A Lower Bound on the Number of Iterations of Long-Step Primal-Dual Linear Programming Algorithms

Michael J. Todd*   Yinyu Ye †

January, 1994; revised July, 1995

*School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853. E-mail: miketodd@cs.cornell.edu. Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

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

Abstract

Recently, Todd has analyzed in detail the primal-dual affine-scaling method for linear programming, which is close to what is implemented in practice, and proved that it may take at least $n^{1/3}$ iterations to improve the initial duality gap by a constant factor. He has also showed that this lower bound holds for some polynomial variants of primal-dual interior-point methods, which restrict all iterates to certain neighborhoods of the central path. In this paper, we further extend his result to long-step primal-dual variants that restrict the iterates to a wider neighborhood. This neighborhood seems the least restrictive one to guarantee polynomiality for primal-dual path-following methods, and the variants are also even closer to what is implemented in practice.

Key words: Linear Programming, Primal-Dual Interior-Point Algorithms, Lower Bounds.

Running Header: Complexity Lower Bound for Primal-Dual Algorithms

1 Introduction

In the last eight years, there has been enormous activity in optimization in the field of interior-point methods for linear programming and extensions. This explosion of research was instigated by the work of Karmarkar [7], who provided a polynomial-time algorithm whose extensions and variants have proved to be very efficient in solving large-scale linear programming problems; see, e.g., Lustig et al. [12]. For an overview of such methods the reader is referred to [4, 5, 20].

The most effective interior-point methods computationally are primal-dual methods, and these are variants of polynomial-time algorithms having the best theoretical complexity also. Several of these methods, such as path-following algorithms (see, e.g., Kojima et al. [10], Monteiro and Adler [16]), or potential-reduction methods (see Kojima et al. [11]), require $O(n^{1/2}t)$ iterations to attain an additional t digits of accuracy in a problem with n inequality constraints. On the other hand, computational experience with sophisticated primal-dual interior-point codes suggests that the number of iterations necessary to find an "exact" optimal solution grows much more slowly with the dimension n (see Lustig et al. [12] and Mehrotra and Ye [14]). There is thus a gap between observed performance and theoretical bounds.

There have been attempts to study the "expected" number of iterations theoretically. Some of these analyses are not rigorous as in the case of the simplex method; instead of assuming a random problem held fixed throughout the iterations, they make a probabilistic assumption about the data at a particular iteration, analyze the performance at that iteration, and hence make heuristic estimates of the "typical" behaviour of interior-point algorithms. Nevertheless, these studies indicate behaviour closer to what is observed in practice; see Nemirovsky [17] and Mizuno et al. [15]. More rigorously, one can study how small t needs to be to terminate such an algorithm and find an exact optimal solution. For some randomly generated problems, it has been shown that the expected total number of iterations to generate an exact solution is bounded above by $O(n^{1/2}\log n)$ – see Anstreicher et al. [2] and Ye [22].

There have also been efforts to look for *lower* bounds on the number of iterations required; see Anstreicher [1], Bertsimas and Luo [3], Ji and Ye [6], Powell [18], and Sonnevend et al. [19]. One important recent result is due to Todd [21], who obtains a bound of at least $n^{1/3}$ iterations to achieve a constant factor decrease in the duality gap. The algorithm he studies is the primal-dual affine-scaling algorithm, which is close to methods used in practical implementations. He allows almost any reasonable step size rule, such as going 99.5% of the way to the boundary of the feasible region, again as used in practical codes; such step size rules definitely do not lead to iterates lying close to the central path. The weakness of the primal-dual affine-scaling algorithm is that no polynomiality or even global convergence has been established for it, except for the case of very small step sizes, and practical experiments indicate that the algorithm alone may not perform well.

Todd also shows that his lower bound extends to other polynomial primal-dual interior-point methods that use directions including some centering component if the iterates are restricted to a certain neighborhood of the central path. This neighborhood uses the infinity norm and is narrower than the one frequently used in practice, which is based on the one-sided infinity norm.

Many practical algorithms have started to use this wider neighborhood to keep the iterates from approaching the boundary too closely prematurely. Kojima et al. first used this neighborhood in their early primal-dual method [9], which is an $O(nt)$-iteration polynomial algorithm. Mizuno et al. [15] further show that, while maintaining polynomiality, this wider neighborhood can be so large that it covers almost all the feasible region, so that the algorithm allows longer steps. Currently, this neighborhood is probably the least restrictive one guaranteeing polynomiality of primal-dual methods. Experimental results obtained using this wider neighborhood seem comparable to other state-of-the-art computational results (Lustig et al. [13]).

The simulation results conducted by Todd [21] indicate that his lower bound might extend to such long-step primal-dual variants. In this paper, we provide a proof for this conjecture. That is, to reduce the primal-dual gap by a constant factor requires at least $\Omega(n^{1/3})$ iterations for such a method, compared to an upper bound (to obtain t extra digits of accuracy) of $O(n^{1/2}t)$ iterations for a method using a very narrow neighborhood or a primal-dual potential function, or $O(nt)$ iterations for a long-step algorithm. This result holds when the centering parameter γ, defined below, either is a constant in $(0,1)$ or takes a nonnegative value possibly depending on n and k lying in $[0,\bar{\gamma}]$ for some constant $\bar{\gamma}$ in $(0,1)$. We can also allow periodic full centering steps. In the former case, we can further show that, to reduce the duality gap by a factor of n (which seems necessary for termination at an exact optimal solution), at least $\Omega(n^{1/3}(\log n)^{2/3})$ iterations are required, compared to an upper bound (on the average number of iterations to obtain an exact solution) of $O(n^{1/2} \log n)$.

Section 2 describes the algorithm we consider. In Section 3, we establish properties of a long sequence of steps in the case that γ is constant; Section 4 extends this to the case that γ may depend on n and k. Finally, Section 5 provides the main result: for n sufficiently large, the number of steps to decrease the duality gap by a factor of 3 for some problem with n inequalities is $\Omega(n^{1/3})$.

2 Primal-Dual Interior-Point Algorithms

Consider the primal linear programming problem in standard form

$$
(P) \qquad \begin{array}{lrcl} \min_x & c^\top x & & \\ & Ax & = & b \\ & x & \geq & 0, \end{array}
$$

with dual problem

$$
(D) \qquad \begin{array}{lrcl} \max_{y,s} & b^\top y & & \\ & A^\top y + s & = & c \\ & s & \geq & 0, \end{array}
$$

where $A \in I\!R^{m\times n}$, $b \in I\!R^m$, and $c \in I\!R^n$ are the data, and $x, s \in I\!R^n$, and $y \in I\!R^m$ the variables. For any x feasible in (P) and (y, s) feasible in (D), it is easy to see that the duality gap is

$$
c^\top x - b^\top y = x^\top s \geq 0;
$$

the strong duality theorem of linear programming states that x and (y, s) are optimal if and only if $x^\top s = 0$.

