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

TECHNICAL REPORT NO. 1050

March 1993

Revised June 1994

A Lower Bound on the Number of Iterations of Primal-Dual Interior-Point Methods for Linear Programming

by

Michael J. Todd

[1]Research supported in part by NSF, AFOSR, and ONR through NSF Grant DMS-8920550.

A Lower Bound on the Number of Iterations of Primal-Dual Interior-Point Methods for Linear Programming *

Michael J. Todd
School of Operations Research
and Industrial Engineering
Engineering and Theory Center
Cornell University
Ithaca, NY 14853

March, 1993

*Research supported in part by NSF, AFOSR, and ONR through NSF Grant DMS-8920550.

Abstract

We describe some elements of interior-point methods for linear programming. In contrast to Dantzig's well-known simplex method, these algorithms generate a sequence of points in the relative interior of the feasible region. They have some remarkable properties: certain of these methods have very attractive theoretical upper bounds on the number of iterations required ($O(\sqrt{n}t)$ iterations for problems with n inequalities to attain a precision of t additional digits), while others can be highly effective in solving large-scale problems. We analyze in detail one of these algorithms (the primal-dual affine-scaling method) that is very close to what is implemented in practice, and show that it may take at least $n^{1/3}$ iterations to improve the initial duality gap by a factor of twenty. We also discuss how far this analysis can be extended to other primal-dual interior-point methods.

One unusual feature of our approach is that we do not construct bad examples explicitly. Instead, our viewpoint is more like that of information-based complexity in nonlinear programming; we reveal to the algorithm at each iteration a bad pair of search directions, which may depend on the previous iterations, but we show that all our directions are consistent with some initial data for the linear programming problem.

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

Running Header: A Lower Bound for Interior-Point Algorithms

1 Introduction

For over forty years, Dantzig's simplex method has been the standard solution algorithm for linear programming problems, which seek to minimise a linear function of several variables subject to linear equations and inequalities [5]. In practice, its performance remains excellent – the number of iterations required is usually a small multiple of the number of equality constraints in a standard-form problem, even though there are examples (see for instance Klee and Minty [10]) needing an exponential number of iterations for many particular simplex pivoting rules. This exponential gap has been explained to some extent by theoretical analyses showing that the expected number of iterations required by some simplex variants on random problems generated from certain probability distributions is polynomial in the dimensions of the problem instance; for a survey, see Borgwardt [4].

In the last eight years, there has been enormous activity in optimisation in the field of interior-point methods for linear programming and extensions – the bibliography of Kranich [13] lists 1303 items. This explosion of research was instigated by the work of Karmarkar [9], 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., Bixby et al. [3] and Lustig et al. [15]. For an overview of such methods the reader is referred to [6, 7, 27].

The most effective interior-point methods computationally are primal-dual methods, and these are variants of polynomial-time algorithms having the best complexity theoretically also. The latter methods, either path-following methods (see, e.g., Kojima et al. [11], Monteiro and Adler [19], and Gonzaga [7]) or potential-reduction methods (see Kojima et al. [12]), require $O(\sqrt{n}t)$ iterations to attain an additional t digits of accuracy in a problem with n inequality constraints, given a sufficiently "centered" pair of initial primal and dual strictly feasible points. On the other hand, computational experience with sophisticated primal-dual interior-point codes suggests that the number of iterations necessary grows much more slowly with the dimension n. Early papers cited an almost constant number of iterations to solve a range of small to reasonably large problems, while the results of Lustig, Marsten, and Shanno on problems with n up to two million suggested that the growth was logarithmic in n [14]. Once again there is an exponential gap between observed performance and theoretical bounds, even though the

latter are now polynomial. We seek to investigate this gap.

There have been attempts to study the "expected" number of iterations theoretically. 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. Nemirovsky [21] for a Karmarkar-like method and Mizuno et al. [17] for a primal-dual wide-neighbourhood method derive "anticipated" bounds growing only logarithmically with n.

Instead, we seek here to understand whether the theoretical upper bounds are close to tight; perhaps a better analysis would yield worst-case bounds nearer to what is observed in practice. Thus we look for *lower* bounds on the number of iterations required. Such bounds have been investigated before, mainly for Karmarkar's original projective-scaling method. Thus Anstreicher [1] showed that $\Omega(\ln(n))$ iterations might be necessary to obtain a fixed improvement in the objective function value. Ji and Ye [8] improved Anstreicher's analysis and obtained a bound of $\Omega(n)$ iterations from a starting point quite close to the boundary. Powell [24] (see also [23]) also derived a lower bound of $\Omega(n)$ iterations for a discretisation of a semi-infinite problem, again using a starting point quite close to the boundary. These results suggest that Karmarkar's result (an upper bound of $O(nt)$ iterations) may be essentially tight.

Very recently, Bertsimas and Luo [2] considered algorithms reducing the Tanabe-Todd-Ye primal-dual potential function, and showed that the $O(\sqrt{n}t)$-iteration bound is tight by proving a similar lower bound. The algorithms they consider are "primal-or-dual" methods that are not symmetric between the primal and dual, and update an iterate in just one of these problems at each iteration.

All these papers construct a specific problem on which the algorithm performs particularly poorly; none addresses the currently popular methods used in implementations – symmetric primal-dual algorithms.

By contrast, Sonnevend et al. [26] discuss a particular problem for which the primal-dual central trajectory has large "total curvature," which shows that primal-dual algorithms that follow this trajectory closely will require $\Omega(n^{1/3})$ iterations. However, most practical algorithms use much longer step

sizes, and their iterates do not stay close to the central path.

In this paper we also obtain a bound of at least $n^{1/3}$ iterations to obtain a constant factor decrease in the duality gap. The algorithm we study is the primal-dual affine-scaling algorithm, which is very close to the methods used in practical implementations. We also allow 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 trajectory. However, to give the algorithms every benefit, we start at points on the central path. Note also that, since our results are lower bounds, they apply also to more general methods, e.g. those that allow infeasible iterates, and more general problems, e.g. convex quadratic programming, as long as the algorithms reduce to the studied method for feasible starts and linear programming instances.

We also discuss how far our results extend to other primal-dual interior-point methods that use directions including some centering component. Many practical algorithms include such a component to a small degree to keep the iterates from approaching the boundary too closely prematurely. Our discussion indicates that the lower bounds we obtain for the affine-scaling method can often be expected to hold for other algorithms also.

In contrast to previous constructions of lower bounds, we do not give explicit problems that cause the algorithm to take many iterations. Instead, our approach is much closer to that used in analyses of the informational complexity of nonlinear optimisation (see Nemirovsky and Yudin [22]), where an oracle is assumed to generate information about the problem instance at each iteration. In this view, the process can be viewed as a game between the oracle, which tries to generate information about the problem instance as unhelpfully as possible, and the algorithm, which uses this information as efficiently as possible and tries to ask the oracle questions which will severely limit the set of possible problem instances. (Think of the bisection algorithm to determine a zero of a continuous function of a single variable.) This is an unnatural view to take of linear programming, where the problem instance is determined by the finitely many real numbers in the data. But interior-point methods, in determining their steps at each iteration, use remarkably little of this information; the search directions are just certain projections of appropriate vectors, and a wide range of problem instances will lead to the same pair of search directions. Thus our approach is to generate, for a long sequence of iterations, a pair of singularly unhelpful search directions which

lead to little improvement. Then we show that there is a problem instance that will give rise to exactly these search directions. (Of course it is necessary that no other part of the algorithm "look at" the data; hence we require that the step size rule depend only on the current iterates and search directions.)

The paper is organised as follows. In the next section, we describe the class of algorithms under discussion, show how the duality gap is reduced at each iteration, and consider possible step size rules. One particularly important algorithm in this class is the so-called primal-dual affine-scaling method, to which we confine ourselves for most of the paper. The section ends with a theorem giving sufficient conditions for a sequence of pairs of search directions to arise from some problem instance.

