Systems
Optimization
Laboratory

An Adaptive Primal-Dual Method for Linear Programming

by
Florian Jarre and Michael A. Saunders

TECHNICAL REPORT SOL 91-1†

January 1991

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022

An Adaptive Primal-Dual Method for Linear Programming

by
Florian Jarre and Michael A. Saunders

TECHNICAL REPORT SOL 91-1†

January 1991

†This work was supported by a research grant from the Deutsche Forschungsgemeinschaft, and in part by the U.S. National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do NOT necessarily reflect the views of the above sponsors.

AN ADAPTIVE PRIMAL-DUAL METHOD FOR LINEAR PROGRAMMING

Florian Jarre* and Michael A. Saunders

Technical Report SOL 91-1†
January 1991

Abstract

A simple analysis for an adaptive primal-dual method for linear programming is given. Starting from a pair of primal and dual feasible points near the path of centers, the method maintains a worst-case complexity of $O(\sqrt{n}\log\frac{1}{\epsilon})$ iterations to reduce the initial duality gap by a factor of ϵ. In contrast to other interior-point algorithms that share the same complexity ([13, 4] and many others), the algorithm proposed here allows an acceleration of the rate of convergence (up to a complexity of $O(\sqrt[4]{n}\log\frac{1}{\epsilon})$) if the problem is "well behaved".

Key words: linear programming, interior-point method, theoretical complexity, orthogonal projection

1. INTRODUCTION

We discuss interior-point algorithms that apply one iteration of Newton's method to a set of nonlinear equations involving a parameter $\mu > 0$, and then reduce μ by a certain amount before repeating the process. For $\mu = 0$ the solution of the set of nonlinear equations coincides with the solution of a given linear program. The reduction of μ is of the form $\mu := \mu(1-\delta)$ for some "steplength" $\delta \in (0,1)$, where the focus is on the size of δ. The method of centers (e.g. [13]) is typical of "short-step" methods in which $\delta \leq 1/\sqrt{n}$ at every iteration, where n is the number of variables. Roughly speaking, the shortness of the steplength $\delta \leq 1/\sqrt{n}$ allows a proof that the Newton iterates remain strictly feasible, but for numerical implementations this rate of convergence is too slow when n is large. A "long-step" method is one for which $\delta > 1/\sqrt{n}$. (Typically δ is a constant independent of n.)

In the recent past a number of investigations have been presented that analyze long-step interior-point methods for linear programming (see [3, 6, 11, 14]). Their goal is to generalize the existing proofs of convergence of certain interior-point algorithms to wider classes of algorithms and to justify the correctness of existing

†This work was supported by a research grant from the Deutsche Forschungsgemeinschaft, and by the U.S. National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

*On leave from Institut für Angewandte Mathematik, University of Würzburg, 8700 Würzburg, (West) Germany.

implementations of interior-point algorithms that yield very fast convergence for (most) numerical examples. In allowing long steps, however, the theoretical complexity of these investigations degrades. This apparently paradoxical behavior is due to the fact that for large steps the Newton iterates may no longer be feasible, and therefore the method has to use damped Newton iterations with a damping factor $\alpha \in (0,1)$. Loosely speaking, the iterates may not be able to "follow" such a long step δ in the parameter μ. The goal of the present paper is to improve the theoretical complexity. We will consider an adaptive choice of the steplength δ such that the full Newton iterate may always be used.

By a thorough analysis of the integral over the curvature of the path of centers it was shown in [16] that for certain subclasses of linear programs the theoretical complexity could be reduced below $O(\sqrt{n}\log\frac{1}{\epsilon})$ iterations, and in [12] a probabilistic analysis showed that the "anticipated" number of iterations could be reduced below $O(\sqrt{n}\log\frac{1}{\epsilon})$, but in a general worst-case analysis this complexity is (still) the state of the art. The analysis presented here is particularly simple. It also illuminates why the theoretical complexity could not be improved so far by any analysis that focuses on the worst case in a single iteration, rather than examining a sequence of iterations.

In Sections 2.1–2.4 we present a proof of convergence for the method of centers. The result is well known (see e.g. [7]), but the analysis is new and allows us to explain in Section 3.1 where the factor $\sqrt{n}$ comes from, and to analyze a new adaptive method in Section 3.2 that automatically chooses large steps if Newton's method for finding the center converges "well", and takes short steps otherwise.

