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

TECHNICAL REPORT NO. 1135

September 1995

ON THE INTERPLAY AMONG ENTROPY, VARIABLE METRICS AND POTENTIAL FUNCTIONS IN INTERIOR-POINT ALGORITHMS

by

Levent Tunçel[1] and Michael J. Todd[2]

[1]Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada. Research supported in part by an operating grant from NSERC of Canada.
[2]School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801. Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

ON THE INTERPLAY AMONG ENTROPY, VARIABLE METRICS AND POTENTIAL FUNCTIONS IN INTERIOR-POINT ALGORITHMS

Levent Tunçel* Michael J. Todd†

September 1995

Abstract

We are motivated by the problem of constructing a primal-dual barrier function whose Hessian induces the (theoretically and practically) popular symmetric primal and dual scalings for linear programming problems. Although this goal is impossible to attain, we show that the primal-dual entropy function may provide a satisfactory alternative. We study primal-dual interior-point algorithms whose search directions are obtained from a potential function based on this primal-dual entropy barrier. We provide polynomial iteration bounds for these interior-point algorithms. Then we illustrate the connections between the barrier function and a reparametrization of the central path equations. Finally, we consider the possible effects of more general reparametrizations on infeasible-interior-point algorithms.

Keywords: linear programming, interior-point algorithms, primal-dual, entropy, potential function.

AMS 1991 Classification: 90C05, 65K05

1 Introduction

Some of the most intriguing concepts involved in the make up of an interior-point algorithm are scaling, variable metrics and potential functions (see, for instance, Karmarkar's seminal paper

*Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada. Research supported in part by an operating grant from NSERC of Canada.

†School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801 U.S.A. Research supported in part by NSF, AFOSR and ONR through NSF Grant DMS-8920550.

[8]). The choice of the search direction is closely tied in with these three fundamental concepts. Usually, a potential-reduction algorithm uses a scaled projected steepest descent direction which is related to the Newton direction; see the surveys Gonzaga [5] and [18]. On the other hand, the choice of the scaling matrix can usually be motivated by considering the Hessian of the barrier which is implicit in the potential function (although the first use of scaling in the affine-scaling method of Dikin [2] was developed without reference to potential or barrier functions). Related to this point, the Hessian of the barrier generally defines the variable metric used (see Bayer and Lagarias [1], Freund and Todd [6], Nesterov and Nemirovskii [14] and [17], [18]). Even though the primal (or dual) barriers possess the aforementioned properties, the primal-dual setting seems more complicated (see Freund and Todd [6] and [17], [18]).

Almost all potential-reduction algorithms use the traditional barrier

$$-\sum_{j=1}^{n} \ln(x_j)$$

for the nonnegative orthant and its extension (for the primal-dual setting)

$$-\sum_{j=1}^{n} \ln(x_j) - \sum_{j=1}^{n} \ln(s_j),$$

where x and s denote a primal and dual strictly feasible solution respectively. In the context of general optimization one of the most commonly used barrier functions is the entropy function

$$\sum_{j=1}^{n} x_j \ln(x_j).$$

Its use has also been suggested for linear programming problems; see, for instance, Erlander [3].

In this paper, we investigate some primal-dual interior-point algorithms motivated by the barrier

$$B(x,s) \quad := \quad \sum_{j=1}^{n} x_j s_j \ln\left(x_j s_j\right).$$

The second derivative of B with respect to x variables gives the (symmetric primal-dual) primal scaling and the second derivative of B with respect to s yields the (symmetric primal-dual) dual scaling. We also present another motivation for $B(x,s)$ in terms of a reparametrization of the Karush-Kuhn-Tucker (KKT) conditions for optimality for LP problems. We conclude with a general study of a family of reparametrizations of KKT conditions and their ramifications for the underlying infeasible-interior-point algorithms.

Finally, we remark that the barrier $B(x,s)$ has close ties to Polyak's work on modified barrier functions (see [15]).

