\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}}) \in \mathcal{N}(2\beta)\}.$$

Next, the P-C iteration generates a point inside the smaller neighborhood $\mathcal{N}(\beta)$ by taking a unit step along the corrector direction $(\Delta x^{\mathrm{c}}, \Delta y^{\mathrm{c}}, \Delta s^{\mathrm{c}})$ at the point w^{a}; that is, it computes the point $(x^+, y^+, s^+) \equiv (x^{\mathrm{a}}, y^{\mathrm{a}}, s^{\mathrm{a}}) + (\Delta x^{\mathrm{c}}, \Delta y^{\mathrm{c}}, \Delta s^{\mathrm{c}}) \in \mathcal{N}(\beta)$. The successive repetition of this iteration leads to the so-called Mizuno–Todd–Ye (MTY) P-C algorithm (see [11]).

Our method is very similar to the algorithm of [11] except that it sometimes replaces the AS step by the LLS step described in the next subsection. The insertion of the LLS step in the above MTY P-C algorithm guarantees that the modified method has the finite termination property. Hence, the LLS step can be viewed as a termination procedure which is performed only when some "not-so-likely-to-occur" conditions are met. Moreover, the LLS step is taken only when it yields a point with a smaller duality gap than the one obtained from the AS step as described above.

In the remainder of this subsection, we discuss some properties of the P-C iteration and the primal-dual AS direction. For a proof of the next two propositions, we refer the reader to [11].

PROPOSITION 2.4 (predictor step). *Suppose that $w = (x, y, s) \in \mathcal{N}(\beta)$ for some constant $\beta \in (0, 1/2]$. Let $\Delta w^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ denote the AS direction at w^{a} and let α_{a} be the stepsize computed according to (12). Then the following statements hold:*

(a) *the point $w + \alpha\Delta w^{\mathrm{a}}$ has normalized duality gap $\mu(\alpha) = (1 - \alpha)\mu$ for all $\alpha \in \Re$;*
(b) *$\alpha_{\mathrm{a}} \geq \sqrt{\beta/n}$, and hence $\mu(\alpha_{\mathrm{a}})/\mu \leq 1 - \sqrt{\beta/n}$.*

PROPOSITION 2.5 (corrector step). *Suppose that $w = (x, y, s) \in \mathcal{N}(2\beta)$ for some constant $\beta \in (0, 1/4]$, and let $(\Delta x^{\mathrm{c}}, \Delta y^{\mathrm{c}}, \Delta s^{\mathrm{c}})$ denote the corrector step at w. Then $w + \Delta w^{\mathrm{c}} \in \mathcal{N}(\beta)$. Moreover, the (normalized) duality gap of $w + \Delta w^{\mathrm{c}}$ is the same as that of w.*

For the purpose of future comparison with the LLS step, we mention the following alternative characterization of the primal-dual AS direction whose verification is straightforward:

(13) $$\Delta x^{\mathrm{a}} \equiv \operatorname*{argmin}_{p \in \Re^n}\{\|\delta(x + p)\|^2 : Ap = 0\},$$

(14) $$(\Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}}) \equiv \operatorname*{argmin}_{(r, q) \in \Re^m \times \Re^n}\{\|\delta^{-1}(s + q)\|^2 : A^T r + q = 0\},$$

where $\delta \equiv \delta(w)$. For a search direction $(\Delta x, \Delta y, \Delta s)$ at a point (x, y, s), the quantity

$$
(Rx, Rs) \equiv \left(\frac{\delta(x + \Delta x)}{\sqrt{\mu}}, \frac{\delta^{-1}(s + \Delta s)}{\sqrt{\mu}} \right) = \left(\frac{x^{1/2}s^{1/2} + \delta\Delta x}{\sqrt{\mu}}, \frac{x^{1/2}s^{1/2} + \delta^{-1}\Delta s}{\sqrt{\mu}} \right)
$$
(15)

appears quite often in our analysis. We refer to it as the *residual* of $(\Delta x, \Delta y, \Delta s)$. Note that if $(Rx^{\mathrm{a}}, Rs^{\mathrm{a}})$ is the residual of $(\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$, then

$$
(16) \qquad Rx^{\mathrm{a}} = -\frac{1}{\sqrt{\mu}}\delta^{-1}\Delta s^{\mathrm{a}}, \qquad Rs^{\mathrm{a}} = -\frac{1}{\sqrt{\mu}}\delta\Delta x^{\mathrm{a}},
$$

and

$$
(17) \qquad Rx^{\mathrm{a}} + Rs^{\mathrm{a}} = \frac{x^{1/2}s^{1/2}}{\sqrt{\mu}},
$$

due to the fact that $(\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ satisfies the first equation in (11) with $\sigma = 0$. The following quantity is used in the test to determine when the LLS step should be used in place of the AS step:

$$
(18) \qquad \varepsilon_{\infty}^{\mathrm{a}} \equiv \max_{i} \left\{ \min \left\{ |Rx_i^{\mathrm{a}}|, |Rs_i^{\mathrm{a}}| \right\} \right\}.
$$

We end this section by providing some estimates involving the residual of the AS direction.

LEMMA 2.6. *Suppose that* $w = (x, y, s) \in \mathcal{N}(\beta)$ *for some* $\beta \in (0, 1/4]$. *Then, for all* $i = 1, \ldots, n$, *we have*

$$
\max \left\{ |Rx_i^{\mathrm{a}}|, |Rs_i^{\mathrm{a}}| \right\} \geq \frac{\sqrt{1 - \beta}}{2} \geq \frac{1}{4}.
$$

Proof. Assume for contradiction that for some $i \in \{1, \ldots, n\}$, $\max \left\{ |Rx_i^{\mathrm{a}}|, |Rs_i^{\mathrm{a}}| \right\} < \sqrt{1 - \beta}/2$. Then, using (17), we obtain the following contradiction:

$$
\frac{x_i^{1/2}s_i^{1/2}}{\sqrt{\mu}} = Rx_i^{\mathrm{a}} + Rs_i^{\mathrm{a}} \leq |Rx_i^{\mathrm{a}}| + |Rs_i^{\mathrm{a}}| < \sqrt{1 - \beta} \leq \frac{x_i^{1/2}s_i^{1/2}}{\sqrt{\mu}}. \qquad \square
$$

2.4. The LLS step. In this subsection we describe the other type of step used in our algorithm, namely the LLS step. This step was first introduced by Vavasis and Ye in [26].

Let $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ and a partition $(J_1, \ldots, J_p)$ of the index set $\{1, \ldots, n\}$ be given and define $\delta \equiv \delta(w)$. The primal LLS direction $\Delta x^{\mathrm{ll}} = (\Delta x_{J_1}^{\mathrm{ll}}, \ldots, \Delta x_{J_p}^{\mathrm{ll}})$ at w with the respect to J is defined recursively according to the order $\Delta x_{J_p}^{\mathrm{ll}}, \ldots, \Delta x_{J_1}^{\mathrm{ll}}$ as follows. Assume that the components $\Delta x_{J_p}^{\mathrm{ll}}, \ldots, \Delta x_{J_{k+1}}^{\mathrm{ll}}$ have been determined. Let $\Pi_{J_k} : \Re^n \to \Re^{J_k}$ denote the projection map defined as $\Pi_{J_k}(u) = u_{J_k}$ for all $u \in \Re^n$. Then $\Delta x_{J_k}^{\mathrm{ll}} \equiv \Pi_{J_k}(L_k^x)$, where L_k^x is given by

$$
\begin{aligned}
L_k^x &\equiv \operatorname*{Argmin}_{p \in \Re^n} \left\{ \|\delta_{J_k}(x_{J_k} + p_{J_k})\|^2 : p \in L_{k-1}^x \right\} \\
(19) \quad &= \operatorname*{Argmin}_{p \in \Re^n} \left\{ \|\delta_{J_k}(x_{J_k} + p_{J_k})\|^2 : p \in \mathrm{Ker}(A), \ p_{J_i} = \Delta x_{J_i}^{\mathrm{ll}} \ \ \forall i = k+1, \ldots, p \right\},
\end{aligned}
$$

with the convention that $L_0^x = \mathrm{Ker}(A)$. The slack component $\Delta s^{\mathrm{ll}} = (\Delta s_{J_1}^{\mathrm{ll}}, \ldots, \Delta s_{J_p}^{\mathrm{ll}})$ of the dual LLS direction $(\Delta y^{\mathrm{ll}}, \Delta s^{\mathrm{ll}})$ at w with the respect to J is defined recursively

as follows. Assume that the components $\Delta s^{\mathrm{ll}}_{J_1}, \ldots, \Delta s^{\mathrm{ll}}_{J_{k-1}}$ have been determined. Then $\Delta s^{\mathrm{ll}}_{J_k} \equiv \Pi_{J_k}(L^s_k)$, where L^s_k is given by

$$
\begin{aligned}
L^s_k &\equiv \operatorname*{Argmin}_{q \in \Re^n} \{\|\delta^{-1}_{J_k}(s_{J_k} + q_{J_k})\|^2 : q \in L^s_{k-1}\} \\
(20) \\
&= \operatorname*{Argmin}_{q \in \Re^n} \{\|\delta^{-1}_{J_k}(s_{J_k} + q_{J_k})\|^2 : q \in \operatorname{Im}(A^T), \ q_{J_i} = \Delta s^{\mathrm{ll}}_{J_i} \quad \forall i = 1, \ldots, k-1\},
\end{aligned}
$$

with the convention that $L^s_0 = \operatorname{Im}(A^T)$. Finally, once Δs^{ll} has been determined, the component Δy^{ll} is determined from the relation $A^T \Delta y^{\mathrm{ll}} + \Delta s^{\mathrm{ll}} = 0$.

Note that (13) and (14) imply that the AS direction is a special LLS direction, namely the one with respect to the only partition in which $p = 1$. Clearly, the LLS direction at a given $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ depends on the partition $J = (J_1, \ldots, J_p)$ used.

A partition $J = (J_1, \ldots, J_p)$ is said to be *ordered* at a point $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ if $\max(\delta_{J_i}) \leq \min(\delta_{J_{i+1}})$ for all $i = 1, \ldots, p-1$. In this case, the gap of J, denoted by $\mathrm{gap}(J)$, is defined as

$$
\mathrm{gap}(J) = \min_{1 \leq i \leq p-1} \left\{ \frac{\min(\delta_{J_{i+1}})}{\max(\delta_{J_i})} \right\} = \frac{1}{\max_{1 \leq i \leq p-1} \left(\|\delta_{J_i}\|_\infty \|\delta^{-1}_{J_{i+1}}\|_\infty \right)} \geq 1,
$$

with the convention that $\mathrm{gap}(J) = \infty$ if $p = 1$.

The LLS step used by our algorithm is computed with respect to a specific partition which is ordered at the current iterate $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$. We now describe the construction of this ordered partition. First, with the aid of the AS direction at w, we compute the bipartition (B, N) of $\{1, \ldots, n\}$ according to

$$
(21) \qquad B \equiv \{i : |Rs^{\mathrm{a}}_i| \leq |Rx^{\mathrm{a}}_i|\}, \qquad N \equiv \{i : |Rs^{\mathrm{a}}_i| > |Rx^{\mathrm{a}}_i|\}.
$$

Note that this definition and (18) imply that

$$
(22) \qquad \varepsilon^{\mathrm{a}}_\infty = \max \{\|Rx^{\mathrm{a}}_N\|_\infty, \|Rs^{\mathrm{a}}_B\|_\infty\}.
$$

Next, an order $(i_1, \ldots, i_n)$ of the index variables is chosen such that $\delta_{i_1} \leq \cdots \leq \delta_{i_n}$. Then the first block of consecutive indices in the n-tuple $(i_1, \ldots, i_n)$ lying in the same set B or N are placed in the first layer $\mathcal{J}_1$, the next block of consecutive indices lying in the other set is placed in $\mathcal{J}_2$, and so on. As an example, assume that $(i_1, i_2, i_3, i_4, i_5, i_6, i_7) \in B \times B \times N \times B \times B \times N \times N$. In this case, we have $\mathcal{J}_1 = \{i_1, i_2\}$, $\mathcal{J}_2 = \{i_3\}$, $\mathcal{J}_3 = \{i_4, i_5\}$, and $\mathcal{J}_4 = \{i_6, i_7\}$. A partition obtained according to the above construction is clearly ordered at w. We refer to it as an *ordered* (B, N)-*partition* and denote it by $\mathcal{J} = \mathcal{J}(w)$. The LLS step with respect to an ordered (B, N)-partition is sometimes used as a replacement for the primal-dual AS direction in the predictor step of our algorithm.