2. THE PROBLEM AND A SIMPLE METHOD

The problem under consideration is the linear program

$$\min_{x\in S} c^T x, \quad S = \{x \geq 0 \mid Ax = b\}, \tag{P}$$

where $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. For brevity we assume that the relative interior of the feasible set S is nonempty and bounded, and that the rows of A are linearly independent. The dual problem to (P) is given by

$$\max_{y\in\mathbb{R}^m} \{b^T y \mid A^T y \leq c\}. \tag{D}$$

2.1. Some Known Theory

Following the motivation given in [8], the "penalized" Lagrangian corresponding to (P) with a logarithmic barrier term for the inequality constraints is given by

$$L_\mu(x,y) = c^T x - y^T(Ax - b) - \mu\sum_{j=1}^{n}\ln x_j,$$

where $\mu > 0$. The necessary (and sufficient) conditions for a stationary point of L_μ are

$$F_\mu(x,y,z) := \begin{pmatrix} Ax - b \\ A^T y + z - c \\ Xz - \mu e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad x, z > 0, \tag{2.1}$$

where $X = \mathrm{diag}(x_1, \ldots, x_n)$, $e = (1, \ldots, 1)^T \in \mathbb{R}^n$, and z is an auxiliary variable. (In the following a small letter and the same letter capitalized—like x and X—will always stand for a vector and its corresponding diagonal matrix.) The unique solution $x(\mu), y(\mu), z(\mu)$ of (2.1) is the *analytic center* as defined in [15]. It is obvious that $x(\mu)$ is primal feasible and $y(\mu)$ is dual feasible. From (2.1) it immediately follows that the duality gap between $x = x(\mu)$ and $y = y(\mu)$ is bounded by $n\mu$:

$$n\mu = \|\mu e\|_1 = \|Xz\|_1 = \|X(c - A^T y)\|_1$$

$$\leq x^T(c - A^T y) = c^T x - (Ax)^T y = c^T x - b^T y.$$

As first proved in [9], $x(\mu)$ and $y(\mu)$ converge to primal and dual optimal solutions as μ tends to zero. (See also [1].)

2.2. The Newton Step for Finding the Center

Suppose that $x > 0$, $z > 0$ and y are given such that $Ax = b$ and $A^T y + z = c$. For $\mu > 0$ we define the residual

$$r := Xz - \mu e,$$

so that $F_\mu(x,y,z) = (0, 0, r^T)^T$. The Newton step for finding the analytic center $x(\mu)$, $y(\mu)$, $z(\mu)$ is then given by

$$\begin{aligned} A\Delta x &= 0, \\ A^T\Delta y + \Delta z &= 0, \\ Z\Delta x + X\Delta z &= -r. \end{aligned} \tag{2.2}$$

Note that if we were able to solve the (nonlinear) system (2.2′) where in the last equation of (2.2) the term $\Delta X \Delta z$ is added on the left-hand side, then we could find the exact center as $(x + \Delta x,\ y + \Delta y,\ z + \Delta z)$, since (2.2′) is then equivalent to (2.1).

Define the positive diagonal matrix D by $D^2 := XZ^{-1}$. Then (2.2) can be solved via

$$\begin{aligned} q &= DX^{-1}r, \\ \Delta y &= (AD^2A^T)^{-1}ADq, \\ \Delta x &= D^2A^T\Delta y - Dq, \\ \Delta z &= -D^{-1}q - D^{-2}\Delta x. \end{aligned} \tag{2.3}$$

This becomes obvious as we observe that (2.3) is equivalent to (2.4)–(2.6). First,

$$DA^T\Delta y = \Pi_R q \tag{2.4}$$

(since DA^T has maximal rank), where Π_R is the orthogonal projection onto the range $R(DA^T) = \{y \mid y = DA^T w \text{ for some } w \in \mathbb{R}^m\}$ of DA^T. Let Π_N denote

the orthogonal projection onto the null space $N(AD)$ of AD. Since $N(AD)$ is the orthogonal complement to $R(DA^T)$, i.e. $\Pi_R + \Pi_N = I$, it follows that

$$\Delta x = D(\Pi_R q - q) = -D\Pi_N q, \tag{2.5}$$

and finally,

$$\Delta z = -D^{-1}(q + D^{-1}\Delta x) = -D^{-1}\Pi_R q. \tag{2.6}$$

It is straightforward to verify that (2.4)–(2.6) satisfy (2.2).