2 Development of the Method and Related Entities

We will focus our attention on linear programming problems; in particular, we use the following form of primal and dual problems:

$$
\begin{array}{lllll}
(P) & \text{minimize} & c^T x & & \\
 & & Ax & = & b, \\
 & & x & \geq & 0,
\end{array}
$$

$$
\begin{array}{lllllll}
(D) & \text{maximize} & b^T y & & & & \\
 & & A^T y & + & s & = & c, \\
 & & & & s & \geq & 0,
\end{array}
$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. We assume A has full row rank. $\mathcal{F}_0$ denotes the set of pairs (x,s) of a primal feasible solution satisfying all inequality constraints with strict inequality and the s-part of a dual feasible solution satisfying all inequality constraints with strict inequality; we call such solutions *strictly feasible.* An initial assumption that there exist strictly feasible solutions for both problems is relaxed in Section 4. Given a pair of (primal and dual) feasible solutions x and s, μ denotes the duality gap divided by n or the average complementary slackness, i.e. $\mu = \frac{x^T s}{n}$. We use e to denote the vector of ones of appropriate dimension and $\mathbf{e}$ to denote the base of the natural logarithm. For a given vector v, V denotes the diagonal matrix $\operatorname{diag}(v)$ whose diagonal contains the components of v. Given a function $f : \mathbb{R} \to \mathbb{R}$, we write $f(v)$ for the vector defined by $(f(v))_j = f(v_j)$ and $f(V)$ for the diagonal matrix $\operatorname{diag}(f(v))$. (It would also be possible to use different scalar functions for each component, but we do not consider this possibility below.) Finally, the standard KKT conditions defining the central path are (for $\mu > 0$):

$$\begin{aligned} Ax &= b, \; x > 0; \\ A^T y + s &= c, \; s > 0; \\ Xs &= \mu e. \end{aligned}$$

We propose the following family of potential functions for primal-dual interior-point algorithms, where ρ is a positive parameter:

$$\phi(x, s; \rho) := \rho \ln(\mu) + \sum_{j=1}^{n} \frac{x_j s_j}{\mu} \ln\left(\frac{x_j s_j}{\mu}\right) \tag{1}$$

$$= (\rho - n) \ln(\mu) + \frac{1}{\mu} \sum_{j=1}^{n} x_j s_j \ln(x_j s_j). \tag{2}$$

We propose an interior-point algorithm whose development is based on the potential function obtained by setting ρ equal to n, yielding

$$\psi(x, s) := \phi(x, s; n) = \frac{1}{\mu} \sum_{j=1}^{n} x_j s_j \ln(x_j s_j). \tag{3}$$

Note that by focusing on a fixed value of the duality gap, $x^T s$, we can extract the corresponding barrier and the centering function from the proposed potential function. A measure of centrality corresponding to (1) is

$$\sum_{j=1}^{n} \frac{x_j s_j}{\mu} \ln\left(\frac{x_j s_j}{\mu}\right). \tag{4}$$

We divide this by n and define

$$\delta(x, s) := \sum_{j=1}^{n} \frac{x_j s_j}{n\mu} \ln\left(\frac{x_j s_j}{\mu}\right). \tag{5}$$

We conveniently have $0 \leq \delta(x,s) \leq \ln(n)$ for all strictly feasible (x,s), with $\delta(x,s) = 0$ if and only if (x,s) is on the central path. Whenever it is clear from the context what (x,s) is, we will write δ for $\delta(x,s)$. Note that ψ can also be expressed as $n\,[\ln(\mu) + \delta]$. This shows that, like other potential functions, ψ is unbounded below, and that driving ψ to minus infinity forces μ to zero. However, because δ is bounded, ψ can converge to minus infinity along a trajectory very far from the central path, so an algorithm using ψ will probably require some neighborhood condition to avoid approaching the boundary too closely.

Finally, to illustrate a relationship between δ and the two-norm centering function, an upper bound on δ can easily be obtained via the weighted-arithmetic-geometric mean inequality (or directly as below using the concavity of the logarithm and the facts that $\ln(1+\lambda) \leq \lambda$ and $\sum_{j=1}^{n} \frac{x_j s_j}{n\mu} = 1$):

$$\begin{aligned}
\delta(x,s) &= \sum_{j=1}^{n} \frac{x_j s_j}{n\mu} \ln\left(\frac{x_j s_j}{\mu}\right) \\
&\leq \ln\left[\frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j s_j}{\mu}\right)^2\right] && (6)\\
&= \ln\left[\frac{1}{n}\left\|\frac{Xs}{\mu} - e\right\|_2^2 + 1\right] && (7)\\
&\leq \frac{1}{n}\left\|\frac{Xs}{\mu} - e\right\|_2^2 . && (8)
\end{aligned}$$