We will view the primal-dual algorithms as generating a sequence of pairs (x^k, s^k) lying in

$$\mathcal{F}^0 := \{(x, s) \in I\!R^{2n} : Ax = b, A^\top y + s = c \\ \text{for some } y,\ x > 0,\ s > 0\}.$$

We assume that $\mathcal{F}^0$ is nonempty, which implies that both (P) and (D) have bounded sets of optimal solutions, and that we have an initial $(x^0, s^0) \in \mathcal{F}^0$.

(Note that practical primal-dual algorithms also have the capability of starting with infeasible interior points. However, if initialized with a feasible pair (x^0, s^0), they maintain feasibility throughout, as does the algorithm below. Similarly, more general methods can solve convex quadratic programming or monotone linear complementarity problems, but they reduce to the method below for the special case of linear programming problems. Finally, our analysis assumes that (x^0, s^0) lies on the so-called central path. Thus we are giving our algorithms every advantage, and we still obtain unattractive lower bounds.)

The pair $(\hat{x}, \hat{s})$ will be updated at each iteration by taking a damped Newton step for the system (a perturbation of the Karush-Kuhn-Tucker conditions for (P) and (D))

$$\begin{aligned} Ax &= b, && (2.1)\\ A^\top y + s &= c, && (2.2)\\ XSe &= \hat{\gamma}\hat{\mu}e, && (2.3)\\ x \geq 0, &\qquad s \geq 0, && (2.4) \end{aligned}$$

where e denotes $(1, ..., 1)^\top \in I\!R^n$, X and S the diagonal matrices with $Xe = x$, $Se = s$, $\hat{\mu} := \hat{x}^\top \hat{s}/n$, and $\hat{\gamma}$ is the *centering parameter* in $[0, 1)$. The set of solutions to (2.1)–(2.4), where $\hat{\gamma}\hat{\mu}$ is replaced by a parameter μ varying in $(0, \infty)$, forms the so-called central path for (P) and (D); path-following algorithms are based on approximately following this path as μ decreases to 0. The long-step algorithm requires that all iterates lie in the neighborhood

$$\mathcal{N}(\beta) = \{(x, s) \in \mathcal{F}^0 : \min_{1 \leq j \leq n} (x_j s_j) \geq \mu/\beta, \quad \text{where} \quad \mu = x^\top s/n\} \qquad (2.5)$$

for some constant $\beta > 1$ (see [15]). (By contrast, Todd's analysis [21] of a primal-dual algorithm with centering uses the narrower neighborhood

$$\{(x, s) \in \mathcal{F}^0 : \|XSe - \mu e\|_\infty \leq \chi\mu, \quad \text{where} \quad \mu = x^\top s/n\} \tag{2.6}$$

for some constant $\chi < 1$.)

The Newton direction for (2.1)–(2.3) at $(\hat{x}, \hat{s}) = (x^k, s^k)$ is the solution to

$$\begin{aligned} A\xi &= 0, \\ A^\top \eta + \sigma &= 0, \\ X^k\sigma + S^k\xi &= \gamma^k\mu^k e - X^k S^k e, \end{aligned} \tag{2.7}$$

where X^k and S^k are the diagonal matrices with $X^k e = x^k$, $S^k e = s^k$, $\mu^k := (x^k)^\top s^k/n$, and $\gamma^k \in [0, 1)$. (Superscripts are used throughout for iteration indices; nonnegative integer powers are indicated by enclosing their arguments in parentheses.) The form of the solution to (2.7) can be made more apparent by the following scaling. Let

$$D^k := (X^k)^{1/2}(S^k)^{-1/2}, \quad V^k := (X^k)^{1/2}(S^k)^{1/2}, \tag{2.8}$$

$$v := v^k := V^k e, \quad w := w^k := v - \gamma^k\mu^k(V^k)^{-1}e, \tag{2.9}$$

$$\tilde{A} := AD^k, \quad \tilde{\xi} := (D^k)^{-1}\xi, \quad \tilde{\sigma} := D^k\sigma. \tag{2.10}$$

Then (2.7) is equivalent to

$$\begin{aligned} \tilde{A}\tilde{\xi} &= 0, \\ \tilde{A}^\top\eta + \tilde{\sigma} &= 0, \\ \tilde{\sigma} + \tilde{\xi} &= -w, \end{aligned} \tag{2.11}$$

whose solution can be written as

$$\tilde{\xi} = -P_{\tilde{A}}w, \quad \tilde{\sigma} = -(I - P_{\tilde{A}})w, \tag{2.12}$$

where P_M denotes the matrix that projects a vector orthogonally onto the null space of M.

The scaling above corresponds to the change of variables taking x to $(D^k)^{-1}x$ and s to $D^k s$. Note that both x^k and s^k are thus transformed to v. We can now state our generic primal-dual interior-point algorithm:

Algorithm

Choose $(x^0, s^0) \in \mathcal{N}(\beta)$, with β a constant greater than 1, and set $k = 0$.
While $(x^k)^\top s^k > 0$ do

begin

set
$$\mu^k := (x^k)^\top s^k / n$$

and choose
$$\gamma^k \in [0, 1);$$

compute from (2.8)–(2.12)
$$\begin{aligned} \tilde{\xi} &:= \tilde{\xi}^k := -P_{\tilde{A}} w, \quad \tilde{\sigma} := \tilde{\sigma}^k := -(I - P_{\tilde{A}}) w; \\ \xi^k &:= D\tilde{\xi}, \quad \sigma^k := D^{-1} \tilde{\sigma}^k; \end{aligned}$$

choose $\rho_P^k > 0$ and $\rho_D^k > 0$ so that
$$(x^k + \rho_P^k \xi^k, s^k + \rho_D^k \sigma^k) \in \mathcal{N}(\beta);$$

let
$$x^{k+1} := x^k + \rho_P^k \xi^k, \quad s^{k+1} := s^k + \rho_D^k \sigma^k;$$

and set $k := k + 1$.

end

It can be verified that if $\rho_P^k = \rho_D^k = \rho^k$, then
$$(x^{k+1})^\top s^{k+1} = [1 - \rho^k (1 - \gamma^k)] (x^k)^\top s^k. \tag{2.13}$$

Hence to achieve a large decrease in the duality gap, we would like ρ^k to be large and γ^k to be small. As in the analysis of [21], we will show that for a long sequence of iterations we can ensure that the directions are such that maintaining feasibility forces small values of ρ^k.

