A PRIMAL-DUAL VARIANT OF THE IRI-IMAI ALGORITHM FOR LINEAR PROGRAMMING

by

Reha H. Tütüncü

Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.

Research Report No. 98-217
May, 1998

A Primal-Dual Variant of the Iri-Imai Algorithm for Linear Programming

Reha H. Tütüncü *

May 22, 1998

Abstract

A local acceleration method for primal-dual potential-reduction algorithms is introduced. The method developed here uses modified Newton search directions to minimize the Tanabe-Todd-Ye (TTY) potential function, and can be regarded as a primal-dual variant of the Iri-Imai algorithm based on the multiplicative analogue of Karmarkar's potential function. When iterates are close to an optimal solution, the TTY potential function has negative curvature along the generated search directions. Therefore, large reductions in the potential function can be obtained, guaranteeing polynomial and quadratic convergence to nondegenerate solutions.

*Research supported in part by NSF through grant DMS-9706950.

1 Introduction

This paper considers interior-point methods for the solution of linear programming problems given in the following standard form:

$$(LP) \qquad \begin{array}{rcl} \min_x & c^T x & \\ & Ax & = & b \\ & x & \geq & 0, \end{array} \qquad (1)$$

where $A \in I\!R^{m\times n}$, $b \in I\!R^m$, $c \in I\!R^n$ are given, and $x \in I\!R^n$. The matrix A is assumed to have full row rank without loss of generality.

We measure the quality of feasible solutions to the LP above using *potential functions* and focus on potential-reduction methods to solve this problem. Ever since Karmarkar (1984) introduced such a function to prove the polynomial convergence of his algorithm, potential functions have been frequently used as both algorithmic and analytical tools in the study of interior-point methods. Two recent surveys on potential-reduction algorithms by Anstreicher (1996) and Todd (1997a) document these developments.

Currently, one of the most intensely studied aspects of interior-point methods is the development of algorithms with fast local convergence in addition to global polynomial convergence. Most of the developments in this area have focused on path-following methods. These methods generate iterates that are restricted to stay in some neighborhood of the so-called central path to reach the optimal solution. Proofs of superlinear convergence results for path-following algorithms invariably and explicitly use these neighborhood restrictions imposed on the iterates. While such restrictions are not required to establish the polynomial convergence of potential-reduction methods, superlinearly convergent potential-reduction methods *without* neighborhood restrictions are rare. Our focus is on the development of such methods.

The most relevant work on potential-reduction methods for our purposes is the method developed by Iri and Imai. Iri and Imai (1986) develop an interior-point algorithm for linear programming by applying the Newton's method to the multiplicative analogue of the potential function introduced by Karmarkar (1984). They demonstrate that when the potential function parameter is large enough the multiplicative potential function is strictly convex, and therefore, Newton directions are descent directions. Iri and Imai also establish the quadratic convergence of their method for nondegenarete linear programming problems when an exact line search is employed along the Newton directions. In subsequent work, among others, Iri (1993) has established the polynomial global convergence of the Iri-Imai method, and Tsuchiya (1995) showed that the quadratic convergence property holds even for degenerate problems, provided that one uses exact line searches to minimize the potential function along the generated directions. See also Sturm and Zhang (1996). Monteiro and Wright (1995) describe a polynomially and superlinearly convergent affine-scaling method that happens to be consistent with a potential-reduction method, but their method uses neighborhood restrictions.

This paper considers a potential-reduction approach using the Tanabe-Todd-Ye (TTY) primal-dual potential function. The motivation for this work is to improve the local convergence rate of potential-reduction methods using this potential function. We achieve this goal using Newton search directions for the multiplicative analogue of the TTY potential function. Therefore, our approach can be seen as a primal-dual variant of the Iri-Imai method, as indicated in the title of this study.

We interpret our search directions as *modified* Newton directions for the *original* TTY potential function. An asymptotic analysis of these directions reveal that they become directions of negative curvature near a solution, along which one can take large steps. Furthermore, one can obtain quadratic convergence for nondegenerate problems *without* resorting to exact line searches. This technique can be embedded in any polynomially convergent potential-reduction algorithm. Therefore, we are not concerned with the global convergence of an algorithm using solely the directions generated by our technique. Rather, we regard the method developed here as a local acceleration technique for primal-dual potential-reduction algorithms. Specifically, we show that our method can be used in conjunction with the algorithm of Kojima, Mizuno, and Yoshise (1991), retaining the polynomial convergence property of their algorithm and improving its local convergence rate from linear to quadratic.

The paper is organized as follows: Following this introduction, in Section 2, we review some of the primal and primal-dual potential functions for linear programming. We establish the convexity of the multiplicative primal-dual potential function with a large enough parameter and derive our modified Newton search directions in Section 3. Section 4 is devoted to the analysis of the asymptotic behavior of our search directions. After proving our polynomial and quadratic convergence results on nondegenerate problems in Section 5, we conclude in Section 6.

The notation used here is mostly standard. $\|\cdot\|$ (without a subscript) denotes the 2-norm of real vectors; we will use a subscript with the 1-norm. We use subscripts to denote components of vectors or matrices, and superscripts to denote iteration indices. We also use standard order notation: For sequences $\{x^k\}$ and $\{y^k\}$ of real numbers with $y^k > 0$, $x^k = \mathcal{O}(y^k)$ means that the sequence $\{\frac{x^k}{y^k}\}$ is bounded above by a number independent of k. If $x^k > 0$ also, then $x^k = \Theta(y^k)$ means that $x^k = \mathcal{O}(y^k)$ and $y^k = \mathcal{O}(x^k)$. If $\{x^k\}$ is a sequence of matrices, $x^k = \mathcal{O}(y^k)$ and $x^k = \Theta(y^k)$ mean that $x^k_{ij} = \mathcal{O}(y^k)$ and $x^k_{ij} = \Theta(y^k)$, respectively, for each i and j. We will denote primal-only and primal-dual potential functions using lower-case and upper-case letters, respectively. For an n-dimensional vector x, the corresponding capital letter X denotes the $n \times n$ diagonal matrix with $X_{ii} \equiv x_i$.

2 Potential Functions for Linear Programming

In his seminal work, Karmarkar (1984) introduced a potential function to measure the quality of different feasible points. Under some mild assumptions, this function tends to $-\infty$ only along the sequences that approach an optimal solution to the linear programming problem. Therefore, a search for an optimal solution can be done by minimizing this function. The potential function Karmarkar introduced is given below:

$$\phi_\rho(x) := \rho \ln(c^T x) - \sum_{i=1}^{n} \ln x_i, \tag{2}$$

Karmarkar assumes that the problem (1) is given in a particular form, and the optimal objective value is *a priori* known to be equal to zero.

Iri and Imai (1986) consider the multiplicative analogue of the function ϕ:

$$f_\rho(x) := \frac{(c^T x)^\rho}{\prod_{i=1}^n x_i}. \tag{3}$$

They show that $f_\rho(x)$ is strictly convex in the relative interior of the primal feasible region, if the parameter $\rho \geq n+1$. Then, they apply Newton's method for minimization to this strictly convex function. As mentioned in the introduction, Iri and Imai also establish the quadratic convergence of their method for nondegenarete linear programming problems if an exact line search is employed along the Newton directions.

The most succesful interior-point methods are the primal-dual methods. In addition to solving (1), the *primal* problem, these methods simultaneously solve a related *dual* problem. The dual of the linear programming problem given in (1) is:

$$\begin{array}{lllllll} (LD) & \max_{y,s} & b^T y & & & & \\ & & A^T y & + & s & = & c \\ & & & & s & \geq & 0, \end{array} \tag{4}$$

where $y \in \mathbb{R}^m$ and $s \in \mathbb{R}^n$.

Let G^T be a null-space basis matrix for A, that is, G is an $(n-m) \times n$ matrix with rank $n-m$ and it satisfies $AG^T = 0$, $GA^T = 0$. Using this matrix and a vector $d \in \mathbb{R}^n$ satisfying $Ad = b$, one can rewrite (2) in a form that is identical to (1):

$$\begin{array}{lllll} (LD') & \min_s & d^T s & & \\ & & Gs & = & Gc \\ & & s & \geq & 0. \end{array} \tag{5}$$

This form of the dual was discussed, among others, by Gonzaga (1992).