Note that potential functions for feasible-interior-point algorithms usually have two parts: the objective function part and the barrier part. In the primal-dual setting the objective function part usually is $\ln(x^T s)$ or equivalently $\ln(\mu)$. However, we see from (3) that our potential function ψ can be viewed as (barrier part)/(objective function part). So, the corresponding barrier is

$$B(x,s) := \sum_{j=1}^{n} x_j s_j \ln(x_j s_j). \tag{9}$$

Note that the objective function part is the duality gap (divided by n), which is always positive for strictly feasible pairs. Close to optimality, all products $x_j s_j$ are less than one, so the barrier part is negative. Hence the potential function is decreased by decreasing the duality gap for a fixed value of the barrier or vice versa. Next, while we shall use the term barrier for B, it does not enjoy the usual property of approaching $+\infty$ as its argument approaches a point in the boundary of the feasible region; however, its gradient tends to infinity along such a sequence, which has the same repelling effect.

One of the interesting features of this barrier function is the metric induced by its Hessian. We partition the Hessian of B and study its pieces:

$$\frac{\partial^2 B(x,s)}{\partial x^2} := SX^{-1}, \tag{10}$$

$$\frac{\partial^2 B(x,s)}{\partial s^2} := XS^{-1}, \tag{11}$$

$$\frac{\partial^2 B(x,s)}{\partial x \partial s} := \ln(XS) + 2I. \tag{12}$$

The second derivative of B with respect to the x variables gives the (symmetric primal-dual) primal scaling and the second derivative of B with respect to s yields the (symmetric primal-dual) dual scaling. It is easy to see that the Hessian is positive definite on any subspace contained in the subset of $\mathbb{R}^{2n}$ defined by

$$\{(dx, ds) \; : \; dx^T ds \; = \; 0, \; dx^T [\ln(XS)] \, ds \; = \; 0\}.$$

If (x, s) lies on the central path then the set above is the same as

$$\{(dx, ds) \; : \; dx^T ds \; = \; 0\},$$

which includes the subspace of all feasible search directions.

Now we use the proposed potential function (3) to derive search directions. Note that the gradient of ψ with respect to the x-variables is

$$\nabla_x \psi(x,s) \;=\; -\mu^{-1}\left(s - \left[\delta s - S \ln\left(\frac{Xs}{\mu}\right)\right]\right),$$

while with respect to the s-variables it is the same expression with x and s interchanged. Suppose we work in scaled space; given $(x, s) \in \mathcal{F}_0$ we scale vectors in x-space by D^{-1} and vectors in s-space by D, with

$$D := X^{1/2} S^{-1/2}. \tag{13}$$

This scaling corresponds to the parts of the Hessian of B given in (10) and (11). Note that under this scaling, our current point (x, s) is transformed to (v, v), where

$$v := X^{1/2}S^{1/2}e. \tag{14}$$

Then in scaled space, the gradient of the potential function with respect to either the (scaled) primal variables or the (scaled) dual variables at the current point (v, v) is a positive multiple of $-w$, with

$$w := -v + \left[\delta v - V \ln\left(\frac{Vv}{\mu}\right)\right], \tag{15}$$

$$\tag{16}$$

Hence we find that the projected steepest-descent direction for ψ in scaled primal-dual space (up to a positive scalar multiple) is given by $(w_p^T, w_q^T)^T$, where w_p and w_q denote the orthogonal projections of w onto the null space and row space of AD respectively. So in the original spaces, the primal, dx, and dual, ds, search directions are

$$dx := Dw_p, \tag{17}$$

$$ds := D^{-1}w_q. \tag{18}$$

We would like to point out that the search direction has two components: affine-scaling and centering (as is the case for generic interior-point methods). However, the centering direction proposed here is different from the one given by the logarithmic barrier function. Our centering direction (in the scaled space) is a projection of $\delta X^{1/2}S^{1/2}e - X^{1/2}S^{1/2}\ln\left(\frac{Xs}{\mu}\right)$ whereas the centering direction usually used in primal-dual algorithms is a projection of $X^{-1/2}S^{-1/2}e$. On the other hand, when the current iterate is close to the central path, $\delta X^{1/2}S^{1/2}e - X^{1/2}S^{1/2}\ln\left(\frac{Xs}{\mu}\right)$ is approximately $(\delta-1)X^{1/2}S^{1/2}e + \mu X^{-1/2}S^{-1/2}e$ (the first order approximation to $\delta X^{1/2}S^{1/2}e + \ln\left(\mu(XS)^{-1}\right) X^{1/2}S^{1/2}e$), a constant-gap version of the standard centering direction.