It appears that the restriction $\rho_P^k = \rho_D^k$ in (2.13) is a limitation on the algorithm we consider. However, let us suppose that the step sizes chosen are of the form

$$\rho_P^k = \rho^{(k)}(x^k, \xi^k, s^k, \sigma^k) \quad \text{and} \quad \rho_D^k = \rho^{(k)}(s^k, \sigma^k, x^k, \xi^k), \tag{2.14}$$

where the function $\rho^{(k)}$ takes positive values and satisfies the property

$$\rho^{(k)}(\Pi x^k, \Pi\xi^k, \Pi s^k, \Pi\sigma^k) = \rho^{(k)}(x^k, \xi^k, s^k, \sigma^k) \tag{2.15}$$

for every $x^k > 0, \xi^k, s^k > 0, \sigma^k$ and every $n \times n$ permutation matrix Π. Thus the step sizes depend only on the current iterates and search directions (and possibly the iteration number) and are symmetric between primal and dual and between different components. (Step sizes usually chosen, such as those in [15, 8], satisfy these requirements.) Hence if $s^k = \Pi x^k$ and $\sigma^k = \Pi\xi^k$ for some permutation Π with $(\Pi)^2 = I$, (2.14) and (2.15) imply

$$\rho_P^k = \rho_D^k.$$

We also require that the centering parameter γ^k only depend on the current iterates (and possibly the iteration number):

$$\gamma^k = \gamma^{(k)}(x^k, s^k). \tag{2.16}$$

We conclude this section by recalling a result of [21] showing that the directions can be chosen with considerable freedom. Suppose at the kth iteration we have $\tilde{\xi} + \tilde{\sigma} = -w = -w^k$. A sufficient condition for $\tilde{\xi}$ to be $-P_{\tilde{A}}w$ and $\tilde{\sigma}$ to be $-(I - P_{\tilde{A}})w$ is then that $\tilde{\xi}$ lies in the null space and $\tilde{\sigma}$ in the row space of $\tilde{A}$. Removing the scaling, we want ξ^k and σ^k to lie in the null space and row space of A respectively. Thus these two directions must be orthogonal.

Suppose now we are given a sequence of vectors ξ^j and σ^j such that the scaled vectors $\tilde{\xi}^j$ and $\tilde{\sigma}^j$ have $-w^j$ for their sum for each j. Then it is sufficient for them to arise as above as the primal and dual directions in a sequence of iterations for ξ^j and σ^j to lie in the null space and row space of A respectively, for each j. Clearly, it is then necessary for each primal vector to be orthogonal to each dual vector. If there are not too many vectors, this condition is also sufficient (the proof is not difficult):

Theorem 2.1 *[21] Let* $(\xi^j, \sigma^j) \in \mathbb{R}^{2n}$ *be given for* $0 \leq j \leq k$*. A sufficient condition that there exists* $A \in \mathbb{R}^{m \times n}$ *and* $\eta^j \in \mathbb{R}^m$*,* $0 \leq j \leq k$*, such that*

$$A\xi^j = 0,\ A^\top \eta^j + \sigma^j = 0,\ 0 \leq j \leq k,$$

is that

(i) $k + 1 \leq \min\{m, n - m\}$*; and*

(ii) $\Xi := [\xi^0, \ldots, \xi^k]$ *and* $\Sigma := [\sigma^0, \ldots, \sigma^k]$ *satisfy* $\Xi^\top \Sigma = 0$*.*

3 The Inductive Proof: Constant γ

Here we treat the case where γ^k is a constant, say $\bar{\gamma}$, in (0,1) independent of n (and k). (The case $\bar{\gamma} = 0$ was treated in [21], while $\bar{\gamma} = 1$ leads to no reduction in the duality gap by (2.13).) Todd [21] describes a technique for constructing bad directions for several iterations, and we will follow the same arguments. The 0th iteration will "contaminate" the first pair of components of x and s; similarly the $(k-1)$st will "contaminate" the kth pair of components, but all subsequent components of x and s will still be equal. It is convenient therefore to index the components in pairs. We assume that $n =: 2p$ is even, and index the components $1, -1, 2, -2, \ldots, p, -p$. (If n is odd, one component is left over, but the arguments below can easily be modified.)

In addition, we will preserve symmetry in the first $2k$ components of x^k and s^k. Specifically, x_1^k will equal s_{-1}^k, x_{-1}^k will equal s_1^k, and so on. To describe these properties conveniently, we let $\bar{\Pi}$ denote the permutation matrix that switches the jth and $(-j)$th components of each vector in $\mathbb{R}^n$, $1 \leq j \leq p$. We also let S_n^k denote the set of permutation matrices that leave fixed the first k pairs of components of vectors in $\mathbb{R}^n$.

We suppose that $(e, e) \in \mathcal{N}(\beta)$, and let the initial iterates be

$$x^0 = s^0 = e.$$

Assume that at the beginning of the kth iteration we have $(x^k, s^k) \in \mathcal{N}(\beta)$ satisfying symmetry of the pairs,

$$\bar{\Pi} x^k = s^k, \tag{3.1}$$

and equality of the final components,

$$\Pi x^k = x^k,\ \Pi s^k = s^k, \text{ for all } \Pi \in S_n^k. \tag{3.2}$$

We make similar assumptions about the previous search directions. Let $\Xi^k := [\xi^0, \ldots, \xi^{k-1}]$ and $\Sigma^k := [\sigma^0, ..., \sigma^{k-1}]$ be the matrices of previous primal and dual search directions. We assume

$$(\Xi^k)^\top \Sigma^k = 0, \tag{3.3}$$

$$\bar{\Pi}\Xi^k = \Sigma^k, \tag{3.4}$$

and, for each $1 \le j \le k$,

$$\Pi \xi^{j-1} = \xi^{j-1},\ \Pi \sigma^{j-1} = \sigma^{j-1}, \text{ for all } \Pi \in S_n^j. \tag{3.5}$$

Here, (3.3) ensures that the previous search directions are consistent with some matrix A (see Theorem 2.1), while (3.4) maintains the symmetry between pairs of components. Finally, (3.5) shows that the search directions ξ^{j-1} and σ^{j-1} treat all components after the first j pairs equally. Note that, by setting $x^0 = s^0 = e$, (3.1) and (3.2) hold for $k = 0$, while (3.3)–(3.5) hold vacuously.

We next examine the effect of these assumptions on X^k, S^k, D^k and $w = w^k$. From (3.1) and (3.2),

$$\begin{aligned} \bar{\Pi} X^k \bar{\Pi} &= S^k, \quad \text{and} \\ \Pi X^k \Pi^\top &= X^k, \quad \Pi S^k \Pi^\top = S^k, \text{ for all } \Pi \in S_n^k. \end{aligned}$$

We then find

$$\begin{aligned} \bar{\Pi} D^k \bar{\Pi} &= \bar{\Pi}(X^k)^{1/2}\bar{\Pi}\ \bar{\Pi}(S^k)^{-1/2}\bar{\Pi} \\ &= (S^k)^{1/2}(X^k)^{-1/2} = (D^k)^{-1}, \end{aligned} \tag{3.6}$$

and similarly

$$\Pi D^k \Pi^\top = D^k \text{ for all } \Pi \in S_n^k. \tag{3.7}$$

In the same way,

$$
\begin{aligned}
\bar{\Pi} w^k &= \bar{\Pi}(X^k)^{1/2}\bar{\Pi}\ \bar{\Pi}(S^k)^{1/2}\bar{\Pi}\ \bar{\Pi}e - \bar{\Pi}(X^k)^{-1/2}\bar{\Pi}\ \bar{\Pi}(S^k)^{-1/2}\bar{\Pi}\ \bar{\Pi}(\bar{\gamma}\mu^k e) \\
&= (S^k)^{1/2}(X^k)^{1/2}e - \bar{\gamma}\mu^k(S^k)^{-1/2}(X^k)^{-1/2}e = w^k, \qquad (3.8)
\end{aligned}
$$

and

$$
\Pi w^k = w^k, \text{ for all } \Pi \in S_n^k. \qquad (3.9)
$$

Our final assumption is

$$
\frac{\mu^k}{\beta} \le (v_j^k)^2 = x_j^k s_j^k \le \alpha\mu^k, \quad 1 \le |j| \le p = n/2, \qquad (3.10)
$$

where

$$
\alpha := 4\max\{1, \beta(\bar{\gamma})^2\}. \qquad (3.11)
$$

Note that $(1/\sqrt{\beta} + \bar{\gamma}\sqrt{\beta})^2 \le (1 + \bar{\gamma}\sqrt{\beta})^2 \le (2\max\{1, \bar{\gamma}\sqrt{\beta}\})^2 = \alpha$, so that

$$
\alpha \ge \max\{4,\ (\frac{1}{\sqrt{\beta}} + \bar{\gamma}\sqrt{\beta})^2\}. \qquad (3.12)
$$