2.3. Analysis of the Newton Step

The residual after executing the above Newton step (without damping[1]) is given by

$$\begin{aligned} \tilde{r} &= (X + \Delta X)(z + \Delta z) - \mu e = Xz + X\Delta z + Z\Delta x + \Delta X\Delta z - \mu e \\ &= Xz - r + \Delta X\Delta z - \mu e = \Delta X\Delta z = \tilde{\Delta X}\tilde{\Delta z}, \end{aligned} \tag{2.7}$$

where $\tilde{\Delta x} = -D^{-1}\Delta x = \Pi_N q$ and $\tilde{\Delta z} = -D\Delta z = \Pi_R q$. Note that

$$DX^{-1} = \sqrt{X^{-1}Z^{-1}} = (\sqrt{R + \mu I})^{-1} \tag{2.8}$$

(using $Xz = r + \mu e$). From (2.7)–(2.8) and the definition of q in (2.3) we can readily derive the classical convergence results about the primal-dual method as given in [7] or [14]. Assume that $x > 0, z > 0$ and y are given such that we can find a positive μ for which

$$\|r\|_2 \le \beta\mu \tag{2.9}$$

for some $\beta \in [0, \frac{1}{2}]$. (This means we are assuming that x, y, z is "moderately close" to the path of centers.) It follows that

$$\|(R + \mu I)^{-1}\|_2 \le \frac{1}{\mu(1 - \beta)}, \tag{2.10}$$

so that

$$\|q\|_2 \le \beta\sqrt{\mu/(1 - \beta)} =: \tilde{\beta}, \tag{2.11}$$

and

$$\|\tilde{\Delta x}\|_2 = \tilde{\beta}\cos\theta, \qquad \|\tilde{\Delta z}\|_2 = \tilde{\beta}\sin\theta, \tag{2.12}$$

where θ is the angle between q and $\Pi_N q$. Finally by (2.7), (2.12) and (2.11),

$$\|\tilde{r}\|_2 = \|\tilde{\Delta X}\tilde{\Delta z}\|_2 \le \tilde{\beta}^2\cos\theta\sin\theta \le \frac{1}{2(1 - \beta)}\mu\beta^2 \le \mu\beta^2. \tag{2.13}$$

Hence, the relative error $\|Xz - \mu e\|_2/\mu = \|r\|_2/\mu$ is squared after each iteration of Newton's method[2].

[1] Analyzing full Newton steps only is not a severe restriction, since a damped Newton step can be viewed as a full Newton step for a linearly perturbed system of equations (the original system of nonlinear equations to which a linear perturbation is added).

[2] Note that (2.13) and the first equation in (2.7) imply that $(X + \Delta X)(z + \Delta z) > 0$. One way to verify strict feasibility (i.e. $x + \Delta x > 0$ and $z + \Delta z > 0$) is to show that $\|X^{-1}\Delta x\|_2 \le 1$. By (2.12), $\|\tilde{\Delta x}\| \le \tilde{\beta}$, and so $\|X^{-1}\Delta x\|_2 = \|X^{-1}D\tilde{\Delta x}\|_2 = \|(\sqrt{R + \mu I})^{-1}\tilde{\Delta x}\|_2 \le \tilde{\beta}/\sqrt{\mu(1 - \beta)} \le 1$.

2.4. A Model Algorithm

The above analysis allows us to formulate the following classical model algorithm. Let $x^0 > 0, y^0, z^0 > 0$ and μ^0 be given such that $Ax^0 = b$, $A^T y^0 + z^0 = c$, $X^0 z^0 - \mu^0 e = r^0$ and $\|r^0\|_2/\mu^0 \leq \frac{1}{2}$. (Again this assumes that the initial point is moderately close to the path of centers.) Further, let some desired accuracy ϵ be given. Set $k = 0$.

