PRIMAL-DUAL METHODS FOR LINEAR PROGRAMMING*

Philip E. GILL†, Walter MURRAY‡,
Dulce B. PONCELEÓN§ and Michael A. SAUNDERS‡

Technical Report SOL 91-3¶
Revised March 1994

Abstract

Many interior-point methods for linear programming are based on the properties of the logarithmic barrier function. After a preliminary discussion of the convergence of the (primal) projected Newton barrier method, three types of barrier method are analyzed. These methods may be categorized as primal, dual and primal-dual, and may be derived from the application of Newton's method to different variants of the same system of nonlinear equations. A fourth variant of the same equations leads to a new primal-dual method.

In each of the methods discussed, convergence is demonstrated without the need for a nondegeneracy assumption or a transformation that makes the provision of a feasible point trivial. In particular, convergence is established for a primal-dual algorithm that allows a different step in the primal and dual variables and does not require primal and dual feasibility.

Finally, a new method for treating free variables is proposed.

Key words. Linear programming, barrier methods, primal-dual interior methods

1. Introduction

This paper is concerned with barrier-function methods for the solution of linear programs in the standard form

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b, \quad x \geq 0, \end{array} \tag{1.1}$$

where A is an $m \times n$ matrix with $m \leq n$.

Sections 2 to 6 consider the formulation and analysis of *primal* and *dual* barrier methods. As a preliminary, convergence is established for a *primal* barrier algorithm

*Presented at the Second Asilomar Workshop on Progress in Mathematical Programming, February, 1990, Asilomar, California. Also published in *Math. Program.* **70**, pp. 251–277, 1995

†Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA.

‡Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, CA 94305-4022, USA.

§Apple Computer, Inc., 20525 Mariani Avenue, Cupertino, CA 95014, USA.

¶The material contained in this report is based upon research supported by the National Science Foundation Grant DDM-8715153 and the Office of Naval Research Grant N00014-90-J-1242.

in which the iterates lie in the strict interior of the feasible region. Several methods are then proposed that do not require the primal iterates to satisfy $Ax = b$.

Section 7 concerns the class of *primal-dual* methods, which are the main subject of this paper. Convergence is established for methods in which the primal and the dual iterates are not necessarily feasible. The analysis allows methods with a general choice of the initial approximation to variables, and allows liberal control of the barrier parameter.

A number of authors have described primal-dual algorithms that converge in polynomial time (e.g., Kojima, Mizuno and Yoshise [7]; Monteiro and Adler [14]). However, such algorithms are generally *theoretical* and differ from the relatively few primal-dual algorithms that have been implemented for practical problems (e.g., McShane, Monma and Shanno [11], Lustig, Marsten and Shanno [9, 10], Mehrotra [13], and Gill et al. [3]). Two key differences are the assumption that the step taken in the primal and dual spaces are the same and that the initial estimate of the solution is primal and dual feasible. It may be argued that the feasibility assumption is not overly restrictive because the linear program can be transformed into another problem with an identical solution, but a known feasible point. However, this ignores the possibility that the transformed problem may be more difficult to solve than the original. Recently, Zhang [18] has proved global convergence and polynomiality (under additional assumptions) for a primal-dual algorithm that does not require a feasible point.

Kojima, Megiddo and Mizuno [6] have analyzed a primal-dual algorithm that is more similar to implemented algorithms. They define a steplength rule that allows (but does not guarantee) the possibility of different steps in the primal and dual spaces. They assume that the initial point is feasible

The principal algorithms considered here do not require feasible iterates, and different steps may always be taken in the primal and dual spaces. These algorithms may be loosely categorized as primal, dual or primal-dual in order to distinguish between the different approaches. However, all of them are primal-dual in the sense that this term has been used for interior-point methods.

It is not within the scope of this paper to provide a numerical comparison between methods. Our intention is to give the methods a common setting and thereby highlight their similarities and differences. The principal aim is to define and analyze *implementable* algorithms, but for the purposes of analysis, it is necessary to include procedures that are not present in standard implementations—the most notable being the definition of the steplength as the result of a linesearch instead of as a fixed fraction of the largest feasible step. However, the proposed linesearches are simple to implement and do not add significantly to the cost of an iteration. Moreover, the traditional "fixed" steplength usually satisfies the linesearch criteria. The proofs of convergence demonstrate that almost any step can be taken in the dual space. The existence of a wide range of steps for which convergence occurs may explain the robustness of algorithms that do not incorporate a linesearch.

The analysis applies directly to methods for the more general problem in which some variables have upper bounds or are free. In Section 9 a new technique is described that avoids some numerical difficulties that arise when free variables are

treated using certain Schur complements.

The analysis presented here is easily generalized to indefinite quadratic programs (see Ponceleón [16]).

1.1. Notation and Assumptions

Let x^* denote a solution to (1.1) and let $\mathcal{X}^*$ be the set of all solutions. Let S_0 denote the feasible region $S_0 = \{x : Ax = b,\ x \geq 0\}$ and let $\text{int}(S_0)$ denote its strict interior $\text{int}(S_0) = \{x : Ax = b,\ x > 0\}$. Initially we make the following assumptions:

(i) the constraint matrix A has full row rank;

(ii) the feasible region S_0 is compact;

(iii) $\text{int}(S_0)$ is nonempty.

We shall use N to denote the matrix whose columns form a basis for the null space of A (thus $AN = 0$). Occasionally it will be necessary to refer to the i-th element of a sequence of vectors $\{x_i\}$ and the j-th component y_j of a vector y. To distinguish between x_i and y_j we shall use i to denote the i-th member of a sequence of vectors, and j to denote the j-th component of a vector. Unless otherwise stated, $\|\cdot\|$ refers to the vector two-norm or its induced matrix norm. The vector e denotes the vector $(1, 1, \ldots, 1)^T$ whose dimension is determined by the context.

Assumption (iii) is only required by the first algorithm discussed, which is a feasible-point algorithm. In Section 8 we describe some alternatives to Assumptions (ii) and (iii).

2. Primal Barrier Methods

Barrier methods for linear programming generate approximations to both the primal and dual variables at each iteration. We shall use the term *primal method* to refer to a method that generates positive values of the primal variables x, but does not restrict the values of the dual slack variables z. In the first algorithm we assume that the primal variables are feasible, i.e., that $Ax = b$. This assumption is relaxed for the remaining algorithms.

2.1. The Primal Barrier Subproblem

Barrier methods involve *major* and *minor* iterations. Each major iteration is associated with an element of a decreasing positive sequence of barrier parameters $\{\mu_k\}$ such that $\lim_{k\to\infty} \mu_k = 0$. The minor iterations correspond to an iterative process for the solution of the subproblem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & B(x,\mu) \equiv c^T x - \mu \sum_{j=1}^{n} \ln x_j \\ \text{subject to} \quad & Ax = b, \end{aligned} \tag{2.1}$$

which is solved approximately at every major iteration, i.e., for each value of $\mu = \mu_k$. Since $B(x, \mu)$ is a strictly convex function, the compactness of S_0 implies that there exists a unique minimizer $x^*(\mu)$ such that $Ax^*(\mu) = b$ and $x^*(\mu) > 0$.

Barrier methods are based on the fundamental result that $\lim_{\mu \to 0} x^*(\mu) \in \mathcal{X}^*$. For a proof of this result and a general discussion of barrier methods, see Fiacco and McCormick [2] and Wright [17].

The special form of the derivatives of the barrier function makes Newton's method a natural choice for solving problem (2.1). At any given point x, Newton's method defines a search direction Δx such that $x + \Delta x$ continues to satisfy the linear constraints and minimizes a quadratic approximation to the barrier function. The vector Δx is the solution of the quadratic program

$$\begin{array}{ll} \underset{\Delta x}{\text{minimize}} & \frac{1}{2}\Delta x^T H \Delta x + g^T \Delta x \\ \text{subject to} & A\Delta x = 0, \end{array}$$

where $g(x, \mu) = c - \mu X^{-1} e$ and $H(x, \mu) = \mu X^{-2}$ are $\nabla B(x, \mu)$ and $\nabla^2 B(x, \mu)$, the gradient and Hessian of the barrier function, with $X = \text{diag}(x_j)$. If y denotes an approximation to the Lagrange multiplier vector of (2.1) at x associated with the constraints $Ax = b$, the updated multipliers approximations $y + \Delta y$ at $x + \Delta x$ satisfy

$$K \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = \begin{pmatrix} -g + A^T y \\ 0 \end{pmatrix}, \qquad \text{where} \qquad K \equiv \begin{pmatrix} H & A^T \\ A & \end{pmatrix}. \tag{2.2}$$

We shall refer to this system of equations as the *KKT system* and to the matrix K as the *KKT matrix*.

2.2. The Projected Newton Barrier Method

The formulation of the barrier subproblem (2.1) and the calculation of $x^*(\mu)$ by Newton's method was first embodied in the *projected Newton barrier method* of Gill et al. [4]. The method requires the specification of two positive sequences: a bounded sequence $\{\delta_k\}$ that determines the accuracy of the solutions of (2.1) and a decreasing sequence of barrier parameters $\{\mu_k\}$ such that $\lim_{k \to \infty} \mu_k = 0$.