We shall see the importance of assumption (3.10) later. However, it is important to remark that the left-hand inequality holds automatically from our choice of step sizes in the algorithm, whereas the right-hand inequality is an assumption which we will have to establish later. For reasonable practical values such as $\beta = 1000$ and $\bar{\gamma} \le 1/50$, we see that α is 4.

We have thus made the following **Inductive Hypothesis**:

> *At the beginning of the kth iteration, the iterates (x^k, s^k) and the previous search directions Ξ^k and Σ^k satisfy (3.1)–(3.5) and (3.10).*

We now make the scaling D^k to primal and dual iterates and directions:

$$
\begin{aligned}
\tilde{x}^k &:= (D^k)^{-1}x^k = v^k, \\
\tilde{s}^k &:= D^k s^k = v^k, \\
\tilde{\Xi}^k &:= (D^k)^{-1}\Xi^k, \quad \text{and} \\
\tilde{\Sigma}^k &:= D^k\Sigma^k.
\end{aligned}
$$

We note that $\tilde{\Xi}^k$ and $\tilde{\Sigma}^k$ satisfy

$$(\tilde{\Xi}^k)^\top \tilde{\Sigma}^k = 0$$

from (3.3), and (3.4) and (3.6) show that they satisfy

$$\bar{\Pi}\tilde{\Xi}^k = \tilde{\Sigma}^k.$$

From (3.5) we deduce $\Pi\Xi^k = \Xi^k$ and $\Pi\Sigma^k = \Sigma^k$ for all $\Pi \in S_n^k$, so (3.7) shows

$$\Pi\tilde{\Xi}^k = \tilde{\Xi}^k, \quad \Pi\tilde{\Sigma}^k = \tilde{\Sigma}^k,$$

for all such Π's.

Todd [21] shows that we can choose the next directions in the form

$$\tilde{\xi}^k = -\frac{1}{2}(w^k + (\epsilon_1, \epsilon_{-1}, \ldots, \epsilon_{k+1}, \epsilon_{-k-1}, 0, \ldots, 0)^\top), \tag{3.13}$$

$$\tilde{\sigma}^k = -\frac{1}{2}(w^k - (\epsilon_1, \epsilon_{-1}, \ldots, \epsilon_{k+1}, \epsilon_{-k-1}, 0, \ldots, 0)^\top), \tag{3.14}$$

where $\epsilon_j + \epsilon_{-j} = 0$, $1 \leq j \leq k+1$. Note that $\bar{\Pi}\tilde{\xi}^k = \tilde{\sigma}^k$ and $\Pi\tilde{\xi}^k = \tilde{\xi}^k$, $\Pi\tilde{\sigma}^k = \tilde{\sigma}^k$, for all $\Pi \in S_n^{k+1}$.

Theorem 3.1 *[21] As long as $k < p = n/2$ and the inductive hypothesis holds, we can choose $\tilde{\xi}^k$ and $\tilde{\sigma}^k$ as in (3.13)–(3.14) so that $\Xi := [\tilde{\Xi}^k, \tilde{\xi}^k]$ and $\Sigma := [\tilde{\Sigma}^k, \tilde{\sigma}^k]$ satisfy $\Xi^\top\Sigma = 0$.*

(Note that this result holds even if γ^k is not constant.)

Theorem 3.1 establishes almost all of the inductive hypothesis with k replaced by $k+1$. Indeed, the new matrices of past search directions are

$$\Xi^{k+1} = D^k\Xi, \quad \Sigma^{k+1} = (D^k)^{-1}\Sigma,$$

where Ξ and Σ are defined in the theorem. Then (3.3)–(3.5) hold for $k+1$ by the conclusions of the theorem, the inductive hypothesis for k, and the form (3.13)–(3.14) of the new directions.

It only remains to show that (3.1), (3.2), and (3.10) hold for $k+1$. If we assume that (2.14) and (2.15) hold, then (3.1) for k and $\bar{\Pi}\xi^k = \sigma^k$ with

$(\bar{\Pi})^2 = I$ show that $\rho_P^k = \rho_D^k$, and hence that $\bar{\Pi}x^{k+1} = s^{k+1}$. Similarly, (3.2) and $\Pi\xi^k = \xi^k$, $\Pi\sigma^k = \sigma^k$ for all $\Pi \in S_n^{k+1}$ show that (3.2) holds for $k+1$.

We now turn to (3.10). Note that the left-hand inequality of (3.10) holds from the restriction $(x^{k+1}, s^{k+1}) \in \mathcal{N}(\beta)$ defined in (2.5). Thus, we focus on establishing the right-hand inequality. We first prove a lemma.

Lemma 3.1 *There exist a j, $1 \le |j| \le k+1$, such that*

$$\epsilon_j \ge (1-\bar{\gamma})\sqrt{\frac{n\mu^k}{2k+2}}$$

Proof. Since $\tilde{\xi}^k$ and $\tilde{\sigma}^k$ (and hence $2\tilde{\xi}^k$ and $2\tilde{\sigma}^k$) are orthogonal, using (3.13) and (3.14) yields

$$\begin{aligned}
\sum_{1\le|j|\le k+1} (\epsilon_j)^2 &= (w^k)^\top w^k \\
&= \|v^k - \gamma^k\mu^k(V^k)^{-1}e\|^2 \\
&= \|v^k\|^2 - 2n\gamma^k\mu^k + (\gamma^k\mu^k)^2\|(V^k)^{-1}e\|^2 \\
&\ge n\mu^k - 2n\gamma^k\mu^k + (\gamma^k\mu^k)^2(n/\mu^k) \\
&= n\mu^k(1-\gamma^k)^2 = n\mu^k(1-\bar{\gamma})^2,
\end{aligned}$$

where we used $\|v^k\|^2 = n\mu^k$ and, from the Cauchy-Schwartz inequality, $\|v^k\|^2\|(V^k)^{-1}e\|^2 \ge n^2$.