$\mathcal{F}$ and $\mathcal{F}^0$ denote the primal-dual feasible region and its relative interior:

$$\begin{aligned} \mathcal{F} &:= \{(x,s) : Ax = b, Gs = Gc, (x,s) \geq 0\} \\ \mathcal{F}^0 &:= \{(x,s) : Ax = b, Gs = Gc, (x,s) > 0\} \end{aligned}$$

We will assume that $\mathcal{F}^0$ is non-empty and a point $(x^0, s^0) \in \mathcal{F}^0$ is available. This assumption is not restrictive; any LP can be embedded in an artificial problem with a known point in the relative interior of its feasible region; see, e.g., Ye, Todd, and Mizuno (1994). Furthermore, certain solutions to this artificial problem will either give the optimal solution to the original LP, or reveal that it is either infeasible or unbounded.

The main algorithmic tool in this paper is a primal-dual variant of Karmarkar's potential function ϕ_ρ. This function was introduced by Tanabe (1988), and Todd and Ye (1990) independently and has been a very useful tool in construction and analysis of efficient interior-point algorithms for linear programming and linear complementarity problems. For the primal-dual pair of problems (1) and (3), this function is defined as

$$\Phi_\rho(x,s) \;:=\; \rho \ln(x^T s) - \sum_{i=1}^{n} \ln(x_i s_i), \text{ for every } (x,s) > 0. \tag{6}$$

Using a primal-dual update, Kojima, Mizuno, and Yoshise (1991) showed that when $\rho \geq n + \sqrt{n}$, Tanabe-Todd-Ye (TTY) potential function can be reduced by at least 0.2 from any feasible point (x, y, s) with $(x, s) > 0$. This guaranteed constant reduction in the potential function leads to an algorithm with $\mathcal{O}((\rho - n)\ln\frac{1}{\epsilon})$ complexity. They use the direction that solves the following system to always achieve such a reduction:

$$\begin{array}{rcrcrcl} & & A\Delta x & & & = & 0 \\ A^T\Delta y & & & + & \Delta s & = & 0 \\ & & S\Delta x & + & X\Delta s & = & \dfrac{x^T s}{\rho'} e - XSe, \end{array} \tag{7}$$

where $X = \text{diag}(x)$, $S = \text{diag}(s)$, e is a vector of ones of appropriate dimension, and ρ' is a parameter related to ρ. Kojima, Mizuno, and Yoshise choose $\rho' = \rho$. Ye et al. (1993) generalized their result to the case where ρ' is not restricted to equal ρ and can vary from one iteration to the other. If ρ' is chosen carefully, it is possible to maintain the guarantee of constant reduction in every iteration and to improve the practical performance. Kojima, Mizuno, and Yoshise also provide an interpretation of the direction (7): When $\rho' = \rho$, the solution to the system (7) can be regarded as a scaled and projected steepest descent direction for the TTY potential function Φ_ρ.

The steepest descent interpretation of the direction (7) also indicates that an algorithm using this direction is not likely to have better than linear convergence, since it only uses first order information on the function Φ_ρ. In the vicinity of an optimal solution to the linear programming problem under consideration, the potential function Φ_ρ tends to $-\infty$, and for faster convergence one needs to proceed along directions of negative curvature. In the next section, we will define a modified Newton direction for minimizing Φ_ρ that we will later show to be a direction of negative curvature around an optimal solution. We end this section by evaluating the gradient and the Hessian of Φ_ρ:

$$\nabla\Phi_\rho = \begin{bmatrix} \dfrac{\rho}{x^T s} s - X^{-1} e \\ \dfrac{\rho}{x^T s} x - S^{-1} e \end{bmatrix}, \tag{8}$$

and

$$\nabla^2\Phi_\rho = \begin{bmatrix} X^{-2} & \\ & S^{-2} \end{bmatrix} + \frac{\rho}{x^T s}\begin{bmatrix} & I \\ I & \end{bmatrix} - \frac{\rho}{(x^T s)^2}\begin{bmatrix} s \\ x \end{bmatrix}\begin{bmatrix} s \\ x \end{bmatrix}^T. \tag{9}$$

3 Modified Newton Directions for Potential-Reduction Methods

A potential-reduction method to solve (1) and (3) using the TTY function is a method to solve the following (nonconvex) nonlinear programming problem:

$$\begin{array}{lrcl} \min_{x,s} & \Phi_\rho(x,s) & & \\ & Ax & = & b \\ & Gs & = & Gc \\ & x, \quad s & \geq & 0. \end{array} \tag{10}$$

In a sequential quadratic programming framework to solve this problem, one has a quadratic model for the function Φ_ρ around the feasible point (x, s) such as

$$M(\Phi_\rho(x+\Delta x, s+\Delta s)) := \Phi_\rho(x,s) + [\nabla\Phi_\rho(x,s)]^T \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix}^T Q(x,s) \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix},$$

where $Q(x,s)$ is a matrix related to the Hessian $\nabla^2\Phi_\rho$ and is often chosen to be positive definite. Then a search direction is found by solving the following local problem:

$$\begin{array}{lrcl} \min_{x,s} & M(\Phi_\rho(x+\Delta x, s+\Delta s)) & & \\ & A\Delta x & = & 0 \\ & G\Delta s & = & 0 \end{array} \tag{11}$$

Let Z be a null space basis for the constraint matrix of (11). In the remainder of this paper, we will use the following convenient form for Z:

$$Z = \begin{bmatrix} G^T & \\ & A^T \end{bmatrix} \tag{12}$$

If Q is chosen so that Z^TQZ is positive definite, the solution to (11) is found by solving the following linear system:

$$(Z^TQ(x,s)Z) \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} = -Z^T\nabla\Phi_\rho(x,s). \tag{13}$$

The Hessian matrix in (9) is generally indefinite. So, we are not even guaranteed to have a descent direction if we use the (unmodified) reduced Newton direction (i.e., if we choose $Q(x,s) = \nabla^2\Phi_\rho(x,s)$). Instead, we will consider a Q matrix that is a rank-1 update of the Hessian $\nabla^2\Phi_\rho(x,s)$:

$$Q(x,s) = \nabla^2\Phi_\rho(x,s) + \nabla\Phi_\rho(x,s)\nabla\Phi_\rho^T(x,s). \tag{14}$$

Below, we show that this rank-1 update of the Hessian will guarantee that $Z^TQ(x,s)Z$ is positive definite if $\rho \geq 2n+1$. This modification to the Hessian is somewhat unusual; a multiple of the identity is preferred in most circumstances. We now give a justification of this choice:

Consider the multiplicative analogue of the Tanabe-Todd-Ye primal-dual potential function:

$$F_\rho(x,s) \quad := \quad \frac{(x^Ts)^\rho}{\prod_{i=1}^n x_i s_i} = \exp\{\Phi_\rho(x,s)\}, \text{ for every } (x,s) > 0. \tag{15}$$

Since $F_\rho(x,s)$ is a monotone transformation of $\Phi_\rho(x,s)$, minimizing one of these functions is equivalent to minimizing the other. Along a sequence of points (x^k, s^k) where $\Phi_\rho(x^k, s^k)$ tends to $-\infty$, $F_\rho(x^k, s^k)$ tends to zero. Note that,

$$\nabla F_\rho(x,s) \;=\; F_\rho(x,s)\nabla\Phi_\rho(x,s), \tag{16}$$
$$\nabla^2 F_\rho(x,s) \;=\; F_\rho(x,s)(\nabla^2\Phi_\rho(x,s) + \nabla\Phi_\rho(x,s)[\nabla\Phi_\rho(x,s)]^T). \tag{17}$$

Our choice for the Q matrix is precisely the second term on the right-hand-side of (17). Our motivation for considering $F_\rho(x,s)$ is the fact that this function is strictly convex on the relative interior of the feasible region for large enough ρ:

Theorem 3.1 *The function $F_\rho(x,s)$ is strictly convex on $\mathcal{F}^0$ if $\rho \geq 2n+1$.*