Note that an ordered (B, N)-partition is not uniquely determined since there can be more than one n-tuple $(i_1, \ldots, i_n)$ satisfying $\delta_{i_1} \leq \cdots \leq \delta_{i_n}$. This situation happens exactly when there are two or more indices i with the same value for δ_i. If these tying indices do not all belong to the same set B or N, then there will be more than one way to generate an ordered (B, N)-partition $\mathcal{J}$.

We say that the bipartition (B, N) is *regular* if there do not exist $i \in B$ and $j \in N$ such that $\delta_i = \delta_j$. Observe that there exists a unique ordered (B, N)-partition if and only if (B, N) is regular. When (B, N) is not regular, our algorithm avoids the computation of an ordered (B, N)-partition and hence of any LLS direction with respect to such a partition. Thus, there is no ambiguity in our algorithm.

2.5. Algorithm and the main convergence result. In this subsection, we describe our algorithm and state the main result of this paper which guarantees the convergence of the method in a strong sense. More specifically, we establish an iteration-complexity bound for our method which depends only on the constraint matrix A. This bound is exactly the same as the one obtained in Vavasis and Ye [26].

P-C LAYERED ALGORITHM.

Let $0 < \beta \leq 1/4$, $\varepsilon_0 > 0$, and $w^0 \in \mathcal{N}(\beta)$ be given. Set $k = 0$.

1. Set $w = w^k$ and compute the AS direction $(\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ at w;
2. Compute the quantities $\varepsilon_\infty^{\mathrm{a}}$ and α_{a} as in (18) and (12), and the bipartition (B, N) according to (21);
3. If $\varepsilon_\infty^{\mathrm{a}} > \varepsilon_0$ or (B, N) is not regular, then set $w \leftarrow w + \alpha_{\mathrm{a}} \Delta w^{\mathrm{a}}$ and go to step 7;
4. Otherwise, determine the ordered (B, N)-partition $\mathcal{J} = (\mathcal{J}_1, \ldots, \mathcal{J}_r)$ and compute the LLS step $\Delta w^{\mathrm{l}} = (\Delta x^{\mathrm{l}}, \Delta y^{\mathrm{l}}, \Delta s^{\mathrm{l}})$ at w with respect to $\mathcal{J}$;
5. Let $w^{\mathrm{l}} = w + \alpha_{\mathrm{l}} \Delta w^{\mathrm{l}}$, where $\alpha_{\mathrm{l}} \equiv \sup\{\alpha \in [0, 1] : w + \alpha \Delta w^{\mathrm{l}} \in \mathcal{N}(2\beta)\}$;
6. If $\mu(w^{\mathrm{l}}) < (1 - \alpha_{\mathrm{a}})\mu$, then set $w \leftarrow w^{\mathrm{l}}$, else set $w \leftarrow w + \alpha_{\mathrm{a}} \Delta w^{\mathrm{a}}$;
7. If $\mu(w) = 0$, then **stop**; in this case w is an optimal solution;
8. Compute the corrector step Δw^{c} at w and set $w \leftarrow w + \Delta w^{\mathrm{c}}$;
9. Set $w^{k+1} = w$, increment k by 1 and go to step 1.

End

We now make a few comments about the above algorithm. Step 2 followed by step 8 is a standard P-C iteration of the type described in subsection 2.3. This iteration is always performed in those iterations for which $\varepsilon_\infty^{\mathrm{a}} > \varepsilon_0$ or (B, N) is not regular. In the other iterations, the algorithm performs either a standard P-C iteration or a layered-corrector iteration, depending on which of the two iterations gives the lowest reduction of the duality gap. This test is performed in step 6 since the term $(1 - \alpha_{\mathrm{a}})\mu$ is the normalized duality gap obtained when the AS step is taken (see Proposition 2.4(a)).

The following convergence theorem is the main result of the paper.

THEOREM 2.7. *The P-C layered algorithm described above finds a primal-dual optimal solution $(x^\infty, s^\infty, y^\infty)$ of problems (1) and (2) satisfying strict complementarity (i.e., $x^\infty + s^\infty > 0$) in at most $\mathcal{O}(n^{3.5} \log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ iterations. In particular, if $\varepsilon_0 = \Omega(1/n^\tau)$ for some constant τ, then the iteration-complexity bound reduces to $\mathcal{O}(n^{3.5} \log(\bar{\chi}_A + n))$.*

3. Basic tools. In this section we introduce the basic tools that will be used in the proof of Theorem 2.7. The analysis heavily relies on the notion of crossover events due to Vavasis and Ye [26]. Subsection 3.1 below gives the definition of a crossover event which is slightly different than the one used in [26] and discusses some of its properties. In subsection 3.2, we state an approximation result that provides an estimation of the closeness between the LLS direction with respect to a partition J of $\{1, \ldots, n\}$ and the AS direction. Subsection 3.3 reviews from a different perspective an important result from [26], namely Lemma 17 of [26], that essentially guarantees the occurrence of crossover events. Since this result is stated in terms of the residual of an LLS step, the use of the approximation result of subsection 3.2 between the AS and LLS steps allows us to obtain a similar result stated in terms of the residual of the AS direction.

3.1. Crossover events. In this subsection we discuss the notion of crossover event which plays a fundamental role in our convergence analysis.

Definition. For two indices $i, j \in \{1, \ldots, n\}$ and a constant $\mathcal{C} \geq 1$, a $\mathcal{C}$-*crossover event* for the pair (i, j) is said to occur on the interval $(\nu', \nu]$ if

(23)
$$\text{there exists } \nu_0 \in (\nu', \nu] \text{ such that } \frac{s_j(\nu_0)}{s_i(\nu_0)} \leq \mathcal{C}$$
$$\text{and } \frac{s_j(\tilde{\nu})}{s_i(\tilde{\nu})} > \mathcal{C} \quad \forall \tilde{\nu} \leq \nu'.$$

Moreover, the interval $(\nu', \nu]$ is said to contain a $\mathcal{C}$-crossover event if (23) holds for some pair (i, j).

Hence, the notion of a crossover event is independent of any algorithm and is a property of the central path only. Note that in view of (3), condition (23) can be reformulated into an equivalent condition involving only the primal variable. For our purposes, we will use only (23).

We have the following simple but crucial result about crossover events.

PROPOSITION 3.1. *Let $\mathcal{C} > 0$ be a given constant. There can be at most $n(n-1)/2$ disjoint intervals of the form $(\nu', \nu]$ containing $\mathcal{C}$-crossover events.*

The notion of $\mathcal{C}$-crossover events can be used to define the notion of $\mathcal{C}$-crossover events between two iterates of the P-C layered algorithm as follows. We say that a $\mathcal{C}$-crossover event occurs between two iterates w^k and w^l, $k < l$, generated by the P-C layered algorithm if the interval $(\mu(w^l), \mu(w^k)]$ contains a $\mathcal{C}$-crossover event. Note that in view of Proposition 3.1, there can be at most $n(n-1)/2$ intervals of this type. We will show in the remainder of this paper that there exists a constant $\mathcal{C} > 0$ with the following property: for any index k, there exists an index $l > k$ such that $l - k = \mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ and a $\mathcal{C}$-crossover event occurs between the iterates w^k and w^l of the P-C layered algorithm. Proposition 3.1 and a simple argument then show that the P-C layered algorithm must terminate within $\mathcal{O}(n^{3.5} \log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ iterations.

3.2. Relation between the LLS and AS directions. In this subsection, we describe how the LLS step provides a good approximation of the AS direction, a result that will be important in our convergence analysis. Another result along this direction has also been obtained by Vavasis and Ye [28]. However, our result is more general and better suited for the development of this paper.

The approximation result below can be proved using the projection decomposition techniques developed in [22]. However, we give a simpler proof using instead the techniques developed in [15]. The result essentially states that the larger the gap of J is, the closer the AS direction and the LLS direction with respect to J will be to one another.

THEOREM 3.2. *Let $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ and an ordered partition $J = (J_1, \ldots, J_p)$ at w be given. Define $\delta \equiv \delta(w)$, and let $\Delta w^{\mathrm{a}} = (\Delta x^{\mathrm{a}}, \Delta y^{\mathrm{a}}, \Delta s^{\mathrm{a}})$ and $\Delta w^{\mathrm{ll}} = (\Delta x^{\mathrm{ll}}, \Delta y^{\mathrm{ll}}, \Delta s^{\mathrm{ll}})$ denote the AS direction at w and the LLS direction at w with respect to J, respectively. If the gap of J satisfies $\mathrm{gap}(J) \geq 4p\,\bar{\chi}_A$, then*

$$\max\left\{ \left\| \delta(\Delta x^{\mathrm{a}} - \Delta x^{\mathrm{ll}}) \right\|_\infty, \left\| \delta^{-1}(\Delta s^{\mathrm{a}} - \Delta s^{\mathrm{ll}}) \right\|_\infty \right\} \leq \frac{12\sqrt{n\mu}\,\bar{\chi}_A}{\mathrm{gap}(J)}.$$

In particular, if $(Rx^{\mathrm{ll}}, Rs^{\mathrm{ll}})$ denote the residual for the LLS direction Δw^{ll}, then

$$\max\left\{ \left\| Rx^{\mathrm{a}} - Rx^{\mathrm{ll}} \right\|_\infty, \left\| Rs^{\mathrm{a}} - Rs^{\mathrm{ll}} \right\|_\infty \right\} \leq \frac{12\sqrt{n}\,\bar{\chi}_A}{\mathrm{gap}(J)}.$$

Proof. Using the characterization (13) of Δx^{a} and the definition (19) of Δx^{ll}, we see that the vectors $d^0 = (d_1^0, \ldots, d_p^0) \equiv (\delta_{J_p} \Delta x_{J_p}^{\mathrm{a}}, \ldots, \delta_{J_1} \Delta x_{J_1}^{\mathrm{a}})$ and $\tilde{d}^0 = (\tilde{d}_1^0, \ldots, \tilde{d}_p^0)$ $\equiv (\delta_{J_p} \Delta x_{J_p}^{\mathrm{ll}}, \ldots, \delta_{J_1} \Delta x_{J_1}^{\mathrm{ll}})$ satisfy the assumptions of Theorem 6.1 with $g = 0$, F_{p+1-i} $= A_{J_i}$, $h_{p+1-i} = (\delta x)_{J_i} = (x^{1/2} s^{1/2})_{J_i}$, and $z_{p+1-i} = \delta_{J_i}^{-1}$ for all $i = 1, \ldots, p$. Hence, by the conclusion of Theorem 6.1, we conclude that

$$\left\| \delta (\Delta x^{\mathrm{a}} - \Delta x^{\mathrm{ll}}) \right\|_\infty \leq \frac{12 \, \bar{\chi}_F \left\| x^{1/2} s^{1/2} \right\|}{\mathrm{gap}(J)} = \frac{12 \, \bar{\chi}_A \sqrt{n\mu}}{\mathrm{gap}(J)}.$$