Hence some $|\epsilon_j|$ is at least $(1-\bar{\gamma})\sqrt{n\mu^k/(2k+2)}$. Recalling that $\epsilon_j + \epsilon_{-j} = 0$ and letting $\epsilon_j \ge 0$, we have the desired result. □

Let j be the index singled out in the above lemma and let $\rho_P^k = \rho_D^k =: \rho^k =: 2\bar{\rho}^k$. Below, we give two bounds on $\bar{\rho}^k$. One is a function of $\alpha, \bar{\gamma}, k$, and n showing $\bar{\rho}^k$ must be very small for $k \le O(n^{1/3})$; this is used in deriving the lower bound on the number of iterations to achieve a constant decrease in the duality gap in Section 5. The other bound is $\bar{\gamma}$ for $k \le O(n)$, and is used to prove that the inductive hypothesis continues to hold for many iterations. To prove these bounds, first note that

$$\tilde{x}_j^+ = v_j^k + \bar{\rho}^k(-v_j^k + \bar{\gamma}\mu^k/v_j^k - \epsilon_j) > 0;$$

thus we must have, from the above lemma (as long as the denominator is positive, as is proved below),

$$
\begin{aligned}
\bar{\rho}^k &\leq \frac{v_j^k}{\epsilon_j + v_j^k - \bar{\gamma}\mu^k/v_j^k} \\
&= \frac{1}{\epsilon_j/v_j^k + 1 - \bar{\gamma}\mu^k/(v_j^k)^2} \\
&\leq \frac{1}{(1-\bar{\gamma})\sqrt{n/(2k+2)}(\sqrt{\mu^k}/v_j^k) + 1 - \bar{\gamma}(\sqrt{\mu^k}/v_j^k)^2}. \qquad (3.15)
\end{aligned}
$$

By the inductive hypothesis,

$$1/\sqrt{\alpha} \leq \sqrt{\mu^k}/v_j^k \leq \sqrt{\beta}.$$

Let us assume that

$$2k + 2 \leq \frac{(1-\bar{\gamma})^2 n}{4\beta(\bar{\gamma})^2}. \qquad (3.16)$$

Then, considering the denominator in (3.15) as a quadratic function of $\sqrt{\mu^k}/v_j^k$, we see from (3.16) that its maximizer lies above $\sqrt{\beta}$. Hence, the denominator satisfies

$$(1-\bar{\gamma})\sqrt{\frac{n}{(2k+2)}}\frac{\sqrt{\mu^k}}{v_j^k} + 1 - \bar{\gamma}(\frac{\sqrt{\mu^k}}{v_j^k})^2 \geq (1-\bar{\gamma})\sqrt{\frac{n}{(2k+2)}}\frac{1}{\sqrt{\alpha}} + 1 - \frac{\bar{\gamma}}{\alpha},$$

which is positive since $\bar{\gamma} < 1 < 4 \leq \alpha$. Thus,

$$\bar{\rho}^k \leq \frac{1}{(1-\bar{\gamma})\sqrt{\frac{n}{\alpha(2k+2)}} + 1 - \bar{\gamma}/\alpha} \leq \frac{1}{(1-\bar{\gamma})(\sqrt{\frac{n}{\alpha(2k+2)}} + 1)} \leq \frac{\sqrt{\alpha(2k+2)}}{(1-\bar{\gamma})\sqrt{n}}. \qquad (3.17)$$

Let us further assume that

$$2k + 2 \leq \frac{(\bar{\gamma})^2(1-\bar{\gamma})^2 n}{\alpha}. \qquad (3.18)$$

This implies that

$$\bar{\rho}^k \leq \bar{\gamma}. \qquad (3.19)$$

We are now ready to prove that the right-hand inequality of (3.10) holds for $k+1$. We see that

$$\begin{aligned} x_j^{k+1}s_j^{k+1} &= ([D^k]^{-1}x^{k+1})_j(D^k s^{k+1})_j \\ &= (v_j^k + \rho^k \tilde{\xi}_j^k)(v_j^k + \rho^k \tilde{\sigma}_j^k) \end{aligned}$$

for each j, $1 \le |j| \le p$, where $\rho^k := \rho_P^k = \rho_D^k$. Recall $\bar{\rho}^k := \rho^k/2$. Then, from (3.13) and (3.14),

$$\begin{aligned} x_j^{k+1}s_j^{k+1} &= ([1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\bar{\gamma}\mu^k(v_j^k)^{-1} + \bar{\rho}^k\epsilon_j)([1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\bar{\gamma}\mu^k(v_j^k)^{-1} - \bar{\rho}^k\epsilon_j) \\ &= ([1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\bar{\gamma}\mu^k(v_j^k)^{-1})^2 - (\bar{\rho}^k\epsilon_j)^2 \\ &\le ([1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\bar{\gamma}\mu^k(v_j^k)^{-1})^2 \end{aligned}$$

for $1 \le |j| \le k+1$, while

$$x_j^{k+1}s_j^{k+1} = ([1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\bar{\gamma}\mu^k(v_j^k)^{-1})^2$$

for $k+1 < |j| \le p$. Note that

$$\mu^{k+1} = (1 - 2\bar{\rho}^k(1-\bar{\gamma}))\mu^k.$$

Since $0 < \bar{\rho}^k \le \bar{\gamma} < 1$ by (3.19), we obtain

$$1 - 2\bar{\rho}^k(1-\bar{\gamma}) > (1-\bar{\rho}^k)^2 > 0.$$

Thus, it suffices to show

$$([1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\bar{\gamma}\mu^k(v_j^k)^{-1})^2 \le \alpha(1 - 2\bar{\rho}^k(1-\bar{\gamma}))\mu^k,$$

or

$$[1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\bar{\gamma}\mu^k(v_j^k)^{-1} \le \sqrt{\alpha(1 - 2\bar{\rho}^k(1-\bar{\gamma}))\mu^k},$$

or

$$[1-\bar{\rho}^k](v_j^k/\sqrt{\mu^k}) + \bar{\rho}^k\bar{\gamma}(\sqrt{\mu^k}/v_j^k) \le \sqrt{\alpha}\sqrt{1 - 2\bar{\rho}^k(1-\bar{\gamma})}.$$

Note again from the inductive hypothesis that

$$1/\sqrt{\beta} \le v_j^k/\sqrt{\mu^k} \le \sqrt{\alpha},$$

and $[1-\bar\rho^k](v_j^k/\sqrt{\mu^k})+\bar\rho^k\bar\gamma(\sqrt{\mu^k}/v_j^k)$ is convex in $v_j^k/\sqrt{\mu^k}$. Hence, it is sufficient to show

$$[1-\bar\rho^k]\sqrt{\alpha}+\bar\rho^k\bar\gamma/\sqrt{\alpha}\le\sqrt{\alpha}\sqrt{1-2\bar\rho^k(1-\bar\gamma)} \tag{3.20}$$

and

$$[1-\bar\rho^k]/\sqrt{\beta}+\bar\rho^k\bar\gamma\sqrt{\beta}\le\sqrt{\alpha}\sqrt{1-2\bar\rho^k(1-\bar\gamma)}. \tag{3.21}$$