Proof:
To show the strict convexity of F_ρ on $\mathcal{F}^0$, all one needs to do is to show that when $\rho \geq 2n+1$ the reduced Hessian matrix $Z^T\nabla^2 F_\rho(x,s)Z$ is positive definite, where Z is as defined in (12). Below, x^{-1} and s^{-1} denote $X^{-1}e$ and $S^{-1}e$, respectively. Using (8), (9), (16) and (17),

$$\begin{aligned}
&\frac{Z^T\nabla^2 F_\rho(x,s)Z}{F_\rho(x,s)} \\
&= \begin{bmatrix} GX^{-2}G^T & \\ & AS^{-2}A^T \end{bmatrix} + \frac{\rho}{x^Ts}\begin{bmatrix} & GA^T \\ AG^T & \end{bmatrix} \\
&\quad -\frac{\rho}{(x^Ts)^2}\begin{bmatrix} Gs \\ Ax \end{bmatrix}\begin{bmatrix} Gs \\ Ax \end{bmatrix}^T + \begin{bmatrix} \frac{\rho}{x^Ts}Gs - Gx^{-1} \\ \frac{\rho}{x^Ts}Ax - As^{-1} \end{bmatrix}\begin{bmatrix} \frac{\rho}{x^Ts}Gs - Gx^{-1} \\ \frac{\rho}{x^Ts}Gx - Gs^{-1} \end{bmatrix}^T \\
&= \begin{bmatrix} GX^{-2}G^T & \\ & AS^{-2}A^T \end{bmatrix} + \frac{\rho(\rho-1)}{(x^Ts)^2}\begin{bmatrix} Gs \\ Ax \end{bmatrix}\begin{bmatrix} Gs \\ Ax \end{bmatrix}^T \\
&\quad -\frac{\rho}{x^Ts}\left(\begin{bmatrix} Gs \\ Ax \end{bmatrix}\begin{bmatrix} Gx^{-1} \\ As^{-1} \end{bmatrix}^T + \begin{bmatrix} Gx^{-1} \\ As^{-1} \end{bmatrix}\begin{bmatrix} Gs \\ Ax \end{bmatrix}^T\right) + \begin{bmatrix} Gx^{-1} \\ As^{-1} \end{bmatrix}\begin{bmatrix} Gx^{-1} \\ As^{-1} \end{bmatrix}^T \\
&= H_1 + H_2,
\end{aligned}$$

where
$$H_1 = \begin{bmatrix} GX^{-1} & \\ & AS^{-1} \end{bmatrix}(I_{2n} - \frac{1}{\rho-1}e_{2n}e_{2n}^T)\begin{bmatrix} X^{-1}G^T & \\ & S^{-1}A^T \end{bmatrix}, \tag{18}$$

and

$$H_2 = \eta\eta^T \text{ where } \eta = \begin{bmatrix} \frac{\sqrt{\rho(\rho-1)}}{x^T s} Gs - \frac{\rho}{\sqrt{\rho(\rho-1)}} GX^{-1}e \\ \frac{\sqrt{\rho(\rho-1)}}{x^T s} Ax - \frac{\rho}{\sqrt{\rho(\rho-1)}} AS^{-1}e \end{bmatrix}. \tag{19}$$

If $\rho > 2n+1$ ($\rho = 2n+1$), H_1 is positive definite (positive semidefinite), since the only nonzero eigenvalue of $e_{2n}e_{2n}^T$ is $2n$. Positive semidefiniteness of H_2 is evident. Consequently, if $\rho > 2n+1$, the matrix $\frac{Z^T\nabla^2 F_\rho(x,s)Z}{F_\rho(x,s)} = H_1 + H_2$ is positive definite.

Next, we analyze the case when $\rho = 2n+1$. The arguments above indicate that H_1, H_2, and therefore, $H_1 + H_2$, are all positive semidefinite even when $\rho = 2n+1$. Assume that a vector $\begin{bmatrix}\alpha^T & \beta^T\end{bmatrix}^T$ with $\alpha \in \Re^{n-m}$ and $\beta \in \Re^m$ satisfies

$$\begin{bmatrix}\alpha^T & \beta^T\end{bmatrix} (H_1 + H_2) \begin{bmatrix}\alpha \\ \beta\end{bmatrix} = 0.$$

From the positive semidefiniteness of H_1 and H_2, it follows that

$$\begin{bmatrix}\alpha^T & \beta^T\end{bmatrix} H_i \begin{bmatrix}\alpha \\ \beta\end{bmatrix} = 0, \; i = 1,2.$$

Let $u = X^{-1}G^T\alpha$ and $v = S^{-1}A^T\beta$. Then, the equations above can be written as:

$$\|(u,v)\|^2 = \frac{1}{2n}(\sum_{j=1}^n u_j + \sum_{j=1}^n v_j)^2$$

and

$$\sum_{j=1}^n (u_j + v_j)x_j s_j = \frac{\sum_{j=1}^n x_j s_j}{2n} \sum_{j=1}^n (u_j + v_j).$$

The first of the above equations and the Cauchy-Schwartz inequality imply that all u_j's and v_j's are equal, say, to μ. Then, the second of the above equations indicates that this holds only if $\mu = 0$, which holds if and only if the vector $\begin{bmatrix}\alpha^T & \beta^T\end{bmatrix}^T$ is a zero vector since A and G have full row ranks. Therefore, $H_1 + H_2$ is positive definite even when $\rho = 2n+1$.

To summarize, we have shown that the matrix $\frac{Z^T\nabla^2 F_\rho(x,s)Z}{F_\rho(x,s)}$ is positive definite for all $\rho \geq 2n+1$. Now the strict convexity of F_ρ on $\mathcal{F}^0$ for $\rho \geq 2n+1$ follows immediately, since F_ρ is positive on its domain. □

As mentioned above, Iri and Imai (1986) show that the multiplicative analogue of the primal-only potential function is convex if the potential function parameter $\rho \geq n+1$. This raises a question about the tightness of the condition $\rho \geq 2n+1$ in Theorem 3.1. It turns out that this bound on ρ is tight in the sense that one can find problem instances for which the function F_ρ ceases to be convex if $\rho < 2n+1$. Below, we provide a class of LP instances where F_ρ is not convex when $\rho = 2n - \delta$ with some small $\delta > 0$. The instances we found for values of ρ closer to $2n+1$ are somewhat more complicated, so we do not include them here.

Consider problem instances with n variables and a single primal constraint. Assume that $\rho = 2n - \delta$ with $\delta > 0$. The class of LP instances in consideration are given by:

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}, \qquad G = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ & & \vdots & & \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

$$b = 1 \qquad c = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T$$

Note that $x = x(\varepsilon) = \begin{bmatrix} 1-(n-1)\varepsilon & \varepsilon & \cdots & \varepsilon \end{bmatrix}^T$ and $s = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$ is a strictly feasible primal-dual pair of solutions, as long as $0 < \varepsilon < \frac{1}{n}$. On the other hand, for each $0 < \delta < 1$, there exists a positive $\varepsilon < \frac{1}{n}$ such that the direction $\begin{bmatrix} -\varepsilon & -\varepsilon & \cdots & -\varepsilon & 1 \end{bmatrix}^T$ is a direction of negative curvature for F_ρ at $x(\varepsilon)$ and s. Therefore, $F_\rho(\cdot,\cdot)$ is not convex.

Gonzaga (1991) proves that the reduced Hessian matrix of Karmarkar's potential function ϕ_ρ (in addition to its multiplicative counterpart) is positive definite for ρ greater than $n-1$, for points on the central path. He assumes that a compactness condition holds for the feasible set. A primal-dual analogue of this result is not possible: For points on the central path, the reduced Hessian matrix of the TTY function is guaranteed to be positive definite only for small enough values of ρ, e.g., for ρ *less* than $\frac{n}{2}$. Small values of ρ are not useful, because they give too much importance to the barrier terms and polynomial convergence can not be achieved.

We will assume that $\rho > 2n+1$ in the rest of the paper. Let d_{RN} be the reduced Newton direction for $F_\rho(x,s)$ at some strictly feasible point (x,s). That is, d_{RN} is a solution of the following system:

$$(Z^T \nabla^2 F_\rho(x,s) Z) d_{RN} = -Z^T \nabla F_\rho(x,s) \tag{20}$$

which is equivalent to

$$\left(Z^T(\nabla^2 \Phi_\rho(x,s) + \nabla \Phi_\rho(x,s) \nabla^T \Phi_\rho(x,s)) Z\right) d_{RN} = -Z^T \nabla \Phi_\rho(x,s). \tag{21}$$

In the remainder of this paper, we describe and analyze algorithms that make use of the full space direction Zd_{RN}. Since $\nabla^2 F_\rho(x,s)$ is positive definite on $\mathcal{F}^0$, the direction Zd_{RN} is guaranteed to be a descent direction for F_ρ and, thus, for Φ_ρ. We will see in Section 5 that Zd_{RN} is also guaranteed to be a direction of negative curvature for Φ_ρ near a solution.

