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

TECHNICAL REPORT NO. 1067

September 1993

SCALING, SHIFTING AND WEIGHTING IN INTERIOR-POINT METHODS[1]

by

Michael J. Todd

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

Scaling, Shifting and Weighting in Interior-point Methods *

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

September, 1993

Dedicated to George B. Dantzig on his 80th birthday.

Abstract

We examine certain questions related to the choice of scaling, shifting and weighting strategies for interior-point methods for linear programming. One theme is the desire to make trajectories to be followed by algorithms into straight lines if possible to encourage fast convergence. While interior-point methods in general follow curves, this occurrence of straight lines seems appropriate to honor George Dantzig's contributions to linear programming, since his simplex method can be seen as following either a piecewise-linear path in n-space or a straight line in m-space (the simplex interpretation).

Prologue

Some six years ago, I bought a small sailboat, which was named Dantzig in view of its ability to follow piecewise-linear paths joining extreme points of a convex set (Cayuga Lake) in trying to attain a certain goal. At times, this is the most efficient method of progress, while in other situations a smooth curve through the interior is preferable.

Key words: Linear programming, interior-point methods.

Running Header: Scaling, shifting, and weighting.

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

1 Introduction

Interior-point methods for linear programming arose from the seminal work of Karmarkar [14]. By now there is a vast literature on the subject (see the extensive bibliography of Kranich [16], numbering over 1300 items), treating a wide variety of algorithms and their analysis. Certain common themes underlie these methods: the idea of moving in a direction of scaled projected steepest descent and the concepts of barrier function and central trajectories (see, e.g., Gonzaga [11] and Todd [23]). Here we study some aspects of the choice of a scaling matrix and of shifting and weighting strategies in barrier methods.

In section 2 we show that a choice of weights that leads to the primal-dual affine-scaling direction can be viewed as an approximation to weights (depending on the optimal solution and hence unknown) that yield a straight-line trajectory from the current solution to the optimal one. Section 3 demonstrates that a primal or dual barrier method that does not update the barrier parameter but merely adjusts weights can still yield convergence even without using shifts in the logarithmic barrier terms. This is related to convergence proofs by Polyak [21] and Powell [22] for Polyak's modified barrier method, and more general strategies for shifting and weighting studied by Gill et al. [12]. While in the unshifted case the main iterates converge to a point in the interior of the feasible region, an auxiliary sequence of iterates defined by straight-line extrapolations from a fixed interior point does converge to an optimal solution; indeed, the auxiliary sequence coincides with that generated by the modified barrier method.

Finally, in Section 4 we observe that the symmetric primal-dual scaling matrix can be motivated as an approximate Hessian of a new primal-dual barrier due to L. Tuncel [25].

Our focus here is on basic concepts rather than practical implementations. Readers interested in algorithms with good performance in practice, along with appropriate choices of starting points etc. and efficient implementations of the linear algebra required, are urged to consult the excellent papers of Lustig, Marsten and Shanno [17] and Mehrotra [19].

Throughout, we consider the primal problem in standard form:

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

and its dual (with explicit slacks s)

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

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. We assume that

$$
F^0(P) := \{x \in \mathbb{R}^n : Ax = b,\ x > 0\}
$$

and

$$
F^0(D) := \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^\top y + s = c,\ s > 0\}
$$

are nonempty. We use e to denote the vector of ones in $\mathbb{R}^n$, and for any lower-case letter representing a vector in $\mathbb{R}^n$ (e.g. x, s^*), the corresponding upper-case letter (e.g. X, S^*) represents the diagonal matrix of order n with the components of the vector on its diagonal.

2 Weighted trajectories

Several authors have introduced weights into the basic logarithmic barrier function. For the primal problem (P), the standard barrier

$$
\psi_\mu(x) := c^\top x - \mu \sum_j \ln x_j \tag{2.1}
$$

is replaced by

$$
\psi_{\mu,w}(x) := c^\top x - \mu \sum_j w_j \ln x_j; \tag{2.2}
$$

here $\mu > 0$ is the barrier parameter and the w_j's are positive weights. The weights are used to define trajectories passing through every strictly feasible point, or alternatively to allow starting path-following methods at initial points that are far from the central trajectories. See, for example, Megiddo [18] and Freund [9].