To show (3.20) we see that

$$\phi(\bar\rho):=\frac{\bar\rho\bar\gamma}{\sqrt{1-2\bar\rho(1-\bar\gamma)}-(1-\bar\rho)}$$

is monotonically increasing for $0<\bar\rho\le\bar\gamma$, and

$$\phi(\bar\gamma)=\sqrt{1-2\bar\gamma(1-\bar\gamma)}+1-\bar\gamma<2\le\alpha,$$

which proves (3.20) (recall (3.19)). Similarly,

$$\psi(\bar\rho):=\frac{\bar\rho\bar\gamma}{\sqrt{1-2\bar\rho(1-\bar\gamma)}}$$

is monotonically increasing for $0<\bar\rho\le\bar\gamma$, and we have

$$\begin{aligned}
\frac{(1-\bar\rho)/\sqrt{\beta}+\bar\rho\bar\gamma\sqrt{\beta}}{\sqrt{1-2\bar\rho(1-\bar\gamma)}} &= \frac{(1-\bar\rho)/\sqrt{\beta}}{\sqrt{1-2\bar\rho(1-\bar\gamma)}}+\frac{\bar\rho\bar\gamma\sqrt{\beta}}{\sqrt{1-2\bar\rho(1-\bar\gamma)}}\\
&\le 1/\sqrt{\beta}+\frac{\bar\rho\bar\gamma\sqrt{\beta}}{\sqrt{1-2\bar\rho(1-\bar\gamma)}}\\
&\le 1/\sqrt{\beta}+\frac{(\bar\gamma)^2\sqrt{\beta}}{\sqrt{(1-\bar\gamma)^2+(\bar\gamma)^2}}\\
&\le 1/\sqrt{\beta}+\bar\gamma\sqrt{\beta}\le\sqrt{\alpha},
\end{aligned}$$

which proves (3.21). Thus (3.10) remains true for $k+1$ as long as $k\le O(n)$ given by (3.16) and (3.18). We therefore have

Theorem 3.2 *Suppose the inductive hypothesis holds for $k<p$, and that at the kth iteration of the generic primal-dual algorithm, γ^k is chosen as the constant $\bar\gamma\in(0,1)$ and the step sizes by some rule satisfying (2.14) and (2.15). Then search directions can be chosen as in (3.13)–(3.14) such that the inductive hypothesis remains true for $k+1$ as long as k satisfies (3.16) and (3.18).* $\square$

4 The Inductive Proof: γ Possibly Depending on k and/or n

In the above analysis, we made assumption (3.10) where α is fixed. Thus, we require k, the iteration count, to satisfy relations (3.16) and (3.18). For constant $\beta > 1$ and constant $0 < \bar{\gamma} < 1$ (independent of n), these relations are satisfied when n is large enough and $k \leq n^{1/3}$. However, in practice γ^k is often chosen as $1/\sqrt{n}$ or $1/n$ (see, e.g., [12]), and may also vary with k. Then, relation (3.18) is no longer satisfied for $k = n^{1/3}$, and α cannot be fixed. In this section, we analyze the case of $0 \leq \gamma^k \leq \bar{\gamma}$ where $\bar{\gamma}$ is a constant in (0,1), and we assume (3.16) still holds for $\bar{\gamma}$, so that it holds also with γ^k. Actually we need the slightly stronger assumption that

$$2k + 2 \leq \frac{(1-\bar{\gamma})^2 n}{4\max\{1, \beta(\bar{\gamma})^2\}}. \tag{4.1}$$

We make the following hypothesis:

$$\frac{\mu^k}{\beta} \leq (v_j^k)^2 = x_j^k s_j^k \leq \alpha^k \mu^k, \quad 1 \leq |j| \leq p = n/2, \tag{4.2}$$

where

$$\alpha^0 = \max\{1, \beta(\bar{\gamma})^2\}, \quad \text{and} \quad \alpha^{k+1} = \frac{\alpha^k}{1 - 2\sqrt{\alpha^k(2k+2)/n}}, \quad k \geq 0.$$

We claim that, although α^k gradually increases, with $\alpha := 4\max\{1, \beta(\bar{\gamma})^2\} \geq 4$ as in (3.11), we have

$$\alpha^{k+1} \leq \alpha \tag{4.3}$$

whenever

$$k + 1 \leq \delta n^{1/3}, \tag{4.4}$$

where δ is some constant in $(0, 1)$. Indeed, let us suppose that $n \geq 125$, and define $\delta := 25/(32\alpha) \leq 25/128$. Then, as long as $\alpha^k \leq \alpha$ and $k + 1 \leq \delta n^{1/3}$, we find

$$1 - 2\sqrt{\alpha^k(2k+2)/n} \geq 1 - 2\sqrt{\alpha(2k+2)/n} \geq 1 - 2\sqrt{\alpha[25/(16\alpha)]n^{-2/3}} \geq 1/2 \tag{4.5}$$

and thus

$$
\begin{aligned}
\frac{1}{1-2\sqrt{\alpha^k(2k+2)/n}} &\le \frac{1}{1-2\sqrt{\alpha(2k+2)/n}} \\
&= 1+\frac{2\sqrt{\alpha(2k+2)/n}}{1-2\sqrt{\alpha(2k+2)/n}} \\
&\le 1+\frac{2\sqrt{\alpha[25/(16\alpha)]n^{-2/3}}}{1/2} \\
&= 1+5n^{-1/3}. \qquad (4.6)
\end{aligned}
$$

Hence, as long as $\alpha^0, \alpha^1, \ldots, \alpha^k \le \alpha$ and $k+1 \le \delta n^{1/3}$,

$$\alpha^{k+1} \le \max\{1, \beta(\bar{\gamma})^2\}(1+5n^{-1/3})^{k+1},$$

and therefore (using (4.4))

$$
\begin{aligned}
\ln(\alpha^{k+1}) &\le \ln(\max\{1,\beta(\bar{\gamma})^2\}) + (k+1)\ln(1+5n^{-1/3}) \\
&\le \ln(\max\{1,\beta(\bar{\gamma})^2\}) + 5\delta \\
&\le \ln(\max\{1,\beta(\bar{\gamma})^2\}) + 125/128 \\
&\le \ln(4\max\{1,\beta(\bar{\gamma})^2\}) = \ln(\alpha).
\end{aligned}
$$

Then (4.3) follows by induction.

We have thus made the following **Inductive Hypothesis**:

At the beginning of the kth iteration, the iterates (x^k, s^k) and the previous search directions Ξ^k and Σ^k satisfy (3.1)–(3.5) and (4.2).

Again, we generate the same directions as before, and Theorem 3.1 establishes almost all of the inductive hypothesis with k replaced by $k+1$; we need only show that (4.2) holds for $k+1$. Note that the left-hand inequality of (4.2) holds from the restriction $(x^{k+1}, s^{k+1}) \in \mathcal{N}(\beta)$. Thus, we need to establish the right-hand inequality. Note that Lemma 3.1 remains true.