Now let G be an $(n - m) \times n$ full row rank matrix such that $AG^T = 0$. Clearly, we have $\mathrm{Ker}(A) = \mathrm{Im}(G^T)$, and hence $\bar{\chi}_A = \bar{\chi}_G$ in view of Proposition 2.2(d). Using the characterization (14) of Δs^{a} and the definition (20) of Δs^{ll}, we see that the vectors $d^0 = (d_1^0, \ldots, d_p^0) \equiv (\delta_{J_1}^{-1} \Delta s_{J_1}^{\mathrm{a}}, \ldots, \delta_{J_p}^{-1} \Delta s_{J_p}^{\mathrm{a}})$ and $\tilde{d}^0 = (\tilde{d}_1^0, \ldots, \tilde{d}_p^0) \equiv$ $(\delta_{J_1}^{-1} \Delta s_{J_1}^{\mathrm{ll}}, \ldots, \delta_{J_p}^{-1} \Delta s_{J_p}^{\mathrm{ll}})$ satisfy the assumptions of Theorem 6.1 with $g = 0$, $G_i = F_{J_i}$, $h_i = (\delta^{-1} s)_{J_i} = (x^{1/2} s^{1/2})_{J_i}$, and $z_i = \delta_{J_i}$ for all $i = 1, \ldots, p$. Hence, by the conclusion of Theorem 6.1, we conclude that

$$\left\| \delta^{-1} (\Delta s^{\mathrm{a}} - \Delta s^{\mathrm{ll}}) \right\|_\infty \leq \frac{12 \, \bar{\chi}_F \left\| x^{1/2} s^{1/2} \right\|}{\mathrm{gap}(J)} = \frac{12 \, \bar{\chi}_A \sqrt{n\mu}}{\mathrm{gap}(J)}.$$

Hence, the first inequality of the theorem follows. The second inequality follows immediately from the first one and the definition of residual of a direction $(\Delta x, \Delta y, \Delta s)$. $\quad\square$

In view of the above result, the AS direction can be well approximated by LLS directions with respect to ordered partitions J which have large gaps. The LLS direction with $p = 1$, which is the AS direction, provides the perfect approximation to the AS direction itself. However, this kind of trivial approximation is not useful for us due to the need of keeping the "spread" of some layers J_k under control. For an ordered partition J at w, the spread of the layer J_k, denoted by $\mathrm{spr}(J_k)$, is defined as

$$\mathrm{spr}(J_k) \equiv \frac{\max(\delta_{J_k})}{\min(\delta_{J_k})} \quad \forall k = 1, \ldots, p.$$

We now describe a special ordered partition introduced by Vavasis and Ye [26] which plays a crucial role in our analysis. Given a point $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ and a parameter $\bar{g} \geq 1$, this partition, which we refer to as the *VY $\bar{g}$-partition* at w, is defined as follows. Let $(i_1, \ldots, i_n)$ be an ordering of $\{1, \ldots, n\}$ such that $\delta_{i_1} \leq \cdots \leq \delta_{i_n}$, where $\delta = \delta(w)$. For $k = 2, \ldots, n$, let $r_k \equiv \delta_{i_k} / \delta_{i_{k-1}}$ and define $r_1 \equiv \infty$. Let $k_1 < \cdots < k_p$ be all the indices k such that $r_k > \bar{g}$ for all $j = 1, \ldots, p$. The VY $\bar{g}$-partition J is then defined as $J = (J_1, \ldots, J_p)$, where $J_q \equiv \{i_{k_q}, i_{k_q+1}, \ldots, i_{k_{q+1}-1}\}$ for all $q = 1, \ldots, p$. More generally, given a subset $I \subset \{1, \ldots, n\}$, we can similarly define the *VY $\bar{g}$-partition* of I at w by taking an ordering $(i_1, \ldots, i_m)$ of I satisfying $\delta_{i_1} \leq \cdots \leq \delta_{i_m}$, where $m = |I|$, defining the ratios $r_1, \ldots, r_m$ as above, and proceeding exactly as in the construction above to obtain the partition $J = (J_1, \ldots, J_p)$ of I.

It is easy to see that the following result holds for the partition J described in the previous paragraph.

PROPOSITION 3.3. *Given a subset $I \subseteq \{1, \ldots, n\}$, a point $w \in \mathcal{P}^{++} \times \mathcal{D}^{++}$, and a constant $\bar{g} \geq 1$, the VY $\bar{g}$-partition $J = (J_1, \ldots, J_p)$ of I at w satisfies $\mathrm{gap}(J) > \bar{g}$ and $\mathrm{spr}(J_q) \leq \bar{g}^{|J_q|} \leq \bar{g}^n$ for all $q = 1, \ldots, p$.*

3.3. Relation between crossover events and search directions. Using Lemma 17 of [26], we derive in this section an upper bound on the number of iterations needed to guarantee the occurrence of a crossover event which depends on the size of the residual of the LLS step and the stepsize at the initial iterate. Under suitable conditions, we derive with the aid of Theorem 3.2 another upper bound on the number of iterations needed to guarantee the occurrence of a crossover event which depends only on the size of the residual of the AS direction at the initial iterate.

Even though Lemma 17 of Vavasis and Ye [26] is stated and proved in a very advanced stage of their paper, one does not need to go through the whole material preceding it. In order to fully understand this result, it is recommended that one read only the material of section 4 of [26], followed by Lemma 16 and finally Lemma 17.

LEMMA 3.4. *Let $w = (x, y, s) \in \mathcal{N}(\beta)$ for some $\beta \in (0, 1)$ and an ordered partition $J = (J_1, \ldots, J_p)$ at w be given. Let $\delta \equiv \delta(w)$, $\mu = \mu(w)$, and (Rx^{ll}, Rs^{ll}) denote the residual of the LLS direction $(\Delta x^{ll}, \Delta y^{ll}, \Delta s^{ll})$ at w with respect to J. Then the following statements hold for every $q = 1, \ldots, p$:*

(a) *There exists $i \in J_1 \cup \cdots \cup J_q$ such that*

$$s_i(\nu) \geq \frac{\sqrt{\mu} \, \|Rs^{ll}_{J_q}\|_\infty \min(\delta_{J_q})}{n^{1.5}\bar{\chi}_A} \quad \forall \nu \in (0, \mu].$$

(b) *There exists $j \in J_q \cup \cdots \cup J_p$ such that*

$$x_j(\nu) \geq \frac{\sqrt{\mu} \, \|Rx^{ll}_{J_q}\|_\infty}{n^{1.5}\bar{\chi}_A \max(\delta_{J_q})} \quad \forall \nu \in (0, \mu].$$

(c) *For any $\mathcal{C}_q \geq (1 + \beta)\operatorname{spr}(J_q)/(1 - \beta)^2$ and for any $\mu' \in (0, \mu)$ such that*

$$\frac{\mu'}{\mu} \leq \frac{\|Rx^{ll}_{J_q}\|_\infty\|Rs^{ll}_{J_q}\|_\infty}{n^3\mathcal{C}_q^2\bar{\chi}_A^2},$$

the interval $(\mu', \mu]$ contains a $\mathcal{C}_q$-crossover event.

Proof. Noting that our definition of δ is the one used in [26] divided by $\sqrt{\mu}$, we easily see that statements (a) and (b) follow directly from Lemma 17 of [26]. We now prove (c). Let i and j be as in statements (a) and (b). First note that by Proposition 2.1 we have

$$\frac{s_i(\mu)}{s_j(\mu)} \leq \frac{1 + \beta}{(1 - \beta)^2}\frac{\delta_i}{\delta_j} \leq \frac{1 + \beta}{(1 - \beta)^2}\frac{\max(\delta_{J_q})}{\min(\delta_{J_q})} = \frac{1 + \beta}{(1 - \beta)^2}\operatorname{spr}(J_q) \leq \mathcal{C}_q.$$

Now, by (b) and (3), we have

$$\frac{1}{s_j(\nu)} \geq \frac{\sqrt{\mu} \, \|Rx^{ll}_{J_q}\|_\infty}{\nu n^{1.5}\bar{\chi}_A \max(\delta_{J_q})} \quad \forall \nu \in (0, \mu].$$

Using the last relation, the relation in (a), the fact that J is an ordered partition for w, and the conditions on $\mathcal{C}_q$ and μ', we obtain for every $\nu \in (0, \mu']$ that

$$\frac{s_i(\nu)}{s_j(\nu)} \geq \frac{\mu \, \|Rx^{ll}_{J_q}\|_\infty\|Rs^{ll}_{J_q}\|_\infty}{\nu \, n^3\bar{\chi}_A^2 \operatorname{spr}(J_q)} > \frac{\mu \, \|Rx^{ll}_{J_q}\|_\infty\|Rs^{ll}_{J_q}\|_\infty}{\mu' \, n^3\bar{\chi}_A^2 \, \mathcal{C}_q} \geq \mathcal{C}_q.$$

We have thus shown that a crossover event for the pair (i, j) occurs in the interval $(\nu', \nu]$. □

An immediate consequence of Lemma 3.4(c) which has implications in the analysis of the P-C layered algorithm is as follows.

LEMMA 3.5. *Let $w = (x, y, s) \in \mathcal{N}(\beta)$ for some $\beta \in (0, 1/4]$ and an ordered partition $J = (J_1, \ldots, J_p)$ at w be given. Define $\delta \equiv \delta(w)$ and $\mu = \mu(w)$, and let $(Rx^{\mathrm{ll}}, Rs^{\mathrm{ll}})$ denote the residual of the LLS direction $(\Delta x^{\mathrm{ll}}, \Delta y^{\mathrm{ll}}, \Delta s^{\mathrm{ll}})$ at w with respect to J. Then, for every $q \in \{1, \ldots, p\}$ and every $\mathcal{C}_q \geq (1 + \beta)\mathrm{spr}(J_q)/(1 - \beta)^2$, the following statements hold:*

(a) *The P-C layered algorithm started from the point w will either generate an iterate $\hat{w}$ with a $\mathcal{C}_q$-crossover event occurring between w and $\hat{w}$ or terminate in*

$$(24) \quad \mathcal{O}\left(\sqrt{n}\left(\log(\bar{\chi}_A + n) + \log \mathcal{C}_q + \log\left(\frac{\mu_+/\mu}{\|Rx^{\mathrm{ll}}_{J_q}\|_\infty \|Rs^{\mathrm{ll}}_{J_q}\|_\infty}\right)\right)\right)$$

iterations, where μ_+ is the normalized duality gap attained immediately after the first iteration.

(b) *If, in addition,*

$$(25) \qquad \mathrm{gap}(J) \geq \max\left\{4\, n\bar{\chi}_A\,, \frac{24\sqrt{n}\bar{\chi}_A}{\varepsilon^{\mathrm{a}}_{J_q}}\right\},$$

where $\varepsilon^{\mathrm{a}}_{J_q} \equiv \min\{\|Rx^{\mathrm{a}}_{J_q}\|_\infty\,, \|Rs^{\mathrm{a}}_{J_q}\|_\infty\}$, then (24) is bounded above by

$$(26) \qquad \mathcal{O}\left(\sqrt{n}\left(\log(\bar{\chi}_A + n) + \log \mathcal{C}_q + \log(\varepsilon^{\mathrm{a}}_{J_q})^{-1}\right)\right).$$

Proof. To prove (a), it is sufficient to show that a $\mathcal{C}_q$-crossover event will occur if the algorithm does not terminate in a number of iterations bounded above by (24). Lemma 3.4(c) guarantees that a $\mathcal{C}_q$-crossover event occurs between w and another iterate $\hat{w}$ whenever

$$(27) \qquad\qquad \frac{\mu(\hat{w})}{\mu(w)} \leq \frac{\|Rx^{\mathrm{ll}}_{J_q}\|_\infty \|Rs^{\mathrm{ll}}_{J_q}\|_\infty}{n^3 \mathcal{C}_q^2 \bar{\chi}_A^2}.$$

Observe that the duality gap is reduced by a factor of μ_+/μ in the first iteration and by a factor of at least $1 - \sqrt{\beta/n}$ in subsequent iterations due to Proposition 2.4(b). Thus, an iterate $\hat{w}$ satisfying (27) will be generated in at most $N_0 + 1$ iterations, where N_0 is the smallest integer satisfying

$$\log\left(\frac{\mu_+}{\mu}\right) + N_0 \log\left(1 - \sqrt{\frac{\beta}{n}}\right) \leq \log\left[\frac{\|Rx^{\mathrm{ll}}_{J_q}\|_\infty \|Rs^{\mathrm{ll}}_{J_q}\|_\infty}{n^3 \mathcal{C}_q^2 \bar{\chi}_A^2}\right].$$

The first part of the lemma now immediately follows by rearranging this inequality and using the fact that $\log(1 + x) < x$ for any $x > -1$.