The left hand side matrix of the equation (21) consists of a block diagonal matrix and a rank-two matrix. Therefore this system can be solved efficiently using the Sherman-Morrison formula. Although the derivation is somewhat tedious, the resulting directions are relatively simple. Below, these directions will be presented formally, but first we introduce some notation. We will make use of two orthogonal projection matrices:

$$\Xi = \Xi(x) \; := \; X^{-1} G^T (G X^{-2} G^T)^{-1} G X^{-1}, \tag{22}$$

and

$$\Sigma = \Sigma(s) \ := \ S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}. \tag{23}$$

We note that Ξ and Σ are orthogonal projection matrices into the range spaces of $X^{-1}G^T$ and $S^{-1}A^T$, respectively. These range spaces are the same as the null spaces of AX and GS, respectively. Consequently,

$$\Xi \ = \ I - XA^T(AX^2A^T)^{-1}AX, \tag{24}$$

and

$$\Sigma \ = \ I - SG^T(GS^2G^T)^{-1}GS. \tag{25}$$

The analysis of the search direction defined in (21) relies heavily on an accurate estimation of the matrices Ξ and Σ, and of the orthogonal projections of the vectors e and XSe using these matrices. To this end, we define

$$\beta_1 \ := \ \frac{e^T XS\,(\Xi + \Sigma)\,XSe}{(x^Ts)^2}, \tag{26}$$

$$\beta_2 \ := \ \frac{e^T XS\,(\Xi + \Sigma)\,e}{(x^Ts)}, \tag{27}$$

$$\beta_3 \ := \ e^T\,(\Xi + \Sigma)\,e, \tag{28}$$

$$\Delta \ := \ (\rho\beta_1 - 1)(\rho - \beta_3 - 1) + \rho(1 - \beta_2)^2. \tag{29}$$

Above, Δ is the determinant of the 2×2 matrix encountered when using the Sherman-Morrison formula to invert the left-hand-side matrix in (21).

Proposition 3.1 *Let d_{RN} be the unique solution to (21). Then,*

$$d_{RN} \ = \ -\begin{bmatrix} (GX^{-2}G^T)^{-1}GX^{-1} \\ (AS^{-2}A^T)^{-1}AS^{-1} \end{bmatrix} \left(\frac{\rho(1-\beta_2)}{\Delta(x^Ts)}XSe + \frac{\rho\beta_1 - 1}{\Delta}e\right). \tag{30}$$

Now let $(\Delta x^T, \Delta s^T)^T = Zd_{RN}$. Then, Δx and Δs satisfy the following equations:

$$X^{-1}\Delta x = -\frac{\rho(1-\beta_2)}{\Delta(x^Ts)}\Xi(XSe) - \frac{\rho\beta_1 - 1}{\Delta}\Xi e, \tag{31}$$

$$S^{-1}\Delta s = -\frac{\rho(1-\beta_2)}{\Delta(x^Ts)}\Sigma(XSe) - \frac{\rho\beta_1 - 1}{\Delta}\Sigma e. \tag{32}$$

Proof:
Equation (30) can be verified by direct multiplication. Equations (31) and (32) follow from (30). Using (24) and (25), it is easily seen that Δx and Δs lie in the null spaces of A and G, respectively. □

Note that the bulk of the work in computing Δx and Δs from (31) and (32) is in the factorization of the matrices AX^2A^T and $AS^{-2}A^T$, which is twice the work required for an iteration of most interior-point methods. Once these factorizations are completed, all that is needed is a few backsolves, and matrix-vector products, which are all of lower order of complexity. We end this section with an evaluation of the change in the duality gap along the direction defined above:

Proposition 3.2 *Let Δx and Δs be defined by (31) and (32), and let $(x^+, s^+) = (x, s) + \alpha(\Delta x, \Delta s)$. Then,*

$$(x^+)^T s^+ \;=\; (x^T s)\left(1 - \frac{\alpha}{\Delta}(\rho\beta_1 - \beta_2)\right). \tag{33}$$

Proof:

$$\begin{aligned}
(x^+)^T s^+ &= (e + \alpha X^{-1}\Delta x)^T XS(e + \alpha S^{-1}\Delta s) \\
&= x^T s - \alpha e^T XS\left(\frac{\rho(1-\beta_2)}{\Delta(x^T s)}(\Xi + \Sigma)XSe + \frac{\rho\beta_1 - 1}{\Delta}(\Xi + \Sigma)e\right) \\
&= x^T s\left(1 - \frac{\alpha}{\Delta}(\rho(1-\beta_2)\beta_1 + (\rho\beta_1 - 1)\beta_2)\right) \\
&= (x^T s)\left(1 - \frac{\alpha}{\Delta}(\rho\beta_1 - \beta_2)\right).
\end{aligned}$$

The second equality above uses the identity $\Delta x^T \Delta s = 0$ and the the third equality uses the definitions (26) and (27). □

4 Analysis of the Search Directions

4.1 Symmetry and Invariance Issues

The search directions given by (31) and (32) are new for interior-point methods, although they are linear combinations of certain directions used by methods using primal-only or dual-only scaling. Our primal direction uses a primal-only scaling and our dual direction uses a dual-only scaling for the respective projection matrices. This fact may seem unpleasant, since the motivation for the search direction was a *primal-dual* potential function. Despite this seeming asymmetry in the projection matrices, the directions we use possess primal-dual symmetry, as well as scale invariance. These two important and desirable properties of search directions in primal-dual interior-point methods were studied by Todd, Toh, and Tütüncü (1996). Also, see papers by Todd (1997b) and Tunçel (1997) for axiomatic descriptions of "desirable" search directions.

The primal-dual symmetry of the search directions is related to the symmetry of the primal and dual linear programming problems, that is, the fact that the dual of the dual problem is the primal problem itself. If one were to treat the dual problem as the primal, and vice versa, the search directions given by certain algorithms may be different from what one would obtain without performing such a switch. If the directions are invariant under this switch of the primal and dual problems, they are called *primal-dual symmetric.* From our construction, it is easy to see that the directions defined in (31) and (32) are primal-dual symmetric. Somewhat paradoxically, the fact that the primal direction uses a primal-only scaling and the dual direction uses a dual only scaling establishes that the directions would not change if one were to swap the primal and dual problems and variables.

Scale invariance refers to the invariance of the search directions under a change of the measuring units used to describe the primal and dual variables. If the primal variable vector x is scaled using a diagonal scaling matrix D, so that new variable vector is defined by $\tilde{x} := D^{-1}x$, then the dual variable vector s is scaled to $\tilde{s} := Ds$, and the data is updated in the following manner:

$$\begin{aligned} A \to \tilde{A} &:= AD, \quad & G \to \tilde{G} &:= GD^{-1}, \\ c \to \tilde{c} &:= Dc, \quad & d \to \tilde{d} &:= D^{-1}d. \end{aligned}$$

Let $(\Delta x, \Delta s)$ be the search direction found by a particular method using the unscaled variables and data. If the search direction $(\tilde{\Delta x}, \tilde{\Delta s})$ found using the same method but the scaled data and scaled variables satisfy that

$$\tilde{\Delta x} = D^{-1}\Delta x \quad \text{and} \quad \tilde{\Delta s} = Ds, \tag{34}$$

then this particular method of finding search directions is said to be scale-invariant. A search direction given by a scale-invariant method is called a scale-invariant direction. The scale-invariance of our search directions immediately follow from the scale-invariance of Newton search directions. However, a direct argument using the equations (31) and (32) is simple and instructive, so we include it here: Since $GX^{-1} = \tilde{G}\tilde{X}^{-1}$ and $AS^{-1} = \tilde{A}\tilde{S}^{-1}$ regardless of the scaling matrix D, the orthogonal projection matrices Ξ and Σ that appear in the equations (31) and (32) are invariant under this scaling transformation. Similarly, the scalars β_j, $j = 1, 2, 3$ and Δ are not altered with the scaling. Therefore the search directions for the scaled problem satisfy (34), and our search directions are scale invariant.