1. Perform one Newton step via (2.2) to obtain x^{k+1}, y^{k+1}, z^{k+1}.
2. Decrease μ^k to $\mu^{k+1} := \mu^k(1 - \frac{1}{4\sqrt{n}})$.
3. Set $k = k + 1$.
4. If $\mu^k \leq \epsilon/n$ then stop, else go to Step 1.

Since $\|r^{k+1}\|_2 \leq \|\tilde{r}^k\|_2 + \frac{1}{4\sqrt{n}}\|e\|_2 \leq \frac{1}{4} + \frac{\sqrt{n}}{4\sqrt{n}} \leq \frac{1}{2}$ (with $\tilde{r}^k$ as in (2.13)) it follows by induction that all residuals r^k fulfill (2.9) with $\beta = \frac{1}{2}$, and hence (by the results of Section 2.2) the method is well defined. Since the complementarity parameter μ converges linearly to zero (Step 2), the method terminates after $O(\sqrt{n}\log\frac{n\mu^0}{\epsilon})$ iterations. This result is well known [7].

3. IMPROVEMENTS

We shall now examine possible improvements of the convergence analysis. For this purpose, it is useful to analyze which of the estimates leading to (2.13) are sharp, and which ones can possibly be tightened.

Note that the relations (2.7), (2.8) and (2.12) are exact. However the bounds (2.10) and (2.11) could also be established under the weaker assumption

$$\|r\|_\infty \leq \beta\mu. \tag{2.9'}$$

(It is interesting to note that for any feasible x, y, z one can always find some $\mu > 0$ such that (2.9') is satisfied with $\beta = 1$.) Also, the bound (2.13) is based on the inequality $\|St\|_2 \leq \|s\|_2\|t\|_2$, which also holds in terms of the infinity norm: $\|St\|_\infty \leq \|s\|_\infty\|t\|_\infty$. Finally, to guarantee feasibility of $x + \Delta x, z + \Delta z$, the inequalities $\|\tilde{\Delta x}\|_\infty$, $\|\tilde{\Delta z}\|_\infty \leq \tilde{\beta}$ would suffice in place of (2.12)[3]. Unfortunately, an estimate corresponding to (2.12) in terms of the infinity norm only holds in the form

$$\|\tilde{\Delta x}\|_\infty \leq \sqrt{n}\tilde{\beta}\cos\theta, \qquad \|\tilde{\Delta z}\|_\infty \leq \sqrt{n}\tilde{\beta}\sin\theta. \tag{2.12'}$$

We "lose" a factor of $\sqrt{n}$ both times. Stating (2.13) in terms of the infinity norm would thus "lose" a factor of n.

[3] An analysis in terms of the infinity norm would be particularly interesting, since by the update in Step 2 of the model algorithm the residual r^{k+1} is obtained from $\tilde{r}^k$ by $r^{k+1} = \tilde{r}^k + (\mu^k - \mu^{k+1})e$, and the infinity norm of e is smaller than its 2 norm: $\|e\|_\infty = \frac{1}{\sqrt{n}}\|e\|_2$.

3.1. Understanding the Analysis

The purpose of the following analysis is to shed some light on this apparent "incompatibility" of norms. Assume for the moment that for some $\mu = \mu^k$ the exact center x, y, z is given. (A "central" analysis is particularly interesting since it has been observed (see e.g. [2]) that at the center all the search directions suggested for interior-point algorithms since Karmarkar [5] are equivalent.) At the center, D, r and q reduce to

$$D^2 = XZ^{-1} = \frac{1}{\mu}X^2, \qquad r = (\mu^k - \mu^{k+1})e =: \delta\mu e, \qquad q = \delta\sqrt{\mu}e, \tag{3.1}$$

where $\delta \in (0,1)$. Using the Newton step for finding $x(\mu^{k+1}), y(\mu^{k+1}), z(\mu^{k+1})$ via (2.2) we obtain (using (2.5))