We now prove (b). We will show that (24) is bounded above by (26) when (25) holds. By Theorem 3.2 and (25), it follows that

$$\max\left\{\left\|Rx^{\mathrm{a}} - Rx^{\mathrm{ll}}\right\|_\infty, \left\|Rs^{\mathrm{a}} - Rs^{\mathrm{ll}}\right\|_\infty\right\} \leq \frac{12\sqrt{n}\,\bar{\chi}_A}{\mathrm{gap}(J)} \leq \frac{\varepsilon^{\mathrm{a}}_{J_q}}{2}.$$

Hence, we have

$$\min\left\{\|Rx_{J_q}^{\mathrm{ll}}\|_\infty, \|Rs_{J_q}^{\mathrm{ll}}\|_\infty\right\}$$
$$\geq \min\left\{\left\|Rx_{J_q}^{\mathrm{a}}\right\|_\infty - \left\|Rx^{\mathrm{a}} - Rx^{\mathrm{ll}}\right\|_\infty, \left\|Rs_{J_q}^{\mathrm{a}}\right\|_\infty - \left\|Rs^{\mathrm{a}} - Rs^{\mathrm{ll}}\right\|_\infty\right\}$$
$$\geq \min\left\{\left\|Rx_{J_q}^{\mathrm{a}}\right\|_\infty, \left\|Rs_{J_q}^{\mathrm{a}}\right\|_\infty\right\} - \frac{\varepsilon_{J_q}^{\mathrm{a}}}{2} = \varepsilon_{J_q}^{\mathrm{a}} - \frac{\varepsilon_{J_q}^{\mathrm{a}}}{2} = \frac{\varepsilon_{J_q}^{\mathrm{a}}}{2}.$$

Using this estimate in (24) together with the fact that $\mu_+/\mu \leq 1$, we conclude that (24) is bounded above by (26). $\quad\square$

4. Convergence analysis of the P-C layered algorithm. In this section, we give the proof of Theorem 2.7.

Lemma 3.5 gives a good idea of the effort which will be undertaken in this section, namely, to show that for each $w \in \mathcal{N}(\beta)$ there exist an ordered partition $J = (J_1, \ldots, J_p)$ and an index $q = 1, \ldots, p$ such that the sum of two last logarithms in (24) can be bounded above by $\mathcal{O}(n\log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$. The analysis of this claim will be broken into two cases, namely (i) $\varepsilon_\infty^{\mathrm{a}} \geq \varepsilon_0$ and (ii) $\varepsilon_\infty^{\mathrm{a}} \leq \varepsilon_0$, where $\varepsilon_\infty^{\mathrm{a}}$ is given by (18). The first result below considers the case $\varepsilon_\infty^{\mathrm{a}} \geq \varepsilon_0$ for which the VY $\bar{g}$-partition at w is quite suitable. We introduce the following global constants which will be used in the remainder of this paper:

$$(28) \qquad \mathcal{C} \equiv \frac{(1+\beta)}{(1-\beta)^2}\,\bar{g}^n, \qquad \bar{g} \equiv 24\,\bar{\chi}_A\,\sqrt{n}\,\max\left\{\varepsilon_0^{-1}, \frac{4(1+2\beta)\sqrt{n}}{\beta - 2\beta^2}\right\}.$$

LEMMA 4.1. *Suppose that* $w = (x, y, s) \in \mathcal{N}(\beta)$ *for some* $\beta \in (0, 1/4]$ *and that* $\varepsilon_\infty^{\mathrm{a}} \geq \varepsilon_0$ *for some constant* $\varepsilon_0 > 0$. *Then the P-C layered algorithm started from the point* w *will either generate an iterate* $\hat{w}$ *with a* $\mathcal{C}$-*crossover event occurring between* w *and* $\hat{w}$ *or terminate in* $\mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ *iterations.*

Proof. The assumption that $\varepsilon_\infty^{\mathrm{a}} \geq \varepsilon_0$ and definition (18) imply the existence of an index $i = 1, \ldots, n$ such that $\min\{|Rx_i^{\mathrm{a}}|, |Rs_i^{\mathrm{a}}|\} \geq \varepsilon_0$. Now let $J = (J_1, \ldots, J_p)$ be a VY $\bar{g}$-partition at w, and let J_q be the layer containing the index i above. Clearly, we have

$$(29) \qquad\qquad \varepsilon_{J_q}^{\mathrm{a}} \equiv \min\{\|Rx_{J_q}^{\mathrm{a}}\|_\infty, \|Rs_{J_q}^{\mathrm{a}}\|_\infty\} \geq \varepsilon_0.$$

Using the above inequality, the fact that $\mathrm{gap}(J) \geq \bar{g}$, and (28), we easily see that (25) holds. Since by Proposition 3.3 the spread of every layer of a VY $\bar{g}$-partition at w is bounded above by $\bar{g}^n$, we conclude that $\mathrm{spr}(J_q) \leq \bar{g}^n$. Hence, we may set $\mathcal{C}_q = \mathcal{C} \equiv (1+\beta)\bar{g}^n/(1-\beta)^2$ in Lemma 3.5, from which it follows that

$$(30) \qquad\qquad \log(\mathcal{C}_q) = \mathcal{O}(n\log\bar{g}) = \mathcal{O}(n\log(\bar{\chi}_A + n + \varepsilon_0^{-1})),$$

where the last equality is due to (28). The result now follows from Lemma 3.5(b) by noting that (26) is $\mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ in view of (29) and (30). $\quad\square$

We now consider the case in which $\varepsilon_\infty^{\mathrm{a}} \leq \varepsilon_0$ and show that a $\mathcal{C}$-crossover also happens within $\mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ iterations of the P-C layered algorithm (if it does not terminate). From now on, we let $\mathcal{J} = (\mathcal{J}_1, \ldots, \mathcal{J}_r)$ denote an ordered (B, N)-partition at w. We will split the analysis of this case into two subcases, namely (i) $\mathrm{gap}(\mathcal{J}) \leq \bar{g}$ and (ii) $\mathrm{gap}(\mathcal{J}) \geq \bar{g}$. The next result takes care of the case in which $\mathrm{gap}(\mathcal{J}) \leq \bar{g}$, without assuming anything about $\varepsilon_\infty^{\mathrm{a}}$.

LEMMA 4.2. *Suppose that* $w = (x, y, s) \in \mathcal{N}(\beta)$ *for some* $\beta \in (0, 1/4]$. *Let* $\bar{g}$ *and* $\mathcal{C}$ *be the constants defined in (28). Let* $\mathcal{J} = (\mathcal{J}_1, \ldots, \mathcal{J}_r)$ *be an ordered* (B, N)-*partition*

at w, where (B, N) is the bipartition defined in (21), and assume that $\mathrm{gap}(\mathcal{J}) < \bar{g}$. Then the P-C layered algorithm started from the point w will either generate an iterate $\hat{w}$ with a $\mathcal{C}$-crossover event occurring between w and $\hat{w}$ or terminate in $\mathcal{O}\left(n^{1.5} \log(\bar{\chi}_A + n + \varepsilon_0^{-1})\right)$ iterations.

Proof. Assume that $\mathrm{gap}(\mathcal{J}) < \bar{g}$. Let $J = (J_1, \ldots, J_p)$ be a VY $\bar{g}$-partition at w. Using the assumption that $\mathrm{gap}(\mathcal{J}) < \bar{g}$, it is easy to see that there exist two indices i, j of different types, say $i \in B$ and $j \in N$, both lying in some layer J_q of J. By Lemma 2.6 and the definition of (B, N) given in (21), it follows that $|Rx_i^\mathrm{a}| \geq 1/4$ and $|Rs_j^\mathrm{a}| \geq 1/4$, and hence that

$$(31) \qquad \varepsilon_{J_q}^\mathrm{a} \equiv \min\{\|Rx_{J_q}^\mathrm{a}\|_\infty, \|Rs_{J_q}^\mathrm{a}\|_\infty\} \geq \frac{1}{4}.$$

Using this inequality and the fact that $\mathrm{gap}(J) \geq \bar{g} \geq 96\bar{\chi}_A n$, where the last inequality is due to (28), we easily see that (25) holds. Since by Proposition 3.3 the spread of every layer of a VY $\bar{g}$-partition at w is bounded above by $\bar{g}^n$, we conclude that $\mathrm{spr}(J_q) \leq \bar{g}^n$. Hence, we may set $\mathcal{C}_q = \mathcal{C} \equiv (1 + \beta)\bar{g}^n/(1 - \beta)^2$ in Lemma 3.5, from which it follows that (30) holds. The result now follows from Lemma 3.5(b) by noting that (26) is $\mathcal{O}(n^{1.5} \log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ in view of (30) and (31). $\square$

The next result considers the case in which $\mathrm{gap}(\mathcal{J}) \geq \bar{g}$ and derives an upper bound on the number of iterations for a $\mathcal{C}$-crossover event to occur. As in Lemma 3.5, nothing is assumed about $\varepsilon_\infty^\mathrm{a}$.

LEMMA 4.3. *Suppose that* $w = (x, y, s) \in \mathcal{N}(\beta)$ *for some* $\beta \in (0, 1/4]$. *Let* $\bar{g}$ *and* $\mathcal{C}$ *be the constants defined in* (28). *Let* $\mathcal{J} = (\mathcal{J}_1, \ldots, \mathcal{J}_r)$ *be the* (B, N)-partition at w, *where* (B, N) *is the bipartition defined in* (21), *and assume that* $\mathrm{gap}(\mathcal{J}) \geq \bar{g}$. *Let* $(Rx^\mathrm{l}, Rs^\mathrm{l})$ *denote the residual of the LLS direction at* w *with respect to* $\mathcal{J}$. *Then the P-C layered algorithm started from the point* w *will either generate an iterate* $\hat{w}$ *with a* $\mathcal{C}$-crossover event occurring between w and $\hat{w}$ or terminate in

$$(32) \qquad \mathcal{O}\left(n^{1.5} \log\left(\bar{\chi}_A + n + \varepsilon_0^{-1}\right) + \sqrt{n} \, \log\left(\frac{\mu_+/\mu}{\varepsilon_\infty^\mathrm{l}}\right)\right)$$

iterations, where μ_+ *is the normalized duality gap attained immediately after the first iteration, and*

$$\varepsilon_\infty^\mathrm{l} \equiv \max\left\{\left\|Rx_N^\mathrm{l}\right\|_\infty, \left\|Rs_B^\mathrm{l}\right\|_\infty\right\}.$$

Proof. Assume without loss of generality that $\varepsilon_\infty^\mathrm{l} = \|Rx_N^\mathrm{l}\|_\infty$; the case in which $\varepsilon_\infty^\mathrm{l} = \|Rs_B^\mathrm{l}\|_\infty$ can be proved similarly. Then $\varepsilon_\infty^\mathrm{l} = |Rx_i^\mathrm{l}|$ for some $i \in N$. Let $\mathcal{J}_t$ be the layer of $\mathcal{J}$ containing the index i and note that

$$(33) \qquad \varepsilon_\infty^\mathrm{l} = |Rx_i^\mathrm{l}| = \|Rx_{\mathcal{J}_t}^\mathrm{l}\|_\infty \leq \|Rx_{\mathcal{J}_t}^\mathrm{l}\|.$$