We summarize the above discussion in the following proposition:

Proposition 4.1 *The search direction given by (31) and (32) is primal-dual symmetric and scale invariant.*

Most search directions for interior-point methods coincide when they are defined for points on the central path. For example, centering and affine-scaling directions are identical upto a scalar constant when they are defined for central points. Next, we demonstrate that the directions (31) and (32) share this property:

Proposition 4.2 *Let (x, s) be a point on the central path $\mathcal{C}$. Then, the search direction given by (31) and (32) is a scalar multiple of the primal-dual affine scaling direction.*

Proof:
The point (x, s) on the central path $\mathcal{C}$ satisfies $XS = \mu I$ for some $\mu > 0$. One can verify that $\Xi + \Sigma = I$ for such (x, s). Therefore, $\beta_2 = 1$, $\beta_3 = n$, and

$$S\Delta x + X\Delta s = -\frac{1}{\rho - n - 1} XSe.$$

□

4.2 Estimation of the Projection Matrices

In this subsection, we analyze the asymptotic behavior of the projection matrices $\Xi(x^k)$ and $\Sigma(s^k)$ on nondegenerate linear programs. Our analysis here is similar to the superlinear convergence analysis of Zhang, Tapia, and Dennis (1992).

Before we present our first result, we state the nondegeneracy assumption formally:

Assumption 1 *There is a (unique) nondegenerate pair of optimal solutions (x^*, s^*) for the primal-dual pair of problems (1) and (3).*

Without loss of generality we may assume that the first m elements of x^* are positive. Let

$$A = [B|N] \tag{35}$$

be the corresponding partition of the constraint matrix for the LP, where B is an $m \times m$ nonsingular matrix. B and N will also be used to denote the index sets corresponding to the indices of columns of matrices B and N; no confusion should occur.

Lemma 4.1 *Assume that the iterates (x^k, s^k) generated by some (feasible) interior-point algorithm converge to (x^*, s^*) described in Assumption 1. Then the matrix $\Xi(x^k) + \Sigma(s^k)$ converges to the identity matrix.*

Proof:
Let us denote the subvectors of x^k and s^k corresponding to the partition (35) by x_B^k, x_N^k, and s_B^k, s_N^k, respectively, and let $X_B^k = \mathrm{diag}(x_B^k)$, etc. We use formulas (24) for $\Xi(x^k)$ and (23) for $\Sigma(s^k)$.

$$\begin{aligned}
&\Xi(x^k) + \Sigma(s^k) \\
&= I - [BX_B^k \;\; NX_N^k]^T(B(X_B^k)^2B^T + N(X_N^k)^2N^T)^{-1}[BX_B^k \;\; NX_N^k] \\
&\quad +[B(S_B^k)^{-1} \;\; N(S_N^k)^{-1}]^T(B(S_B^k)^{-2}B^T + N(S_N^k)^{-2}N^T)^{-1}[B(S_B^k)^{-1} \;\; N(S_N^k)^{-1}] \\
&= I - \begin{bmatrix} \tilde{I}_x^k & \tilde{I}_x^k R_x^k \\ (R_x^k)^T \tilde{I}_x^k & (R_x^k)^T \tilde{I}_x^k R_x^k \end{bmatrix} + \begin{bmatrix} \tilde{I}_s^k & \tilde{I}_s^k R_s^k \\ (R_s^k)^T \tilde{I}_s^k & (R_s^k)^T \tilde{I}_s^k R_s^k \end{bmatrix}
\end{aligned}$$

where

$$R_x^k = (X_B^k)^{-1}B^{-1}NX_N^k, \qquad R_s^k = S_B^k B^{-1} N (S_N^k)^{-1}, \tag{36}$$

$$\tilde{I}_x^k = (I + R_x^k (R_x^k)^T)^{-1}, \qquad \tilde{I}_s^k = (I + R_s^k (R_s^k)^T)^{-1}. \tag{37}$$

Since X_B^k and S_N^k are bounded away from zero and X_N^k and S_B^k converge to zero, R_x^k and R_s^k both converge to zero matrices, and $\tilde{I}_x^k$ and $\tilde{I}_s^k$ both converge to the $m \times m$ identity matrix. Thus, the lemma follows. □

In order to analyze the asymptotic behavior of the directions given in (31) and (32) we need to estimate the scalars β_i, $i = 1, 2, 3$ and Δ. For this purpose, a tighter estimation of the matrices $\Xi(x^k)$ and $\Sigma(s^k)$ is required. Since $R_x^k \to 0$ we have that

$$\tilde{I}_x^k = I_m - R_x^k(R_x^k)^T + R_x^k\mathcal{O}(\|R_x^k\|^2)(R_x^k)^T = I_m - R_x^k\mathcal{O}(1)(R_x^k)^T.$$

Therefore,

$$\Xi(x^k) = \begin{bmatrix} 0 & 0 \\ 0 & I_{n-m} \end{bmatrix} - \begin{bmatrix} 0 & R_x^k \\ (R_x^k)^T & 0 \end{bmatrix} + E_x^k \tag{38}$$

where

$$E_x^k = \begin{bmatrix} R_x^k & 0 \\ 0 & (R_x^k)^T \end{bmatrix} \begin{bmatrix} \mathcal{O}(1) & \mathcal{O}(1)(R_x^k)^T \\ R_x^k\mathcal{O}(1) & \mathcal{O}(1) \end{bmatrix} \begin{bmatrix} (R_x^k)^T & 0 \\ 0 & R_x^k \end{bmatrix}.$$

Similarly,

$$\Sigma(s^k) = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & R_s^k \\ (R_s^k)^T & 0 \end{bmatrix} - E_s^k \tag{39}$$

where

$$E_s^k = \begin{bmatrix} R_s^k & 0 \\ 0 & (R_s^k)^T \end{bmatrix} \begin{bmatrix} \mathcal{O}(1) & \mathcal{O}(1)(R_s^k)^T \\ R_s^k\mathcal{O}(1) & \mathcal{O}(1) \end{bmatrix} \begin{bmatrix} (R_s^k)^T & 0 \\ 0 & R_s^k \end{bmatrix}.$$

4.3 Estimation of the Projected Vectors

As in the proof of Lemma 4.1, let x_B^k, x_N^k, and s_B^k, s_N^k denote the subvectors of x^k and s^k corresponding to the partition (35). Let p_1 and p_2 be nonnegative integers, and let N_i and B_j denote the ith and jth components of the index sets N and B, respectively. Then, for any i and j,

$$(x_{N_i}^k)^{p_1}(s_{B_j}^k)^{p_2} = \mathcal{O}\left(((x^k)^T s^k)^{(p_1+p_2)}\right).$$

Therefore, if H is a matrix such that $H = \mathcal{O}(1)$, the following relation holds:

$$(X_N^k)^{p_1} H (S_B^k)^{p_2} = \mathcal{O}\left(((x^k)^T s^k)^{(p_1+p_2)}\right). \tag{40}$$

Using (40), we make the following straightforward observations:

$$\begin{bmatrix} 0 & R_x^k \\ (R_x^k)^T & 0 \end{bmatrix} X^k S^k e = \begin{bmatrix} \left((X_B^k)^{-1}B^{-1}NS_N^k\right)(X_N^k)^2 e \\ X_N^k(N^TB^{-T})S_B^k e \end{bmatrix} = \mathcal{O}\left(((x^k)^T s^k)^2\right),$$

$$\begin{bmatrix} 0 & R_x^k \\ (R_x^k)^T & 0 \end{bmatrix} e = \begin{bmatrix} \left((X_B^k)^{-1}B^{-1}N\right) X_N^k e \\ X_N^k\left(N^TB^{-T}(X_B^k)^{-1}\right) e \end{bmatrix} = \mathcal{O}\left((x^k)^T s^k\right),$$

$$\begin{bmatrix} 0 & R_s^k \\ (R_s^k)^T & 0 \end{bmatrix} X^k S^k e = \begin{bmatrix} S_B^k(B^{-1}N)X_N^k e \\ \left((S_N^k)^{-1}N^TB^{-T}X_B^k\right)(S_B^k)^2 e \end{bmatrix} = \mathcal{O}\left(((x^k)^T s^k)^2\right),$$

$$\begin{bmatrix} 0 & R_s^k \\ (R_s^k)^T & 0 \end{bmatrix} e = \begin{bmatrix} S_B^k\left(B^{-1}N(S_N^k)^{-1}\right) e \\ \left((S_N^k)^{-1}N^TB^{-T}\right) S_B^k e \end{bmatrix} = \mathcal{O}\left((x^k)^T s^k\right).$$