$$\Delta x = -\delta X \Pi_N e. \tag{3.2}$$

What is the largest possible steplength δ that keeps the next iterate strictly feasible? Clearly, $x + \Delta x > 0$ is maintained iff $\delta\Pi_N e < e$. As mentioned in (2.12'), the orthogonal projection Π_N may increase some components of e by $O(\sqrt{n})$. A simple example to illustrate this worst case is $A = (-\sqrt{n}, 1, \ldots, 1) \in \mathbb{R}^{1\times(n+1)}$, $x = z = c = e = (1, \ldots, 1)^T \in \mathbb{R}^{n+1}$, $y = 0$ and $b = n - \sqrt{n}$. These vectors satisfy (2.1) with $\mu = 1$, i.e. x, y, z form a center, and $\Pi_R e = \frac{n-\sqrt{n}}{2n}(-\sqrt{n}, 1, \ldots, 1)^T$, and hence $\Pi_N e = \frac{n+\sqrt{n}}{2n}(\sqrt{n}, 1, \ldots, 1)^T$. The infinity norms of $\Pi_R e$ and $\Pi_N e$ are larger than $\|e\|_\infty$ by a factor of (nearly) $\sqrt{n}/2$.

We emphasize that in this situation the search direction, even when secured by a linesearch, does not allow a "long" step[4]! This is particularly interesting since all interior-point methods (also the "long-step" methods) generate the same search direction at this point. To guarantee feasibility of $x + \Delta x$ in the worst case for a fixed steplength δ we therefore need $\delta < 2/\sqrt{n}$, almost as in the model algorithm. Thus the factor $\sqrt{n}$ can be interpreted as the "magnification" of certain components of a vector during its orthogonal projection, rather than a result of "incompatibility" of the norms or a result of our inability to prove the best possible bound.

At this point we may ask for a geometrical interpretation of the orthogonal projection and of the fact that it sometimes magnifies certain components of the vector e, i.e. that $\|\Pi_N e\|_\infty = O(\sqrt{n}) \gg 1 = \|e\|_\infty$. Given a point x, y, z on the path of centers, it is well known that $-X\Pi_N e$ is the direction of the tangent (in the primal space) to $x(\mu)$. (To see this, simply differentiate (2.1) with respect to μ and observe that the resulting equation is exactly (2.2) with $r = -e$ and with the derivatives x', z' in place of Δx, Δz.) Now, if some components of $\Pi_N e$ are $\gg 1$ (the case $\ll -1$ is similar with z in place of x) then the linear approximation $x(\mu(1-\delta)) \approx x - \delta\Delta x$ to the path of centers has a short "trust region", since even for small δ (any $\delta > 1/\|\Pi_N e\|_\infty$) the points $x - \delta\Delta x$ are not feasible, i.e. not near

[4]The conventional notion of short and long steps is somewhat confusing. Of course one may decrease the complementarity parameter μ by a long step at this point (and this is the general understanding of a long step). What this example really shows is that the variables x, y, z may not be able to follow such a long step; i.e. a damped Newton step is required, and the actual complementarity gap is only reduced by a small amount.

the path of centers. This implies that the second derivative of $x(\mu)$ must be large near μ^k.

Similar considerations also hold for "non-central" points $Xz = \mu w$ where $w \neq e$ is a positive weight vector. For estimates about the curvature of $x(\mu)$ we refer to [16].

3.2. An Adaptive Method

If the path of centers is "well behaved" we do not anticipate the worst case—that the infinity norm of the residual is increased by a factor of $O(\sqrt{n})$—in each step (see [12]). The hope that the estimates used in the model algorithm are too pessimistic in the average case suggests the following *adaptive* choice of the steplength.

Let us again take $\beta = \frac{1}{2}$ as in the model algorithm. Given x^k, y^k, z^k and $\mu = \mu^k$ such that

$$\|r^k\|_2 = \|X^k z^k - \mu e\|_2 \leq \frac{\mu}{2},$$

we wish to determine $\tilde{\mu} := \mu^{k+1}$ as small as possible such that the residual r^{k+1} following the Newton step for finding $x(\tilde{\mu}), y(\tilde{\mu}), z(\tilde{\mu})$ satisfies the analogous bound

$$\|r^{k+1}\|_2 = \|X^{k+1} z^{k+1} - \tilde{\mu} e\|_2 \leq \frac{\tilde{\mu}}{2}, \tag{3.3}$$