Now let $\mathcal{I} = (\mathcal{I}_1, \ldots, \mathcal{I}_p)$ be the VY $\bar{g}$-partition of $\mathcal{J}_t$ at w and consider the ordered partition $\mathcal{J}'$ defined as

$$\mathcal{J}' \equiv (\mathcal{J}_1, \ldots, \mathcal{J}_{t-1}, \mathcal{I}_1, \ldots, \mathcal{I}_p, \mathcal{J}_{t+1}, \ldots, \mathcal{J}_r).$$

Let $(Rx^{\mathrm{ll}}, Rs^{\mathrm{ll}})$ denote the residual of the LLS direction at w with respect to $\mathcal{J}'$. Using the definition of the LLS step, it is easy to see that $Rx_{\mathcal{J}_j}^\mathrm{l} = Rx_{\mathcal{J}_j}^{\mathrm{ll}}$ for all $j = t+1, \ldots, r$. Moreover, we have $\|Rx_{\mathcal{J}_t}^\mathrm{l}\| \leq \|Rx_{\mathcal{J}_t}^{\mathrm{ll}}\|$ since $\|Rx_{\mathcal{J}_t}^\mathrm{l}\|$ is the optimal

value of the least squares problem which determines the $\Delta x^l_{\mathcal{J}_t}$-component of the LLS step with respect to $\mathcal{J}$, whereas $\|Rx^{ll}_{\mathcal{J}_t}\|$ is the objective value at a certain feasible solution for the same least squares problem. Hence, for some $q \in \{1, \ldots, p\}$ we have

$$(34) \qquad \|Rx^{ll}_{\mathcal{I}_q}\|_\infty = \|Rx^{ll}_{\mathcal{J}_t}\|_\infty \geq \frac{1}{\sqrt{|\mathcal{J}_t|}} \|Rx^{ll}_{\mathcal{J}_t}\| \geq \frac{1}{\sqrt{n}} \|Rx^{ll}_{\mathcal{J}_t}\| \geq \frac{1}{\sqrt{n}} \|Rx^l_{\mathcal{J}_t}\|.$$

Combining (33) and (34), we then obtain

$$(35) \qquad\qquad\qquad\qquad \|Rx^{ll}_{\mathcal{I}_q}\|_\infty \geq \frac{1}{\sqrt{n}} \varepsilon^l_\infty.$$

Let us now bound the quantity $\|Rs^{ll}_{\mathcal{I}_q}\|_\infty$ from below. Using triangle inequality for norms, Lemma 2.6, Theorem 3.2, and the fact that $\mathrm{gap}(\mathcal{J}') \geq \bar{g} \geq 96\bar{\chi}_A n$, where the second inequality is due to (28), we obtain

$$(36) \quad \|Rs^{ll}_{\mathcal{I}_q}\|_\infty \geq \|Rs^a_{\mathcal{I}_q}\|_\infty - \|Rs^{ll}_{\mathcal{I}_q} - Rs^a_{\mathcal{I}_q}\|_\infty \geq \frac{1}{4} - \frac{12\sqrt{n}\,\bar{\chi}_A}{\mathrm{gap}(\mathcal{J}')} \geq \frac{1}{4} - \frac{1}{8} \geq \frac{1}{8}.$$

Also note that by (28) and Proposition 3.3 we have

$$(37) \qquad\qquad\qquad \mathcal{C} = \frac{1+\beta}{(1-\beta)^2} \bar{g}^n \geq \frac{1+\beta}{(1-\beta)^2} \mathrm{spr}(\mathcal{I}_q)$$

and

$$(38) \qquad\qquad\qquad \log \mathcal{C} = \mathcal{O}\left(n \log\left(\bar{\chi}_A + n + \varepsilon_0^{-1}\right)\right).$$

Hence, from Lemma 3.5(a) with $J = \mathcal{J}'$ and $\mathcal{C}_q = \mathcal{C}$ and the estimates (35)–(38), it follows that the P-C layered algorithm started from w will find an iterate $\hat{w}$ with a $\mathcal{C}$-crossover event occurring between w and $\hat{w}$ in

$$\mathcal{O}\left(n^{1.5} \log\left(\bar{\chi}_A + n + \varepsilon_0^{-1}\right) + \sqrt{n} \log\left(\frac{\mu_+/\mu}{\|Rx^l_{\mathcal{I}_q}\|_\infty \|Rs^l_{\mathcal{I}_q}\|_\infty}\right)\right)$$

$$= \mathcal{O}\left(n^{1.5} \log\left(\bar{\chi}_A + n + \varepsilon_0^{-1}\right) + \sqrt{n} \log\left(\frac{\mu_+/\mu}{\varepsilon^l_\infty}\right)\right)$$

iterations.   □

Our goal now will be to estimate the second logarithm that appears in the iteration-complexity bound (32). It is exactly in this estimation process that we will need to assume that $\varepsilon^a_\infty \leq \varepsilon_0$. Under this condition, we know that the duality gap reduction μ^+/μ obtained in the first iteration from w is the smaller between the two duality gap reductions obtained by taking an AS step and an LLS step. Hence, μ^+/μ is majorized by the duality gap reduction obtained from an LLS step from w. Lemma 4.6 below provides an estimation of the duality gap reduction obtained from an LLS step. The two lemmas that precede it, namely Lemmas 4.4 and 4.5, are just technical results which are used in its proof.

LEMMA 4.4. *Let $w = (x, y, s) \in \mathcal{P}^{++} \times \mathcal{D}^{++}$ be given and assume that $\|xs - \nu e\| \leq \tau\nu$ for some constants $\tau \in (0,1)$ and $\nu > 0$. Then $(1 - \tau/\sqrt{n})\nu \leq \mu(w) \leq (1 + \tau/\sqrt{n})\nu$ and $w \in \mathcal{N}(\tau/(1-\tau))$.*

Proof. We have

$$\left|\mu(w) - \nu\right| = \left|\frac{x^T s - n\nu}{n}\right| = \left|\frac{e^T(xs - \nu e)}{n}\right| \le \frac{\|e\| \, \|xs - \nu e\|}{n} \le \frac{\tau}{\sqrt{n}}\nu,$$

from which the two inequalities of the lemma follow. Since $\tilde{\nu} = \mu(w)$ is the constant which minimizes $\|xs - \tilde{\nu}e\|$, we have

$$\|xs - \mu(w)e\| \le \|xs - \nu e\| \le \tau\nu \le \frac{\tau}{1 - \tau/\sqrt{n}}\,\mu(w) \le \frac{\tau}{1 - \tau}\,\mu(w),$$

showing that $w \in \mathcal{N}(\tau/(1 - \tau))$. $\quad\square$

The following lemma is well known (see [4] or [8], for example).

LEMMA 4.5. *Let $\{w^k\} = \{(x^k, y^k, s^k)\}$ be a sequence of points in $\mathcal{P}^{++} \times \mathcal{D}^{++}$ such that $\lim_{k\to\infty} \mu_k = 0$ and, for some $\gamma > 0$, $x^k s^k \ge \gamma\mu_k e$ for all k, where $\mu_k \equiv \mu(w^k)$. Then every accumulation point $w^\infty = (x^\infty, y^\infty, s^\infty)$ of the sequence $\{w^k\}$ is a primal-dual optimal solution of (1) and (2) satisfying the strict complementarity condition, namely $(x^\infty)^T s^\infty = 0$ and $x^\infty + s^\infty > 0$.*

The following lemma gives an estimate of the duality gap reduction obtained by taking an LLS step.

LEMMA 4.6. *Suppose that $w \in \mathcal{N}(\beta)$ for some $\beta \in (0, 1/2)$. Let $J = (J_1, \ldots, J_p)$ be an ordered partition at w, and let $\Delta w^{ll} = (\Delta x^{ll}, \Delta y^{ll}, \Delta s^{ll})$ denote the LLS direction at w with respect to J. Define*

$$(39) \qquad \varepsilon_\infty^{ll} \equiv \max\left\{\left\|Rx_N^{ll}\right\|_\infty, \left\|Rs_B^{ll}\right\|_\infty\right\},$$

$$\alpha_{ll} \equiv \sup\left\{\alpha \in [0, 1] : w + \alpha\Delta w^{ll} \in \mathcal{N}(2\beta)\right\},$$

where (Rx^{ll}, Rs^{ll}) is the residual of Δw^{ll}. Then the following statements hold:

(a) *If $\mathrm{gap}(J) > \max\{4p\bar{\chi}_A, 24\sqrt{n}\bar{\chi}_A\}$, then $x^T\Delta s^{ll} + s^T\Delta x^{ll} < 0$, and hence $\mu(w + \alpha\Delta w^{ll})$ is a strictly decreasing affine function of α.*

(b) *If $\mathrm{gap}(J) \ge 96n\bar{\chi}_A/\eta$, where $\eta \equiv (\beta - 2\beta^2)/(1 + 2\beta)$, then*

$$\frac{\mu(w + \alpha_{ll}\Delta w^{ll})}{\mu(w)} \le \frac{4\sqrt{n}\,\varepsilon_\infty^{ll}(\varepsilon_\infty^{ll} + 4)}{\eta}.$$

Proof. We first show (a). From the first equation in (11), we easily see that $s^T\Delta x^a + x^T\Delta s^a = -n\mu$, where $\mu \equiv \mu(w)$. Using this fact, the definition of the residual of a direction, Theorem 3.2, and the assumption that $\mathrm{gap}(J) > \max\{4p\,\bar{\chi}_A, 24\sqrt{n}\bar{\chi}_A\}$, we obtain

$$s^T\Delta x^{ll} + x^T\Delta s^{ll} = s^T\Delta x^a + x^T\Delta s^a + s^T\left(\Delta x^{ll} - \Delta x^a\right) + x^T\left(\Delta s^{ll} - \Delta s^a\right)$$

$$= -n\mu + \mu\left(\frac{x^{1/2}s^{1/2}}{\sqrt{\mu}}\right)^T\left[\left(Rx^{ll} - Rx^a\right) + \left(Rs^{ll} - Rs^a\right)\right]$$

$$\le -n\mu + \mu\sqrt{n}\left\|\frac{x^{1/2}s^{1/2}}{\sqrt{\mu}}\right\|\left(\left\|Rx^{ll} - Rx^a\right\|_\infty + \left\|Rs^{ll} - Rs^a\right\|_\infty\right)$$

$$\le -n\mu + \mu n\frac{24\sqrt{n}\bar{\chi}_A}{\mathrm{gap}(J)} = -n\mu\left(1 - \frac{24\sqrt{n}\bar{\chi}_A}{\mathrm{gap}(J)}\right) < 0,$$

from which (a) follows.