Algorithm PFP (Model Primal Feasible-Point Algorithm)

Compute x_0 such that $Ax_0 = b$, $x_0 > 0$;
Set $k = 0$, $i = 0$ and $i_k = 0$;
while not *converged* **do**
  Set $\mu = \mu_k$;
  while $\|N^T g(x_i, \mu)\| > \delta_k \mu$ **do**
    Find x_{i+1} such that
    $B(x_{i+1}, \mu) < B(x_i, \mu)$, $x_{i+1} > 0$ and $Ax_{i+1} = b$;
    Set $i = i + 1$;
  end do;
  Set $k = k + 1$, $i_k = i$;
end do

Each member of the subsequence $\{x_{i_k}\}$ corresponds to an approximate minimizer of the subproblem (2.1) defined by μ_k. We shall refer to the consecutive indices of the sequence of minor iterations i_{k-1}, $i_{k-1}+1$, ..., i_k as $\mathcal{I}_k$.

Since $\lim_{k\to\infty}\mu_k = 0$, it follows that $\lim_{k\to\infty}\|x_{i_k} - x_k^*\| = 0$, where x_k^* is the nearest point to x_{i_k} in $\mathcal{X}^*$. The main difficulty lies in generating the sequence of minor iterates $\{x_i,\ i \in \mathcal{I}_k\}$ so that the condition $\|N^T g(x_i,\mu)\| \le \delta_k\mu_k$ is eventually satisfied. This issue is addressed in the next section.

The precise form of the termination condition for the minor iterations is not crucial. The only requirement is that the condition be satisfied in a neighborhood of $x^*(\mu_k)$ that shrinks to the point $x^*(\mu_k)$ as $k \to \infty$.

The form of Algorithm PFP is general enough to allow almost all primal barrier methods defined in the literature. For example, if the elements of the sequence $\{\delta_k\}$ are suitably large and Newton's method is used to define the minor iterates, $\|N^T g(x_i,\mu_k)\|$ is reduced at each step and only a single iteration need be performed before μ is adjusted. (For more discussion of this type of strategy see Gonzaga [5].) In practice, the sequence $\{\delta_k\}$ need only be defined implicitly. For example, if $\|N^T g(x_i,\mu_k)\|/\mu_k$ increases above some threshold, either the reduction in μ may be moderated or δ_k may be chosen so that additional minor iterations are performed.

Similarly, the choice of the next μ can be influenced by either an explicit δ_k or the amount of progress made on the final minor iteration. We emphasize that in practice it is unnecessary to reduce μ_k below some prescribed small value related to the required final accuracy of the approximate solution.

The form of the projected Newton barrier method and its associated convergence proof are easily extended to the case of a nonlinear nonconvex objective function. In the case of nonconvex quadratic programming, Ponceleón [16] has shown that it may be preferable not to reduce μ at every iteration.

2.3. Convergence for the Primal Subproblem

In this section we show that the sequence $\{x_i,\ i = i_{k-1}, \ldots\}$ generated by Newton's method with $\delta_k = 0$ converges to $x^*(\mu_k)$. It follows that for $\delta_k > 0$ the number of minor iterations required to satisfy the termination condition $\|N^T g(x_i,\mu)\| < \delta_k\mu$ is finite.

Throughout this section we shall use the notation

$$\mu = \mu_k, \qquad B(x) = B(x,\mu), \qquad g(x) = g(x,\mu), \qquad H(x) = H(x,\mu),$$

to refer to quantities associated with the k-th subproblem.

The feasible set S_0 is compact by assumption. Given a positive constant θ and a feasible vector w such that $w \ge \theta e$, let $\Omega(w,\mu)$ denote the level set

$$\Omega(w,\mu) = \{x : B(x) \le B(w)\}.$$

We have in mind w being the first minor iterate $x_{i_{k-1}}$ associated with μ, and θ being the smallest component of w. Every subsequent minor iterate will lie in $S_0 \cap \Omega(w,\mu)$.

The essential element of the proof is the demonstration that the KKT matrix is bounded and has a bounded condition number at every point in $S_0 \cap \Omega(w,\mu)$. By

assumption, A is bounded and has a bounded condition number. It follows that K will also have this property if H is bounded and has a bounded condition number. The latter properties in H follow from the next lemma, which shows that $\{(x_i)_j\}$ is bounded above and bounded away from zero.

Lemma 2.1. *Let θ be a positive constant and w a given vector such that $w \geq \theta e$ and $Aw = b$. There exist positive constants σ_X and τ_X, independent of x, such that $\sigma_X e \leq x \leq \tau_X e$ for all $x \in S_0 \cap \Omega(w, \mu)$.*

Proof. Let $\widehat{S}$ denote the set $S_0 \cap \Omega(w, \mu)$. $\widehat{S}$ is compact since it is the intersection of the two closed sets S_0 and Ω, and it is a subset of the bounded set S_0. Since $\widehat{S}$ is compact, there exists a constant τ_X such that $x_j \leq \tau_X$ for $x \in \widehat{S}$. The definition of $\widehat{S}$ implies that every $x \in \widehat{S}$ gives $B(x) \leq B(w)$. It follows that for all $x \in \widehat{S}$,

$$c^T x - \mu \sum_{j=1}^{n} \ln x_j \leq c^T w - \mu \sum_{j=1}^{n} \ln w_j.$$

Therefore for each j,

$$-\mu \ln x_j \leq c^T w - \mu \sum_{j=1}^{n} \ln w_j - c^T x + \mu \sum_{r \neq j} \ln x_r.$$

Since $\widehat{S}$ is compact, the quantities $\omega = \max\{|c^T x| : x \in \widehat{S}\}$ and $\hat{\beta} = \max\{\ln x_j : x \in \widehat{S}\}$ are bounded. Similarly, if $\theta > 0$, the quantity $\beta = \max\{\hat{\beta}, -\ln \theta\}$ is also bounded and $-\mu \ln x_j \leq 2\omega + 2n\mu\beta$, or equivalently,

$$x_j \geq e^{-2(n\beta + \omega/\mu)} > 0,$$

as required. ∎

Corollary 2.1. *Let x be any element of $S_0 \cap \Omega(w, \mu)$. Let $H(x) = \mu X^{-2}$ where $X = \operatorname{diag}(x_j)$. Then there exist positive constants σ_H and τ_H, independent of x, such that for all vectors u,*

$$\sigma_H \|u\|^2 \leq u^T H(x) u \leq \tau_H \|u\|^2. \quad ∎$$

Lemma 2.2. *At every element of the sequence $\{x_i,\ i \in \mathcal{I}_k\}$ generated by Algorithm* PFP*, the matrix K is bounded and has a bounded condition number.* ∎

We now show that the sequence $\{x_i\}$ generated by Newton's method converges to $x^*(\mu)$, which implies that the condition $\|N^T g(x_i)\| \leq \delta_k \mu$ will be satisfied in a finite number of iterations.

The iterates of the projected Newton barrier method satisfy $x_{i+1} = x_i + \alpha_i \Delta x_i$, where Δx_i is a search direction defined by (2.2) and α_i is a scalar steplength that produces a sufficient decrease in $B(x)$ along Δx_i. Here we use the *Goldstein-Armijo* algorithm (sometimes known as the *backtracking* algorithm) to determine

the steplength, although any of the standard steplength algorithms would be suitable (see, e.g., Ortega and Rheinboldt [15]).

The Goldstein-Armijo algorithm defines a steplength of the form $\alpha_i = \bar{\alpha}_i/2^{j_i-1}$, where $\bar{\alpha}_i$ is an initial step and j_i ($j_i \geq 1$) specifies the largest fraction of $\bar{\alpha}_i$ such that

$$B(x_i + \alpha_i \Delta x_i) \leq B(x_i) + \eta \alpha_i \Delta x_i^T g(x_i), \tag{2.3}$$

for some preassigned constant η ($0 < \eta \leq \frac{1}{2}$). The initial step $\bar{\alpha}_i$ is the smaller of the unit step and a fixed fraction of β_i, the distance to the boundary. Thus, $\beta_i = \min_r \{-(x_i)_r/(\Delta x_i)_r : (\Delta x_i)_r < 0\}$ and the initial step $\bar{\alpha}_i$ is

$$\bar{\alpha}_i = \min \{1, \omega \beta_i\},$$

where the constant ω ($0 < \omega < 1$) is typically 0.9 or 0.99.

Theorem 2.1. *If $\{x_i\}$ is the infinite sequence generated by Newton's method with a Goldstein-Armijo linesearch, then* $\lim_{i\to\infty} \|x_i - x^*(\mu)\| = 0$.

Proof. Without loss of generality we assume that $k = 0$. Let $\widehat{S} = S_0 \cap \Omega(x_0, \mu)$.

The following inductive argument shows that every x_i lies in $\widehat{S}$. Assume that $x_i \in \widehat{S}$. The corollary to Lemma 2.1 states that there exists a positive constant σ_H such that

$$\Delta x_i^T H(x_i) \Delta x_i \geq \sigma_H \|\Delta x_i\|^2. \tag{2.4}$$

From (2.2) and (2.4) we have

$$\Delta x_i^T g(x_i) = -\Delta x_i^T H(x_i) \Delta x_i \leq -\sigma_H \|\Delta x_i\|^2. \tag{2.5}$$

Since $B(x)$ is convex and $B(x_i + \alpha \Delta x_i) \to +\infty$ as $\alpha \to \beta_i$, there exists a unique positive α^* such that

$$B(x_i + \alpha^* \Delta x_i) = B(x_i) + \eta \alpha^* \Delta x_i^T g(x_i). \tag{2.6}$$

This implies that there exists at least one strictly feasible point $x_i + \alpha_i \Delta x_i$ that satisfies the sufficient decrease condition (2.3). From (2.3) and (2.5) we obtain

$$B(x_i) - B(x_{i+1}) \geq -\eta \alpha_i \Delta x_i^T g(x_i) \geq \eta \sigma_H \alpha_i \|\Delta x_i\|^2, \tag{2.7}$$

so that $B(x_{i+1}) < B(x_i)$ and x_{i+1} lies in $\widehat{S}$. By induction, $x_i \in \widehat{S}$ for all i.