Section 3 provides the main tool, showing inductively that a sequence of particularly unpleasant search directions might occur. In Section 4, we demonstrate that these search directions do lead to very slight reduction of the duality gap, and hence obtain our lower bound for the primal-dual affine-scaling algorithm.

In Section 5, we argue that, as long as the iterates satisfy a certain condition, the lower bound remains valid for other primal-dual methods with a centering component in the search directions. We verify that this condition seems to hold for many such algorithms by making some computational tests in MATLAB [18].

We must mention that these results are mainly theoretical. For n less than a few billion, there is little difference between logarithmic growth in n and growth like $n^{1/3}$. Indeed, a good fit to the iteration counts in [14] can be obtained using $n^{1/3}$ instead of $\ln(n)$. The correlation coefficient in [14] using a logarithmic fit was $R^2 = 0.979$; using a fit to $n^{1/3}$ gives instead $R^2 = 0.952$.

2 Primal-Dual Interior-Point Algorithms

We consider the primal linear programming problem in standard form

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

with dual problem

$$(D) \qquad \begin{aligned} \max_{y,s} \quad & b^T y \\ A^T y + s &= c \\ s &\geq 0, \end{aligned}$$

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

$$c^T x - b^T y = x^T s \geq 0; \tag{2.1}$$

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

As long as (P) is feasible, (2.1) shows that the objective function value of (y, s) is determined from s alone, and we will view our algorithms as generating a sequence of pairs (x^k, s^k) lying in

$$\begin{aligned} \mathcal{F}^0 := \{(x, s) \in \mathbb{R}^{2n} : Ax = b, A^T y + s = c \\ \text{for some } y,\ x > 0,\ s > 0\}; \end{aligned} \tag{2.2}$$

the strict inequalities are the reason for the name "interior-point algorithms." It is easy to update y^k along with s^k so that (y^k, s^k) remains feasible in (D), but we will suppress this for simplicity.

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 $(x^0, s^0) \in \mathcal{F}^0$.

The pair (x^k, s^k) will be updated at each iteration by taking a damped perturbed Newton step for the optimality conditions

$$Ax = b, \tag{2.3}$$

$$A^T y + s = c, \tag{2.4}$$

$$XSe = 0, \tag{2.5}$$

$$x \geq 0, \qquad s \geq 0, \tag{2.6}$$

where e denotes $(1,...,1)^T \in \mathbb{R}^n$ and X and S the diagonal matrices with $Xe = x$, $Se = s$. The damping is carried out to maintain positivity in $\{(x^k, s^k)\}$. Frequently, in order to encourage the possibility of a full step, while maintaining positivity, a damped Newton step is taken for the perturbed system

$$\begin{aligned} Ax &= b, && (2.7)\\ A^T y + s &= c, && (2.8)\\ XSe &= \gamma\mu e, && (2.9)\\ x \geq 0, &\quad s \geq 0, && (2.10) \end{aligned}$$

where $\mu = (x^k)^T s^k/n$ and $0 \leq \gamma < 1$. The set of solutions to (2.7)–(2.10), for $0 < \gamma < \infty$, forms the so-called central trajectory for (P) and (D); path-following algorithms are based on approximately following this path as γ decreases to 0.

The Newton direction for (2.7)–(2.9) at (x^k, s^k) is the solution to

$$\begin{aligned} A\xi &= 0, && (2.11)\\ A^T\eta + \sigma &= 0, && (2.12)\\ X^k\sigma + S^k\xi &= \gamma\mu e - X^k S^k e. && (2.13) \end{aligned}$$

(Superscripts are used throughout for indices; nonnegative integer powers are indicated by enclosing their arguments in parentheses.) The form of the solution to (2.11)–(2.13) can be made more apparent by the following scaling. Let

$$\begin{aligned} D^k &:= (X^k)^{1/2}(S^k)^{-1/2}, \quad V^k := (X^k)^{1/2}(S^k)^{1/2}, && (2.14)\\ v &:= v^k := V^k e, \quad w := v - \gamma\mu(V^k)^{-1}e, && (2.15)\\ \tilde{A} &:= AD^k, \quad \tilde{\xi} := (D^k)^{-1}\xi, \quad \tilde{\sigma} := D^k\sigma. && (2.16) \end{aligned}$$

Then (2.11)–(2.13) is equivalent to

$$\begin{aligned} \tilde{A}\tilde{\xi} &= 0, && (2.17)\\ \tilde{A}^T\eta + \tilde{\sigma} &= 0, && (2.18)\\ \tilde{\sigma} + \tilde{\xi} &= -w, && (2.19) \end{aligned}$$

whose solution can be written as

$$\tilde{\xi} = -P_{\tilde{A}} w, \quad \tilde{\sigma} = -(I - P_{\tilde{A}}) w, \tag{2.20}$$

where P_M denotes the matrix that projects a vector orthogonally onto the null space of M. For future reference, we note that

$$P_M = I - \bar{M}^T (\bar{M}\bar{M}^T)^{-1} \bar{M}, \tag{2.21}$$

where the rows of $\bar{M}$ form a basis for the row 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. If $\gamma = 0$, i.e., if we do not perturb (2.5) to (2.9), then $w = v$ and our steps in scaled space are then $\tilde{\xi} = -P_{\tilde{A}} v$, $\tilde{\sigma} = -(I - P_{\tilde{A}}) v$.

We can now state our generic primal-dual interior-point algorithm:

Algorithm

Choose $(x^0, s^0) \in \mathcal{F}^0$ and set $k = 0$.
While $(x^k)^T s^k > \tau$ do

begin

choose $0 \leq \gamma^k \leq 1$;

set

$$\begin{aligned}
\mu^k &:= (x^k)^T s^k / n; && (2.22)\\
D &:= D^k := (X^k)^{1/2} (S^k)^{-1/2}, \quad \tilde{A} := AD; && (2.23)\\
V &:= V^k := (X^k)^{1/2} (S^k)^{1/2}; && (2.24)\\
v &:= v^k := Ve, \quad w := w^k := v - \gamma^k \mu^k V^{-1} e; && (2.25)\\
\tilde{x} &:= \tilde{x}^k := D^{-1} x^k = v, \quad \tilde{s} := \tilde{s}^k := Ds^k = v; && (2.26)
\end{aligned}$$

compute

$$\begin{aligned}
\tilde{\xi} &:= \tilde{\xi}^k := -P_{\tilde{A}} w, \quad \tilde{\sigma} := \tilde{\sigma}^k := -(I - P_{\tilde{A}}) w; && (2.27)\\
\xi^k &:= D\tilde{\xi}, \quad \sigma^k = D^{-1} \tilde{\sigma}^k; && (2.28)
\end{aligned}$$

choose $\rho_P^k > 0$ and $\rho_D^k > 0$ so that

$$x^{k+1} := x^k + \rho_P^k \xi^k > 0, \quad s^{k+1} := s^k + \rho_D^k \sigma^k > 0; \tag{2.29}$$

and set $k := k + 1$.

end

This general framework includes most primal-dual methods. For instance, short-step path-following algorithms (e.g., [11, 19]) arise by taking $\gamma^k = 1 - \kappa/\sqrt{n}$ for some fixed κ and $\rho_P^k = \rho_D^k = 1$ for all k; the version of OB1 described in [15] chooses $\gamma^k = 1/n$ or $1/\sqrt{n}$ and ρ_P^k and ρ_D^k equal to .995 of their maximum values. Of particular interest to us are the primal-dual affine-scaling methods, which choose $\gamma^k = 0$ at each iteration, with various choices for the step sizes ρ_P^k and ρ_D^k [20, 16, 28]. The parameter τ in the algorithm is the stopping criterion; for theoretical purposes, we assume it is zero (or sufficiently small) in the analysis.