Note that, the matrices in parentheses in the middle terms of the four equations above are all $\mathcal{O}(1)$ matrices, which justifies the second equality in each of these four systems. We continue by evaluating asymptotic values of some relevant vector-matrix-vector products:

$$\begin{aligned} e^T X^k S^k \begin{bmatrix} 0 & R_x^k \\ (R_x^k)^T & 0 \end{bmatrix} X^k S^k e &= 2e^T S_B^k\left(B^{-1}NS_N^k\right)(X_N^k)^2 e \\ &= \mathcal{O}\left(((x^k)^T s^k)^3\right), \\ e^T \begin{bmatrix} 0 & R_x^k \\ (R_x^k)^T & 0 \end{bmatrix} X^k S^k e &= \mathcal{O}\left(((x^k)^T s^k)^2\right), \\ e^T \begin{bmatrix} 0 & R_x^k \\ (R_x^k)^T & 0 \end{bmatrix} e &= \mathcal{O}\left((x^k)^T s^k\right), \end{aligned}$$

and similarly;

$$\begin{aligned} e^T X^k S^k \begin{bmatrix} 0 & R_s^k \\ (R_s^k)^T & 0 \end{bmatrix} X^k S^k e &= \mathcal{O}\left(((x^k)^T s^k)^3\right), \\ e^T \begin{bmatrix} 0 & R_s^k \\ (R_s^k)^T & 0 \end{bmatrix} X^k S^k e &= \mathcal{O}\left(((x^k)^T s^k)^2\right), \\ e^T \begin{bmatrix} 0 & R_s^k \\ (R_s^k)^T & 0 \end{bmatrix} e &= \mathcal{O}\left((x^k)^T s^k\right). \end{aligned}$$

The identities above also imply that

$$\begin{aligned} e^T X^k S^k (E_x^k - E_s^k) X^k S^k e &= \mathcal{O}\left(((x^k)^T s^k)^4\right), \\ e^T (E_x^k - E_s^k) X^k S^k e &= \mathcal{O}\left(((x^k)^T s^k)^3\right), \\ e^T (E_x^k - E_s^k) e &= \mathcal{O}\left(((x^k)^T s^k)^2\right). \end{aligned}$$

Now we are ready to estimate β_i, $i = 1, 2, 3$:

Lemma 4.2 *Under the assumptions of Lemma 4.1 we have the following:*

$$\beta_1 = \frac{\|X^kS^ke\|_2^2}{\|X^kS^ke\|_1^2} + \mathcal{O}\left((x^k)^Ts^k\right), \tag{41}$$

$$\beta_2 = 1 + \mathcal{O}\left((x^k)^Ts^k\right), \tag{42}$$

$$\beta_3 = n + \mathcal{O}\left((x^k)^Ts^k\right). \tag{43}$$

Proof:
We only derive the result for β_1, derivations for β_2 and β_3 are similar. Using (38) and (39),

$$\begin{aligned}
\beta_1 &= \frac{e^TX^kS^k\left(\Xi^k + \Sigma^k\right)X^kS^ke}{((x^k)^Ts^k)^2}, \\
&= \frac{\sum_{i=1}^n (x_i^ks_i^k)^2}{((x^k)^Ts^k)^2} + \frac{e^TX^kS^k\begin{bmatrix} 0 & R_s^k - R_x^k \\ (R_s^k - R_x^k)^T & 0 \end{bmatrix}X^kS^ke}{((x^k)^Ts^k)^2} \\
&\quad + \frac{e^TX^kS^k(E_x^k - E_s^k)X^kS^ke}{((x^k)^Ts^k)^2} \\
&= \frac{\|X^kS^ke\|_2^2}{\|X^kS^ke\|_1^2} + \mathcal{O}\left((x^k)^Ts^k\right).
\end{aligned}$$

The last equality above follows from the asymptotic results preceding the lemma. □

Note that, for $(x^k, s^k) > 0$,

$$\frac{1}{n} \le \delta(x^k, s^k) := \frac{\|X^kS^ke\|_2^2}{\|X^kS^ke\|_1^2} < 1. \tag{44}$$

The first inequality above follows from the Cauchy-Schwartz inequality and it holds with equality only if (x^k, s^k) is on the central path, i.e., when all $x_i^ks_i^k$ values are equal. Therefore, $\delta(x^k, s^k)$ can be regarded as a measure of centrality. The second inequality is always strict for $(x^k, s^k) > 0$. The important fact for our purposes is that $\delta(x^k, s^k)$ is bounded below and above by positive numbers independent from the iterate (x^k, s^k). Lemma 4.2 also implies that

$$\Delta = (\rho\beta_1 - 1)(\rho - \beta_3 - 1) + \rho(1 - \beta_2)^2 = (\rho\beta_1 - 1)(\rho - n - 1) + \mathcal{O}\left((x^k)^Ts^k\right). \tag{45}$$

If (x^k, s^k) is amall enough $\rho\beta_1 > 1$ and $\Delta > 0$, so we can write,

$$\Delta = (\rho\delta(x^k, s^k) - 1)(\rho - n - 1) + \mathcal{O}\left((x^k)^Ts^k\right) = \Theta(1). \tag{46}$$

The next lemma establishes that the search direction defined by (31) and (32) is similar to the primal-dual affine scaling direction. In the lemma we use 0_B and e_B to denote $|B|$-dimensional vectors of zeros and ones, respectively. The same applies for 0_N and e_N.

Lemma 4.3 *Under the assumptions of Lemma 4.1, the search direction defined by (31) and (32) satisfies the following relations:*

$$X^{-1}\Delta x = -\frac{1}{\rho - n - 1}\begin{bmatrix} 0_B \\ e_N \end{bmatrix} + \mathcal{O}\left((x^k)^T s^k\right), \tag{47}$$

$$S^{-1}\Delta s = -\frac{1}{\rho - n - 1}\begin{bmatrix} e_B \\ 0_N \end{bmatrix} + \mathcal{O}\left((x^k)^T s^k\right). \tag{48}$$

Proof:
Using Lemma 4.2 and (46) we have that

$$\frac{\rho(1-\beta_2)}{\Delta(x^T s)} = \mathcal{O}(1), \text{ and } \frac{\rho\beta_1 - 1}{\Delta} = \frac{1}{\rho - n - 1} + \mathcal{O}\left((x^k)^T s^k\right). \tag{49}$$

Next, using (38), (39), and the results preceding Lemma 4.2 we obtain:

$$\begin{aligned} \Xi X^k S^k e &= \mathcal{O}\left((x^k)^T s^k\right), \\ \Xi e &= \begin{bmatrix} 0_B \\ e_N \end{bmatrix} + \mathcal{O}\left((x^k)^T s^k\right), \\ \Sigma X^k S^k e &= \mathcal{O}\left((x^k)^T s^k\right), \\ \Sigma e &= \begin{bmatrix} e_B \\ 0_N \end{bmatrix} + \mathcal{O}\left((x^k)^T s^k\right). \end{aligned} \tag{50}$$

The statement of the lemma follows directly from (49) and (50). □

5 Polynomial and Quadratic Convergence on Nondegenerate Problems

The search direction introduced in (31) and (32) is a pure Newton direction for the multiplicative potential function F_ρ. Since this function is strictly convex on $\mathcal{F}^0$, this direction is a descent direction for both F_ρ and Φ_ρ. Next, we show that this direction is actually a direction of negative curvature for Φ_ρ when the duality gap $x^T s$ is small:

Proposition 5.1 *Let Δx and Δs be defined by (31) and (32). Then,*

$$\begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix}^T \nabla^2 \Phi_\rho(x,s) \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} = \frac{\Delta - (1-\rho\beta_1)}{\Delta^2}(1-\rho\beta_1),$$

which is negative if $\beta_1 > \frac{1}{\rho}$. If Assumption 1 holds, with small enough $x^T s$ we have $\beta_1 > \frac{1}{\rho}$ and $[\Delta x^T \;\; \Delta s^T]^T$ is a direction of negative curvature for Φ_ρ.