Usually, the weights w_j are fixed and the barrier parameter μ approaches zero. An alternative approach, in which μ is held fixed while the weights are adjusted, is usually applied together with shifts, and is discussed in the next section. Here we provide a new motivation for particular weights that have been studied by Megiddo [18] and Adler and Monteiro [1].

Suppose we have $x^0 \in F^0(P)$ and $(y^0, s^0) \in F^0(D)$. We can then define

$$w = w^0 = X^0 S^0 e \tag{2.3}$$

as our weight vector. Adler and Monteiro [1] have shown that these weights are very closely related to the primal-dual affine-scaling directions, which we now describe.

The optimality conditions for (P) and (D) can be written as

$$\begin{array}{rcll} Ax &=& b & (x \geq 0) \\ A^{\mathsf{T}} y + s &=& c & (s \geq 0) \\ XSe &=& 0, & \end{array} \tag{2.4}$$

where the first two systems ensure feasiblity in the primal and dual respectively, and the last complementary slackness. A Newton step from (x^0, y^0, s^0) for equations (2.4) moves in the direction (d_x, d_y, d_s) defined by

$$\begin{array}{rcl} Ad_x &=& 0 \\ A^{\mathsf{T}} d_y + d_s &=& 0 \\ S^0 d_x + X^0 d_s &=& -X^0 S^0 e. \end{array} \tag{2.5}$$

It is not hard to show that

$$d_x = -DP_{AD}Dc, \tag{2.6}$$

where D is the scaling matrix

$$D := (X^0)^{1/2}(S^0)^{-1/2} \tag{2.7}$$

and P_M denotes the orthogonal projection matrix into the null space of the matrix M. (Efficient methods of computing d_x, d_y and d_s are described in

[17, 19].) In contrast, the primal affine-scaling algorithm of Dikin [5, 6] (see also Barnes [2] and Vanderbei, Meketon and Freedman [26]) moves in the direction

$$\tilde{d}_x = -\tilde{D} P_{A\tilde{D}} \tilde{D} c \tag{2.8}$$

where $\tilde{D}$ is the primal scaling matrix

$$\tilde{D} := X^0. \tag{2.9}$$

Returning to the primal-dual setting, we can use the directions (d_x, d_y, d_s) to motivate the trajectory defined by $(x(\mu), y(\mu), s(\mu))$ solving

$$\begin{array}{rcl} A\dot{x} & = & 0 \\ A^\top \dot{y} + \dot{s} & = & 0 \\ S\dot{x} + X\dot{s} & = & +XSe \end{array} \tag{2.10}$$
$$x(1) = x^0, \quad y(1) = y^0, \quad s(1) = s^0.$$

(We have reversed the sign in the last system so that moving in the direction (d_x, d_y, d_s) corresponds to decreasing μ.) Adler and Monteiro [1] prove:

Theorem 2.1 *Let $(x(\mu), y(\mu), s(\mu))$ solve (2.10). Then, for each $\mu > 0$, $x(\mu)$ is the unique solution to the weighted barrier problem*

$$(WBP) \qquad \min\{\psi_{\mu,w}(x) : Ax = b, x > 0\}$$

where the weights $w = w^0$ are given by (2.3). □

Thus the primal-dual affine-scaling directions are the tangents to the trajectories defined by the weighted barrier function $\psi_{\mu,w}$. In the following we provide a motivation for choosing these weights and hence a possible reason for the efficiency of these directions.

Suppose (P) has a unique optimal solution x^*, and let (y^*, s^*) be an optimal solution to (D) satisfying strict complementarity with x^*, so that $x^* + s^* > 0$. Such a solution exists by a result of A.W. Tucker (see [4], p. 139). If $B := \{j : s_j^* = 0\} = \{j : x_j^* > 0\}$, then it is easy to see that the columns of A indexed by B are linearly independent.