Proposition 2.1 *If $\rho_P^k = \rho_D^k = \rho^k$, then*

$$(x^{k+1})^T s^{k+1} = \{1 - \rho^k(1 - \gamma^k)\}(x^k)^T s^k. \tag{2.30}$$

Proof. Note that $\tilde{\xi}^T \tilde{\sigma} = 0$ ($\tilde{\xi}$ lies in the null space of $\tilde{A}$, while $\tilde{\sigma}$ lies in its row space). It is convenient to write

$$\tilde{x}^+ := (D^k)^{-1} x^{k+1} = \tilde{x} + \rho_P^k \tilde{\xi} \tag{2.31}$$

and

$$\tilde{s}^+ := D^k s^{k+1} = \tilde{s} + \rho_D^k \tilde{\sigma}. \tag{2.32}$$

Then

$$\begin{aligned}
(x^{k+1})^T s^{k+1} &= (\tilde{x}^+)^T \tilde{s}^+ \\
&= (\tilde{x} + \rho_P^k \tilde{\xi})^T (\tilde{s} + \rho_D^k \tilde{\sigma}) \\
&= \tilde{x}^T \tilde{s} + \rho^k (\tilde{x}^T \tilde{\sigma} + \tilde{\xi}^T \tilde{s}) + (\rho^k)^2 \tilde{\xi}^T \tilde{\sigma} \\
&= \tilde{x}^T \tilde{s} + \rho^k (v^T (\tilde{\sigma} + \tilde{\xi}))
\end{aligned}$$

$$
\begin{aligned}
&= \tilde{x}^T\tilde{s} + \rho^k(-v^Tw) \\
&= \tilde{x}^T\tilde{s} + \rho^k(-v^Tv + \gamma^k\mu^k e^Te) \\
&= \tilde{x}^T\tilde{s} + \rho^k(-1+\gamma^k)\tilde{x}^T\tilde{s} \\
&= \{1-\rho^k(1-\gamma^k)\}(x^k)^Ts^k.
\end{aligned}
$$

Here we have used $\mu^k e^Te = \mu^k n = (x^k)^Ts^k = \tilde{x}^T\tilde{s}$. □

Hence to achieve a large decrease in the duality gap, we would like ρ^k to be large and γ^k to be small. Most of the results of this paper concern the affine-scaling algorithm, for which $\gamma^k = 0$ for all k. Then the algorithm will perform poorly if ρ^k is small. 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 Proposition 2.1 is unduly limiting on the algorithm we consider; indeed, practical algorithms of this type allow different step sizes in the primal and dual problems. However, under a very natural symmetry property, the examples we construct will have the property that $\rho_P^k = \rho_D^k$. Specifically, we assume that

$$
\rho_P^k = \rho^{(k)}(x^k, \xi^k) \quad \text{and} \quad \rho_D^k = \rho^{(k)}(s^k, \sigma^k), \tag{2.33}
$$

where the function $\rho^{(k)}$ satisfies the property

$$
\rho^{(k)}(\Pi x^k, \Pi\xi^k) = \rho^{(k)}(x^k, \xi^k) \tag{2.34}
$$

for every x^k, ξ^k and every $n \times n$ permutation matrix Π. Thus the step sizes depend only on the current iterate and search direction (and possibly the iteration number) and are symmetric between primal and dual and between different components. Hence if $s^k = \Pi x^k$ and $\sigma^k = \Pi\xi^k$ for some permutation Π, (2.33) and (2.34) imply

$$
\rho_P^k = \rho_D^k. \tag{2.35}
$$

We assume until Section 5 that

$$
\gamma^k = 0 \quad \text{for all} \quad k, \tag{2.36}
$$

so that we are considering an affine-scaling algorithm.

We describe below a particular step size rule satisfying (2.33) and (2.34). Suppose at some iteration that $\tilde{\xi} = 0$, so that $\tilde{\sigma} = -v$. Then $\rho_P^k = \rho_D^k = 1$ leads to $x^{k+1} = x^k$, $s^{k+1} = 0$, with $(x^{k+1})^\top s^{k+1} = 0$. It follows that x^{k+1} and s^{k+1} are optimal in (P) and (D) and the algorithm stops. But if $\tilde{\xi} = 0$, v lies in the row space of $\tilde{A}$, whence s^k and so c lie in the row space of A. In this case, at the initial iteration we would find v in the row space of $\tilde{A}$, so the algorithm would terminate immediately. Similarly, $\tilde{\sigma} = 0$ at some iteration implies $\tilde{\xi} = -v$, and then $\rho_P^k = \rho_D^k = 1$ leads to $x^{k+1} = 0$, $s^{k+1} = s^k$, and x^{k+1} and s^{k+1} are then again optimal in (P) and (D). If this happens, v lies in the null space of $\tilde{A}$, so $b = \tilde{A}v = 0$. Again, the initial iteration would then find $\tilde{\sigma} = 0$ and the algorithm would terminate immediately.

Let us therefore assume that at each iteration $\tilde{\xi} \neq 0$ and $\tilde{\sigma} \neq 0$. Then $v^\top \tilde{\xi} = -v^\top P_{\tilde{A}} v = -\|\tilde{\xi}\|^2 < 0$ shows that $\tilde{\xi}$ has a negative component, and similarly so does $\tilde{\sigma}$. Then we can write

$$\rho_P^k = \lambda_P^k \min_{j:\xi_j^k<0} \left\{ \frac{x_j^k}{-\xi_j^k} \right\}, \quad \rho_D^k = \lambda_D^k \min_{j:\sigma_j^k<0} \left\{ \frac{s_j^k}{-\sigma_j^k} \right\}, \tag{2.37}$$

for positive λ's, and (2.29) holds iff $\lambda_P^k, \lambda_D^k < 1$. The λ's represent the proportion of the way to the boundary of the feasible region taken by the step. Note that (2.33) and (2.34) hold if $\lambda_P^k = \lambda_D^k = \lambda^k$ depends only on the iteration k. (The code OB1 [15] chooses $\lambda^k = .995$ for all k.)

The only obvious restriction on the directions $\tilde{\xi}$ and $\tilde{\sigma}$ is that they be orthogonal and sum to $-v$. The following example from [17] shows that at least one of ρ_P^k, ρ_D^k could be forced to be very small:

$$\begin{aligned} \tilde{x} &:= \tilde{s} := v := e, \\ \tilde{\xi} &:= -(1+\sqrt{n}, 1, \ldots, 1)^\top/2, \\ \tilde{\sigma} &:= -(1-\sqrt{n}, 1, \ldots, 1)^\top/2. \end{aligned}$$

Then ρ_P^k must be less than $2(1+\sqrt{n})^{-1}$ in order that $\tilde{x}^+$ be positive. In order to force both ρ_P^k and ρ_D^k to be small, we modify the example slightly:

$$\tilde{x} := \tilde{s} := v := e, \tag{2.38}$$

$$\tilde{\xi} := -(1+\sqrt{n/2}, 1-\sqrt{n/2}, 1, \ldots, 1)^\top/2, \tag{2.39}$$

$$\tilde{\sigma} := -(1-\sqrt{n/2}, 1+\sqrt{n/2}, 1, \ldots, 1)^\top/2. \tag{2.40}$$

Then any step size rule satisfying (2.33) and (2.34) must yield $\rho_P^k = \rho_D^k =: \rho^k$, and feasibility demands $\rho^k < 2(1 + \sqrt{n/2})^{-1}$. Proposition 2.1 then gives

$$(x^{k+1})^\top s^{k+1} > \left(1 - \frac{2}{1 + \sqrt{n/2}}\right) (x^k)^\top s^k. \tag{2.41}$$

Suppose that in fact this happens at the initial iteration, with $x^0 = s^0 = e$. It is not hard to construct a matrix A so that (2.39)–(2.40) hold, so that (2.41) implies a very small decrease in the duality gap. Can we continue to choose "bad" directions for a long sequence of iterations? We will show in the next two sections that this is indeed possible. To conclude this section we present a result giving sufficient conditions for a sequence of directions to be produced by the affine-scaling algorithm.