To prove (b), assume that $\mathrm{gap}(J) \geq 96 n \bar{\chi}_A / \eta$. Define $v(\alpha) \equiv (x + \alpha \Delta x^{\mathrm{ll}})(s + \alpha \Delta s^{\mathrm{ll}})$ for all $\alpha \in \Re$. We claim that

$$(40) \quad \|v(\alpha) - (1-\alpha)\mu e\| \leq \frac{2\beta}{1+2\beta}(1-\alpha)\mu \text{ for every } 0 \leq \alpha \leq 1 - \frac{2\sqrt{n}\varepsilon_\infty^{\mathrm{ll}}(\varepsilon_\infty^{\mathrm{ll}} + 4)}{\eta}.$$

Using this claim, (b) can be proved as follows. By Lemma 4.4 with $w = w + \alpha \Delta w^{\mathrm{ll}}$, $\nu = (1-\alpha)\mu$, and $\tau = 2\beta/(1+2\beta)$ we conclude from the claim that for any $0 \leq \alpha \leq 1 - 2\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}(\varepsilon_\infty^{\mathrm{ll}} + 4)/\eta$, we have $w + \alpha \Delta w^{\mathrm{ll}} \in \mathcal{N}(2\beta)$ and

$$(41) \qquad \mu(w + \alpha \Delta w^{\mathrm{ll}}) \leq \left(1 + \frac{2\beta}{\sqrt{n}(1+2\beta)}\right)(1-\alpha)\mu \leq 2(1-\alpha)\mu.$$

By the definition of α_{ll}, we then conclude that $\alpha_{\mathrm{ll}} \geq \alpha_* \equiv 1 - 2\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}(\varepsilon_\infty^{\mathrm{ll}} + 4)/\eta$. Setting $\alpha = \alpha_*$ in (41) and using the fact that $\alpha_{\mathrm{ll}} \geq \alpha_*$ and $\mu(w + \alpha \Delta w^{\mathrm{ll}})$ is a decreasing function of α, we obtain

$$\mu(w + \alpha_{\mathrm{ll}} \Delta w^{\mathrm{ll}}) \leq \mu(w + \alpha_* \Delta w^{\mathrm{ll}}) \leq 2(1-\alpha_*)\mu = \frac{4\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}(\varepsilon_\infty^{\mathrm{ll}} + 4)}{\eta}\mu;$$

that is, (b) holds. In the remainder of the proof, we show that (40) holds. It is easy to see that

$$
\begin{aligned}
(42) \quad v(\alpha) - (1-\alpha)\mu e &= (x + \alpha \Delta x^{\mathrm{ll}})(s + \alpha \Delta s^{\mathrm{ll}}) - (1-\alpha)\mu e \\
&= (1-\alpha)(xs - \mu e) + \alpha h^1 + \alpha(1-\alpha)h^2 + \alpha^2 h^3,
\end{aligned}
$$

where h^1, h^2, and h^3 are vectors in $\Re^n$ defined as

$$(43) \qquad \begin{pmatrix} h_B^1 \\ h_N^1 \end{pmatrix} \equiv \begin{pmatrix} x_B(s_B + \Delta s_B^{\mathrm{ll}}) \\ s_N(x_N + \Delta x_N^{\mathrm{ll}}) \end{pmatrix} = \mu \begin{pmatrix} w_B p_B \\ w_N p_N \end{pmatrix},$$

$$(44) \qquad \begin{pmatrix} h_B^2 \\ h_N^2 \end{pmatrix} \equiv \begin{pmatrix} s_B \Delta x_B^{\mathrm{ll}} \\ x_N \Delta s_N^{\mathrm{ll}} \end{pmatrix} = \mu \begin{pmatrix} w_B q_B \\ w_N q_N \end{pmatrix},$$

$$(45) \qquad \begin{pmatrix} h_B^3 \\ h_N^3 \end{pmatrix} \equiv \begin{pmatrix} \Delta x_B^{\mathrm{ll}}(s_B + \Delta s_B^{\mathrm{ll}}) \\ \Delta s_N^{\mathrm{ll}}(x_N + \Delta x_N^{\mathrm{ll}}) \end{pmatrix} = \mu \begin{pmatrix} p_B q_B \\ p_N q_N \end{pmatrix}.$$

Here the vectors p, q, and w appearing in the second alternative expressions for h^1, h^2, and h^3 are defined as

$$\begin{pmatrix} p_B \\ p_N \end{pmatrix} \equiv \begin{pmatrix} Rs_B^{\mathrm{ll}} \\ Rx_N^{\mathrm{ll}} \end{pmatrix}, \qquad \begin{pmatrix} q_B \\ q_N \end{pmatrix} \equiv \begin{pmatrix} \delta_B \Delta x_B^{\mathrm{ll}}/\sqrt{\mu} \\ \delta_N^{-1}\Delta s_N^{\mathrm{ll}}/\sqrt{\mu} \end{pmatrix}, \qquad w \equiv \frac{x^{1/2}s^{1/2}}{\sqrt{\mu}}.$$

Clearly, we have

$$(46) \qquad \|p\|_\infty = \varepsilon_\infty^{\mathrm{ll}}, \qquad \|w\|_\infty \leq \sqrt{1+\beta} \leq 2, \qquad \|w\| = \sqrt{n}.$$

We will now derive an upper bound for $\|q\|$. Using the definition of $(Rx^{\mathrm{ll}}, Rs^{\mathrm{ll}})$ and (17), we obtain

$$\frac{\delta_B \Delta x_B^{\mathrm{ll}}}{\sqrt{\mu}} = Rx_B^{\mathrm{ll}} - w_B = -Rs_B^{\mathrm{ll}} + (Rx_B^{\mathrm{ll}} - Rx_B^{\mathrm{a}}) + (Rs_B^{\mathrm{ll}} - Rs_B^{\mathrm{a}})$$

and

$$\frac{\delta_N^{-1}\Delta s_N^{\mathrm{ll}}}{\sqrt{\mu}} = Rs_N^{\mathrm{ll}} - w_N = -Rx_N^{\mathrm{ll}} + (Rs_N^{\mathrm{ll}} - Rs_N^{\mathrm{a}}) + (Rx_N^{\mathrm{ll}} - Rx_N^{\mathrm{a}}),$$

from which it follows that

$$q = -p + (Rx^{\mathrm{ll}} - Rx^{\mathrm{a}}) + (Rs^{\mathrm{ll}} - Rs^{\mathrm{a}}).$$

Hence, using the triangle inequality for norms, Theorem 3.2, and the assumption that $\mathrm{gap}(J) \geq 96n\bar{\chi}_A/\eta \geq 4\,p\,\bar{\chi}_A$, we obtain

$$(47) \quad \|q\| \leq \|p\| + \|Rx^{\mathrm{ll}} - Rx^{\mathrm{a}}\| + \|Rs^{\mathrm{ll}} - Rs^{\mathrm{a}}\| \leq \sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}} + \frac{24n\bar{\chi}_A}{\mathrm{gap}(J)} \leq \sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}} + \frac{\eta}{4}.$$

Using (43), (44), (45), (46), and (47), we obtain

$$\begin{aligned}
\|h^1\| &\leq \mu\|w\|\,\|p\|_\infty \;\leq\; \mu\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}, \\
\|h^2\| &\leq \mu\|w\|_\infty\,\|q\| \;\leq\; 2\mu\left(\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}} + \frac{\eta}{4}\right), \\
\|h^3\| &\leq \mu\|p\|_\infty\,\|q\| \;\leq\; \mu\varepsilon_\infty^{\mathrm{ll}}\left(\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}} + \frac{\eta}{4}\right) \;\leq\; \mu\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}\left(\varepsilon_\infty^{\mathrm{ll}} + 1\right).
\end{aligned}$$

Using (42), the triangle inequality for norms, and the three estimates above, we then obtain

$$\begin{aligned}
\|v(\alpha) - (1-\alpha)\mu e\| &\leq (1-\alpha)\|xs - \mu e\| + \alpha\|h^1\| + \alpha(1-\alpha)\|h^2\| + \alpha^2\|h^3\| \\
&\leq (1-\alpha)\left(\|xs - \mu e\| + \|h^2\|\right) + \|h^1\| + \|h^3\| \\
&\leq \left[(1-\alpha)\left(\beta + 2\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}} + \frac{\eta}{2}\right) + \sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}} + \sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}\left(\varepsilon_\infty^{\mathrm{ll}} + 1\right)\right]\mu \\
&\leq \left[\left(\beta + \frac{\eta}{2}\right)(1-\alpha) + \sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}\left(\varepsilon_\infty^{\mathrm{ll}} + 4\right)\right]\mu \\
&\leq (\beta + \eta)(1-\alpha)\mu \;=\; \frac{2\beta}{1+2\beta}(1-\alpha)\mu
\end{aligned}$$

for all $0 \leq \alpha \leq 1 - 2\sqrt{n}\,\varepsilon_\infty^{\mathrm{ll}}(\varepsilon_\infty^{\mathrm{ll}} + 4)/\eta$. Hence, the validity of the claim follows. □

We are now ready to prove the main result of this paper, namely Theorem 2.7.

Proof of Theorem 2.7. Let $\mathcal{C}$ and $\bar{g}$ be the constant defined in (28). We claim that the P-C layered algorithm started from any $w \in \mathcal{N}(\beta)$ either terminates (at step 7) or generates an iterate $\hat{w}$ with a $\mathcal{C}$-crossover event occurring between w and $\hat{w}$ in $\mathcal{O}(n^{1.5}\log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ iterations. Since by Proposition 3.1 there can be at most $n(n+1)/2$ $\mathcal{C}$-crossover events of the above type, we conclude that the P-C layered algorithm must ultimately terminate in $\mathcal{O}(n^{3.5}\log(\bar{\chi}_A + n + \varepsilon_0^{-1}))$ iterations. To show the above claim, let $\mathcal{J} = (\mathcal{J}_1, \ldots, \mathcal{J}_r)$ denote an ordered (B, N)-partition at w, where (B, N) is the bipartition defined in (21). We split the proof into three possible cases: (1) $\varepsilon_\infty^{\mathrm{a}} > \varepsilon_0$, (2) $\mathrm{gap}(\mathcal{J}) \leq \bar{g}$, and (3) $\varepsilon_\infty^{\mathrm{a}} \leq \varepsilon_0$ and $\mathrm{gap}(\mathcal{J}) > \bar{g}$. The claim clearly holds for the first two cases due to Lemmas 4.1 and 4.2. Moreover, Lemma 4.3 implies that the claim also holds in the third case as long as we can show that the quantity $(\mu_+/\mu)/\varepsilon_\infty^{\mathrm{l}}$ appearing in (32) is $\mathcal{O}(\sqrt{n})$. Indeed, let α_{l} be defined as in step 5 of the P-C layered algorithm. Since in case (3) the LLS step is computed and step 6 of the P-C layered algorithm is performed, we must have $\mu_+ \leq \mu(w + \alpha_{\mathrm{l}}\Delta w^{\mathrm{l}})$. Hence,

the second statement of Lemma 4.6 applied to the partition $\mathcal{J}$ and the fact that $\text{gap}(\mathcal{J}) > \bar{g} \geq 96n\bar{\chi}_A/\eta$, where the second inequality is due to (28), imply

$$\frac{\mu_+}{\mu} \leq \frac{\mu(w + \alpha_1 \Delta w^{\mathrm{l}})}{\mu} \leq \frac{4\sqrt{n}\,\varepsilon_\infty^{\mathrm{l}}(\varepsilon_\infty^{\mathrm{l}} + 4)}{\eta}.$$

Hence, we conclude that $(\mu_+/\mu)/\varepsilon_\infty^{\mathrm{l}} = \mathcal{O}(\sqrt{n})$ whenever $\varepsilon_\infty^{\mathrm{l}} \leq 1$. If, on the other hand, $\varepsilon_\infty^{\mathrm{l}} > 1$, then we have $(\mu_+/\mu)/\varepsilon_\infty^{\mathrm{l}} \leq 1$ since $\mu_+/\mu \leq 1$.

It remains to show that when the method terminates at step 7 of the P-C layered algorithm it always finds a strictly complementary optimal solution. Indeed, let $\hat{w}$ be the iterate satisfying the stopping criterion of step 7. Clearly, $\mu(\hat{w}) = 0$ and $\hat{w} = w + \bar{\alpha}\Delta w$ for some $w \in \mathcal{N}(\beta)$, primal-dual feasible direction Δw, and stepsize $\bar{\alpha} > 0$ satisfying the property that $w + \alpha\Delta w \in \mathcal{N}(\beta)$ for all $\alpha \in [0, \bar{\alpha})$. Using Lemma 4.5, we conclude that $\bar{w}$ is a strictly complementary optimal solution. $\quad\square$

5. Concluding remarks. We consider our algorithm from the point of view of scaling-invariance. If one considers the change of variables $x = D\tilde{x}$, where D is a positive diagonal matrix, then the LP problem (1) is equivalent to

$$\min\{(Dc)^T\tilde{x} : AD\tilde{x} = b, \ \tilde{x} \geq 0\}.$$

It turns out that the sequence of points generated by the P-C layered algorithm when applied to (1) does not necessarily correspond (under the transformation $x = D\tilde{x}$) to the one obtained by applying it to the above LP problem. Algorithms with this desirable property are called scaling-invariant. The lack of scaling-invariance of the P-C layered algorithm, as well as the algorithms of Megiddo, Mizuno, and Tsuchiya [10] and Vavasis and Ye [26], is due to the fact that the choice of the layered partition used in the LLS step is not scaling-invariant. The construction of this partition is based on comparing the magnitudes of different components of δ, which per se is not a scaling-invariant quantity.