Let

$$w^* := X^0 S^* e. \tag{2.11}$$

Theorem 2.2 *If $w = w^*$, the unique solution to (WBP) for $0 < \mu \leq 1$ is*

$$x(\mu) := \mu x^0 + (1-\mu)x^*. \tag{2.12}$$

Proof. Since $A^\top y^* + s^* = c$ and $Ax = b$ for all feasible solutions, the objective function $\psi_{\mu,w^*}(x)$ differs by a constant $(b^\top y^*)$ from

$$\tilde{\psi}_{\mu,w^*}(x) := (s^*)^\top x - \mu \sum_j w_j^* \ln x_j,$$

whose gradient is

$$\nabla \tilde{\psi}_{\mu,w^*}(x) = s^* - \mu X^0 S^* X^{-1} e.$$

Let $N = \{1, 2, \ldots, n\} \backslash B$. Since $x_N^* = 0$,

$$x_N(\mu) = \mu x_N^0,$$

which shows that the components of $\nabla \tilde{\psi}_{\mu,w^*}(x(\mu))$ indexed by N vanish. Since $s_B^* = 0$, the components indexed by B are also zero. Moreover, $x(\mu) > 0$ for $0 < \mu \leq 1$ (since $x^0 > 0$ and $x^* \geq 0$) and $Ax(\mu) = b$. It follows that $x(\mu)$ is the unique solution to (WBP) as long as ψ_{μ,w^*} (or equivalently $\tilde{\psi}_{\mu,w^*}$) is strictly convex on $F^0(P)$. But this is a consequence of the fact that the logarithm function is strictly convex and the linear independence of the columns of A_B (so that any nonzero direction in $F^0(P)$ must have a nonzero component with index in N). □

Theorem 2.2 implies that, if we knew an optimal dual slack vector s^*, the weights w^* would result in a straight-line trajectory from the current solution x^0 to the optimal solution x^*. Of course, we do not know s^*; but in a primal-dual algorithm, we have an estimate (y^0, s^0) of the optimal dual solution. It therefore seems highly appropriate to use the weight vector w^0 in (2.3) as an approximation to w^* in (2.11); we can thus hope that the primal-dual affine-scaling direction will point close to the optimal solution.

Of course, a similar analysis motivates the use of w^0 to improve the dual solution, since it approximates

$$w^+ := X^* S^0 e,$$

which also gives a straight-line trajectory to the optimal dual solution under suitable conditions.

To conclude this section, we mention a related result: Polyak [21] has shown that, with optimal weights, a single unconstrained minimization of the modified barrier function yields the exact solution of a linear programming problem.

3 Shifted barriers

Adding shifts to the barrier terms in a logarithmic barrier function can serve several purposes: it allows one to start at a point which is not strictly feasible with respect to the unshifted constraints (Freund [8]); it permits an algorithm to adjust both shifts and weights in an effort to get faster convergence (Gill et al. [12]); and it allows convergence to be achieved with fixed shifts and fixed barrier parameter, only adjusting the weights (Polyak [21] and Powell [22]). The last results are for Polyak's modified barrier method, which also enjoys the other advantages. Here we show that, if a strictly feasible solution is available, a natural choice of shifts is in some sense equivalent to no shifts at all; hence we obtain an unshifted algorithm which only adjusts the weights and yet converges to the optimal solution. For easy comparison with Gill et al. [12], Polyak [21] and Powell [22], we assume that a barrier method is applied to (D). (Polyak and Powell consider a minimization problem with greater-than-or-equal-to inequality constraints, but the changes are obvious.) The standard barrier function is then

$$\theta_\mu(y) := -b^\top y - \mu \sum_j \ln(c_j - a_j^\top y), \tag{3.1}$$

where c_j and a_j are the jth columns of $c^\top$ and A. Gill et al., Polyak, and Powell use instead the weighted and shifted function

$$\theta_{\mu,w,h}(y) := -b^\top y - \mu \sum_j w_j \ln(c_j - a_j^\top y + \mu h_j), \tag{3.2}$$

with $h = e$ for [21] and [22].