Suppose at the kth iteration we have $\tilde{\xi} + \tilde{\sigma} = -v$. A sufficient condition for $\tilde{\xi}$ to be $-P_{\tilde{A}} v$ and $\tilde{\sigma}$ to be $-(I - P_{\tilde{A}})v$ 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.

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

$$A\xi^j = 0,\ A^\top \eta^j + \sigma^j = 0,\ 0 \le j \le k, \tag{2.42}$$

is that

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

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

Proof. Let the columns of $\hat{\Xi}$ and $\hat{\Sigma}$ form bases for the column spaces of Ξ and Σ respectively, and suppose they have $r_\xi \le k+1$ and $r_\sigma \le k+1$ columns. By (ii), the orthogonal complement of the column space of Ξ (which is the same as that of $\hat{\Xi}$) is an $(n - r_\xi)$-dimensional subspace of $\mathbb{R}^n$ containing the r_σ linearly independent columns of $\hat{\Sigma}$ (which span those of Σ). Extend these r_σ vectors to a basis for the subspace, and let the m rows of A consist of m of these basis vectors including the r_σ columns of $\hat{\Sigma}$ ($r_\sigma \le k + 1 \le m \le n - k - 1 \le n - r_\xi$ from (i)). Then (2.42) clearly holds. □

3 The Inductive Step

In this section and the next we demonstrate how the bad behaviour of (2.39)–(2.41) can continue for several iterations. 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 $1, -1, 2, -2, \ldots, p, -p$.

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 $(e, e) \in \mathcal{F}^0$, and let the initial iterates be

$$x^0 = s^0 = e. \tag{3.1}$$

We assume that at the beginning of the kth iteration we have $(x^k, s^k) \in \mathcal{F}^0$ satisfying symmetry of the pairs,

$$\bar{\Pi} x^k = s^k, \tag{3.2}$$

and equality of the final components,

$$\Pi x^k = x^k, \ \Pi s^k = s^k, \text{ for all } \Pi \in S_n^k. \tag{3.3}$$

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

$$(\Xi^k)^T \Sigma^k = 0, \tag{3.4}$$

$$\bar{\Pi} \Xi^k = \Sigma^k, \tag{3.5}$$

and, for each $0 \leq j \leq k$,

$$\Pi \xi^{j-1} = \xi^{j-1}, \ \Pi \sigma^{j-1} = \sigma^{j-1}, \text{ for all } \Pi \in S_n^j. \tag{3.6}$$

Here, (3.4) ensures that the previous search directions are consistent with some matrix A (see Theorem 2.1), while (3.5) maintains the symmetry between pairs of components. Finally, (3.6) 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 $x^0 = s^0 = e$, (3.2) and (3.3) hold for $k = 0$, while (3.4)–(3.6) hold vacuously. Also, note that, if Ξ^k and Σ^k contain the single columns (2.39) and (2.40), (3.4)–(3.6) hold for $k = 1$.

We next examine the effect of these assumptions on X^k, S^k, D^k and $v = v^k$. From (3.2) and (3.3),

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

We then find

$$\begin{aligned} \bar{\Pi} D^k \bar{\Pi} &= \bar{\Pi}(X^k)^{1/2}\bar{\Pi}\ \bar{\Pi}(S^k)^{-1/2}\bar{\Pi} \\ &= (S^k)^{1/2}(X^k)^{-1/2} = (D^k)^{-1}, \end{aligned} \tag{3.7}$$

and similarly

$$\Pi D^k \Pi^\top = D^k \text{ for all } \Pi \in S_n^k. \tag{3.8}$$

In the same way,

$$\begin{aligned} \bar{\Pi} v^k &= \bar{\Pi}(X^k)^{1/2}\bar{\Pi}\ \bar{\Pi}(S^k)^{1/2}\bar{\Pi}\ \bar{\Pi} e \\ &= (S^k)^{1/2}(X^k)^{1/2} e = v^k, \end{aligned} \tag{3.9}$$

and

$$\Pi v^k = v^k, \text{ for all } \Pi \in S_n^k. \tag{3.10}$$

Our final assumption is

$$\max\{v_j^k, v_{-j}^k : 1 \le j \le k\} \le v_{k+1}^k = v_{-k-1}^k = \ldots = v_p^k = v_{-p}^k. \tag{3.11}$$

(The equalities follow from (3.10).) We shall see the importance of this assumption later. We have thus made the

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 (3.2)–(3.6) and (3.11).

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

$$\begin{aligned} \tilde{x}^k &:= (D^k)^{-1}x^k = v^k, && (3.12)\\ \tilde{s}^k &:= D^k s^k = v^k, && (3.13)\\ \tilde{\Xi}^k &:= (D^k)^{-1}\Xi^k, \quad \text{and} && (3.14)\\ \tilde{\Sigma}^k &:= D^k\Sigma^k. && (3.15) \end{aligned}$$

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

$$(\tilde{\Xi}^k)^\top\tilde{\Sigma}^k = 0 \tag{3.16}$$

from (3.4), and (3.5) and (3.7) show that they satisfy

$$\bar{\Pi}\tilde{\Xi}^k = \tilde{\Sigma}^k. \tag{3.17}$$

From (3.6) we deduce $\Pi\Xi^k = \Xi^k$ and $\Pi\Sigma^k = \Sigma^k$ for all $\Pi \in S_n^k$, so (3.8) shows

$$\Pi\tilde{\Xi}^k = \tilde{\Xi}^k, \quad \Pi\tilde{\Sigma}^k = \tilde{\Sigma}^k, \tag{3.18}$$

for all such Π's.

The purpose of this section is to show that we can choose the next directions in the form

$$\begin{aligned} \tilde{\xi}^k &= -\frac{1}{2}(v^k + (\epsilon_1, \epsilon_{-1}, \ldots, \epsilon_{k+1}, \epsilon_{-k-1}, 0, \ldots, 0)^\top), && (3.19)\\ \tilde{\sigma}^k &= -\frac{1}{2}(v^k - (\epsilon_1, \epsilon_{-1}, \ldots, \epsilon_{k+1}, \epsilon_{-k-1}, 0, \ldots, 0)^\top), && (3.20) \end{aligned}$$

where $\epsilon_j + \epsilon_{-j} = 0$, $1 \le j \le k+1$. 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 *As long as $k+1 < p = n/2$ and the inductive hypothesis holds, we can choose $\tilde{\xi}^k$ and $\tilde{\sigma}^k$ as in (3.19)–(3.20) so that $\Xi := [\tilde{\Xi}^k, \tilde{\xi}^k]$ and $\Sigma := [\tilde{\Sigma}^k, \tilde{\sigma}^k]$ satisfy $\Xi^\top\Sigma = 0$.*

Proof. For simplicity, we write P_ξ for $P_{(\tilde{\Xi}^k)^\top}$ and P_σ for $P_{(\tilde{\Sigma}^k)^\top}$. Set

$$\begin{aligned} \tilde{\xi}^{(1)} &:= -P_\sigma v^k, \quad \tilde{\xi}^{(2)} := -(I - P_\xi)v^k, \\ \tilde{\sigma}^{(1)} &:= -P_\xi v^k, \quad \tilde{\sigma}^{(2)} := -(I - P_\sigma)v^k. \end{aligned}$$

Then $(\tilde{\Sigma}^k)^\top \tilde{\xi}^{(1)} = 0$ and, since $\tilde{\xi}^{(2)}$ lies in the column space of $\tilde{\Xi}^k$, (3.4) shows that $(\tilde{\Sigma}^k)^\top \tilde{\xi}^{(2)} = 0$. Similarly, $(\tilde{\Xi}^k)^\top \tilde{\sigma}^{(1)} = (\tilde{\Xi}^k)^\top \tilde{\sigma}^{(2)} = 0$. Thus, if we set $\tilde{\xi}^{(3)} := \frac{1}{2}(\tilde{\xi}^{(1)} + \tilde{\xi}^{(2)})$ and $\tilde{\sigma}^{(3)} := \frac{1}{2}(\tilde{\sigma}^{(1)} + \tilde{\sigma}^{(2)})$, we have

$$(\tilde{\Sigma}^k)^\top \tilde{\xi}^{(3)} = (\tilde{\Xi}^k)^\top \tilde{\sigma}^{(3)} = 0. \tag{3.21}$$