We see that if the current iterate is on the central path then we have $\delta = 0$ and $w = -X^{1/2}S^{1/2}e$ (i.e., we obtain the primal-dual affine-scaling directions). Otherwise, as follows from Lemma 3.1(b), the projections of $\delta X^{1/2}S^{1/2}e - X^{1/2}S^{1/2}\ln\left(\frac{Xs}{\mu}\right)$ form a constant-gap centering direction. This fact can also be easily verified by noting that

$$v^T\left[\delta v - V \ln\left(\frac{Vv}{\mu}\right)\right] = 0; \tag{19}$$

hence, the two components of w are orthogonal.

Another interpretation for the new search direction can be given via an alternative description of the central path (for a more general treatment see Section 4 below). Recall the standard KKT conditions defining the central path:

$$\begin{aligned} Ax &= b, \ x > 0; \\ A^T y + s &= c, \ s > 0; \\ Xs &= \mu e. \end{aligned}$$

We can also replace the last vector equation by its equivalent:

$$\ln(Xs) = \ln(\mu e).$$

Suppose we apply the Newton method to this new system in order to approximate a point on the central path corresponding to the parameter μ_+. Then it is easy to see that the Newton direction coincides with our search direction if μ_+ is chosen such that

$$\ln\left(\frac{\mu}{\mu_+}\right) = 1 - \delta(x, s). \tag{20}$$

So, if (x, s) is on the central path then $\ln\left(\frac{\mu}{\mu_+}\right) = 1$ (i.e. $\mu_+ = \mu/\mathbf{e}$); while if (x, s) is less centered then the search direction is more conservative. We note that μ_+ may even be larger than μ.

3 Convergence Results

In this section global convergence properties of an algorithm based on the new search direction are discussed. The search direction is a projected steepest-descent direction for the proposed potential function, whereas the step size selection rule attempts to keep all the iterates in the following neighborhood of the central path:

$$\mathcal{N}_E(\beta) \;:=\; \left\{(x,s)\in\mathcal{F}_0 : \frac{1}{2}-\beta \;\le\; \ln\left(\frac{x_js_j}{\mu}\right) \;\le\; \frac{1}{2}+\beta, \quad \text{for all } j\right\}, \tag{21}$$

where $\beta \ge \frac{1}{2}$ is a constant. Note that the neighborhood is compatible with the reparametrization of the central path given by $f(\cdot) = \ln(\cdot)$. For small values of β this neighborhood is similar to the infinity-norm neighborhood while for large values β it looks more like the one-sided infinity-norm (wide) neighborhood, and almost includes the whole feasible region (see, e.g., Mizuno, Todd, and Ye [11]). We propose to pick the largest step size possible so that the iterates lie on the boundary of the corresponding neighborhood of the central path.

First, we establish some basic properties of the iterates based on the new search directions. We define

$$x(\alpha) \;:=\; x+\alpha dx, \tag{22}$$

$$s(\alpha) \;:=\; s+\alpha ds. \tag{23}$$

We have the following identities.

Lemma 3.1 *We have*
(a) $x_j(\alpha)s_j(\alpha) = x_js_j\left\{(1-\alpha)+\alpha\left[\delta(x,s)-\ln\left(\frac{x_js_j}{\mu}\right)\right]+\alpha^2\frac{(w_p)_j(w_q)_j}{x_js_j}\right\}.$
(b) $x(\alpha)^Ts(\alpha) = (1-\alpha)x^Ts.$

Proof. We simply use the definition of the search directions and orthogonality of w_p and w_q. □

From Lemma 3.1(b) we see that the length of the search direction is such that the duality gap decreases exactly as in the primal-dual affine-scaling algorithm [12]. We will show that if we have a starting solution (x^0, s^0) in a neighborhood of the central path and at each iteration we take the largest step that will keep the next iterate in the same neighborhood, we can guarantee polynomial-time convergence.

Clearly,

$$\max_j |(w_p)_j(w_q)_j| \;\le\; \|w_p\|_2\|w_q\|_2 \;\le\; \|w\|_2^2.$$

Now we bound $\|w\|_2^2$, assuming (x,s) lies in a neighborhood of the central path.

Lemma 3.2 *Given $\beta \geq \frac{1}{2}$, let $C(\beta) := \frac{5}{4} + \beta + \beta^2$. Then*

$$(x,s) \in \mathcal{N}_E(\beta) \text{ implies } \|w\|_2^2 \;\leq\; C(\beta)n\mu.$$

Moreover, if $(x,s) \in \mathcal{N}_E(\beta)$ for $\beta \leq \frac{3}{2}$ then $\delta(x,s) \;<\; 1$.