An interesting open problem is whether there exists a scaling-invariant algorithm whose complexity depends only on m, n, and $\bar{\chi}_A$. Note that if such an algorithm exists, its complexity will in fact depend only on m, n, and the quantity $\inf\{\bar{\chi}_{AD} : D \in \mathcal{D}\}$.

As in [26] and [10], we developed our algorithm for LP problems in which a well-centered interior feasible solution is given in advance. General LP problems can also be solved by the same algorithm applied to a suitably constructed artificial LP problem, and the resulting computational complexity can be shown to be the same as the one obtained in this paper. We refer the reader to section 10 of [26] and section 5 of [10] for more details.

6. Appendix. In this section we give the proof of Theorem 3.2. We start by stating the following result which yields Theorem 3.2 almost as an immediate consequence.

THEOREM 6.1. *Let $g \in \Re^m$, $F_i \in \Re^{m \times n_i}$, $h_i \in \Re^{n_i}$, $z_i \in \Re_{++}^{n_i}$, $i = 1, \ldots, l$, be given and assume that $g \in \text{Im}([F_1, \ldots, F_l])$. Define $d^0 = (d_1^0, \ldots, d_l^0) \in \Re^{n_1} \times \cdots \times \Re^{n_l}$ as*

$$(48) \quad (d_1^0, \ldots, d_l^0) \equiv \operatorname*{argmin}_{(d_1,\ldots,d_l) \in \Re^{n_1} \times \ldots \times \Re^{n_l}} \left\{ \sum_{i=1}^{l} \|d_i - h_i\|^2 : \sum_{i=1}^{l} F_i Z_i d_i = g \right\},$$

and define $\tilde{d}^0 = (\tilde{d}_1^0, \ldots, \tilde{d}_l^0) \in \Re^{n_1} \times \cdots \times \Re^{n_l}$ recursively starting from $k = 1$ upwards as

$$\tilde{d}_k^0 \equiv \operatorname*{argmin}_{\tilde{d}_k \in \Re^{n_k}} \left\{ \|\tilde{d}_k - h_k\|^2 : F_k Z_k \tilde{d}_k = g - \sum_{i=1}^{k-1} F_i Z_i \tilde{d}_i^0 + \operatorname{Im}(\tilde{F}_{k+1}) \right\}$$

for every $k = 1, \ldots, l-1$, where $Z_k \equiv \operatorname{Diag}(z_k)$ and $\tilde{F}_k \equiv [F_k, \ldots, F_l] \in \Re^{m \times (n_k + \cdots + n_l)}$. If the quantity $\Delta \equiv \max\{\Delta_i : i = 1, \ldots, l-1\}$, where $\Delta_i \equiv \|z_i\|_\infty \|z_{i+1}^{-1}\|_\infty$ for all $i = 1, \ldots, l-1$, satisfies $\bar{\chi}_F \Delta \le 1/\sqrt{2}$, then

$$(49) \qquad \|d^0 - \tilde{d}^0\|_\infty \le 4\bar{\chi}_F \Delta (1 + 4\bar{\chi}_F \Delta)^{l-2} \|d^0 - h\|,$$

where $h \equiv (h_1, \ldots, h_l)$ and $F = \tilde{F}_1$. In particular, if $g = 0$ and $4\bar{\chi}_F \Delta \le 1/l$, then

$$(50) \qquad \|d^0 - \tilde{d}^0\|_\infty \le 12\bar{\chi}_F \Delta \|h\|.$$

The proof of Theorem 6.1 will be given at the end of this section after some preliminary results are derived. Note that when $g = 0$ in Theorem 6.1 the point d^0 is the projection of h onto the null space of the matrix $[F_1 Z_1, \ldots, F_l Z_l] \in \Re^{m \times (n_1 + \cdots + n_l)}$ and the point $\tilde{d}^0$ is the layered projection of h onto the null space of $[F_1 Z_1, \ldots, F_l Z_l]$ according to the partition of variables $(z_1, \ldots, z_l)$.

The proof of Theorem 6.1 will be done by induction on the number l. A crucial step in this induction proof is the validity of certain proximity bounds for the case in which $l = 2$. Hence, as a preliminary step we will derive a special result for the case in which $l = 2$.

PROPOSITION 6.2. Let $g \in \Re^m$, $F_i \in \Re^{m \times n_i}$, $h_i \in \Re^{n_i}$, $z_i \in \Re_{++}^{n_i}$, $i = 1, 2$, be given and assume that $g \in \operatorname{Im}([F_1, F_2])$. Consider the points $d^0 = (d_1^0, d_2^0)$ and $\tilde{d}^0 = (\tilde{d}_1^0, \tilde{d}_2^0)$ determined as

$$(51) \qquad (d_1^0, d_2^0) \equiv \operatorname*{argmin}_{d} \{ \|d_1 - h_1\|^2 + \|d_2 - h_2\|^2 : F_1 Z_1 d_1 + F_2 Z_2 d_2 = g \},$$

$$(52) \qquad \tilde{d}_1^0 \equiv \operatorname*{argmin}_{d_1} \{ \|d_1 - h_1\|^2 : F_1 Z_1 d_1 \in g + \operatorname{Im}(F_2) \},$$

$$(53) \qquad \tilde{d}_2^0 \equiv \operatorname*{argmin}_{d_2} \{ \|d_2 - h_2\|^2 : F_2 Z_2 d_2 = g - F_1 Z_1 \tilde{d}_1^0 \},$$

where $Z_1 \equiv \operatorname{Diag}(z_1)$ and $Z_2 \equiv \operatorname{Diag}(z_2)$. Let $\Delta \equiv \|z_1\|_\infty \|(z_2)^{-1}\|_\infty$ and assume that $\bar{\chi}_F \Delta \le 1/\sqrt{2}$, where $F \equiv [F_1, F_2]$. Then the following estimates of the proximity between d^0 and $\tilde{d}^0$ hold:

$$\|d_1^0 - \tilde{d}_1^0\| \le 4\bar{\chi}_F \Delta \|d_2^0 - h_2\|, \qquad \|d_2^0 - \tilde{d}_2^0\| \le 4\bar{\chi}_F^2 \Delta^2 \|d_2^0 - h_2\|.$$

Before giving the proof of the above proposition, we first state and prove the following result which characterizes the displacements $\delta_1^0 \equiv d_1^0 - \tilde{d}_1^0$ and $\delta_2^0 \equiv d_2^0 - \tilde{d}_2^0$ as optimal solutions of certain optimization problems.

LEMMA 6.3. Let g, F_i, Z_i, $i = 1, 2$, be as defined in Proposition 6.2. Then the following statements hold:

(a) The vector $\delta_2^0 \equiv d_2^0 - \tilde{d}_2^0$ is the unique optimal solution of the problem

$$(54) \qquad \begin{aligned} &\operatorname{minimize}_{\delta_2} && \tfrac{1}{2}\|\delta_2\|^2 \\ &\text{subject to} && F_2 Z_2 \delta_2 = -F_1 Z_1 \delta_1^0. \end{aligned}$$

(b) *The pair (δ_1^0, d_2^0), where $\delta_1^0 \equiv d_1^0 - \tilde{d}_1^0$, is the unique optimal solution of the problem*

(55)
$$
\begin{aligned}
minimize_{(\delta_1, d_2)} \quad & \tfrac{1}{2}\|\delta_1\|^2 + \tfrac{1}{2}\|d_2 - h_2\|^2 \\
subject\ to \quad & F_1 Z_1 \delta_1 + F_2 Z_2 d_2 = g - F_1 Z_1 \tilde{d}_1^0.
\end{aligned}
$$

Proof. We first show (a). Since d^0 and $\tilde{d}_2^0$ are optimal solutions of (51) and (53), respectively, we have

(56)
$$
\begin{pmatrix} d_1^0 - h_1 \\ d_2^0 - h_2 \end{pmatrix} \in \mathrm{Im}\begin{pmatrix} Z_1 F_1^T \\ Z_2 F_2^T \end{pmatrix}, \qquad F_1 Z_1 d_1^0 + F_2 Z_2 d_2^0 = g,
$$

(57)
$$
\tilde{d}_2^0 - h_2 \in \mathrm{Im}(Z_2 F_2^T), \qquad F_1 Z_1 \tilde{d}_1^0 + F_2 Z_2 \tilde{d}_2^0 = g,
$$

and hence

(58)
$$
d_2^0 - \tilde{d}_2^0 \in \mathrm{Im}(Z_2 F_2^T), \qquad F_2 Z_2 \delta_2^0 = -F_1 Z_1 \delta_1^0.
$$

This shows that $\delta_2^0 = d_2^0 - \tilde{d}_2^0$ satisfies the optimality conditions for problem (54). Since (54) is a strictly convex quadratic program, its optimal solution is unique and hence (a) follows. We next show (b). Since $\tilde{d}_1^0$ is the optimal solution of (52), we have

$$
\begin{pmatrix} \tilde{d}_1^0 - h_1 \\ 0 \end{pmatrix} \in \mathrm{Im}(ZF^T),
$$

which, together with (56) and the definition of δ_1^0, yields

(59)
$$
\begin{pmatrix} \delta_1^0 \\ d_2^0 - h_2 \end{pmatrix} \in \mathrm{Im}(ZF^T), \qquad F_1 Z_1 \delta_1^0 + F_2 Z_2 d_2^0 = g - F_1 Z_1 \tilde{d}_1^0.
$$

This shows that (δ_1^0, d_2^0) satisfies the optimality conditions for (55). Since (55) is a strictly convex quadratic program, its optimal solution is unique and hence (b) holds.   □

Using the above lemma, we now give a proof of Proposition 6.2.

Proof of Proposition 6.2. By (58), we have that $F_1 Z_1 \delta_1^0 \in \mathrm{Range}(F_2)$. Hence, by Lemma 2.3, there exists a vector v_2^0 such that

(60)
$$
F_2 v_2^0 = F_1 Z_1 \delta_1^0, \qquad \|v_2^0\| \le \bar{\chi}_F \|Z_1 \delta_1^0\| \le \bar{\chi}_F \|z_1\|_\infty \|\delta_1^0\|.
$$

Relation (59) and (60) imply that $F_2[Z_2 d_2^0 + v_2^0] = g - F_1 Z_1 \tilde{d}_1^0$, and hence that the pair $(d_2^0 + Z_2^{-1} v_2^0, 0)$ is feasible for (55). This together with Lemma 6.3(b) implies that

$$
\|d_2^0 - h_2\|^2 + \|\delta_1^0\|^2 \le \|d_2^0 + Z_2^{-1} v_2^0 - h_2\|^2.
$$

Rearranging this expression and using relation (60) and the inequality $\|r\|^2 - \|u\|^2 \le \|r - u\|\,\|r + u\|$ for any $r, u \in \Re^p$, we obtain

$$
\begin{aligned}
\|\delta_1^0\|^2 &\le \left\{ \|d_2^0 + Z_2^{-1} v_2^0 - h_2\|^2 - \|d_2^0 - h_2\|^2 \right\} \\
&\le \|Z_2^{-1} v_2^0\|\,\| 2(d_2^0 - h_2) + Z_2^{-1} v_2^0 \| \\
&\le \|z_2^{-1}\|_\infty \|v_2^0\| \left\{ 2\|d_2^0 - h_2\| + \|z_2^{-1}\|_\infty \|v_2^0\| \right\} \\
&\le \bar{\chi}_F \Delta \|\delta_1^0\| \left\{ 2\|d_2^0 - h_2\| + \bar{\chi}_F \Delta \|\delta_1^0\| \right\},
\end{aligned}
$$

from which it follows that

$$(61) \qquad \|\delta_1^0\| \leq \frac{2\,\bar{\chi}_F\,\Delta\,\|d_2^0 - h_2\|}{1 - \bar{\chi}_F^2\,\Delta^2} \leq 4\bar{\chi}_F\,\Delta\,\|d_2^0 - h_2\|,$$

where the last inequality is due to the assumption that $\bar{\chi}_F\Delta \leq 1/\sqrt{2}$. The first relation in (60) implies that $-Z_2^{-1}v_2^0$ is a feasible solution of problem (54). Hence, by Lemma 6.3(a), the second relation in (60), and relation (61), it follows that

$$\|\delta_2^0\| \leq \|Z_2^{-1}v_2^0\| \leq \|z_2^{-1}\|_\infty \|v_2^0\| \leq \bar{\chi}_F\Delta\|\delta_1^0\| \leq 4\bar{\chi}_F^2\,\Delta^2\,\|d_2^0 - h_2\|. \qquad \square$$

We are now ready to give the proof of Theorem 6.1.