Also, from the definitions,

$$\tilde{\xi}^{(3)} + \tilde{\sigma}^{(3)} = -v^k. \tag{3.22}$$

Let $\hat{\Xi}^k$ have as columns a column basis for $\tilde{\Xi}^k$. Then from (3.17) we deduce that $\hat{\Sigma}^k = \bar{\Pi}\hat{\Xi}^k$ has as columns a column basis for $\tilde{\Xi}^k$. Using (2.21) we find

$$\begin{aligned} \bar{\Pi}\tilde{\xi}^{(1)} &= -\bar{\Pi}v^k + \bar{\Pi}\hat{\Sigma}^k((\hat{\Sigma}^k)^\top \hat{\Sigma}^k)^{-1}(\hat{\Sigma}^k)^\top v^k \\ &= -\bar{\Pi}v^k + (\bar{\Pi}\hat{\Sigma}^k)((\bar{\Pi}\hat{\Sigma}^k)^\top(\bar{\Pi}\hat{\Sigma}^k))^{-1}(\bar{\Pi}\hat{\Sigma}^k)^\top \bar{\Pi}v^k. \end{aligned}$$

But $\bar{\Pi}v^k = v^k$ by (3.9) and $\bar{\Pi}\hat{\Sigma}^k = \hat{\Xi}^k$ so $\bar{\Pi}\tilde{\xi}^{(1)} = -P_\xi v^k = \tilde{\sigma}^{(1)}$. Similarly, $\bar{\Pi}\tilde{\xi}^{(2)} = \tilde{\sigma}^{(2)}$, and thus

$$\bar{\Pi}\tilde{\xi}^{(3)} = \tilde{\sigma}^{(3)}. \tag{3.23}$$

An analogous argument using (3.10) and (3.18) shows that $\Pi\tilde{\xi}^{(1)} = \tilde{\xi}^{(1)}$, and similarly $\Pi\tilde{\xi}^{(2)} = \tilde{\xi}^{(2)}$ and $\Pi\tilde{\sigma}^{(i)} = \tilde{\sigma}^{(i)}$, $i = 1, 2$, for all $\Pi \in S_n^k$, so that

$$\Pi\tilde{\xi}^{(3)} = \tilde{\xi}^{(3)}, \quad \Pi\tilde{\sigma}^{(3)} = \tilde{\sigma}^{(3)} \text{ for all } \Pi \in S_n^k. \tag{3.24}$$

We can now almost choose $\tilde{\xi}^k = \tilde{\xi}^{(3)}$, $\tilde{\sigma}^k = \tilde{\sigma}^{(3)}$. Indeed, (3.22), (3.23) and (3.24) show that they have the correct form (3.19) and (3.20) (with $\epsilon_{k+1} = \epsilon_{-k-1} = 0$), while (3.4) and (3.21) show that we only need $(\tilde{\xi}^{(3)})^\top \tilde{\sigma}^{(3)} = 0$ to satisfy the conclusions of the theorem. Now by their definitions,

$$
\begin{aligned}
(\tilde{\xi}^{(1)})^T\tilde{\sigma}^{(1)} &= (v^k)^T P_\sigma P_\xi v^k, \\
(\tilde{\xi}^{(1)})^T\tilde{\sigma}^{(2)} &= (v^k)^T P_\sigma (I - P_\sigma) v^k = 0, \\
(\tilde{\xi}^{(2)})^T\tilde{\sigma}^{(1)} &= (v^k)^T (I - P_\xi) P_\xi v^k = 0, \\
(\tilde{\xi}^{(2)})^T\tilde{\sigma}^{(2)} &= (v^k)^T (I - P_\xi)(I - P_\sigma) v^k.
\end{aligned}
$$

Since $I - P_\xi = R(\tilde{\Xi}^k)^T$ for some R and $I - P_\sigma = \tilde{\Sigma}^k S$ for some S, (3.4) shows that $(\tilde{\xi}^{(2)})^T\tilde{\sigma}^{(2)} = 0$ also. Now, since $(\tilde{\Xi}^k)^T\tilde{\Sigma}^k = 0$,

$$
P_\sigma P_\xi = P_M =: P_{\xi\sigma} \quad \text{for} \quad M = [\tilde{\Xi}^k, \tilde{\Sigma}^k]^T,
$$

and

$$
\begin{aligned}
(\tilde{\xi}^{(3)})^T\tilde{\sigma}^{(3)} &= \frac{1}{4}(\tilde{\xi}^{(1)})^T\tilde{\sigma}^{(1)} \\
&= \frac{1}{4}(v^k)^T P_{\xi\sigma} v^k \\
&= \frac{1}{4}\|P_{\xi\sigma} v^k\|^2 \geq 0. \qquad (3.25)
\end{aligned}
$$

We then define

$$
\begin{aligned}
\tilde{\xi}^k &:= \tilde{\xi}^{(3)} + (0, \ldots, 0, -\epsilon, +\epsilon, 0, \ldots, 0)^T, \\
\tilde{\sigma}^k &:= \tilde{\sigma}^{(3)} + (0, \ldots, 0, +\epsilon, -\epsilon, 0, \ldots, 0)^T,
\end{aligned}
$$

where the epsilons occur in the $(k+1)$st pair of components and where $\epsilon \geq 0$ is chosen so that $(\tilde{\xi}^k)^T\tilde{\sigma}^k = 0$. (Thus, $\epsilon = (\frac{1}{8})^{1/2}\|P_{\xi\sigma} v^k\|$.) Then $\tilde{\xi}^k$ and $\tilde{\sigma}^k$ are of the correct form (3.19), (3.20) (with $\epsilon_{k+1} \geq 0$), and defining Ξ and Σ as in the theorem, we easily check that $\Xi^T\Sigma = 0$. (Note that each column of $\tilde{\Xi}^k$ has equal entries in the $(k+1)$st pair of components, so $(\tilde{\Xi}^k)^T\tilde{\sigma}^k = (\tilde{\Xi}^k)^T\tilde{\sigma}^{(3)} = 0$, and similarly $(\tilde{\Sigma}^k)^T\tilde{\xi}^k = 0$.) □

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 (3.4)–(3.6) hold for $k+1$ by the conclusions of the theorem, the inductive hypothesis for k, and the form (3.19)–(3.20) of the new directions.

It only remains to show that (3.2), (3.3), and (3.11) hold for $k+1$. If we assume that (2.33) and (2.34) hold, then (3.2) for k and $\bar{\Pi}\xi^k = \sigma^k$ show that $\rho_P^k = \rho_D^k$, and hence that $\bar{\Pi} x^{k+1} = s^{k+1}$. Similarly, (3.3) and $\Pi \xi^k = \xi^k$, $\Pi \sigma^k = \sigma^k$ for all $\Pi \in S_n^{k+1}$ show that (3.3) holds for $k+1$.

We now turn to (3.11). We see that

$$\begin{aligned} v_j^{k+1} &= \{x_j^{k+1} s_j^{k+1}\}^{1/2} = \{([D^k]^{-1} x^{k+1})_j (D^k s^{k+1})_j\}^{1/2} \\ &= \{(v_j^k + \rho^k \tilde{\xi}_j^k)(v_j^k + \rho^k \tilde{\sigma}_j^k)\}^{1/2} \end{aligned}$$

for each j, $1 \leq |j| \leq p$, where $\rho^k := \rho_P^k = \rho_D^k$. Let $\bar{\rho}^k := \rho^k/2$. Then, from (3.19) and (3.20),

$$\begin{aligned} v_j^{k+1} &= \{([1-\bar{\rho}^k] v_j^k + \bar{\rho}^k \epsilon_j)([1-\bar{\rho}^k] v_j^k - \bar{\rho}^k \epsilon_j)\}^{1/2} \\ &= \{(1-\bar{\rho}^k)^2 (v_j^k)^2 - (\bar{\rho}^k \epsilon_j)^2\}^{1/2} \leq (1-\bar{\rho}^k) v_j^k \end{aligned}$$

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

$$v_j^{k+1} = \{([1-\bar{\rho}^k] v_j^k)([1-\bar{\rho}^k] v_j^k)\}^{1/2} = (1-\bar{\rho}^k) v_j^k$$

for $k+1 < |j| \leq p$. Thus (3.11) for k establishes that it remains true for $k+1$.