Since $\widehat{S}$ is compact and $B(x_{i+1}) < B(x_i)$, the sequence $\{B(x_i)\}$ is monotone decreasing and bounded below. Thus $\{B(x_i)\}$ converges, and

$$\lim_{i\to\infty} \{B(x_i) - B(x_{i+1})\} = 0. \tag{2.8}$$

This result and the inequality (2.7) imply that $\lim_{i\to\infty} \alpha_i \|\Delta x_i\|^2 = 0$.

Let I_1 denote the subsequence of iterations for which the initial step $\bar{\alpha}_i$ produces a sufficient decrease in B. In this case, (2.7) yields

$$B(x_i) - B(x_{i+1}) \geq \eta \sigma_H \bar{\alpha}_i \|\Delta x_i\|^2 \quad \text{for all} \quad i \in I_1,$$

and it follows that $\lim_{i\to\infty} \Delta x_i = 0$ for all $i \in I_1$ if $\bar{\alpha}_i$ is always greater than a positive constant $\alpha_{\min}$. Consider the Taylor-series expansion of $B(x_i + \alpha_i^* \Delta x_i)$ in (2.6). Then

$$B(x_i + \alpha_i^* \Delta x_i) = B(x_i) + \alpha_i^* \Delta x_i^T g(x_i) + \tfrac{1}{2}\alpha_i^{*2} \Delta x_i^T H(\widehat{x}_i) \Delta x_i, \tag{2.9}$$

where $\widehat{x}_i = x_i + \theta \alpha_i^* \Delta x_i$ for some $0 \le \theta \le 1$. Combining (2.6) and (2.9) gives

$$\alpha^* = 2(\eta - 1)\frac{\Delta x_i^T g(x_i)}{\Delta x_i^T H(\widehat{x}_i) \Delta x_i} = 2(1 - \eta)\frac{\Delta x_i^T H(x_i) \Delta x_i}{\Delta x_i^T H(\widehat{x}_i) \Delta x_i}. \tag{2.10}$$

Since x_i, $\widehat{x}_i \in \widehat{S}$, the corollary of Lemma 2.1 implies that $\alpha^* \ge 2(1 - \eta)\sigma_H/\tau_H$. Finally, since $\beta_i > \alpha^*$ and $\bar{\alpha}_i = \min\{1,\ \omega\beta_i\}$, we have

$$\bar{\alpha}_i \ge \alpha_{\min} > 0, \quad \text{where} \quad \alpha_{\min} = \min\left\{1,\ 2\omega(1 - \eta)\frac{\sigma_H}{\tau_H}\right\}. \tag{2.11}$$

Since $\bar{\alpha}_i$ is bounded away from zero by $\alpha_{\min}$, (2.7) gives

$$B(x_i) - B(x_{i+1}) \ge \eta \alpha_{\min} \sigma_H \|\Delta x_i\|^2, \quad \text{for all} \quad i \in I_1.$$

and $\lim_{i\to\infty} \Delta x_i = 0$ for all $i \in I_1$.

Now consider the subsequence I_2 such that $\alpha_i < \bar{\alpha}_i$ and $x_i + 2\alpha_i \Delta x_i \notin \widehat{S}$ (i.e., the penultimate step of the linesearch gives a point lying outside $\widehat{S}$). Then

$$2\alpha_i \ge \alpha^* \ge 2(1 - \eta)\frac{\sigma_H}{\tau_H},$$

and similar arguments to those above give $\lim_{i\to\infty} \Delta x_i = 0$ for $i \in I_2$.

Finally, consider the subsequence I_3 such that $\alpha_i < \bar{\alpha}_i$ and $x_i + 2\alpha_i \Delta x_i \in \widehat{S}$. Then

$$B(x_i + 2\alpha_i \Delta x_i) > B(x_i) + \eta 2\alpha_i \Delta x_i^T g(x_i). \tag{2.12}$$

The Taylor-series expansion of $B(x_i + 2\alpha_i \Delta x_i)$ gives

$$B(x_i + 2\alpha_i \Delta x_i) = B(x_i) + 2\alpha_i \Delta x_i^T g(x_i) + 2\alpha_i^2 \Delta x_i^T H(\bar{x}_i) \Delta x_i,$$

where $\bar{x}_i = x_i + \xi 2\alpha_i \Delta x_i$ for some $0 \le \xi \le 1$. Substituting this expression for $B(x_i + 2\alpha_i \Delta x_i)$ in (2.12) gives

$$|\Delta x_i^T g(x_i)| \le \frac{\alpha_i}{1 - \eta} \Delta x_i^T H(\bar{x}_i) \Delta x_i. \tag{2.13}$$

Since $\widehat{S}$ is convex, $\bar{x}_i \in \widehat{S}$ and it follows from the corollary to Lemma 2.1 that there exists a constant τ_H such that

$$\Delta x_i^T H(\bar{x}_i) \Delta x_i \le \tau_H \|\Delta x_i\|^2.$$

Combining this inequality with (2.13) gives

$$|\Delta x_i^T g(x_i)| \le \frac{\alpha_i}{1 - \eta} \Delta x_i^T H(\bar{x}_i) \Delta x_i \le \frac{\tau_H}{1 - \eta} \alpha_i \|\Delta x_i\|^2.$$

Since $\lim_{i\to\infty} \alpha_i \|\Delta x_i\|^2 = 0$, we obtain

$$\lim_{i\to\infty} \Delta x_i^T g(x_i) = 0 \quad \text{for all} \quad i \in I_3.$$

It follows from (2.5) that $\lim_{i\to\infty} \Delta x_i = 0$ for $i \in I_3$.

From (2.2) we have

$$N^T H(x_i) N \Delta x_{N_i} = -N^T g(x_i),$$

where $\Delta x_i = N \Delta x_{N_i}$. Since $N^T H(x_i) N$ is bounded and has a bounded condition number, it follows from $\lim_{i\to\infty} \Delta x_i = 0$ that $\lim_{i\to\infty} N^T g(x_i) = 0$. Since $x^*(\mu)$ is the unique feasible point for which $N^T g(x^*(\mu)) = 0$, we have $\lim_{i\to\infty} \|x_i - x^*(\mu)\| = 0$ as required. ∎

3. Getting Feasible

There are various ways to eliminate the requirement that x_0 is a strictly feasible point.

3.1. An Artificial Variable

A common approach is to introduce an additional variable, or set of variables, and minimize a composite objective function. For example, given $x_0 > 0$, consider the transformed problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n,\, \xi\in\mathbb{R}}{\text{minimize}} \quad & c^T x + \rho\xi \\ \text{subject to} \quad & Ax + \xi u = b, \quad x \geq 0, \quad \xi \geq -1, \end{aligned}$$

where $u = (b - Ax_0)/\|b - Ax_0\|$ and ρ is a positive scalar. The initial value of ξ is $\|b - Ax_0\|$, so a strictly feasible point for the transformed problem is known. If a step would make ξ negative during an iteration, a shorter step is taken to make $\xi = 0$. Once ξ is zero, it is eliminated from the problem.

This type of approach has a number of associated difficulties, including the introduction of a dense column in the matrix of constraints. The principal difficulty lies in the choice of the parameter ρ. The value of ρ must be chosen large enough that the original and transformed problems have the same solution, yet small enough that the infeasibilities do not dominate the objective and make the method behave like a two-phase algorithm. It is particularly difficult to choose a suitable value for ρ for large problems that are poorly scaled. Although automatic scaling procedures exist, they cannot always be assured of success. In practice, ρ is altered dynamically—for example, the value chosen to reduce the initial infeasibilities can be increased whenever the infeasibilities are not being reduced sufficiently rapidly. If the constraints are feasible but have no strict interior, ρ is present at *all* iterations and is likely to become large, whatever adjustment strategy is used.

3.2. A Merit Function

The method of Section 2.1 may be generalized so that Δx is the solution of the quadratic program

$$\begin{array}{ll} \underset{\Delta x}{\text{minimize}} & \frac{1}{2}\Delta x^T H \Delta x + g^T \Delta x \\ \text{subject to} & A\Delta x = b - Ax \end{array}$$

and satisfies

$$K \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = \begin{pmatrix} -g + A^T y \\ b - Ax \end{pmatrix}. \tag{3.1}$$

Progress may now be measured by means of a *merit function* that balances the aims of minimizing $B(x,\mu)$ and reducing some norm of $Ax-b$. For example, one possible merit function is

$$M(x,\rho) = B(x,\mu) + \rho\|Ax - b\|_1,$$

where ρ is chosen suitably large. It can be shown that the direction Δx defined by (3.1) is a descent direction for $M(x,\rho)$. (For a proof in the context of a related method, see Section 7.) Convergence follows from arguments similar to those used in the discussion of the feasible-point algorithm.

It would appear that this approach also depends on choosing a good value for the parameter ρ. However, Δx is independent of ρ, and so ρ affects only the magnitude of the steplength. Moreover, it is relatively easy to adjust ρ dynamically. For example, given a steplength that provides a sufficient decrease in the pure barrier function, ρ can be increased *a posteriori* if necessary to give sufficient decrease in $M(x,\rho)$.

Primal-infeasible methods will not be discussed further as we prefer the primal-dual approach discussed in the next section. However, the merit function $M(x,\rho)$ will be reconsidered in Section 7 in conjunction with a primal-dual method.