Proof of Theorem 6.1. During the proof, we refer to $\tilde{d}^0$ as the l-layer point associated with problem (48). We prove the inequality (49) by induction on l. Using Proposition 6.2 and noting that $\bar{\chi}_F\Delta \leq 1/\sqrt{2}$ by assumption, we obtain

$$\|d^0 - \tilde{d}^0\|_\infty \leq \max\{\|d_1^0 - \tilde{d}_1^0\|,\, \|d_2^0 - \tilde{d}_2^0\|\} \leq 4\bar{\chi}_F\Delta\|d_2^0 - h_2\| \leq 4\bar{\chi}_F\Delta\|d^0 - h\|,$$

from which we conclude that (49) holds for $l = 2$. Assume now that $l \geq 3$ and inequality (49) holds for $l - 1$. Consider the solution $(p_2^0, \ldots, p_l^0)$ of the problem

$$(62) \qquad (p_2^0, \ldots, p_l^0) \equiv \operatorname*{argmin}_{(p_2,\ldots,p_l)} \left\{ \sum_{i=2}^l \|p_i - h_i\|^2 : \sum_{i=2}^l F_i Z_i p_i = g - F_1 Z_1 \tilde{d}_1^0 \right\}$$

and note that $\tilde{d}_1^0$ and $(p_2^0, \ldots, p_l^0)$ are the optimal solutions of problems (52) and (53) in which F_1, F_2, z_1, and z_2 in Proposition 6.2 are identified with F_1, $\tilde{F}_2$, z_1, and $(z_2, \ldots, z_l)$, respectively. Hence, it follows from Proposition 6.2 that

$$(63) \qquad\qquad\qquad \|d_1^0 - \tilde{d}_1^0\| \leq 4\bar{\chi}_F\Delta\|d^0 - h\|,$$
$$(64) \qquad\qquad \|(d_2^0 - p_2^0, \ldots, d_l^0 - p_l^0)\| \leq 4\bar{\chi}_F^2\Delta^2\|d^0 - h\|.$$

Note also that $(\tilde{d}_2^0, \ldots, \tilde{d}_l^0)$ is the $(l-1)$-layer point associated with the problem (62). Hence, it follows from the induction hypothesis, i.e., that inequality (49) holds for $l - 1$, that

$$\|(p_2^0 - \tilde{d}_2^0, \ldots, p_l^0 - \tilde{d}_l^0)\|_\infty \leq 4\bar{\chi}_F\Delta(1 + 4\bar{\chi}_F\Delta)^{l-3}\|(p_2^0 - h_2, \ldots, p_l^0 - h_l)\|.$$

Using the triangle inequality for norms and (64), we obtain

$$\|(p_2^0 - h_2, \ldots, p_l^0 - h_l)\| \leq \|(d_2^0 - p_2^0, \ldots, d_l^0 - p_l^0)\| + \|(d_2^0 - h_2, \ldots, d_l^0 - h_l)\|$$
$$\leq 4\bar{\chi}_F^2\Delta^2\|d^0 - h\| + \|d^0 - h\| \;=\; (4\bar{\chi}_F^2\Delta^2 + 1)\|d^0 - h\|.$$

Combining the two last inequalities yields

$$\|(p_2^0 - \tilde{d}_2^0, \ldots, p_l^0 - \tilde{d}_l^0)\|_\infty \leq 4\bar{\chi}_F\Delta(1 + 4\bar{\chi}_F\Delta)^{l-3}(4\bar{\chi}_F^2\Delta^2 + 1)\|d^0 - h\|.$$

Using the triangle inequality for norms again, the last inequality, and (64), we obtain

$$\|(d_2^0 - \tilde{d}_2^0, \ldots, d_l^0 - \tilde{d}_l^0)\|_\infty \leq \|(d_2^0 - p_2^0, \ldots, d_l^0 - p_l^0)\|_\infty + \|(p_2^0 - \tilde{d}_2^0, \ldots, p_l^0 - \tilde{d}_l^0)\|_\infty$$
$$\leq \left[4\bar{\chi}_F^2\Delta^2 + 4\bar{\chi}_F\Delta(1 + 4\bar{\chi}_F\Delta)^{l-3}(4\bar{\chi}_F^2\Delta^2 + 1)\right]\|d^0 - h\|$$
$$\leq 4\bar{\chi}_F\Delta(1 + 4\bar{\chi}_F\Delta)^{l-3}(\bar{\chi}_F\Delta + 4\bar{\chi}_F^2\Delta^2 + 1)\|d^0 - h\|$$
$$\leq 4\bar{\chi}_F\Delta(1 + 4\bar{\chi}_F\Delta)^{l-3}(1 + (1 + 2\sqrt{2})\bar{\chi}_F\Delta)\|d^0 - h\|$$
$$\leq 4\bar{\chi}_F\Delta(1 + 4\bar{\chi}_F\Delta)^{l-2}\|d^0 - h\|.$$

The last inequality together with (63) implies that inequality (49) holds for l. It then follows by an induction argument that inequality (49) holds for any l.

We now prove that (50) holds when $g = 0$ and $4\bar{\chi}_F\Delta \leq 1/l$. Indeed, when $g = 0$, (48) implies that the vector d^0 is the orthogonal projection of h onto a subspace. Hence, $\|d^0 - h\| \leq \|h\|$. Also, $4\bar{\chi}_F\Delta \leq 1/l$ implies that $(1 + 4\bar{\chi}_F\Delta)^{l-2} \leq (1 + 1/l)^l \leq e \approx 2.718$. Substituting these two bounds into (49), we obtain (50).   □

REFERENCES

[1] I. I. DIKIN AND V. I. ZORKALCEV, *Iterative Solution of Mathematical Programming Problems: Algorithms for the Method of Interior Points*, Nauka, Novosibirsk, USSR, 1980 (in Russian).

[2] C. C. GONZAGA, *Path-following methods for linear programming*, SIAM Rev., 34 (1992), pp. 167–224.

[3] C. C. GONZAGA AND H. J. LARA, *A note on properties of condition numbers*, Linear Algebra Appl., 261 (1997), pp. 269–273.

[4] O. GÜLER AND Y. YE, *Convergence behavior of interior-point algorithms*, Math. Programming, 60 (1993), pp. 215–228.

[5] N. KARMARKAR, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.

[6] M. KOJIMA, S. MIZUNO, AND A. YOSHISE, *A primal-dual interior-point algorithm for linear programming*, in Progress in Mathematical Programming, Interior-Point and Related Methods, N. Megiddo, ed., Springer-Verlag, New York, 1989, pp. 29–47.

[7] M. KOJIMA, S. MIZUNO, AND A. YOSHISE, *A polynomial-time algorithm for a class of linear complementarity problems*, Math. Programming, 44 (1989), pp. 1–26.

[8] L. MCLINDEN, *The complementarity problem for maximal monotone multifunction*, in Variational Inequalities and Complementarity Problems, R.W. Cottle, F. Giannessi, and J.L. Lions, ed., John Wiley, New York, 1980, pp. 251–270.

[9] N. MEGIDDO, *Pathways to the optimal set in linear programming*, in Progress in Mathematical Programming, Springer-Verlag, New York, Berlin, 1989, pp. 131–158.

[10] N. MEGIDDO, S. MIZUNO, AND T. TSUCHIYA, *A modified layered-step interior-point algorithm for linear programming*, Math. Programming, 82 (1998), pp. 339–355.

[11] S. MIZUNO, M. J. TODD, AND Y. YE, *On adaptive-step primal-dual interior-point algorithms for linear programming*, Math. Oper. Res., 18 (1993), pp. 964–981.

[12] S. MIZUNO, N. MEGIDDO, AND T. TSUCHIYA, *A linear programming instance with many crossover events*, J. Complexity, 12 (1996) pp. 474–479.

[13] R. D. C. MONTEIRO AND I. ADLER, *Interior path-following primal-dual algorithms. Part* I: *Linear programming*, Math. Programming, 44 (1989), pp. 27–41.

[14] R. D. C. MONTEIRO AND I. ADLER, *Interior path-following primal-dual algorithms. Part* II: *Convex quadratic programming*, Math. Programming, 44 (1989), pp. 43–66.

[15] R. D. C. MONTEIRO AND T. TSUCHIYA, *Global convergence of the affine scaling algorithm for convex quadratic programming*, SIAM J. Optim., 8 (1998), pp. 26–58.

[16] J. RENEGAR, *A polynomial-time algorithm, based on Newton's method, for linear programming*, Math. Programming, 40 (1988), pp. 59–93.

[17] G. W. STEWART, *On scaled projections and pseudoinverses*, Linear Algebra Appl., 112 (1989), pp. 189–193.

[18] K. TANABE, *Centered Newton method for mathematical programming*, in System Modeling and Optimization, M. Iri and K. Yajima, eds., Springer-Verlag, Berlin, 1988, pp. 197–206.

[19] É. TARDOS, *A strongly polynomial algorithm to solve combinatorial linear programs*, Oper. Res., 34 (1986), pp. 250–256.

[20] M. J. TODD, *A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm*, Oper. Res., 38 (1990), pp. 1006–1018.

[21] M. J. TODD, L. TUNÇEL, AND Y. YE, *Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems*, Math. Programming, 90 (2001), pp. 59–69.

[22] T. Tsuchiya, *Global convergence property of the affine scaling method for primal degenerate linear programming problems*, Math. Oper. Res., 17 (1992), pp. 527–557.

[23] L. Tunçel, *Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard*, Math. Program., 86 (1999), pp. 219–223.

[24] L. Tunçel, *On the condition numbers for polyhedra in Karmarkar's form*, Oper. Res. Lett., 24 (1999), pp. 149–155.

[25] R. J. Vanderbei and J. C. Lagarias, *I. I. Dikin's Convergence Result for the Affine-Scaling Algorithm*, Contemp. Math. 114, AMS, Providence, RI, 1990, pp. 109–119.

[26] S. Vavasis and Y. Ye, *A primal-dual accelerated interior-point method whose running time depends only on A*, Math. Programming, 74 (1996), pp. 79–120.

[27] S. Vavasis and Y. Ye, *A simplification to "A primal-dual interior-point method whose running time depends only on the constraint matrix,"* in High Performance Optimization, Appl. Optim. 33, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, pp. 233–243.

[28] S. Vavasis and Y. Ye, *On the relationship between layered least squares and affine scaling steps*, in The Mathematics of Numerical Analysis, Lectures in Appl. Math. 32, AMS, Providence, RI, 1996, pp. 857–865.