We therefore have

Theorem 3.2 *Suppose the inductive hypothesis holds for $k < p$, and that at the kth iteration of the primal-dual affine-scaling algorithm, the step sizes are chosen by some rule satisfying (2.33) and (2.34). Then search directions can be chosen such that the inductive hypothesis remains true for $k+1$.* □

4 The Main Result

By Theorem 3.2, we can continue generating search directions satisfying the inductive hypothesis for $p = n/2$ 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 affine-scaling algorithm applied to (P) starting at $x^0 = s^0 = e$.

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 (3.19) and (3.20). We suppose the step sizes are chosen by (2.33) and (2.34) so that $\rho_P^k = \rho_D^k =: \rho^k =: 2\bar{\rho}^k$. We show that $\bar{\rho}^k$ must be very small, and apply Proposition 2.1.

Let β denote $v_{k+1}^k = v_{-k-1}^k = \ldots = v_p^k = v_{-p}^k$. Then

$$(v^k)^\top v^k \geq 2(p-k)(\beta)^2.$$

Now $\tilde{\xi}^k$ and $\tilde{\sigma}^k$ (and hence $2\tilde{\xi}^k$ and $2\tilde{\sigma}^k$) are orthogonal, and using (3.19) and (3.20) this yields

$$\sum_{1 \leq |j| \leq k+1} (\epsilon_j)^2 = (v^k)^\top v^k \geq 2(p-k)(\beta)^2.$$

Hence some $|\epsilon_j|$ is at least $\sqrt{\frac{p-k}{k+1}}\beta$. Recalling that $\epsilon_j + \epsilon_{-j} = 0$, we have

$$\epsilon_j \geq \sqrt{\frac{n-2k}{2k+2}}\beta \tag{4.1}$$

for some $1 \leq |j| \leq k+1$. Since

$$\tilde{x}_j^+ = (1-\bar{\rho}^k)v_j^k - \bar{\rho}^k \epsilon_j > 0,$$

we have

$$\bar{\rho}^k < \frac{v_j^k}{\epsilon_j + v_j^k} \leq \frac{\beta}{\epsilon_j + \beta} \leq \frac{1}{1 + \sqrt{\frac{n-2k}{2k+2}}}, \tag{4.2}$$

where we have used $v_j^k \leq \beta$ from (3.11).

For $k = 0$, the bound (4.2) is like that which led to (2.41), giving a reduction of only a factor of $1 - O(n^{-1/2})$ in the duality gap. As k increases, the bound in (4.2) increases. However, suppose K is the first iteration for which

$$(x^K)^\top s^K \leq \exp(-t)(x^0)^\top s^0.$$

Then, using Proposition 2.1,

$$\prod_{k=0}^{K-1} (1 - 2\bar{\rho}^k) \leq \exp(-t),$$

whence

$$\sum_{k=0}^{K-1} \ln(1 + \frac{2\bar{\rho}^k}{1 - 2\bar{\rho}^k}) = -\sum_{k=0}^{K-1} \ln(1 - 2\bar{\rho}^k) \geq t,$$

and, using $\ln(1 + \theta) \leq \theta$, we obtain

$$\sum_{k=0}^{K-1} \left(\frac{2}{1/\bar{\rho}^k - 2} \right) \geq t.$$

Next, substituting the bound (4.2) on $\bar{\rho}^k$, we reach

$$\sum_{k=0}^{K-1} \left(\frac{2\sqrt{2}\sqrt{k+1}}{\sqrt{n-2k} - \sqrt{2k+2}} \right) \geq t. \tag{4.3}$$

Now let us assume that

$$K \leq n/50,$$

so that, for $k \leq K-1$, $\sqrt{2k+2} \leq \sqrt{2K} \leq \sqrt{n}/5$ and $\sqrt{n-2k} \geq \sqrt{n-2K} \geq \sqrt{24n/25} \geq 19\sqrt{n}/20$. Thus the denominators in (4.3) are always at least $3\sqrt{n}/4$, and we deduce that

$$\sum_{k=1}^{K} \sqrt{k} \geq \frac{1}{2\sqrt{2}} \cdot \frac{3}{4} \cdot \sqrt{n}t.$$

But

$$\sum_{k=1}^{K} \sqrt{k} \leq \int_1^{K+1} \sqrt{\theta} d\theta \leq \frac{2}{3}(K+1)^{3/2},$$

whence

$$K + 1 \geq (\frac{3}{2} \cdot \frac{1}{2\sqrt{2}} \cdot \frac{3}{4} \cdot \sqrt{n}t)^{2/3} \geq .54 n^{1/3} t^{2/3}. \tag{4.4}$$

Combining this analysis with Theorems 2.1 and 3.2, we arrive at

Theorem 4.1 *Consider a primal-dual affine-scaling algorithm that uses a step size rule satisfying (2.33) and (2.34) at each iteration. Also, suppose* $n/50 \leq m \leq 49n/50$. *Then there is an instance of* (P), *with* $A \in \mathbb{R}^{m \times n}$, $b = Ae \in \mathbb{R}^m$, *and* $c = A^T y + e \in \mathbb{R}^n$ *for any* $y \in \mathbb{R}^m$, *such that to decrease the duality gap by a factor of* $\exp(t)$, *starting with* $x^0 = e$ *and* $s^0 = e$, *the algorithm requires at least*

$$\min\{n/50,\ .54n^{1/3}t^{2/3} - 1\}$$

iterations.

(Observe that for large n and moderate t, the second term attains the minimum. Hence for n at least 1000, it takes no less than $n^{1/3} - 1$ iterations to achieve the modest reduction of a factor of exp(3) – about 20 – and at least $2n^{1/3} - 1$ iterations to achieve a reduction of a factor of exp(8) – about 3000.)

We conclude this section with a few remarks. Note that we have to be given the algorithm's step size rule before we can construct the bad example, and this must satisfy (2.33) and (2.34). Thus the rule cannot use any information about the problem instance besides the current iterate and search direction. (It could use some properties of previous iterates and search directions, as long as it is symmetrical enough to yield $\rho_P^k = \rho_D^k$.) But it cannot inspect A globally, because A is not yet known – only after constructing bad search directions for $\Omega(n^{1/3})$ iterations do we provide the matrix A that produced them. Subject to this minor restriction, any primal-dual affine-scaling algorithm requires $n^{1/3}$ iterations to achieve even a modest constant factor reduction in the duality gap.

5 Extensions and Discussion

In this section we investigate how the analysis of Sections 3 and 4 can be extended to more general primal-dual interior-point algorithm using a centering component. Thus, at some or all iterations, γ^k in (2.25) is nonzero, and $\tilde{\xi}^k$ and $\tilde{\sigma}^k$ are projections of $w^k \neq v^k$, where we recall that

$$w^k := v^k - \gamma^k \mu^k V^{-1} e = v^k - \gamma^k ((v^k)^T v^k / n) V^{-1} e. \tag{5.1}$$

Like the step sizes, we allow γ^k to depend on k and the current iterates x^k and s^k, but not on any other information in A, b, or c.