3.3. Newton's Method Applied to the Optimality Conditions

Since $B(x,\mu)$ is strictly convex, $x^*(\mu)$ is the only strictly feasible constrained stationary point for problem (2.1). This suggests an alternative method for finding $x^*(\mu)$ based on using Newton's method for *nonlinear equations* to find the stationary point of the Lagrangian $L(x,y) = B(x,\mu) - y^T(Ax-b)$. Since the gradient of $L(x,y)$ is zero at $x^*(\mu)$, we obtain the $n+m$ nonlinear equations

$$\nabla L(x,y) = \begin{pmatrix} c - \mu X^{-1}e - A^T y \\ Ax - b \end{pmatrix} = 0, \tag{3.2}$$

whose Jacobian is closely related to the KKT matrix K. The KKT system (3.1) defines a descent direction for $\|\nabla L\|$, and a steplength may be chosen to achieve a sufficient reduction in $\|\nabla L\|$. As in Algorithm PFP, this merit function ensures that x_j cannot be arbitrarily close to its bound.

This approach is now extended to obtain the algorithms of principal interest in this paper.

4. A Primal Primal-Dual Method

Following common practice, we introduce a third vector of variables $z = c - A^T y$ and solve the nonlinear equations $f_P(z, x, y) = 0$, where

$$f_P(z, x, y) \equiv \begin{pmatrix} \bar{f} \\ \widehat{f} \\ r \end{pmatrix} = \begin{pmatrix} z - \mu X^{-1} e \\ c - A^T y - z \\ Ax - b \end{pmatrix}. \tag{4.1}$$

When it is necessary to consider the full vector of variables z, x and y, the vector v will denote the $(2n + m)$-vector $(z,\ x,\ -y)$. The symbols $f_P(z, x, y)$ and $f_P(v)$ will be used interchangeably for f_P, depending on the point of emphasis. The Newton direction $\Delta v = (\Delta z,\ \Delta x,\ -\Delta y)$ satisfies the linear system

$$J_P \Delta v = -f_P, \qquad \text{where} \qquad J_P = \begin{pmatrix} I & \mu X^{-2} & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}. \tag{4.2}$$

Apart from the last block of columns being multiplied by -1, J_P is the Jacobian of the nonlinear equations (4.1). We shall refer to J_P as the Jacobian.

The directions Δx and Δy from (4.2) are identical to those defined by the KKT system (3.1), and to those associated with (3.2). However, for the nonlinear equations $\nabla L(x, y) = 0$ and $f_P(z, x, y) = 0$, the steplength is chosen to produce a sufficient decrease in $\|\nabla L\|$ and $\|f_P\|$ respectively. In the latter case, the sufficient decrease condition for the Goldstein-Armijo linesearch is

$$\|f_P(v_i + \alpha_i \Delta v)\| \leq (1 - \eta\alpha_i)\|f_P(v_i)\|, \tag{4.3}$$

which is easily tested.

Since the residuals $\widehat{f}$ and r are linear in x, y and z, they are simply related to their values in the previous iteration. Suppose that r and $\widehat{f}$ are nonzero at iteration i. After a step of Newton's method with steplength α_i, we have

$$r_{i+1} = (1 - \alpha_i) r_i \quad \text{and} \quad \widehat{f}_{i+1} = (1 - \alpha_i)\widehat{f}_i. \tag{4.4}$$

At the first iteration, $\|z_0\|$ and $\|y_0\|$ are bounded and x_0 is bounded away from zero, which implies that the Jacobian is bounded and has a bounded condition number. It follows that $\alpha_0 > 0$. Hence the relations (4.4) imply that $r_i = \gamma_i r_0$ for some scalar γ_i such that $0 \leq \gamma_i < \bar{\gamma} < 1$. If a unit step is taken at any iteration, $\widehat{f}$ and r will be zero in all subsequent iterations.

The complete algorithm is as follows.

Algorithm PPD (Model Primal Primal-Dual Algorithm)

Set v_0, with $x_0 > 0$ and $z_0 > 0$;
Set $k = 0$, $i = 0$ and $i_k = 0$;
while not *converged* **do**
  Set $\mu = \mu_k$;
  while $\|f_P(v_i, \mu)\| > \delta_k \mu$ **do**
    Find v_{i+1} such that
    $\|f_P(v_{i+1}, \mu)\| < \|f_P(v_i, \mu)\|$ and $x_{i+1} > 0$;
    Set $i = i + 1$;
  end do;
  Set $k = k + 1$, $i_k = i$;
end do

4.1. Convergence

Convergence is established by showing that J_P is bounded and has a bounded condition number for every barrier subproblem. In contrast to Algorithm PFP, whose iterates lie in S_0, it is not obvious that the primal iterates of Algorithm PPD are bounded. This property is verified in the next lemma and then used to show that the primal-dual iterates $\{v_i\}$ also lie in a compact set.

Lemma 4.1. *Let τ_r denote a positive constant. If the feasible region S_0 is compact, then so is the set*

$$S_A = \{x : x \geq 0,\ \|Ax - b\| \leq \tau_r\}\,.$$

Proof. Since S_A is closed, it only remains to be shown that S_A is bounded. If S_A is not bounded there must exist an $\bar{x} \in S_A$ and a unit vector u such that $\bar{x} + \gamma u \in S_A$ for all $\gamma > 0$. If $Au \neq 0$, then $\|A(\bar{x} + \gamma u)\| > \tau_r$ for some γ sufficiently large. Similarly, if some element of u is negative, then $\bar{x} + \gamma u \not\geq 0$ for some γ sufficiently large. It follows that if S_A is unbounded there must exist u such that $Au = 0$, $u \geq 0$ and $\|u\| = 1$. The existence of such a vector contradicts the assumption that S_0 is compact. ∎

Lemma 4.2. *Let r_0 denote the residual $r_0 = Ax_0 - b$, with $x_0 > 0$. Define the set*

$$S_{\bar{\gamma}} = \{(z, x, y) : x \geq 0,\ Ax - b = \gamma r_0 \quad \textit{for some} \quad 0 \leq \gamma \leq \bar{\gamma}\}\,,$$

where $\bar{\gamma}$ is given and satisfies $0 \leq \bar{\gamma} < 1$. Also define the level set

$$\Gamma(\tau_f, \mu) = \{(z, x, y) : \|f_P(z, x, y)\| \leq \tau_f\}\,.$$

Then $\widehat{S} = S_{\bar{\gamma}} \cap \Gamma(\tau_f, \mu)$ is compact.

Proof. Throughout this proof we shall assume that (z, x, y) is a vector in $\widehat{S}$. From the definition of $S_{\bar{\gamma}}$ we have $\|Ax - b\| \leq \|r_0\|$ and it follows from Lemma 4.1 that the x is bounded. It remains to be shown that the y and z components are bounded.

Note that the components of both $\bar{f}$ and $\widehat{f}$ are bounded since they are components of the bounded vector f_P.

Consider the equations $\bar{f} = z - \mu X^{-1}e$ of (4.1). Premultiplying $\bar{f}$ by x^T and using the fact that both x and $\bar{f}$ are bounded, it follows that a constant τ_1 exists such that

$$x^T z = x^T \bar{f} + \mu n < \tau_1. \tag{4.5}$$

Also, since $x \geq 0$ it follows from (4.1) that

$$z_j > \bar{f}_j > -\tau_2 \tag{4.6}$$

for some positive constant τ_2.

If x^T is now applied to the second equation of (4.1), $\widehat{f} = c - A^T y - z$, we obtain

$$x^T \widehat{f} = x^T c - x^T A^T y - x^T z = x^T c - (b^T + r^T)y - x^T z.$$

Simple rearrangement and the definition of r from (4.1) gives

$$-(b^T + r^T)y = x^T \widehat{f} + x^T z - x^T c. \tag{4.7}$$

It follows from (4.5) and the bounds on $\widehat{f}$ and x that

$$-(b^T + r^T)y < \tau_3. \tag{4.8}$$

Similarly, using $x = x_0$ in (4.7) gives

$$(b^T + r_0^T)y = -x_0^T \widehat{f} - x_0^T z + x_0^T c \leq -x_0^T \widehat{f} - \sum_{J_-} (x_0)_j z_j + x_0^T c,$$

where J_- is the set of indices of the *negative* components of z. (Recall that the components of x_0 are positive.) It follows from (4.6) that

$$(b^T + r_0^T)y < \tau_4. \tag{4.9}$$

Using (4.8) and the assumption that $r = \gamma r_0$ gives

$$-(b^T + \gamma r_0^T)y < \tau_3. \tag{4.10}$$

Combining (4.9) and (4.10) and using the inequality $0 \leq \gamma < \bar{\gamma} < 1$ gives

$$-b^T y \leq \frac{\tau_3 + \gamma \tau_4}{1 - \gamma} < \tau_5.$$

Using (4.7) with $x = x^*(\mu)$ gives

$$x^*(\mu)^T z = x^*(\mu)^T c - x^*(\mu)^T \widehat{f} - b^T y. \tag{4.11}$$

Since $x_j^*(\mu) > 0$ and $\|x^*(\mu)\|$ is bounded (see Lemma 2.1), all the terms on the right-hand side of this expression are bounded, with $x^*(\mu)^T z < \tau_6$ for some positive constant τ_6. Lemma 2.1 also implies the existence of positive constants σ_X and τ_X

such that $\sigma_X \leq x_j^*(\mu) \leq \tau_X$. It follows from the lower bound on $x_j^*(\mu)$ and (4.11) that $\|z\|$ is bounded, with

$$z_j < (\tau_6 + n\tau_X\tau_2)/\sigma_X.$$

Since A has full row rank, the bounds on $\|\widehat{f}\|$ and $\|z\|$ in the equation $\widehat{f} = c - A^T y - z$ imply that $\|y\|$ is bounded, as required. ∎

Lemma 4.3. *If $v \in \widehat{S}$ then J_P is bounded and has a bounded condition number.*

Proof. It is enough to show that x_j is bounded away from zero if $v \in \widehat{S}$. We have from (4.1) that $z_j - \bar{f}_j = \mu/x_j$. Hence

$$|z_j| + \|\bar{f}\| \geq \mu/x_j \quad \text{or equivalently} \quad x_j \geq \frac{\mu}{|z_j| + \|\bar{f}\|}.$$

It follows from Lemma 4.2 that there exists a positive constant τ_Z such that $|z_j| < \tau_Z$ for all $v \in \widehat{S}$, and by assumption, $\|\bar{f}\| \leq \tau_f$. Hence, $x_j \geq \mu/(\tau_Z + \tau_f) > 0$.