where $x^{k+1} = x^k + \Delta x^k$ and $z^{k+1} = z^k + \Delta z^k$ are obtained from x^k and z^k by (2.2) with $r = \tilde{r}^k := X^k z^k - \tilde{\mu} e$. If we set $\tilde{\mu} := \mu(1 - \delta)$ (for some $\delta = \delta^k > 0$), then it follows that $\tilde{r}^k = r^k + \delta\mu e$. The component of $\Delta x^k = D\Pi_N DX^{-1}\tilde{r}^k$ resulting from r^k is often referred to as the *centering direction* (it brings the iterate closer to the center $x(\mu)$), while the component resulting from δe is aimed at reducing the duality gap and is usually referred to as the *affine scaling direction*. By (2.7) the next residual is given by $r^{k+1} = \tilde{\Delta X}^k \tilde{\Delta z}^k$, where

$$\tilde{\Delta x}^k = \Pi_N DX^{-1}\tilde{r}^k, \qquad \tilde{\Delta z}^k = \Pi_R DX^{-1}\tilde{r}^k, \tag{3.4}$$

(where $x = x^k$, $z = z^k$, $D^2 = XZ^{-1}$). If we define $q^k = DX^{-1}r^k$, $s^k = DX^{-1}e$ and

$$v_1 = \Pi_N q^k, \qquad v_2 = q^k - v_1, \qquad v_3 = \Pi_N s^k, \qquad v_4 = s^k - v_3,$$

then

$$\tilde{\Delta x}^k = v_1 + \delta v_3, \qquad \tilde{\Delta z}^k = v_2 + \delta v_4.$$

This shows that (3.3) is equivalent to

$$\|r^{k+1}\|_2^2 = \|(V_1 + \delta V_3)(v_2 + \delta v_4)\|_2^2 \leq \frac{\tilde{\mu}^2}{4} = \frac{\mu^2 - 2\delta\mu + \delta^2}{4}, \tag{3.5}$$

which in turn is equivalent to

$$p(\delta) := a_0 + a_1\delta + a_2\delta^2 + a_3\delta^3 + a_4\delta^4 \leq 0, \tag{3.6}$$

where

$$\begin{aligned} a_0 &= v_1^2 v_2^2 - \mu^2/4, \\ a_1 &= 2(v_1 v_3 v_2^2 + v_2 v_4 v_1^2 + \mu/4), \\ a_2 &= v_2^2 v_3^2 + v_1^2 v_4^2 + 4 v_1 v_2 v_3 v_4 - 1/4, \\ a_3 &= 2(v_1 v_3 v_4^2 + v_2 v_4 v_3^2), \\ a_4 &= v_3^2 v_4^2, \end{aligned} \tag{3.7}$$

and each "product" of four vectors "*abcd*" stands for the sum $\sum_{j=1}^{n}(a_j b_j c_j d_j)$.

The maximal permissible steplength δ is therefore given by the largest positive zero of p that is less than one[5]. (This $\delta = \delta^k$ can be computed in constant time.) Thus we obtain a simple modification of the model algorithm by changing Step 2 to

2'. $\mu^{k+1} := \mu^k(1 - \delta^k)$.

3.3. Achievable Gains

The analysis of the model algorithm in Section 2.3 has shown that a steplength of $\delta = \mu/(4\sqrt{n})$ is possible in the worst case, i.e. that $p(\delta) < 0$ for $\delta \in (0, \mu/(4\sqrt{n}))$. We now examine how long the step might be on average. We will call a vector v *well balanced* if

$$\|v\|_\infty \leq \frac{\kappa}{\sqrt{n}}\|v\|_2$$

for some small constant $\kappa > 1$. This excludes the case that v has only a few large components.

Let us observe first that the vector s^k is well balanced, since $1/\sqrt{2} \leq s_j^k \leq \sqrt{2}$ from the estimates (2.8) and (2.9).

To find how large we may expect δ to be in the "best case" we assume that the projections $v_3 = \Pi_N s^k$ and $v_4 = s^k - v_3$ are also well balanced, which essentially excludes the case that some of the components of s^k are magnified by more than a constant factor during the orthogonal projection Π_N. This allows a tight bound on the size of the coefficients of $p(\delta)$ by the following Lemma.