Let j be the index singled out in the lemma and let $\rho_P^k = \rho_D^k =: \rho^k =: 2\bar{\rho}^k$. Note that

$$\tilde{x}_j^+ = v_j^k + \bar{\rho}^k(-v_j^k + \gamma^k\mu^k/v_j^k - \epsilon_j) > 0.$$

Hence, using Lemma 3.1, relation (3.15), the inductive hypothesis

$$1/\sqrt{\alpha^k} \le \sqrt{\mu^k}/v_j^k \le \sqrt{\beta}$$

and assumption (4.1), we find that the denominator in (3.15) (with γ^k replacing $\bar{\gamma}$) satisfies

$$(1-\gamma^k)\sqrt{\frac{n}{(2k+2)}}\frac{\sqrt{\mu^k}}{v_j^k} + 1 - \gamma^k(\frac{\sqrt{\mu^k}}{v_j^k})^2 \ge (1-\gamma^k)\sqrt{\frac{n}{(2k+2)}}\frac{1}{\sqrt{\alpha^k}} + 1 - \frac{\gamma^k}{\alpha^k},$$

which is positive since $\gamma^k \le \bar{\gamma} < 1 \le \alpha^k$. Thus, as in (3.17),

$$\bar{\rho}^k \le \frac{1}{(1-\gamma^k)\sqrt{\frac{n}{\alpha^k(2k+2)}} + 1 - \gamma^k/\alpha^k} \le \frac{\sqrt{\alpha^k(2k+2)}}{(1-\gamma^k)\sqrt{n}}. \tag{4.7}$$

This implies that, as long as k satisfies (4.4),

$$\bar{\rho}^k \le \frac{\sqrt{\alpha(2k+2)}}{(1-\gamma^k)\sqrt{n}} \le 1, \tag{4.8}$$

where the last inequality follows from (4.1).

We are now ready to prove the right-hand inequality of (4.2) for $k+1$. Following the same reasoning as in the previous section, we only need to show

$$([1-\bar{\rho}^k]v_j^k + \bar{\rho}^k\gamma^k\mu^k(v_j^k)^{-1})^2 \le \alpha^{k+1}(1-2\bar{\rho}^k(1-\gamma^k))\mu^k,$$

or

$$[(1-\bar{\rho}^k)(v_j^k/\sqrt{\mu^k}) + \bar{\rho}^k\gamma^k(\sqrt{\mu^k}/v_j^k)]^2 \le \alpha^{k+1}(1-2\bar{\rho}^k(1-\gamma^k))$$

for each j, $1 \le |j| \le p$. Using (4.5) and (4.8), we find that, as long as k satisfies (4.4),

$$1 - 2\bar{\rho}^k(1-\gamma^k) \ge 1 - 2\sqrt{\alpha(2k+2)/n} \ge 1/2.$$

Note again from the inductive hypothesis that

$$1/\sqrt{\beta} \le v_j^k/\sqrt{\mu^k} \le \sqrt{\alpha^k},$$

and $(1-\bar\rho^k)(v_j^k/\sqrt{\mu^k}) + \bar\rho^k\gamma^k(\sqrt{\mu^k}/v_j^k)$ is convex in $v_j^k/\sqrt{\mu^k}$. Hence, it is sufficient to show

$$[(1-\bar\rho^k)\sqrt{\alpha^k} + \bar\rho^k\gamma^k/\sqrt{\alpha^k}]^2 \le \alpha^{k+1}(1-2\bar\rho^k(1-\gamma^k)) \tag{4.9}$$

and

$$[(1-\bar\rho^k)/\sqrt{\beta} + \bar\rho^k\gamma^k\sqrt{\beta}]^2 \le \alpha^{k+1}(1-2\bar\rho^k(1-\gamma^k)). \tag{4.10}$$

From (4.7) and $\gamma^k/\alpha^k \le 1$

$$\frac{(1-\bar\rho^k+\bar\rho^k\gamma^k/\alpha^k)^2\alpha^k}{1-2\bar\rho^k(1-\gamma^k)} \le \frac{\alpha^k}{1-2\sqrt{\alpha^k(2k+2)/n}} = \alpha^{k+1},$$

which proves (4.9). Similarly, we have

$$\frac{[(1-\bar\rho^k)/\sqrt{\beta} + \bar\rho^k\gamma^k\sqrt{\beta}]^2}{1-2\bar\rho^k(1-\gamma^k)} \le \frac{\max\{1/\beta, \beta(\gamma^k)^2\}}{1-2\sqrt{\alpha^k(2k+2)/n}} \le \frac{\max\{1, \beta(\gamma^k)^2\}}{1-2\sqrt{\alpha^k(2k+2)/n}} \le \alpha^{k+1},$$

since $\alpha^k \ge \alpha^0$, which proves (4.10). Thus (4.2) remains true for $k+1$ as long as relations (4.1) and (4.4) hold. We therefore have

Theorem 4.1 *Suppose the inductive hypothesis holds for $k < p$, and that at the kth iteration of the generic primal-dual algorithm, γ^k is chosen in $[0,\bar\gamma]$ for some $\bar\gamma \in (0,1)$ and the step sizes by some rule satisfying (2.14) and (2.15). Then search directions can be chosen as in (3.13)–(3.14) such that the inductive hypothesis remains true for $k+1$ as long as k satisfies (4.1) and (4.4).* □

We needed to assume $\bar\gamma < 1$ above so that (4.4) could hold, and this precludes centering steps with $\gamma^k = 1$. However, suppose we choose $\bar\gamma \in [1/2 + 1/(2\beta), 1)$ and at each iteration choose $\gamma^k \in [0,\bar\gamma]$ and step sizes by a rule as above, *or* $\gamma^k = 1$ and $\rho_P^k = \rho_D^k = 1$ ($\bar\rho^k = 1/2$), i.e., a full centering step. The decision as to whether to take a centering step can only depend on the current iterates x^k and s^k. We claim that once again the inductive hypothesis holds true for $k+1$ as long as k satisfies (4.1) and (4.4). Indeed, in the first case the proof is as above, while in the second we only need to establish (4.9) and (4.10). The first holds since the left-hand side is $[\sqrt{\alpha^k}/2 + 1/(2\sqrt{\alpha^k})]^2 \le \alpha^k$ while the right-hand side is α^{k+1}. The second holds because the left-hand side is $[\sqrt{\beta}/2 + 1/(2\sqrt{\beta})]^2 \le \beta(\bar\gamma)^2 \le \alpha^k$ by our assumption on $\bar\gamma$, and the right-hand side is again α^{k+1}. Thus full centering steps can be incorporated, as long as $\bar\gamma$ is chosen appropriately.

5 Conclusion

By Theorems 3.2 and 4.1, we can continue generating search directions satisfying the inductive hypothesis for $O(n^{1/3})$ iterations (in the case of constant γ, for $O(n)$ iterations). Moreover, for $m < n$, the first $\min\{m, n-m\}$ of these iterations will be consistent with some $m \times n$ matrix A by Theorem 2.1, and hence, if we choose $b = Ae$ and $c = A^\top y + e$ for any y, will be the directions obtained in the generic primal-dual algorithm applied to (P) starting at $x^0 = s^0 = e$.

Note that the case of constant γ is included in the analysis of the previous section, so for most of this section we rely on the arguments and make the stronger assumptions therein.