From Lemma 4.1, x is uniformly bounded above. Since x_j is bounded away from zero, J_P is bounded and the condition number of J_P is bounded. ∎

The proof of the next theorem is similar to that for Theorem 2.1.

Theorem 4.1. *If $\{v_i\}$ is the infinite sequence generated by Newton's method applied to $f_P(v) = 0$, with steplength satisfying the Goldstein-Armijo condition (4.3), then $\lim_{i\to\infty} \|v_i - v^*(\mu)\| = 0$.* ∎

It follows that Newton's method generates a point that satisfies the condition $\|f_P(v_i, \mu)\| \leq \delta_k \mu$ in a finite number of iterations.

5. Summary of Primal Methods

In all the algorithms considered so far (excluding the artificial-variable method of Section 3.1), the search directions for x and y are the same as those given by (4.2). The steplength α may be chosen to reduce one of the following functions:

(i) $M(x, \rho) = B(x, \mu) + \rho\|Ax - b\|_1$ (search in x space).

(ii) $\|c - \mu X^{-1}e - A^T y\|^2 + \|Ax - b\|^2$ (search in x and y space).

(iii) $\|c - z - A^T y\|^2 + \|z - \mu X^{-1}e\|^2 + \|Ax - b\|^2$ (search in x, y and z space).

The only additional restriction on α is the requirement that $x + \alpha\Delta x > 0$. In all cases, approximations in the x, y and z space may be generated even though they are *necessary* only in (iii). Thus, all three methods may be viewed as primal-dual algorithms.

If some steplength other than α is taken along Δz and Δy, a sequence of *auxiliary* y and z values can be generated that approximate y^* and z^*. For this sequence, a different step α_Z in the y and z space is needed to maintain $z > 0$. Since α_Z is not

usually equal to α, a dual feasible point may be found before a primal feasible point (or vice versa). Provided that the step taken in the y space is also α_Z, once a dual feasible point is found, all subsequent approximations will be dual feasible.

One advantage of (ii) and (iii) is that it is not necessary to compute logarithms. Moreover, it is not necessary to define a parameter ρ that balances feasibility and optimality, although it may be advantageous to weight the norms occurring in (ii) and (iii).

6. Dual Methods

The dual of the linear program (1.1) may be written as

$$\begin{aligned} \underset{y,\,z}{\text{minimize}} \quad & -b^T y \\ \text{subject to} \quad & c - A^T y - z = 0, \quad z \geq 0. \end{aligned} \tag{6.1}$$

The dual barrier subproblem is

$$\begin{aligned} \underset{y\in\mathbb{R}^m,\; z\in\mathbb{R}^n}{\text{minimize}} \quad & -b^T y - \mu \sum_{j=1}^{n} \ln z_j \\ \text{subject to} \quad & c - A^T y - z = 0. \end{aligned} \tag{6.2}$$

If Newton's method is applied to this problem the direction Δy is defined from a system similar to (2.2). (The right-hand side is different and $H = (1/\mu)Z^2$, where $Z = \operatorname{diag}(z_j)$.) Given an initial point $(y_0,\ z_0)$ in the strict interior of the dual constraints, an algorithm DFP analogous to PFP may be defined.

Similarly, an algorithm may be formulated based upon the optimality conditions for (6.2):

$$\begin{aligned} x - \mu Z^{-1} e &= 0, \\ c - A^T y - z &= 0, \\ Ax - b &= 0. \end{aligned} \tag{6.3}$$

As noted by Megiddo [12], the solution of these equations is identical to the solution of (4.1). Newton's method applied to (6.3) solves the linear system $J_D \Delta v = -f_D$, where

$$f_D(z,x,y) \equiv \begin{pmatrix} \bar{f} \\ \widehat{f} \\ r \end{pmatrix} = \begin{pmatrix} x - \mu Z^{-1} e \\ c - A^T y - z \\ Ax - b \end{pmatrix} \tag{6.4}$$

and

$$J_D = \begin{pmatrix} \mu Z^{-2} & I & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}.$$

The resulting algorithm, DPD, is identical to PPD except that J_P and f_P are replaced by J_D and f_D, and the z variables are restricted during the linesearch instead of the x variables.

Instead of assuming the primal feasible region is bounded we now assume the dual feasible region is bounded. Bounding the feasible region may be done by adding to the original problem upper bounds on the variables of the form $x_i \leq M$, where M is large. Such bounds have a negligible affect on Algorithm PPD, but would have a catastrophic affect on Algorithm DPD, since M would now appear in the objective. Fortunately, adding simialr bounds on z to the dual problem has a negligible affect on Algorithm DPD. Hence it makes more sense to assume the dual feasible region is bounded. We now need to show that $\|x\|$ generated by Algorithm DPD is bounded. Note that a bound on $\|z\|$ implies a bound on $\|y\|$ when $\widehat{f}$ is bounded and A is full rank.

Lemma 6.1. *Let $\widehat{f}_0$ denote $\widehat{f}_0 = c - A^T y_0 - z_0$. Define the set*

$$S_{\bar{\gamma}} = \left\{(z, x, y) : z > 0,\ \widehat{f} = \gamma \widehat{f}_0 \quad \text{for some} \quad 0 \leq \gamma \leq \bar{\gamma}\right\},$$

where $\bar{\gamma}$ is given and satisfies $0 \leq \bar{\gamma} < 1$. Also define the level set

$$\Gamma(\tau_f, \mu) = \{(z, x, y) : \|f_D(z, x, y)\| \leq \tau_f\}.$$

Then $\widehat{S} = S_{\bar{\gamma}} \cap \Gamma(\tau_f, \mu)$ is compact.

Proof. Throughout the proof we assume $(z, x, y) \in \widehat{S}$. The required result follows if $\|x\|$ is bounded. We now have $\widehat{f}$, $\bar{f}$ and r defined as in (6.4). From (6.4) we get

$$x_i z_i - \mu = z_i \bar{f}_i,$$

which implies

$$x^T z < \tau_1 < \infty. \tag{6.5}$$

It also follows from (6.4) and $z_i > 0$ that

$$x_i > -\tau_2 > -\infty. \tag{6.6}$$

Premultiplying $\widehat{f}$ in (6.4) by x^T and rearranging gives

$$c^T x - \widehat{f}^T x = x^T z + (b + r)^T y < \tau_3 < \infty, \tag{6.7}$$

with the inequality following from (6.5). Using $z = z_0$ and $y = y_0$ in the above equation gives

$$-c^T x + \widehat{f}_0^T x = -x^T z_0 - (b + r)^T y_0 < \tau_4, \tag{6.8}$$

with the inequality following from $z > 0$ and (6.6). Since $\widehat{f} = \gamma \widehat{f}_0$ it follows from (6.7) that

$$c^T x - \gamma \widehat{f}_0^T x < \tau_3. \tag{6.9}$$

Combining (6.8) and (6.7) gives

$$c^T x < \frac{\tau_3 + \gamma \tau_4}{1 - \gamma} < \tau_5. \tag{6.10}$$

We have that $c = A^T y^*(\mu) + z^*(\mu)$. Premultiplying this equation by x^T gives

$$x^T z^*(\mu) = c^T x - b^T y^*(\mu).$$

Since $z_i^*(\mu)$ is bounded away from zero it follows from the bound on $c^T x$ given in (6.10) that $x_i < \infty$, which when combined with the previous lower bound derived on x_i implies $\|x\|$ is bounded as required. ∎

Lemma 6.2. *If $v \in \widehat{S}$ then J_D is bounded and has a bounded condition number.*

Proof. It is enough to show that z_j is bounded away from zero if $v \in \widehat{S}$. We have from (6.4) that $x_j - \bar{f}_j = \mu / z_j$. Hence

$$|x_j| + \|\bar{f}\| \geq \mu / z_j \quad \text{or equivalently} \quad z_j \geq \frac{\mu}{|x_j| + \|\bar{f}\|}.$$

It follows from Lemma 6.1 that there exists a positive constant τ_X such that $|x_j| < \tau_X$ for all $v \in \widehat{S}$, and by assumption, $\|\bar{f}\| \leq \tau_f$. Hence, $z_j \geq \mu/(\tau_X + \tau_f) > 0$.

Since z_j is bounded and bounded away from zero, J_D is bounded and has a bounded condition number. ∎

The proof of the next theorem is similar to that for Theorem 2.1.

Theorem 6.1. *If $\{v_i\}$ is the infinite sequence generated by Newton's method applied to $f_D(v) = 0$, with steplength satisfying the Goldstein-Armijo condition (4.3), then $\lim_{i \to \infty} \|v_i - v^*(\mu)\| = 0$.*

∎

As with Algorithm PPD, an auxiliary positive approximation to x^* may be generated by allowing the primal and dual steplengths to be different.