Lemma

1. If a and b are well balanced, then $abcd \leq \frac{\kappa^2}{n}\|a\|_2\|b\|_2\|c\|_2\|d\|_2$.
2. If a is well balanced, then $abcd \leq \frac{\kappa}{\sqrt{n}}\|a\|_2\|b\|_2\|c\|_2\|d\|_2$.

Proof

1. $\sum a_j b_j c_j d_j \leq \max|a_j| \max|b_j| \sum c_j d_j \leq \max|a_j| \max|b_j| \ \|c\|_2\|d\|_2 \leq \frac{\kappa^2}{n}\|a\|_2\|b\|_2\|c\|_2\|d\|_2$. (The second inequality is the Cauchy-Schwarz inequality, and the third uses that a and b are well balanced.)

[5] It is easy to see that if p is nonnegative in $[0,1]$ we may choose $\delta = 1$ and thus find the solution of (P) in one step. Unfortunately, this will not happen in general, and so we assume that p has a zero in $(0,1)$.

2. Same as above, only in the third inequality we use $\|b\|_\infty \leq \|b\|_2$ instead of $\|b\|_\infty \leq \frac{\kappa}{\sqrt{n}}\|b\|_2$. ∎

By definition of v_i and by (2.8) and (2.10), it follows that $\|v_1\|_2, \|v_2\|_2 \leq \sqrt{\mu/2}$ and $\|v_3\|_2, \|v_4\|_2 \leq \sqrt{2n/\mu}$. Further, inequality (2.13) (with $v_1 = \Delta x$ and $v_2 = \Delta z$) states that $v_1^2 v_2^2 \leq \mu^2/\beta^4 = \mu^2/16$. From these estimates and the lemma we obtain

$$a_0 \leq -\frac{3}{16}\mu^2 < 0, \qquad a_1 \leq 2\mu(\frac{\kappa}{2} + \frac{\kappa}{2} + \frac{1}{4}) = (2\kappa + \frac{1}{2})\mu,$$

$$a_2 \leq 6\kappa^2 - \frac{1}{4}, \qquad a_3 \leq 8\kappa^2 \frac{\sqrt{n}}{\mu}, \qquad a_4 \leq 4\kappa^2 \frac{n}{\mu^2}.$$

In this case we can verify that $p(1/(16\kappa\sqrt[4]{n})) < 0$, so that in the "best case" of the adaptive algorithm we may expect a steplength of $O(1/\sqrt[4]{n})$, while the worst-case bound of $O(1/\sqrt{n})$ on the steplength is maintained regardless.

The factor $\sqrt[4]{n}$ still exists because the iterates remain close to the path of centers (in the 2 norm). If we knew in advance that all projections were well balanced, then the whole analysis could be carried out in terms of the infinity norm as indicated earlier, and the rate of convergence would be independent of n.

4. CONCLUSIONS

The short-step interior-point methods that allow "nice" polynomial-time proofs of convergence for linear programming turn out to be much too slow for practical algorithms. Thus a number of "long-step" methods have been analyzed to date, most of which are aimed at proving the correctness of existing numerical implementations; see in particular [6].

In Section 3.1, however, we have presented a worst-case example in which the iterates x, y, z are not able to "follow" a long step in the reduction of the complementarity parameter μ. The possibility of such worst cases is responsible for the weak proofs of convergence for long-step methods. These proofs not only fail to explain the fast convergence of the implementations that has been observed for *all* numerical examples, but also exhibit a worse theoretical complexity than even the short-step methods.

The adaptive method presented here is intended to close the gap between theoretical and practical complexity. If we were able to show that some percentage of the projections of the vectors s^k are well balanced over a sequence of iterations, then we could indeed reduce the theoretical complexity. This is the goal of future work.

References

[1] I. Adler, R.D.C. Monteiro, "The limiting behavior of the affine scaling trajectories for linear programming problems", Report ESRC 88-9, University of California, Berkeley, CA (1988).

[2] D. den Hertog, C. Roos, "A survey of search directions in interior point methods for linear programming", Report 89-65, Delft University of Technology, The Netherlands (1989).