Let $y(\mu, w, h)$ denote the minimizer of $\theta_{\mu,w,h}$, which exists and is unique under the assumption, made for this section, that $F^0(D)$ is bounded and nonempty. Consider the modified barrier method (Polyak [21]): Let $\mu > 0$, $w^0, h \in I\!R^n$, with $w^0 > 0$ and $h > 0$. For each k, let

$$y^k := y(\mu, w^k, h), \tag{3.3}$$

and let

$$w_j^{k+1} := \mu w_j^k h_j / (c_j - a_j^\top y^k + \mu h_j). \tag{3.4}$$

Then Polyak shows that $\{y^k\}$ converges to an optimal solution to (D), as a special case of a result for nonlinear programming, under some additional assumptions; also, the weights converge to an optimal solution to (P). Powell [22] shows that these assumptions are unnecessary in the case of linear programming. (Actually, both assume that $h = e$, but clearly $y(\mu, w, h)$ also maximizes

$$b^{\top}y + \mu \sum_j w_j \ln(c_j/h_j - (a_j/h_j)^{\top} y + \mu),$$

so we can assume all the shifts are one by rescaling the data A and c.) We will sometimes call the modified barrier method above the *shifted* barrier algorithm, to compare it with the unshifted version (with $h = 0$) introduced below.

How should the shifts h be chosen? Clearly, they should reflect in some way the scaling of the constraints. One possibility is to choose $h_j = \|a_j\|$, so that all constraints are shifted by the same Euclidean distance. This may not be appropriate if, say, the dual constraints are $-1 \leq y_1 \leq 1$, $-10^6 \leq y_2 \leq 10^6$. Let us suppose that some $(\hat{y}, \hat{s}) \in F^0(D)$ is known. Then

$$\hat{h} := \tau \hat{s} = \tau(c - A^{\top}\hat{y}) \tag{3.5}$$

for some $\tau > 0$ seems suitable; each constraint is then shifted by a Euclidean distance proportional to the distance from the constraint to the strictly feasible solution $\hat{y}$.

The result below relates $y(\mu, w, \hat{h})$ to $y(\tilde{\mu}, w) := y(\tilde{\mu}, w, 0)$, the unique minimizer of the unshifted barrier function, with

$$\tilde{\mu} := \frac{\mu}{1 + \mu\tau}. \tag{3.6}$$

Proposition 3.1 *Let μ and the components of $w \in I\!R^n$ be positive. Let $(\hat{y}, \hat{s}) \in F^0(D)$ be arbitrary, and define $\hat{h}$ by (3.5) and $\tilde{\mu}$ by (3.6). Then*

$$y(\tilde{\mu}, w) = \frac{1}{1 + \mu\tau} y(\mu, w, \hat{h}) + \frac{\mu\tau}{1 + \mu\tau}\hat{y} \tag{3.7}$$

is a convex combination of the shifted barrier minimizer $y(\mu, w, \hat{h})$ and $\hat{y}$.

Proof. Under our assumptions, both $\theta_{\tilde{\mu},w,0}$ and $\theta_{\mu,w,\hat{h}}$ are strictly convex functions that attain their minima uniquely at points where their gradients vanish. Thus $\tilde{y} = y(\tilde{\mu}, w)$ uniquely solves

$$-b + \sum_j \frac{\tilde{\mu} w_j}{c_j - a_j^{\top}\tilde{y}} a_j = 0, \tag{3.8}$$

and $y := y(\mu, w, \hat{h})$ uniquely solves

$$-b + \sum_j \frac{\mu w_j}{c_j - a_j^\top y + \mu \hat{h}_j} a_j = 0. \tag{3.9}$$

If we set $\tilde{y} = \frac{1}{1+\mu\tau} y + \frac{\mu\tau}{1+\mu\tau} \hat{y}$, we find

$$\begin{aligned} c_j - a_j^\top \tilde{y} &= \tfrac{1}{1+\mu\tau}(c_j - a_j^\top y) + \tfrac{\mu\tau}{1+\mu\tau}(c_j - a_j^\top \hat{y}) \\ &= \tfrac{1}{1+\mu\tau}(c_j - a_j^\top y + \mu \hat{h}_j), \end{aligned} \tag{3.10}$$

and thus (3.8) follows from (3.9). Since (3.8) has a unique solution, we obtain (3.7). □

(3.7) can be viewed as describing a straight-line trajectory from $\hat{y}$ extending past $\tilde{y}$ along which the shifted barrier minimizer moves as the shifts and the barrier parameter are adjusted suitably.