7. Primal-Dual Methods

7.1. A Primal-Dual Method

Algorithms PPD and DPD both generate a sequence of approximations to $v^*(\mu)$. However, $v^*(\mu)$ also solves the nonlinear system $f_{\text{PD}}(z, x, y) = 0$, where

$$f_{\text{PD}}(z, x, y) \equiv \begin{pmatrix} \bar{f} \\ \widehat{f} \\ r \end{pmatrix} = \begin{pmatrix} Xz - \mu e \\ c - A^T y - z \\ Ax - b \end{pmatrix}. \tag{7.1}$$

Newton's method for these equations leads to the linear system

$$J_{\text{PD}} \Delta v = -f_{\text{PD}}, \qquad \text{where} \qquad J_{\text{PD}} = \begin{pmatrix} X & Z & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}, \tag{7.2}$$

which has been used by Lustig, Marsten and Shanno [9, 10], Mehrotra [13], and Gill et al. [3] (see also Lustig [8]). Methods based on the solution of (7.2) are usually referred to as *primal-dual* algorithms because both x and z are maintained to be positive. It must be stressed that this terminology does not imply any direct connection between (7.2) and the primal-dual form of LP. If the latter is transformed using a barrier function, the resulting optimality conditions involve six sets of variables and two independent systems of equations that are identical to (4.1) and (6.3).

Unlike J_P and J_D, $J_{\rm PD}$ is independent of μ. If α is chosen to maintain sufficient positivity in both x and z, $J_{\rm PD}$ will be a bounded matrix with a bounded condition number. A key feature of these equations is that it is no longer obvious that the Goldstein-Armijo steplength maintaining both $z > 0$ and $x > 0$ is bounded away from zero.

We therefore propose an algorithm that takes a different step in the x and (y, z) spaces and uses $M(x, \rho)$ as a merit function instead of $\|f_P\|$. If σ_Z, τ_Y and τ_Z are preassigned positive constants, let S_Y and S_Z be the sets

$$S_Y = \{y : \|y\| \leq \tau_Y\} \quad \text{and} \quad S_Z = \{z : 0 < \sigma_Z e \leq z \leq \tau_Z e\}.$$

Algorithm PD (Model Primal-Dual Algorithm)

Set v_0, with $x_0 > 0$, $z_0 \in S_Z$ and $y_0 \in S_Y$;
Set $k = 0$, $i = 0$ and $i_k = 0$;
while not *converged* **do**
  Set $\mu = \mu_k$;
  while $\|N^T g(x_i, \mu)\| + \|r\| > \delta_k \mu$ **do**
    Select any $z_{i+1} \in S_Z$ and $y_{i+1} \in S_Y$;
    Solve $J_{\rm PD} \Delta v_i = -f_{\rm PD}$ for Δx_i;
    Find $x_{i+1} = x_i + \alpha_i \Delta x_i$ such that
    $M(x_{i+1}, \rho) < M(x_i, \rho)$ and $x_{i+1} > 0$;
    Set $i = i + 1$;
  end do;
  Set $k = k + 1$, $i_k = i$;
end do

The convergence of Algorithm PD follows directly if it can be shown that (7.2) generates a sequence $\{x_i\}$ converging to $x^*(\mu)$.

Given positive constants τ_r and τ_M, define the level set

$$\bar{S} = \{x : \|Ax - b\| \leq \tau_r, \quad M(x, \rho) \leq \tau_M\}.$$

Similar arguments to those proving Lemma 4.1 show that $\bar{S}$ is compact.

It will be shown that the search direction in the x variables is a descent direction for $M(x, \rho)$. This function is not differentiable at any point at which a component of the residual $r = Ax - b$ is zero. However, the function $M(x, \rho)$ is infinitely differentiable along each search direction Δx generated by Algorithm PD, as the following argument shows. At any iterate only nonzero components of r are included in the definition of $M(x, \rho)$. It follows from (4.4) that if a step causes a nonzero

component of r to become zero then every other component of r will become zero simultaneously. Thus, within any given iteration, every nonzero component of r will remain nonzero and constant in sign. When every component of r becomes zero, $B(x,\mu)$ replaces $M(x,\rho)$ as the merit function.

Lemma 7.1. *If $x \in \bar{S}$ then there exists a positive σ_X, independent of x, such that $x \geq \sigma_X e$.*

Proof. Similar to that of Lemma 2.1. ∎

Lemma 7.2. *Given positive constants τ_r, τ_Y, τ_X, τ_Z, σ_X and σ_Z assume that x, y and z satisfy $\|r\| = \|Ax - b\| \leq \tau_r$, $\|y\| < \tau_Y$, $\sigma_X e < x < \tau_X e$ and $\sigma_Z e < z < \tau_Z e$. Then there exist constants ρ, γ ($\gamma > 0$) and β ($\beta \geq 1$) such that*

$$\Delta x^T \nabla M(x,\rho) \leq -\gamma \|N^T g\|^2 - \beta \|r\|_1,$$

where Δx is defined by (7.2).

Proof. Elimination of Δz from (7.2) yields the KKT system

$$\begin{pmatrix} ZX^{-1} & A^T \\ A & \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = - \begin{pmatrix} c - \mu X^{-1} e - A^T y \\ Ax - b \end{pmatrix}. \tag{7.3}$$

It follows from the assumptions that $\|\Delta x\|$ is bounded. Observe that the right-hand side of (7.3) is identical to that of (3.1). It follows from (7.3) that

$$H_N \Delta x_N = -N^T (g + ZX^{-1} A^T \Delta x_A), \tag{7.4}$$

where $g = \nabla B(x,\mu)$, $H_N = N^T Z X^{-1} N$, $\Delta x = N \Delta x_N + A^T \Delta x_A$, and

$$A A^T \Delta x_A = -r. \tag{7.5}$$

From the definition of the merit function $M(x,\rho)$ of Section 3, we have

$$\Delta x^T \nabla M(x,\rho) = \Delta x_N^T N^T (g + \rho A^T \bar{e}) + \Delta x_A^T A (g + \rho A^T \bar{e}),$$

where $\bar{e}$ is a vector with components of magnitude one and $\text{sign}(\bar{e}_i) = \text{sign}(r_i)$. Define $u \equiv (AA^T)^{-1} A (I - X^{-1} Z N H_N^{-1} N^T) g$. Substituting for Δx_N from (7.4) and Δx_A from (7.5) gives

$$\begin{aligned} \Delta x^T \nabla M(x,\rho) &= -g^T N H_N^{-1} N^T g - r^T u - \rho r^T \bar{e}, \\ &\leq -\gamma \|N^T g\|^2 - \beta \|r\|_1, \end{aligned}$$

where $\gamma > 0$ is the reciprocal of the largest eigenvalue of H_N, $\beta \geq 1$, and ρ is chosen such that

$$\rho = \max \left\{ 1 - \frac{r^T u_M}{r^T \bar{e}},\ 0 \right\},$$

with u_M the vector u evaluated at the point $x \in \bar{S}$ for which $r^T u$ has its minimum value. ∎

Lemma 7.3. *Let σ_Z, τ_Y and τ_Z be preassigned positive constants. Consider sequences $\{z_i\}$ and $\{y_i\}$ such that $\sigma_Z e \leq z_i \leq \tau_Z e$ and $\|y_i\| \leq \tau_Y$. Let $\{x_i\}$ denote the sequence $x_0 > 0$ and $x_{i+1} = x_i + \alpha_i \Delta x_i$, where Δx_i is defined by (7.2) and α_i is computed using a Goldstein-Armijo linesearch on $M(x, \rho)$ with the requirement that $x_{i+1} > 0$. If ρ is sufficiently large (but bounded) then $\Delta x_i \to 0$ and $x_i \to x^*(\mu)$.*

Proof. Since $\{x_i\}$ lies in a compact set, it follows that x_i is bounded for all i. Moreover, since x_i lies in $\bar{S}$, there exists a positive σ_X such that $x_i \geq \sigma_X e$ for all i. Every element of the sequence $\{x_i\}$ satisfies the assumptions of Lemma 7.2 and we have

$$\Delta x_i^T \nabla M(x_i, \rho) \leq -\gamma \|N^T g(x_i)\|^2 - \beta \|r(x_i)\|_1,$$

where $\gamma > 0$ and $\beta \geq 1$. It follows from Lemma 7.2 that $\{M(x_i, \rho)\}$ is a monotone decreasing sequence. Since $\{x_i\} \in \bar{S}$, it follows that $\{M(x_i, \rho)\}$ must converge and the sufficient decrease condition from the linesearch gives

$$\lim_{i \to \infty} \alpha_i \Delta x_i^T \nabla M(x_i, \rho) = \lim_{i \to \infty} \alpha_i (\gamma \|N^T g(x_i)\|^2 + \beta \|r(x_i)\|_1) = 0.$$

If β_i is the largest feasible step along Δx, then $M(x + \alpha \Delta x, \rho) \to +\infty$ as $\alpha \to \beta_i$ and it follows that there exists a unique positive α^* such that

$$M(x_i + \alpha^* \Delta x_i, \rho) = M(x_i, \rho) + \mu \alpha^* \Delta x_i^T \nabla M(x_i, \rho).$$

Using similar arguments to those of Theorem 2.1, it follows that the steplengths $\{\alpha_i\}$ are bounded away from zero by a positive constant independent of x_i and

$$\|N^T g(x_i)\| \to 0 \quad \text{and} \quad \|r(x_i)\|_1 \to 0.$$

The required results now follow from the uniqueness of x^*. ∎

Lemma 7.4. *If the assumptions and definitions of Lemma 7.3 hold then*

$$\lim_{i \to \infty} y_i + \Delta y_i = y^*(\mu) \quad \text{and} \quad \lim_{i \to \infty} z_i + \Delta z_i = z^*(\mu).$$

Proof. It follows from (7.3), Lemma 7.3 and the optimality conditions of (2.1) that

$$\lim_{i \to \infty} y_i + \Delta y_i = y^*(\mu).$$

From (7.2) we have $Z_i(x_i + \Delta x_i) = -X_i \Delta z_i + \mu e$. Since $x_i \to x^*(\mu)$ and $\Delta x_i \to 0$, we have

$$\lim_{i \to \infty} z_i + \Delta z_i = \mu X_i^*(\mu)^{-1} e = z^*(\mu),$$

where $X_i^*(\mu) = \text{diag}((x_i^*(\mu))_j)$. ∎

This result shows that even for arbitrary choices of $\{z_i\}$ and $\{y_i\}$, approximations to $y^*(\mu)$ and $z^*(\mu)$ may be obtained. It is not necessary for either $y^*(\mu)$ to lie in S_Y or $z^*(\mu)$ to lie in S_Z; for example, σ_Z and τ_Z may be fixed at one.