We again assume that the algorithm is initiated with $x^0 = s^0 = e$, and that at iteration k, the current iterate $(x^k, s^k) \in \mathcal{F}^0$ satisfies (3.2) and (3.3) and the past search directions Ξ^k and Σ^k satisfy (3.4)–(3.6). As in Section 3, these assumptions imply that D^k and v^k satisfy (3.7)–(3.10), and we deduce from (5.1) that

$$\bar{\Pi} w^k = w^k \quad \text{and} \quad \Pi w^k = w^k \quad \text{for all } \Pi \in S_n^k. \tag{5.2}$$

It may be unrealistic to assume (3.11) when there is a centering component in the search direction; let us instead assume that

$$\alpha \max\{v_j^k, v_{-j}^k : 1 \le j \le k\} \;\le\; v_{k+1}^k = v_{-k-1}^k = \ldots = v_p^k = v_{-p}^k, \tag{5.3}$$

where $0 < \alpha \le 1$.

With this new inductive hypothesis, we can check that Theorem 3.1 remains true where now, in (3.19) and (3.20), v^k is replaced with w^k. We call the resulting expressions (3.19)′ and (3.20)′. Moreover, Theorem 3.2 is still valid, except that the inequality in (5.3) may not be true for $k+1$.

We now turn to the effect of the kth iteration on the duality gap as in the previous section. Our tool is again Proposition 2.1. To simplify the notation, we suppress the dependence on k. If $\gamma = 1$ (a so-called *centering* or *constant-cost centering* step), then the proposition shows that the duality gap is unchanged.

If $\gamma^k = 1 - \kappa/\sqrt{n}$ for some fixed κ and $\rho^k = 1$, as in short-step path-following methods, we see

$$(x^{k+1})^T s^{k+1} = (1 - \kappa/\sqrt{n})(x^k)^T s^k,$$

and hence $\Omega(n^{1/2})$ steps will be necessary to achieve a constant factor reduction in the duality gap. Our interest is in algorithms using small but possibly nonzero values of γ, perhaps interspersed with centering steps. Thus we assume that

$$0 \leq \gamma \leq \bar{\gamma} < \frac{1}{2}, \tag{5.4}$$

where $\bar{\gamma}$ does not depend on the iteration number.

Again, let $\rho_P^k = \rho_D^k =: \rho^k =: 2\bar{\rho}^k$. We wish to show that $\bar{\rho}^k$ is small, so that the reduction in the duality gap (no better than the factor $1 - 2\bar{\rho}^k$) is also small. As in Section 4, we find, for $1 \leq |j| \leq k+1$,

$$\begin{aligned} \tilde{x}_j^+ &= \tilde{x}_j + 2\bar{\rho}^k \tilde{\xi}_j \\ &= v_j + \bar{\rho}^k(-v_j + \gamma\mu v_j^{-1} - \epsilon_j) \end{aligned}$$

using (5.1) and (3.19)′. Hence, if there is such a j with $(v_j)^2 + \epsilon_j v_j - \gamma\mu > 0$, we have

$$\bar{\rho}^k < \frac{(v_j)^2}{(v_j)^2 + \epsilon_j v_j - \gamma\mu}. \tag{5.5}$$

Thus we wish to show that there exists a j such that the denominator of (5.5) is large.

Since $\tilde{\xi}^k$ and $\tilde{\sigma}^k$ are orthogonal, we find as before

$$\sum_{1 \leq |j| \leq k+1} (\epsilon_j)^2 = w^\top w = (1-2\gamma)v^\top v + (\gamma\mu)^2 \sum_{1 \leq |i| \leq p} v_i^{-2},$$

where the last equation follows from the definition (5.1) of w. Hence

$$\sum_{1 \leq |j| \leq k+1} ((\epsilon_j)^2 - (\gamma\mu)^2 v_j^{-2}) \geq (1-2\gamma)n\mu,$$

and there is therefore some j, $1 \leq |j| \leq k+1$, with

$$(\epsilon_j)^2 \geq (\gamma\mu)^2 v_j^{-2} + \frac{(1-2\gamma)n}{2k+2}\mu$$

or

$$v_j|\epsilon_j| \geq \gamma\mu \left(1 + \frac{1-2\gamma}{(\gamma)^2} \cdot \frac{(v_j)^2}{\mu} \cdot \frac{n}{2k+2}\right)^{1/2}. \tag{5.6}$$

Since $\epsilon_j + \epsilon_{-j} = 0$, we may assume that $\epsilon_j > 0$, and hence the denominator in (5.5) is positive.

We now use (5.3) to get a bound on $(v_j)^2/\mu$. Indeed

$$n\mu = v^T v \geq (n-2k)(v_p)^2 \geq (n-2k)(\alpha)^2(v_j)^2,$$

so that

$$\delta := \frac{(v_j)^2}{\mu} \leq \frac{n}{n-2k}\alpha^{-2} \leq 2\alpha^{-2} \tag{5.7}$$

as long as $k \leq n^{1/3} - 1$ so that $n - 2k \geq n/2$. Combining (5.5) and (5.6), and using δ as in (5.7), we obtain

$$\begin{aligned} \bar{\rho}^k < \frac{\delta}{\delta + \gamma\left[\left(1 + \frac{1-2\gamma}{(\gamma)^2} \cdot \frac{n}{2k+2} \cdot \delta\right)^{1/2} - 1\right]} &= \frac{\delta}{\delta + \gamma[(1+\nu\delta)^{1/2} - 1]} \\ &=: \frac{1}{f(\delta)}, \end{aligned}$$

where $\nu := \frac{1-2\gamma}{(\gamma)^2} \cdot \frac{n}{2k+2} > 0$ using (5.4).

It is easy to show that $f(\delta)$ is monotonically decreasing for $\delta > 0$, and thus we can obtain a valid bound on $\bar{\rho}^k$ by substituting the bound on δ in (5.7). Hence we find, for $k \leq n^{1/3} - 1$,

$$\begin{aligned} \bar{\rho}^k &< \frac{2}{2 + \gamma(\alpha)^2\left[\left(1 + \frac{2(1-2\gamma)}{(\alpha\gamma)^2} \cdot \frac{n}{2k+2}\right)^{1/2} - 1\right]} \\ &\leq \frac{2}{\gamma(\alpha)^2\left(1 + \frac{2(1-2\gamma)}{(\alpha\gamma)^2} \cdot \frac{n}{2k+2}\right)^{1/2}} \\ &\leq \frac{2}{\gamma(\alpha)^2\left(\frac{2(1-2\gamma)}{(\alpha\gamma)^2} \cdot \frac{n}{2k+2}\right)^{1/2}} \\ &= \frac{2}{(1-2\gamma)^{1/2}\alpha}\sqrt{\frac{k+1}{n}} \\ &\leq \frac{2}{(1-2\bar{\gamma})^{1/2}\alpha}\sqrt{\frac{k+1}{n}}, \end{aligned} \tag{5.8}$$

where we have used the uniform bound (5.4) on γ.

We therefore have

Theorem 5.1 *Suppose at the kth iteration, with $k \leq n^{1/3} - 1$, the inductive hypothesis holds, with condition (5.3) replacing (3.11). Suppose that the search directions are given by (3.19)′ and (3.20)′. Then, if the step size rule satisfies (2.33) and (2.34), and if either $\gamma^k = 1$ or $0 \leq \gamma^k \leq \bar{\gamma} < \frac{1}{2}$, then*

$$(x^{k+1})^\top s^{k+1} \geq \left(1 - \frac{4}{(1-2\bar{\gamma})^{1/2}\alpha} n^{-1/3}\right)(x^k)^\top s^k.$$

Moreover, if (5.3) holds for all such iterations, then $\Omega(n^{1/3})$ iterations are required to obtain a constant factor decrease in the duality gap. □

The weakness of this result, in contrast to Theorem 4.1, is that we must *assume* that the iterates satisfy (5.3) at each iteration, whereas this could be *proved* by induction (with $\alpha = 1$) in the case of the affine-scaling algorithm. One way to assure (5.3) is to require that all iterates satisfy the fairly weak centering condition that they lie in

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

for some $0 < \chi < 1$ (see [17]). Then all $x_j s_j'$s lie within a factor of $(1+\chi)/(1-\chi) \leq (1-\chi)^{-2}$ of each other, and (5.3) holds for $\alpha = 1-\chi$. However, this centering condition is not imposed for most practical implementations. In order to see whether this assumption is reasonable in practice, we have made a number of computational runs with $n = 10^6$ using MATLAB [18].