From (3.4), (3.6), and (3.10), we obtain

$$\begin{aligned} w_j^{k+1} &= \mu w_j^k \hat{h}_j / (c_j - a_j^\top y^k + \mu \hat{h}_j) \\ &= [\tfrac{\mu}{1+\mu\tau} w_j^k \hat{h}_j] / [\tfrac{1}{1+\mu\tau}(c_j - a_j^\top y^k + \mu \hat{h}_j)] \\ &= \tilde{\mu} w_j^k \hat{h}_j / (c_j - a_j^\top \tilde{y}^k) \end{aligned} \tag{3.11}$$

if $y^k = y(\mu, w^k, \hat{h})$ and $\tilde{y}^k = y(\tilde{\mu}, w^k)$. Thus the update formula for the algorithm with shifts $\hat{h}$ corresponds to a simple formula for the following unshifted algorithm: Let $\tilde{\mu} > 0$, $w^0 \in \mathbb{R}^n$ with $w^0 > 0$, and $\hat{h} = \tau \hat{s} = \tau(c - A^\top \hat{y})$ for some $(\hat{y}, \hat{s}) \in F^0(D)$ and some $0 < \tau < 1/\tilde{\mu}$. For each k, let

$$\tilde{y}^k := y(\tilde{\mu}, w^k) \tag{3.12}$$

and let

$$w_j^{k+1} := \tilde{\mu} w_j^k \hat{h}_j / (c_j - a_j^\top \tilde{y}^k). \tag{3.13}$$

Define also the auxiliary iterates

$$y^k := \frac{1}{1 - \tilde{\mu}\tau} (\tilde{y}^k - \tilde{\mu}\tau \hat{y}). \tag{3.14}$$

Corollary 3.2 *In the unshifted algorithm above, the auxiliary iterates are exactly those generated by the shifted algorithm initiated with $\mu = \tilde{\mu}/(1 - \tilde{\mu}\tau)$, w^0 and $\hat{h}$. Hence $\{y^k\}$ converges to an optimal solution to (D).*

Proof. By induction. Suppose both algorithms have the same weights w^k at the start of iteration k. It is easy to see that $\tilde{\mu} = \mu/(1+\mu\tau)$. Hence by Proposition 3.1, $y(\mu, w^k, \hat{h})$ and $\tilde{y}^k$ are related by (3.7). Solve for $y(\mu, w^k, \hat{h})$, and using the definition of μ, we find that $y(\mu, w^k, \hat{h}) = y^k$ given by (3.14). Then (3.11) shows that the two update formulae, (3.4) and (3.13), generate the same w^{k+1}. $\square$

As in the modified barrier method, the Hessian of the barrier function in our unshifted barrier algorithm does not become increasingly ill-conditioned as the iterations proceed; here this is because the main iterates $\{\tilde{y}^k\}$ do not approach the boundary, while the barrier parameter is fixed. In the nonlinear case, the shifts are necessary (as in the augmented Lagrangian method–see, e.g., [7]) so that the iterates can approach the boundary while the barrier (or penalty) parameter remains fixed, in order that the algorithm can gain information about the objective and constraint functions near the optimal point. In the linear case, this is no longer necessary; our algorithm's main iterates $\{\tilde{y}^k\}$ do not approach the boundary, and yet the auxiliary iterates (which do not affect the method) still converge to an optimal solution.

We remark that the shifts $\hat{h}$ are not used in generating the sequence $\{\tilde{y}^k\}$ directly, but just in the updates of the weights. Hence we can postpone choosing $\hat{h}$ until $\tilde{y}^0$ is computed, and then use $\tilde{y}^0$ for $\hat{y}$ to define $\hat{h}$. In this form, the unshifted algorithm requires only $\tilde{\mu}$, τ and w^0 for its initialization.

One of the referees has pointed out that Jensen and Polyak (in the concluding section of [13]) discuss another interesting relationship between the modified (shifted) barrier method and the unshifted method. Namely, the iterates of the modified barrier method for a linear programming problem can alternatively be generated by successive minimization of a weighted logarithmic barrier function for the *dual* problem, followed by appropriate adjustments of the weights. In the dual barrier function there are no shifts in the barrier terms, but there is a shift in the objective function.