Proof:

$$\begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix}^T \nabla^2 \Phi_\rho(x,s) \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} = \Delta x^T X^{-2} \Delta x + \Delta s^T S^{-2} \Delta s - \frac{\rho}{(x^T s)^2}(s^T \Delta x + x^T \Delta s).$$

Since,

$$\Delta x^T X^{-2}\Delta x + \Delta s^T S^{-2} \Delta s = \|X^{-1}\Delta x\|^2 + \|S^{-1}\Delta s\|^2 = \frac{\rho^2(1-\beta_2)^2\beta_1 + 2\rho(1-\beta_2)(\rho\beta_1 - 1)\beta_2 + (\rho\beta_1 - 1)^2\beta_3}{\Delta^2},$$

and

$$s^T \Delta x + x^T \Delta s = -\frac{x^T s}{\Delta}(\rho\beta_1 - \beta_2),$$

after some simplification we get

$$\begin{aligned} \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix}^T \nabla^2 \Phi_\rho(x,s) \begin{bmatrix} \Delta x \\ \Delta s \end{bmatrix} &= \frac{\Delta - (1-\rho\beta_1)}{\Delta^2}(1-\rho\beta_1) \\ &= \frac{(\rho\beta_1 - 1)(\rho - \beta_3) + \rho(1-\beta_2)^2}{(\rho\beta_1 - 1)(\rho - \beta_3 - 1) + \rho(1-\beta_2)^2}(1-\rho\beta_1). \end{aligned}$$

Recall that $\beta_3 = e^T \Xi e + e^T \Sigma e = e^T \Xi^2 e + e^T \Sigma^2 e$, which is the sum of squares of the norms of two different orthogonal projections of the vector e. Each of these squared norms are bounded above by n, so $\beta_3 \le 2n$ and $\rho - \beta_3 - 1 > 0$. This implies that the right-hand-side of the above equation is negative when $(1 - \rho\beta_1) < 0$.

From (41) we have that $\beta_1 = \delta(x,s) + \mathcal{O}(x^T s)$ for nondegenerate problems, and from (44) we have $\delta(x,s) \ge 1/n$. Since $\rho > 2n+1$, we conclude that $(1-\rho\beta_1) < 0$ is negative when $x^T s$ is small. Thus, the statement of the proposition follows. □

Polynomial convergence of potential-reduction algorithms for linear programming problems is established by guaranteeing a significant reduction in the potential function in each iteration. As stated before, Kojima, Mizuno, and Yoshise (1991) proved such a result for an algorithm using the TTY potential function Φ_ρ, assuming that $\rho \ge n + \sqrt{n}$. Their algorithm, which we will call the KMY algorithm, converges linearly to an optimal solution. Here we will argue that, we can embed our technique within their algorithm, obtaining a variant that retains the polynomial convergence property and also exhibits quadratic convergence to nondegenerate solutions.

Let us describe this variant formally. We assume that an LP problem that satisfies Assumption 1 is given. Also, let ρ be an $\mathcal{O}(n)$ parameter greater than $2n+1$, and let $\varepsilon > 0$. Finally, let (x^0, s^0) be a strictly feasible solution for this LP such that $\Phi_\rho\left((x^0)^T s^0\right) = \mathcal{O}(n \ln \frac{1}{\varepsilon})$.

Step	**1.**	Let k=0.
Step	**2.**	Compute the search direction $(\Delta x, \Delta s)$ from (31) and (32) and choose a step size α^k such that $(x^+, s^+) := (x^k, s^k) + \alpha^k(\Delta x, \Delta s) > 0$.
Step	**3.**	<u>If</u> ($\Phi_\rho(x^+, s^+) - \Phi_\rho(x, s) \leq -0.2$) <u>then</u> let $(x^{k+1}, s^{k+1}) := (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$, $k = k + 1$. Go to step 2. <u>else</u> compute $(\Delta x, \Delta s)$ from (7) and choose the step size prescribed in Kojima, Mizuno, Yoshise (1991). Let $(x^{k+1}, s^{k+1}) := (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$, $k = k + 1$. Go to step 2. <u>End if</u>

Figure 1: A quadratically convergent variant of KMY algorithm

The algorithm in Figure 1 is reminiscent of the algorithms using fast and safe steps to achieve superlinear convergence. Polynomial convergence of this algorithm follows from the polynomial convergence of the KMY algorithm. To show that quadratic convergence can be obtained, we need to demostrate that appropriate step sizes can be found in Step 2 of the algorithm such that a sufficient decrease in the potential function Φ_ρ, and that a quadratic reduction in the duality gap can be achieved simultaneously.

A delicate balance in the selection of the step sizes is required, as the larger steps necessary for quadratic convergence may not guarantee reduction in the potential function; barrier terms may increase faster than the duality gap term decreases. Let $\alpha^k_{\max}$ be the largest feasible step size in iteration k:

$$\alpha^k_{\max} = \max\{\alpha : x^k + \alpha \Delta x^k \geq 0, s^k + \alpha \Delta s^k \geq 0\}$$

If Δx and Δs are given by (31) and (32), from Lemma 4.3, we have

$$\alpha^k_{\max} = (\rho - n - 1) + \mathcal{O}\left((x^k)^T s^k\right). \tag{51}$$

We will choose α^k, the step size in iteration k, as $\alpha^k = \tau^k \alpha^k_{\max}$ with $\tau^k = 1 - \Theta\left((x^k)^T s^k\right) < 1$. We first show that when $(x^k)^T s^k$ is small enough, this choice of the step size gives a significant reduction in the TTY potential function, and then establish that it also ensures the quadratic convergence of the duality gap sequence.

Theorem 5.1 *Let Δx and Δs be given by (31) and (32). Let $\alpha^k = \tau^k \alpha^k_{\max}$ with $\tau^k = 1 - \Theta\left((x^k)^T s^k\right) < 1$, and $(x^{k+1}, s^{k+1}) = (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$. Then, under the assumptions* *of Lemma 4.1,*

$$\Phi_\rho(x^{k+1}, s^{k+1}) - \Phi_\rho(x^k, s^k) \;\leq\; (\rho - n)\ln\left(\mathcal{O}\left((x^k)^T s^k\right)\right).$$

Proof:
To simplify the notation we let $(x^+, s^+) = (x^{k+1}, s^{k+1})$, $(x, s) = (x^k, s^k)$, and $\alpha = \alpha^k$. From Proposition 3.2,

$$\begin{aligned}
&\Phi_\rho(x^+, s^+) - \Phi_\rho(x, s) \\
&= \rho \ln\left(\frac{(x^+)^T s^+}{x^T s}\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta x_i}{x_i}\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta s_i}{s_i}\right) \\
&= \rho \ln\left(1 - \frac{\alpha}{\Delta}(\rho\beta_1 - \beta_2)\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta x_i}{x_i}\right) - \sum_{i=1}^n \ln\left(1 + \alpha\frac{\Delta s_i}{s_i}\right).
\end{aligned}$$

Using (41), (42), (45), and (51), we get:

$$\begin{aligned}
&\frac{\alpha}{\Delta}(\rho\beta_1 - \beta_2) \\
&= \frac{\left(1 - \Theta(x^T s)\right)\left((\rho - n - 1) + \mathcal{O}(x^T s)\right)\left(\rho\delta(x, s) - 1 + \mathcal{O}(x^T s)\right)}{(\rho\delta(x, s) - 1)(\rho - n - 1) + \mathcal{O}(x^T s)} \\
&= 1 - \Theta(x^T s).
\end{aligned}$$

Also, from Lemma 4.3 we have,

$$\begin{aligned}
1 + \alpha\frac{\Delta x_i}{x_i} &= \Theta(x^T s) \ \mathrm{i} \in \mathrm{N}, & 1 + \alpha\frac{\Delta s_i}{s_i} &= \Theta(x^T s) \ \mathrm{i} \in \mathrm{B}, \\
1 + \alpha\frac{\Delta x_i}{x_i} &= 1 - \mathcal{O}(x^T s) \ \mathrm{i} \in \mathrm{B}, & 1 + \alpha\frac{\Delta s_i}{s_i} &= 1 - \mathcal{O}(x^T s) \ \mathrm{i} \in \mathrm{N}.
\end{aligned}$$