Specific choices of z_i and y_i may define more efficient algorithms. The primal algorithm of Section 3.2 that uses the merit function $M(x,\rho)$ may be viewed as being equivalent to Algorithm PD with $z_i = \mu X_i^{-1} e$ and $y_{i+1} = y_i + \alpha_i \Delta y_i$. Since $\|Ax-b\|_1$ is implicitly bounded by the linesearch, Lemma 4.1 implies that x_i is bounded. It follows that each $(z_i)_j$ is bounded away from zero, and $z_i \in S_Z$ for suitably small σ_Z.

Alternatively, y and z may be determined from a linesearch. A steplength θ_i in the z and y space can be taken as an approximate solution of the univariate problem

$$\begin{array}{ll} \underset{\theta}{\text{minimize}} & \|f_{\text{PD}}(z_i + \theta\Delta z_i, x_{i+1}, y_i + \theta\Delta y_i)\| \\ \text{subject to} & z_i + \theta\Delta z_i \geq \eta\mu X_{i+1}^{-1} e, \qquad 0 \leq \theta \leq 1, \end{array}$$

where η is some preassigned constant in $(0,1]$.

7.2. Another Primal-Dual Algorithm

A second primal-dual algorithm can be derived by observing that $v^*(\mu)$ solves the system of equations $f_{\text{PDD}}(z,x,y) = 0$, where

$$f_{\text{PDD}}(z,x,y) \equiv \begin{pmatrix} \bar{f} \\ \widehat{f} \\ r \end{pmatrix} = \begin{pmatrix} \mu X^{-1} Z^{-1} e - e \\ c - A^T y - z \\ Ax - b \end{pmatrix}. \tag{7.6}$$

Newton's method for these equations gives the linear system $J_{\text{PDD}} \Delta v = -f_{\text{PDD}}$, where

$$J_{\text{PDD}} = \begin{pmatrix} -\mu Z^{-2} X^{-1} & -\mu X^{-2} Z^{-1} & 0 \\ -I & 0 & A^T \\ 0 & A & 0 \end{pmatrix}.$$

Unlike the primal-dual method of Section 7.1 there are no problems with computing a suitable steplength that maintains x and z positive. Although the Jacobian J_{PDD} appears to be more complicated than J_{PD}, the system for Δx and Δy may be written as

$$\begin{pmatrix} ZX^{-1} & A^T \\ A & \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = - \begin{pmatrix} c - 2z + \dfrac{1}{\mu} Z^2 x - A^T y \\ Ax - b \end{pmatrix}, \tag{7.7}$$

which has a KKT matrix identical to that of the primal-dual method of Section 7.1. Since the direction for Algorithm PD can be computed with little additional effort, another strategy would be to use a linear combination of the two directions. The precise linear combination could be adjusted dynamically and need be specified only after both directions are known.

The right-hand side of (7.7) is identical to that of the KKT system for the dual algorithm. Hence, this algorithm is related to the dual barrier algorithm in the same way that the primal-dual algorithm of Section 7.1 is related to the primal. For example, a merit function based on the dual barrier function and dual constraint violations would allow the calculation of different steps in the primal and dual variables, and it is the step in the primal variables that could be chosen arbitrarily.

Note that any linear combination of the systems (4.1), (6.3), (7.1) and (7.6) can be used to define directions for x, z and y. In particular, any linear combination that includes the primal-dual equations (7.6) (no matter how tiny a proportion) has the property that a suitable steplength exists for which x and z are positive.

8. Alternative Assumptions

The compactness of the feasible region required by assumption (ii) is used to establish that the (generally infeasible) iterates in the x space lie in a compact set. In turn, this result is used to show that the y and z iterates also lie in a compact set. (This result is not obvious since the primal-dual *feasible* space is generally not compact.) Compactness is crucial for the proof that the various KKT systems are bounded and have bounded condition numbers. In this section we weaken the assumptions for Algorithm PFP and Algorithm PD. More specifically the assumption of compactness of the feasible region is replaced by the assumption that the set of solutions $\mathcal{X}^*$ is nonempty and bounded. First we show that, under this weaker assumption and the existence of a strictly feasible point, the strictly feasible iterates of Algorithm PFP lie in a compact region. Existence of a nonempty interior of the feasible region will not be required for the convergence of the primal-dual method.

The next lemma states that if the set of solutions to the LP (1.1) is nonempty and bounded, then the iterates generated by a descent algorithm applied to the (2.1) lie in a compact set.

Lemma 8.1. *Let S_0 denote the set of feasible points for problem (1.1) and assume that $\mathrm{int}(S_0)$ is nonempty. Let $\bar{x}$ be any point in $\mathrm{int}(S_0)$ and let $\Omega(\bar{x},\mu)$ denote the level set*

$$\Omega(\bar{x},\mu) = \{x : B(x,\mu) \le B(\bar{x},\mu)\}.$$

If $\mathcal{X}^*$ the set of solutions of (1.1) is nonempty and bounded, then $S_0 \cap \Omega(\bar{x},\mu)$ is compact.

Proof. If $\bar{x}$ is a minimizer of (2.1) then it is unique and the level set is trivially compact. If $\bar{x}$ is not a minimizer, assume that $S_0 \cap \Omega(\bar{x},\mu)$ is not compact. Consider the set $\mathcal{U}$ of all unit feasible directions of non-ascent for the barrier function $B(\bar{x},\mu)$:

$$\mathcal{U} = \{u : Au = 0, \quad u^T g(\bar{x}) \le 0 \quad \text{and} \quad \|u\| = 1\},$$

where $g(\bar{x}) \equiv g(\bar{x},\mu)$. Since $S_0 \cap \Omega(\bar{x},\mu)$ is not compact by assumption, there must exist $u \in \mathcal{U}$ such that

$$\bar{x} + \gamma u \in S_0, \quad u^T g(\bar{x} + \gamma u) \le 0 \quad \text{for all } \gamma > 0.$$

Any such u must be nonnegative since otherwise there must exist a finite $\bar{\gamma}$ such that $u^T g(\bar{x} + \gamma u) \to +\infty$ as $\gamma \to \bar{\gamma}$. Moreover, since

$$u^T g(\bar{x} + \gamma u) = c^T u - \mu \sum_{j=1}^{n} \frac{u_j}{\bar{x}_j + \gamma u_j}$$

and the barrier term of this expression tends to zero as $\gamma \to +\infty$, we must have $c^T u \leq 0$. The existence of a vector such that $Au = 0$, $u \geq 0$ and $c^T u \leq 0$ contradicts the assumption that $\mathcal{X}^*$ is bounded. ∎

The strict convexity of $B(x, \mu)$ on the compact set $S_0 \cap \Omega(\bar{x}, \mu)$ implies the following corollary.

Corollary 8.1. *The minimizer $x^*(\mu)$ of (2.1) is unique.*

Lemma 8.2. *Let S_0 denote the set of feasible points for problem (1.1) and assume that $\text{int}(S_0)$ is nonempty. Let $\bar{x}$ be any point such that $\bar{x} > 0$, and let $\Omega(\bar{x}, \mu)$ denote its associated level set*

$$\Omega(\bar{x}, \mu) = \{x : M(x, \rho) \leq M(\bar{x}, \rho)\}.$$

If $\mathcal{X}^*$ is nonempty and bounded, and ρ is sufficiently large, then $\Omega(\bar{x}, \mu)$ is compact.

Proof. If ρ is sufficiently large, $x^*(\mu)$ is a minimizer of $M(x, \rho)$. The uniqueness of $x^*(\mu)$ stated by Corollary 8.1 implies that $\Omega(x^*(\mu), \mu)$ is nonempty and bounded. Since $M(x, \rho)$ is convex, it follows from standard convex analysis that every level set of $M(x, \rho)$ is bounded. ∎

This result proves the minor iterates of Algorithm PD lie in a compact set. The assumption that $\text{int}(S_0)$ is nonempty is not required for the following lemma.

Lemma 8.3. *Let $\bar{x}$ be any point such that $\bar{x} > 0$ and let $\Omega(\bar{x}, \mu)$ denote the level set*

$$\Omega(\bar{x}, \mu, \rho) = \{x : M(x, \rho) \leq M(\bar{x}, \rho)\},$$

where

$$M(x, \rho) = B(x, \mu) + \rho \sum_{i=1}^{m} |a_i^T x - b_i|.$$

If $\mathcal{X}^*$ is nonempty and bounded, and the parameter ρ is sufficiently large, then $\Omega(\bar{x}, \mu, \rho)$ is compact.