4 Scaling matrices

As we observed in Section 2, primal-dual methods use as search directions negative scaled projected gradients, there of the objective function but more generally of a barrier or potential function, using the scaling matrix

$$D = (X^k)^{1/2}(S^k)^{-1/2} \tag{4.1}$$

where (x^k, s^k) is the current iterate. Primal methods use directions that can be similarly motivated, using the scaling matix

$$D = X^k. \tag{4.2}$$

Here we provide some justification for the choice (4.1).

It is not hard to motivate (4.2). The simplest such reason is that the transformation $x \longrightarrow \hat{x} := D^{-1}x$ takes the current iterate x^k to $\hat{x}^k = e$, a distance 1 from all constraints, where a ball of radius 1 inscribes the nonnegative orthant and hence the feasible set. A more fundamental reason is that $D^{-2} = (X^k)^{-2}$ is the Hessian of the barrier function

$$\frac{1}{\mu} c^\top x - \sum_j \ln x_j$$

for any $\mu > 0$, and also the Hessian of the "convex part" of the potential function

$$\rho \ln(c^\top x - z) - \sum_j \ln x_j$$

for any $\rho > 0$ and z less than the optimal value of (P). Hence a negative scaled projected gradient can often be regarded as a Newton or Newton-like step. Alternatively, the Hessian matrix of the barrier $-\sum_j \ln x_j$ can be regarded as providing a local metric of the feasible region, possessing many useful properties (Bayer and Lagarias [3], Nesterov and Nemirovsky [20], and Karmarkar [15]). This viewpoint extends to more general linear constraints ([10] and [24]) and to the nonlinear case [20].

Now let us revert to the primal-dual setting. The transformation $x \longrightarrow \hat{x} = D^{-1}x$, $s \longrightarrow \hat{s} = Ds$ takes both x^k and s^k to $\hat{x}^k = \hat{s}^k = v := (X^k)^{1/2}(S^k)^{1/2}e$, around which balls of radius $\min_j v_j$ can be inscribed in the nonnegative orthants, and hence in the feasible regions. This is the largest radius achievable by a diagonal scaling that preserves the scalar product $x^\top s$. We would like a motivation in terms of the Hessian of a primal-dual barrier; is there a function $\psi(x, s)$ with

$$\nabla^2 \psi(x, s) = \begin{pmatrix} X^{-1}S & 0 \\ 0 & S^{-1}X \end{pmatrix}? \tag{4.3}$$

The standard barrier $\psi(x, s) = -\sum_j \ln x_j - \sum_j \ln s_j$ gives rise to

$$\nabla^2 \psi(x, s) = \begin{pmatrix} X^{-2} & 0 \\ 0 & S^{-2} \end{pmatrix}, \tag{4.4}$$

leading to separate scalings for primal and dual, which do not preserve the scalar product $x^\mathsf{T} s$; this leaves unanswered the question of whether (4.3) is achievable.

Unfortunately, the answer is in the negative, by simple arguments—consider the case with $n = 1$. We should add, however, that the diagonal scaling D of the variables x and s that preserves the scalar product $x^\mathsf{T} s$ (i.e., $x \to D^{-1}x, s \to Ds$) and that transforms the Hessian matrix in (4.4) to the most well-conditioned one (or to the closest matrix to the identity in either the Frobenius or ℓ_2-operator norm) is $D := X^{1/2}S^{-1/2}$.

On the other hand, we can approximate (4.3), using an entropy-like barrier introduced by L. Tuncel [25]. Let

$$\psi(x,s) := \sum_j x_j s_j \ln(x_j s_j). \tag{4.5}$$

Then it is easy to obtain

Proposition 4.1 *(Tuncel). Let ψ be given by (4.5) above. Then*

$$\nabla^2\psi(x,s) = \begin{pmatrix} X^{-1}S & \mathrm{diag}(2+\ln(x_j s_j)) \\ \mathrm{diag}(2+\ln(x_j s_j)) & S^{-1}X \end{pmatrix}, \tag{4.6}$$

whose diagonal, positive definite part agrees with the right-hand side of (4.3).
□

It is easy to see that (4.6) itself is not positive definite—take $n = 1$ and $x = s = 1$, for instance—so that using just its diagonal part seems reasonable for a Newton-like method. This result suggests the examination of algorithms using ψ in (4.5) as their barrier, which is the subject of current research.