We now examine how the duality gap changes in the kth iteration, assuming that the inductive hypothesis holds and the search directions are given by (3.13) and (3.14). We again suppose the step sizes are chosen by some rule satisfying (2.14) and (2.15) so that $\rho_P^k = \rho_D^k =: \rho^k =: 2\bar{\rho}^k$. We assume that $n \geq 125$ and that k satisfies (4.1) and (4.4). Then, according to (2.13) and (4.8) we have

$$(x^{k+1})^\top s^{k+1} \geq \left(1 - 2\sqrt{\alpha(2k+2)/n}\right)(x^k)^\top s^k.$$

Now note that, as long as k satisfies (4.4), inequality (4.6) holds, and we have

$$\begin{aligned}
\ln((x^{k+1})^\top s^{k+1}) &\geq \ln((x^k)^\top s^k) - \ln(\frac{1}{1 - 2\sqrt{\alpha(2k+2)/n}}) \\
&\geq \ln((x^k)^\top s^k) - \ln(1 + 5n^{-1/3}) \\
&\geq \ln((x^k)^\top s^k) - 5n^{-1/3}.
\end{aligned}$$

Hence, for $K := \lfloor \min\{\delta n^{1/3}, (1-\bar{\gamma})^2 n/(8\alpha)\}\rfloor$, we have $\ln((x^K)^\top s^K) \geq \ln((x^0)^\top s^0) - 5\delta \geq \ln((x^0)^\top s^0) - 125/128$ and thus $(x^K)^\top s^K \geq (x^0)^\top s^0/3$. Assembling this inequality with Theorems 2.1, 3.1, and 4.1 yields

Theorem 5.1 *Consider the generic primal-dual algorithm that at each iteration chooses γ^k according to (2.16) in $[0, \bar{\gamma}]$ for some $\bar{\gamma} \in (0, 1)$ and employs a step size rule satisfying (2.14) and (2.15) so that the iterates lie in $\mathcal{N}(\beta)$ for some $\beta > 1$. Let*

$$\alpha := 4 \max\{1, \beta(\bar{\gamma})^2\}\}, \quad \delta := 25/(32\alpha),$$

and let

$$K := \lfloor \min\{\delta n^{1/3}, (1 - \bar{\gamma})^2 n/(8\alpha)\} \rfloor.$$

Then, if $m \leq K \leq n - m$, there is an instance of (P), with $A \in I\!R^{m \times n}$, $b = Ae \in I\!R^m$, and $c = A^\top y + e \in I\!R^n$ for any $y \in I\!R^m$, such that to decrease the duality gap by a factor of 3, starting from $x^0 = s^0 = e$, the algorithm requires at least $K = \Omega(n^{1/3})$ iterations.

Moreover, the discussion at the end of the last section shows that this result remains true (with $\bar{\gamma}$ sufficiently large) if periodic full centering steps are allowed, as long as these are triggered only by the current iterates.

Now suppose that γ is constant, independent of n. Then we no longer require (4.4) to establish the inductive claim. As long as k satisfies (3.16) and (3.18) (i.e., $k = O(n)$), the inductive hypothesis for k implies that it holds also for $k + 1$, and that $\bar{\rho}^k$ is bounded by (3.17). But then we can proceed exactly as in [21] (the paragraph including (4.4)) to see that $\Omega(n^{1/3}(t)^{2/3})$ iterations are required to decrease the duality gap by a factor of $\exp(t)$, and hence $\Omega(n^{1/3}(\log n)^{2/3})$ to decrease it by a factor of n.

References

[1] K. M. Anstreicher, On the performance of Karmarkar's algorithm over a sequence of iterations, SIAM J. Optim. 1(1991)22.

[2] K. M. Anstreicher, J. Ji, F. Potra and Y. Ye, Average performance of a self-dual interior-point algorithm for linear programming, in: *Complexity in Numerical Optimization*, ed. P. Pardalos (World Scientific, New Jersey, 1993) p. 1.

[3] D. Bertsimas and X. Luo, On the worst case complexity of potential reduction algorithms for linear programming, Working Paper 3558-93, Sloan School of Management, MIT, Cambridge, MA 02139, USA (1993).

[4] D. Goldfarb and M. J. Todd, Linear programming, in: *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, ed. G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd (North Holland, Amsterdam, The Netherlands, 1989).

[5] C. C. Gonzaga, Path following methods for linear programming, SIAM Rev. 34(1992)167.

[6] J. Ji and Y. Ye, A complexity analysis for interior-point algorithms based on Karmarkar's potential function, SIAM J. Optim. 4(1994)512.

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

[8] M. Kojima, N. Megiddo, and S. Mizuno, A primal-dual infeasible-interior-point algorithm for linear programming, Math. Progr. 61(1993)263.

[9] 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*, ed. N. Megiddo (Springer-Verlag, New York, 1989) p. 29.

[10] M. Kojima, S. Mizuno, and A. Yoshise, A polynomial-time algorithm for a class of linear complementarity problems, Math. Progr. 44(1989)1.

[11] M. Kojima, S. Mizuno, and A. Yoshise, An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems, Math. Progr. 50(1991)331.

[12] I. J. Lustig, R. E. Marsten, and D. F. Shanno, Computational experience with a primal-dual interior point method for linear programming, Lin. Alg. Appl. 152(1991)191.

[13] I. J. Lustig, R. E. Marsten, and D. F. Shanno, Computational experience with a globally convergent primal-dual predictor-corrector algorithm for linear programming, Math. Progr. 66(1994)123.

[14] S. Mehrotra and Y. Ye, On finding an interior point on the optimal face of linear programs, Math. Progr. 62(1993)497.

[15] S. Mizuno, M. J. Todd, and Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming, Math. of Oper. Res. 18(1993)964.

[16] R. D. C. Monteiro and I. Adler, Interior path following primal-dual algorithms: Part I: Linear programming, Math. Progr. 44(1989)27.

[17] A. S. Nemirovsky, An algorithm of the Karmarkar type, Tekhnicheskaya Kibernetika, 1(1987)105, translated in: Sov. J. Comp. Syst. Sci. 25(5)(1987)61.

[18] M. J. D. Powell, On the number of iterations of Karmarkar's algorithm for linear programming, Math. Progr. 62(1993)153.

[19] G. Sonnevend, J. Stoer, and G. Zhao, On the complexity of following the central path of linear programs by linear extrapolation II, Math. Progr. 52(1991)527.

[20] M. J. Todd, Recent developments and new directions in linear programming, in: *Mathematical Programming: Recent Developments and Applications*, ed. M. Iri and K. Tanabe (Kluwer Academic Press, Dordrecht, The Netherlands, 1989), p. 109.

[21] M. J. Todd, A lower bound on the number of iterations of primal-dual interior-point methods for linear programming, in: *Numerical Analysis 1993*, Volume 303 of *Pitman Research Notes in Mathematics*, ed. G. A. Watson and D. F. Griffiths (Longman Press, Burnt Hill, UK, 1994) p. 89.

[22] Y. Ye, Toward probabilistic analysis of interior-point algorithms for linear programming, Math. of Oper. Res. 19(1994)38.