At each iteration, we do not compute the search directions by projections. Instead, we use the form (3.19)′–(3.20)′ of the directions, computing $\epsilon_1^k \ldots \epsilon_k^k$ by the conditions $(\tilde{\Sigma}^k)^\top \tilde{\xi}^k = 0$, and then ϵ_{k+1}^k so that $(\tilde{\xi}^k)^\top \tilde{\sigma}^k = 0$. In theory, $\epsilon_{k+1}^k = (\frac{1}{2}(\tilde{\xi}^{(3)})^\top \tilde{\sigma}^{(3)})^{1/2} = \frac{1}{8^{1/2}}\|P_{\xi\sigma} w^k\|$ is real (see the proof of Theorem 3.1), but we found many times that $(\tilde{\xi}^{(3)})^\top \tilde{\sigma}^{(3)}$ was numerically negative and these runs were aborted. (Again, $\tilde{\xi}^{(3)}$ and $\tilde{\sigma}^{(3)}$ were not computed by projections; they are the vectors of the form (3.19)′ and (3.20)′, with $\epsilon_{k+1}^k = 0$, satisfying $(\tilde{\Sigma}^k)^\top \tilde{\xi}^{(3)} = 0$, $(\tilde{\Xi}^k)^\top \tilde{\sigma}^{(3)} = 0$.) These numerical difficulties were caused by near linear dependence in the columns of $\tilde{\Sigma}^k$ and $\tilde{\Xi}^k$. However the remaining runs (about a fifth of the total) successfully completed 150 iterations ($n^{1/3} = 100$) with the anticipated results.

We tried keeping γ^k fixed at the level $0, .1, \ldots, .9$ (not all satisfying (5.4)), with step sizes chosen by (2.37) with $\lambda_p^k = \lambda_D^k = \lambda$ equal to .1, .2, ..., .9, .95, .99. In these runs, (5.3) was always satisfied with $\alpha = 1$, i.e., (3.11) remained true. We next tried the same values for λ with γ^k fixed at .95, .97, .99 and .995. Even with such a large centering component, (5.3) remained true for $\alpha = .98$.

Finally, we tried predictor-corrector algorithms as in [17] – see also [25, 26]. Here every other iteration used $\gamma^k = 1$ (a centering step), while the remaining iterations used a fixed γ^k with the values mentioned in the previous paragraph. For the noncentering steps, we used the step size rule of the previous paragraph, with the values of λ listed there. For the centering steps, we chose ρ to be the smaller of 1 and that given by strategy above; $\rho = 1$ corresponds to a Newton step for a centering problem, and was chosen in all but two runs. Once again, (5.3) was true for all runs if we set $\alpha = .75$.

It thus appears that the poor behaviour established rigorously for the primal-dual affine-scaling algorithm might also be exhibited by most primal-dual algorithms of the form given in Section 2. We must stress again, however, that this is a theoretical bound on the asymptotic behaviour of such algorithms. For problems with a million variables, $n^{1/3}$ is only 100, and this is not an unreasonable number of iterations, and seems to be in line with what has been observed in practice for such large problems – see Lustig, Marsten and Shanno [15].

References

[1] K. M. Anstreicher. On the performance of Karmarkar's algorithm over a sequence of iterations. *SIAM Journal on Optimization*, 1(1):22–29, 1991.

[2] 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.

[3] R. E. Bixby, J. W. Gregory, I. J. Lustig, R. E. Marsten, and D. F. Shanno. Very large-scale linear programming: A case study in combining interior point and simplex methods. *Operations Research*, 40:885–897, 1992.

[4] K. H. Borgwardt. *The Simplex Method, a Probabilistic Analysis.* Springer-Verlag, Berlin, 1987.

[5] G. B. Dantzig. *Linear Programming and Extensions.* Princeton University Press, Princeton, NJ, 1960.

[6] D. Goldfarb and M. J. Todd. Linear Programming. In G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors, *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, pp. 73–170. North Holland, Amsterdam, The Netherlands, 1989.

[7] C. C. Gonzaga. Path following methods for linear programming. *SIAM Review*, 34(2):167–227, 1992.

[8] J. Ji and Y. Ye. A complexity analysis for interior-point algorithms based on Karmarkar's potential function. Working Paper, Dept. of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, revised 1991.

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

[10] V. Klee and G. J. Minty. How good is the simplex algorithm. In O. Shisha, eds., Inequalities III, pp. 159–175. Academic Press, New York, 1972.

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

[12] M. Kojima, S. Mizuno, and A. Yoshise. An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems. *Mathematical Programming*, 50:331–342, 1991.

[13] E. Kranich. Interior point methods for mathematical programming: A bibliography. Discussion Paper 171, Institute of Economy and Operations Research, FernUniversität Hagen, P.O. Box 940, D–5800 Hagen 1, West–Germany, May 1991. The (actual) bibliography can be accessed electronically by sending e–mail to '*netlib@research.att.com*' with the message '*send intbib from bib.*'

[14] I. J. Lustig, R. E. Marsten, and D. F. Shanno. The primal-dual interior point method on the Cray supercomputer. In T. F. Coleman and Y. Li, editors, *Large-Scale Numerical Optimization*, volume 46 of *SIAM Proceedings in Applied Mathematics*, pp. 70–80. Society of Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 1990.

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

[16] S. Mizuno and A. Nagasawa. A primal-dual affine scaling potential reduction algorithm for linear programming. Research Memorandum 427, The Institute of Statistical Mathematics, Tokyo, Japan, February 1992.

[17] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive step primal-dual interior-point algorithms for linear programming. Technical Report 944, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853–3801, USA, October 1990. To appear in *Mathematics of Operations Research.*

[18] C. B. Moler, J. Little, S. Bangert and S. Kleiman. *Pro-Matlab User's Guide*, MathWorks, Sherborn, MA., 1987.

[19] R. D. C. Monteiro and I. Adler. Interior path following primal-dual algorithms: Part I: Linear programming. *Mathematical Programming*, 44:27–41, 1989.

[20] R. D. C. Monteiro, I. Adler, and M. G. C. Resende. A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. *Mathematics of Operations Research*, 15:191–214, 1990.

[21] A. S. Nemirovsky. An algorithm of the Karmarkar type. *Tekhnicheskaya Kibernetika*, 1:105–118, 1987. Translated in: *Soviet Journal on Computers and System Sciences*, 25(5):61–74, 1987.

[22] A.S. Nemirovsky and D.B. Yudin. *Problem Complexity and Method Efficiency in Optimization*, John Wiley and Sons, Chichester, 1983.

[23] M. J. D. Powell. The complexity of Karmarkar's algorithm for linear programming. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis 1991*, volume 260 of *Pitman Research Notes in Mathematics*, pages 142–163. Longman, Burnt Hill, UK, 1992.

[24] M. J. D. Powell. On the number of iterations of Karmarkar's algorithm for linear programming. Technical Report DAMTP 1991/NA23, Dept. of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK, December 1991.

[25] G. Sonnevend, J. Stoer, and G. Zhao. On the complexity of following the central path by linear extrapolation in linear programming. *Methods of Operations Research*, 62:19–31, 1990.

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

[27] M. J. Todd. Recent developments and new directions in linear programming. In M. Iri and K. Tanabe, editors, *Mathematical Programming: Recent Developments and Applications*, pp. 109–157. Kluwer Academic Press, Dordrecht, The Netherlands, 1989.

[28] L. Tunçel. A note on the primal-dual affine scaling algorithms. Technical Report 92–8, Cornell Computational Optimization Project, Center for Applied Mathematics, Cornell University, Ithaca, NY 14853–3801, USA, May 1992.