Proof. Let u denote a vector such that $\|u\| = 1$. There must exist a scalar $\bar{\gamma} \geq 0$ such that the sign of $a_i^T(\bar{x} + \gamma u) - b_i$ is fixed for all $\gamma \geq \bar{\gamma}$. It follows that the directional derivative of $M(\bar{x} + \gamma u)$ exists for all $\gamma \geq \bar{\gamma}$ and $x + \gamma u \in \Omega(\bar{x}, \mu, \rho)$. Also note that if $\mathcal{X}^*$ is bounded it implies there exists positive constants τ, τ_1 and τ_2 such that $c^T u > \tau$, if $\|Au\|_1 \leq \tau_1$ and $u \geq -\tau_2 e$.

If $\Omega(\bar{x}, \mu.\rho)$ is not compact, there must exist a vector u and $\hat{\gamma}$ such that

$$\theta(\gamma) \equiv u^T \nabla M(\bar{x} + \gamma u) \leq 0 \quad \text{for all } \gamma \geq \hat{\gamma} \geq \bar{\gamma}$$

and $x + \gamma u \in \Omega(\bar{x}, \mu, \rho)$. We must have $u \geq -\tau_2 e$, otherwise there exist a finite β such that $M(\bar{x} + \gamma u) \to +\infty$ as $\gamma \to \beta$. Define $\mathcal{U} \equiv \{u : u \geq -\tau_2 e,\ \|u\| = 1\}$. There must exist $u \in \mathcal{U}$ such that

$$\theta(\gamma) = c^T u - \mu \sum_{j=1}^{n} \frac{u_j}{\bar{x}_j + \gamma u_j} + \rho \sum_{i=1}^{m} |a_i^T u| \leq 0 \quad \text{for all } \gamma \geq \hat{\gamma} \geq \bar{\gamma}. \tag{8.1}$$

If $\|Au\|_1 \geq \tau_1$ then since $c^T u$ is bounded below for $u \in \mathcal{U}$ there exists ρ such that $\theta(\gamma) > 0$ for all $\gamma \geq \bar{\gamma}$. If $\|Au\|_1 \leq \tau_1$ then $c^T u > \tau$ and since for γ sufficiently large $-\mu \sum u_j/(\bar{x}_j + \gamma u_j) \geq -\tau/2$ it follows from (8.1) that there exists $\tilde{\gamma}$ such that $\theta(\gamma) > 0$ for all $\gamma > \tilde{\gamma}$. The required result follows. ∎

Let S denote the set $\{x : x \geq 0,\ \|Ax - b\| \leq \|A\bar{x} - b\|\}$. The points generated by algorithm PD for a fixed value of μ lie in $S \cap \Omega(x_0, \mu, \rho)$. Moreover, $\|Ax_k - b\|$ is strictly montonically decreasing and if ever a unit step is taken then $\|Ax_k - b\| = 0$ for all subsequent k, which makes ρ redundant.

9. The Treatment of Free Variables

If a variable has no upper or lower bounds, the leading diagonal block of the KKT matrix is singular. Since the full KKT matrix remains nonsingular, this has no effect on methods that solve KKT system directly. However, difficulties arise with a Schur complement approach in which the KKT system is solved by forming two smaller systems that involve the inverse of the leading diagonal block.

Here a new method for treating free variables is proposed that can be used in conjunction with the Schur complement approach. To illustrate, consider the calculation of Δv in the primal-dual algorithm of Section 7. For simplicity, assume that x_r is the only free variable. In place of $(\bar{f}_{\text{PD}})_r = z_r x_r - \mu$ we have $(\bar{f}_{\text{PD}})_r = z_r$, with $z_r^*(\mu) = 0$. If the KKT system (7.2) is solved directly, this equation has no effect, but in the partially eliminated system (7.3) ZX^{-1} must be replaced by H, where $H = \text{diag}(h_j)$ with $h_r = 0$ and $h_j = z_j x_j^{-1}$ for $j \neq r$. Clearly, H is singular and the Schur complement $AH^{-1}A^T$ does not exist.

The singularity of H may be circumvented by replacing the equation $z_r = 0$ by another that ensures $z_r \to 0$ as $\mu \to 0$. Consider the two alternatives

$$e^{x_r} z_r = \mu \qquad \text{or} \qquad z_r + \mu x_r = 0,$$

which give $h_r = z_r$ and $h_r = \mu$ respectively. In the first case, $z_r^*(\mu) = \mu e^{-x_r^*(\mu)}$ and we may keep $z_r \geq \sigma_Z > 0$. It follows that H is nonsingular and $AH^{-1}A^T$ exists. In the second case, z_r does not appear in the Jacobian and neither its sign nor its magnitude are important. Similarly, the nonsingularity of J_P is no longer dependent on x_r and x_r can be excluded from the calculation of the step to the boundary.

10. Further Comments

A common practice in interior-point implementations is to define the steplength as some fixed percentage of the maximum feasible step. By contrast, all the algorithms described in this paper require some form of linesearch for the steplength. In practice this requirement has little effect upon the computation time, given the work needed to compute the search direction. Moreover, if $\eta \approx 0$, almost any step is likely to satisfy the Goldstein-Armijo backtracking condition because the linesearch functions are convex and increase rapidly near the boundary of the feasible region. In practice

we have observed that the need to perform a linesearch arises only when there is significant numerical error in the search direction.

Currently the most efficient implementations use a predictor-corrector method to define the search direction (see, e.g., [9, 13]). Such a strategy may be incorporated in the algorithms discussed here. The important point is to be able to fall back on a guaranteed method should the predictor-corrector direction fail to be a suitable descent direction for the merit function. A similar view was adopted by Mehrotra [13].

It has not been our intent to compare the performance of the algorithms considered. All the primal-dual algorithms have very similar theoretical properties, but only the primal-dual algorithm of Section 7.1 has been used in the principal known implementations [9, 10, 13, 3]. The key system of equations appears to be "less nonlinear" than the other three variations. Even so, in the neighborhood of the solution, the Jacobian behaves almost identically to the Jacobians of the other systems (as does the KKT matrix). It is not immediately apparent that this method is inherently superior to the others.

It is likely that the best method will be determined by the method used to solve the linear systems. For example, all the methods may be implemented by solving systems of the form $ADA^T\Delta y = u$, where D is either X^2, Z^{-2} or XZ^{-1}. Suppose that these systems are solved using a conjugate-gradient method in which a preconditioner is based on periodically forming the Cholesky factors of ADA^T. As the iterates converge, the systems using $D = X^2$ should yield better preconditioners because the ratio of consecutive values of any significant diagonal element of D converges to one. When $D = XZ^{-1}$ or $D = Z^{-2}$, the significant diagonal elements correspond to the small components of z. It is not obvious that the ratio of consecutive values of any such diagonal will behave as smoothly. El-Bakery et al. [1] report that when $D = Z^{-2}$ in the predictor-corrector method the elements of D do not behave smoothly.

Our analysis is directed at the linear programming problem. However, the extension of the results to convex programs is relatively straightforward. The more challenging problem is to extend the results to nonconvex problems.

Acknowledgement

The authors are grateful to Joe Shinnerl for his careful reading of the manuscript and to the referees for a number of helpful suggestions.

References

[1] A. S. El-Bakry, R. A. Tapia, and Y. Zhang. A study of indicators for identifying zero variables in interior-point methods. Technical Report TR-91-15, Dept of Mathematical Sciences, Rice University, Houston, TX, 1991.

[2] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* John Wiley and Sons, Inc., New York-London-Sydney, 1968.

[3] P. E. Gill, W. Murray, D. B. Ponceleón, and M. A. Saunders. Solving reduced KKT systems in barrier methods for linear and quadratic programming. Report SOL 91-7, Department of Operations Research, Stanford University, Stanford, CA, 1991.

[4] P. E. Gill, W. Murray, M. A. Saunders, J. Tomlin, and M. H. Wright. On projected Newton methods for linear programming and an equivalence to Karmarkar's projective method. *Math. Program.*, 36:183–209, 1986.

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

[6] M. Kojima, N. Megiddo, and S. Mizuno. Theoretical convergence of large-step primal-dual interior-point algorithms for linear programming. Report RJ 7872, IBM T. J. Watson Research Center, San Jose, CA, 1990.

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

[8] I. J. Lustig. Feasibility issues in an interior-point method for linear programming. Report SOR-88-9, Department of Civil Engineering and Operations Research, Princeton University, 1988.

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

[10] I. J. Lustig, R. E. Marsten, and D. F. Shanno. On implementing Mehrotra's predictor-corrector interior point method for linear programming. *SIAM J. Optim.*, 2:435–449, 1992.

[11] K. A. McShane, C. L. Monma, and D. F. Shanno. An implementation of a primal-dual interior point method for linear programming. *ORSA Journal on Computing*, 1:70–83, 1989.

[12] N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pages 131–158. Springer Verlag, NY, 1989.

[13] S. Mehrotra. On the implementation of a primal-dual method. *SIAM J. Optim.*, 2:575–601, 1992.

[14] R. D. C. Monteiro and I. Adler. Interior path-following primal-dual algorithms. Part II: Convex quadratic programming. *Math. Program.*, 44:43–66, 1989.

[15] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables.* Academic Press, New York, 1970.

[16] D. B. Ponceleón. *Barrier methods for large-scale quadratic programming.* PhD thesis, Department of Computer Science, Stanford University, Stanford, CA, 1990.

[17] M. H. Wright. Interior methods for constrained optimization. In *Acta Numerica, 1992*, pages 341–407. Cambridge University Press, New York, USA, 1992.

[18] Y. Zhang. On the convergence of an infeasible interior-point method for linear programming and other problems. Report 92-07, Dept of Mathematics and Statistics, University of Maryland, Baltimore County, MD, 1992.