References

[1] I. Adler and R. D. C. Monteiro, "Limiting behavior of the affine scaling continuous trajectories for linear programming problems," in: J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pp. 189–211, American Mathematical Society, Providence, RI, 1990.

[2] E. R. Barnes, "A variation on Karmarkar's algorithm for solving linear programming problems," *Mathematical Programming*, 36:174–182, 1986.

[3] D. A. Bayer and J. C. Lagarias, "The nonlinear geometry of linear programming, Part II : Legendre transform coordinates," *Transactions of the American Mathematical Society*, 314(2):527–581, 1989.

[4] G. B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, 1963.

[5] I. I. Dikin, "Iterative solution of problems of linear and quadratic programming," *Doklady Akademii Nauk SSSR*, 174:747–748, 1967 (translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967).

[6] I. I. Dikin, "On the convergence of an iterative process," *Upravlyaemye Sistemi*, 12:54–60, 1974 (in Russian).

[7] R. Fletcher, *Practical Methods of Optimization*, Wiley, Chichester–New York, 1987.

[8] R. M. Freund, "Theoretical efficiency of a shifted barrier function algorithm for linear programming," *Linear Algebra and Its Applications*, 152:19–41, 1991.

[9] R. M. Freund, "Projective transformations for interior point algorithms, a superlinearly convergent algorithm for the w–center problem," *Mathematical Programming*, 58(1993), 385–414.

[10] R. M. Freund and M. J. Todd, "Barrier functions and interior–point algorithms for linear programming with zero–, one–, or two–sided bounds on the variables," Technical Report 1016, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853–3801, USA, July 1992, to appear in *Mathematics of Operations Research.*

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

[12] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, "Shifted barrier methods for linear programming," Technical Report SOL 88–9, Systems Optimization Laboratory, Dept. of Operations Research, Stanford University, Stanford, CA 94305, USA, August 1988.

[13] D. L. Jensen and R. Polyak, "The convergence of a modified barrier method for convex programming," IBM Research Report RC 18570 (# 81288), T. J. Watson Research Center, Yorktown Heights, NY 10598, USA, November 1992, revised March 1994.

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

[15] N. K. Karmarkar, "Riemannian geometry underlying interior–point methods for linear programming," in: J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming*, volume 114 of *Contemporary Mathematics*, pp. 51–75, American Mathematical Society, Providence, RI, 1990.

[16] E. Kranich, "Interior–point methods bibliography," *SIAG/OPT Views–and–News, A Forum for the SIAM Activity Group on Optimization*, 1:11, 1992.

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

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

[19] S. Mehrotra, "On the implementation of a primal–dual interior point method," *SIAM Journal on Optimization*, 2(4):575–601, 1992.

[20] Y. E. Nesterov and A. S. Nemirovsky, *Interior Point Polynomial Algorithms in Convex Programming* SIAM, Philadelphia, USA, 1994.

[21] R. Polyak, "Modified barrier functions (theory and methods)," *Mathematical Programming*, 54:177–222, 1992.

[22] M. J. D. Powell, "Some convergence properties of the modified log barrier method for linear programming," Technical Report DAMTP 1992/NA7, Dept. of Applied Mathematics and Theoretical

Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK, November 1992, revised January 1994.

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

[24] M. J. Todd, "Analysis of interior–point methods for linear programming problems with variable upper bounds," Technical Report 1044, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853–3801, USA, February 1993, to appear in: Susana Gomez and Jean-Pierre Hennart, editors, *Advances in Optimization and Numerical Analysis*, Mathematics and its Applications 275, Kluwer Academic Publishers, 1994.

[25] L. Tuncel, private communication.

[26] R. J. Vanderbei, M. S. Meketon, and B. A. Freedman, "A modification of Karmarkar's linear programming algorithm," *Algorithmica*, 1(4):395–407, 1986.