[3] D. den Hertog, C. Roos, T. Terlaky, "A potential reduction variant of Renegar's short-step path-following method for linear programming", Report 90-14, Delft University of Technology, The Netherlands (1990).

[4] C.C. Gonzaga, "An algorithm for solving linear programming problems in $O(n^3L)$ operations", Memo No. UCB/ERL M87/10, College of Engineering, University of California, Berkeley, CA, (1987).

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

[6] M. Kojima, N. Megiddo, S. Mizuno, "Theoretical convergence of large-step primal-dual interior point algorithms for linear programming", Report RJ 7872 (72532) 12/7/90, IBM Research Division, Yorktown Heights, NY 10598.

[7] M. Kojima, S. Mizuno, A. Yoshise, "A primal-dual interior point algorithm for linear programming", in N. Megiddo, ed., *Progress in Mathematical Programming: Interior-Point and Related Methods*, Springer-Verlag, New York (1989) pp. 29–47.

[8] I.J. Lustig, R.E. Marsten, D.F. Shanno, "On implementing Mehrotra's predictor-corrector interior-point method for linear programming", Report SOR 90-03, Dept. of Civil Eng. and OR, Princeton University, Princeton, NJ 08544 (1990).

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

[10] S. Mehrotra, "On the implementation of a (primal-dual) interior point method", Report 90-03, Dept. of Ind. Engineering and Management Sciences, Northwestern University, Evanston, IL (1990).

[11] S. Mizuno, "An $O(n^3L)$ algorithm using a sequence for a linear complementarity problem", *J. Operations Research Soc. of Japan* 32 (1989).

[12] S. Mizuno, M. Todd, Y. Ye, "Anticipated behavior of long-step algorithms for linear programming," Report No. 24, Department of Industrial Engineering and Management, Tokyo Institute of Technology, Oh-Okayama, Meguro-ku, Tokyo 152, Japan, (1989).

[13] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming", *Mathematical Programming* 40 (1988) 59–93.

[14] C. Roos, J.P. Vial, "Long-steps with the logarithmic penalty barrier function in linear programming", Report 89-44, Delft University of Technology, The Netherlands (1989).

[15] G. Sonnevend, "An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming", *Lecture Notes of Control and Information Sciences* 84 (1986) 866–878.

[16] G. Sonnevend, J. Stoer, G. Zhao, "On the complexity of following the central path by linear extrapolation in linear programs", to appear in U. Rieder and P. Kleinschmidt eds., *Proc. 14 Symp. on Operations Research* (Ulm 1989).

[17] P.M. Vaidya, "An algorithm for solving linear programs which requires $O(((m+n)n^2+(m+n)^{1.5}n)L)$ arithmetic operations", preprint, AT&T Bell Laboratories, Murray Hill, NJ (1987).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 1991	Technical Report

4. TITLE AND SUBTITLE
An Adaptive Primal-Dual Method for Linear Programming

5. FUNDING NUMBERS
N00014-90-J-1242

6. AUTHOR(S)
Florian Jarre and Michael A. Saunders

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Operations Research - SOL
Stanford University
Stanford, CA 94305-4022

8. PERFORMING ORGANIZATION REPORT NUMBER
1111MA

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research - Department of the Navy
800 N. Quincy Street
Arlington, VA 22217

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
SOL 91-1

12a. DISTRIBUTION / AVAILABILITY STATEMENT
UNLIMITED

12b. DISTRIBUTION CODE
UL

13. ABSTRACT *(Maximum 200 words)*

A simple analysis for an adaptive primal-dual method for linear programming is given. Starting from a pair of primal and dual feasible points near the path of centers, the method maintains a worst-case complexity of $O(\sqrt{n}\log\frac{1}{\epsilon})$ iterations to reduce the initial duality gap be a factor of ϵ. In contrast to other interior-point algorithms that share the same the same complexity ([13, 4] and many others), the algorithm proposed here allows an acceleration of the rate of convergence (up to a complexity of $O(\sqrt[4]{n}\log\frac{1}{\epsilon})$) if the problem is "well behaved".

14. SUBJECT TERMS	15. NUMBER OF PAGES
linear programming; interior-point method; theoretical complexity; orthogonal projection.	10 pp.
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED			SAR