From the definition of the order notation, there exists positive real numbers r and R independent from the current iterate (x, s) such that, any number $z = \Theta(x^T s)$ satisfies $r(x^T s) \leq z \leq R(x^T s)$ and any number $z = \mathcal{O}(x^T s)$ satisfies $z \leq R(x^T s)$. Using the asymptotic relations above, we obtain:

$$\begin{aligned}
\Phi_\rho(x^+, s^+) - \Phi_\rho(x, s) &\leq \rho\ln(R(x^T s)) - n\ln(r(x^T s)) - \mathcal{O}(1) \\
&= (\rho - n)\ln(x^T s) + \mathcal{O}(1) \\
&= (\rho - n)\ln\left(\mathcal{O}\left(x^T s\right)\right).
\end{aligned}$$

Therefore, as $x^T s$ tends to zero, the reduction in the potential function tends to $-\infty$. □

Theorem 5.2 *Let Δx and Δs be given by (31) and (32). Let $\alpha^k = \tau^k \alpha^k_{\max}$ with $\tau^k = 1 - \Theta\left((x^k)^T s^k\right) < 1$, and $(x^{k+1}, s^{k+1}) = (x^k, s^k) + \alpha^k(\Delta x, \Delta s)$. Then, under the assumptions of Lemma 4.1,*

$$\frac{(x^{k+1})^T s^{k+1}}{\left((x^k)^T s^k\right)^2} \;=\; \Theta(1).$$

Proof:
As in the proof of Theorem 5.1,

$$\frac{\alpha^k(\rho\beta_1 - \beta_2)}{\Delta} = 1 - \Theta\left((x^k)^T s^k\right).$$

Therefore,

$$\begin{aligned}\frac{(x^{k+1})^T s^{k+1}}{(x^k)^T s^k} &= 1 - \frac{\alpha^k}{\Delta}(\rho\beta_1 - \beta_2) \\ &= \Theta\left((x^k)^T s^k\right).\end{aligned}$$

This last equality indicates that

$$\frac{(x^{k+1})^T s^{k+1}}{((x^k)^T s^k)^2} = \Theta(1).$$

□

Under Assumption 1, there is a unique solution to the primal-dual pair of problems in consideration. The global convergence property of the KMY algorithm indicates that assumptions of Lemma 4.1 hold for our algorithm. By virtue of "safe steps", the duality gap will eventually get small. Therefore, the large decreases in the potential function expressed in Theorem 5.1 as well as the quadratic convergence outlined in Theorem 5.2 will take effect. Therefore, we have:

Corollary 5.1 *Assume that an LP problem that satisfies Assumption 1 is given. Also, let* ρ *be an* $\mathcal{O}(n)$ *parameter greater than* $2n+1$*, and let* $\varepsilon > 0$*. Finally, let* (x^0, s^0) *be a strictly feasible solution for this LP such that* $\Phi_\rho\left((x^0)^T s^0\right) = \mathcal{O}(n \ln \frac{1}{\varepsilon})$*. Then, the algorithm given in Figure 1 finds a solution* (x^k, s^k) *with duality gap less than* ε *in* $\mathcal{O}(n \ln \frac{1}{\varepsilon})$ *iterations. Furthermore, if the step sizes in Step 2 of the algorithm are chosen as in Thereom 5.1, the duality gap sequence* $\{(x^k)^T s^k\}$ *converges to zero quadratically.* □

6 Conclusion

This work was motivated by the following questions: Can primal-dual pure potential-reduction algorithms converge to optimal solutions of linear programs superlinearly? Superlinearly convergent path-following algorithms converge to optimal solutions by staying close to the central path. Is it possible to achieve such convergence without imposing any neighborhood restrictions on the iterates?

To answer these questions, we developed search directions that use second order information on the Tanabe-Todd-Ye primal-dual potential function. Since the Hessian of this function is indefinite in general, a modified Hessian is used in Newton's equation to produce

for LCP. *Math. Programming* **69** 311–333.

Sturm, J.F., S. Zhang (1996). New complexity results for the Iri-Imai method. *Ann. of Oper. Res.* **62** 539–564.

Tanabe, K. (1988). Centered Newton method for mathematical programming. M. Iri and K. Yajima, eds., *Lecture Notes in Control and Information Sciences,* Springer-Verlag, Berlin 197–206.

Todd, M. J. (1997a). Potential-reduction methods in mathematical programming. *Math. Programming* **76** 3–45.

——— (1997b). *On Search Directions in Interior-Point Methods for Semidefinite Programming,* Technical Report 1205, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.

——— , K.C. Toh, R.H. Tütüncü (1996). *On the Nesterov-Todd direction in semidefinite programming,* Technical Report 1154, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY. *SIAM J. Optim.* (to appear).

——— , Y. Ye (1990). A centered projective algorithm for linear programming. *Math. Oper. Res.* **15** 508–529.

Tsuchiya, T. (1995). Quadratic convergence of the Iri-Imai algorithm for degenerate linear programming problems. *J. Optim. Theo. Appl.*, **87** 703–726.

Tunçel, L. (1997). Primal-dual symmetry and scale invariance of interior-point methods for convex optimization. *Math. Oper. Res.* (to appear).

Ye, Y., K. O. Kortanek, J. Kaliski, S. Huang (1993). Near boundary behavior of primal-dual potential-reduction algorithms for linear programming. *Math. Programming* **58** 243–255.

——— , M. J. Todd, S. Mizuno (1994). An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Math. Oper. Res.* **19** 53–67.

Zhang, Y., R. A. Tapia, J. E. Dennis, Jr. (1992) On the superlinear and quadratic convergence of primal-dual interior-point linear programming algorithms. *SIAM J. Optim.* **2** 304–324.

Reha H. Tütüncü : Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA; e-mail: `reha+@andrew.cmu.edu`.

descent directions. The modification we consider is a rank-1 modification using the gradient of the function. We obtain a search direction that is equivalent to a pure Newton direction for the multiplicative analogue of the TTY function, which we show to be convex inside the feasible region when the function parameter is large enough.

Our approach is a primal-dual variant of the Iri-Imai method that uses similar ideas on Karmarkar's primal-only potential function. While Iri and Imai use their search directions from start to end, we prefer to use our directions only to improve local convergence. With this approach, we are able to modify Kojima, Mizuno, and Yoshise's polynomially convergent potential-reduction algorithm so that it converges quadratically to nondegenerate solutions. We believe our approach will attain its full potential once similar convergence results are established for degenerate problems. Our preliminary computational results in this direction are encouraging; degeneracy does not appear to hurt superlinear convergence. We leave a theoretical analysis of this case to a future study.

Finally, we would like comment on an issue that may have confused some readers. Although our search directions are Newton directions for the multiplicative primal-dual potential function F_ρ, the asymptotic step sizes we use to achieve quadratic convergence are much larger than 1, the usual step size for Newton's method. Indeed, our step sizes converge to the constant $\rho-n-1$. For the sake of the argument, assume that ρ is an integer. If one were to extend the definition of the function F_ρ and its derivatives for boundary points by using limits and extended values, it could be verified that first $\rho-n-1$ derivatives of F_ρ are zero operators at the optimal solution of a nondegerate problem. In other words, for a nondegenerate problem, the unique optimal solution is a root of multiplicity $\rho-n-1$ for the nonlinear equation $\nabla F_\rho = 0$. In such cases, the standard Newton's method converges only linearly to the optimal solution, while the quadratic convergence can be recovered by multiplying the Newton step with the multiplicity of the root. This is essentially what we do in our approach.

References

Anstreicher, K. M. (1996). Potential reduction algorithms. T. Terlaky, ed. *Interior point methods of mathematical programming*, Kluwer Academic Publishers, Dordrecht, 125–158.

Gonzaga, C. C. (1991). Large step path-following methods for linear programming, Part II: Potential-reduction method. *SIAM J. Optim.* **1** 280–292.

——— (1992). Path following methods for linear programming. *SIAM Rev.* **34** 167–227.

Iri, M., H. Imai (1986). A multiplicative barrier function method for linear programming. *Algorithmica* **1** 455–82.

——— (1993). A proof of the polynomiality of the Iri-Imai method. *J. Complexity* **9** 269–290.

Karmarkar, N. K. (1984). A new polynomial time algorithm for linear programming. *Combinatorica* **4** 373–395.

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

Monteiro, R. D. C., S. J. Wright (1995). Superlinear primal-dual affine scaling algorithms