Proof. Using (15), we find:

$$\begin{aligned}
\|w\|_2^2 &= v^Tv + \delta^2 v^Tv - 2\delta v^T V \ln\left(\frac{Vv}{\mu}\right) + \left\|V\ln\left(\frac{Vv}{\mu}\right)\right\|_2^2 &(24)\\
&= (1-\delta^2)n\mu + \left\|V\ln\left(\frac{Vv}{\mu}\right)\right\|_2^2 &(25)\\
&= \sum_{j=1}^n x_js_j\left\{1-\delta^2+\left[\ln\left(\frac{x_js_j}{\mu}\right)\right]^2\right\}. &(26)
\end{aligned}$$

The first equality above is justified by the equality (19). Note that when (x,s) lies in $\mathcal{N}_E(\beta)$, we easily have $\left[\ln\left(\frac{x_js_j}{\mu}\right)\right]^2 \leq \frac{1}{4}+\beta+\beta^2$ for all j. So we conclude that

$$\|w\|_2^2 \;\leq\; C(\beta)n\mu.$$

Now, let us fix a positive value for μ and consider maximizing $\delta(x,s)$ subject to $x^Ts/n = \mu$ and $(x,s) \in \mathcal{N}_E(\beta)$. A relaxation (by ignoring the equality constraints of the underlying linear program) of this maximization problem is

$$\begin{aligned}
&\max\ \frac{1}{n}\sum_{j=1}^n t_j\ln(t_j)\\
&\text{s.t.}\ \sum_{j=1}^n t_j \;=\; n,\\
&\lambda \;\leq\; t_j \;\leq\; v,\quad \text{for all } j,
\end{aligned}$$

where $\lambda := \mathrm{e}^{(\frac{1}{2}-\beta)}$, $v := \mathrm{e}^{(\frac{1}{2}+\beta)}$ and t_j represents x_js_j/μ. If $\beta = 1/2$ then $\lambda = 1$ and the unique solution of the above problem yields an objective value of zero. If $\beta > 1/2$ then the problem is that of maximizing a convex function over a polytope; and, for such problems, we know that there exists an extreme point solution that is optimal. Thus, we can assume that at least $(n-1)$ of the variables t_j take values at their bounds, say k at their lower and $n-1-k$ at their upper bounds. One variable might take a value strictly between its bounds. However, since the scalar function $\phi(\tau) := \tau \ln(\tau)/n$ is convex, for any $\alpha \in (0,1)$, if $t_i = \alpha\lambda + (1-\alpha)v$ then

$$\phi(t_i) \;\leq\; \alpha\phi(\lambda) + (1-\alpha)\phi(v). \tag{27}$$

Thus the corresponding objective function value is bounded by $k+\alpha$ times $\phi(\lambda)$ plus $n-k-\alpha$ times $\phi(v)$. Since all t_j's must add up to n, we find that $k+\alpha = n(v-1)/(v-\lambda)$, and substituting this value gives as an upper bound on the optimal value $[(v-1)\phi(\lambda) + (1-\lambda)\phi(v)]/(v-\lambda)$. Now using the values of λ and v above yields

$$\delta(x,s) \;\leq\; \frac{(\frac{1}{2}+\beta)\mathbf{e}^{(\frac{1}{2}+\beta)} - 2\beta\mathbf{e} - (\frac{1}{2}-\beta)\mathbf{e}^{(\frac{1}{2}-\beta)}}{\mathbf{e}^{(\frac{1}{2}+\beta)} - \mathbf{e}^{(\frac{1}{2}-\beta)}}. \tag{28}$$

We note that when $\beta \leq \frac{3}{2}$, we get $\delta < 1$ as required. In fact, since the problem above becomes more relaxed as β increases, it suffices to consider the case where $\beta = \frac{3}{2}$. □

For the sake of simplicity, we analyze the algorithm only for $\beta = \frac{3}{2}$.

Lemma 3.3 *Let $(x,s) \in \mathcal{N}_E(\frac{3}{2})$ and $0 \leq \alpha \leq \frac{1}{12\mathbf{e}n}$. Then $(x(\alpha), s(\alpha)) \in \mathcal{N}_E(\frac{3}{2})$.*

Proof. Let $(x,s) \in \mathcal{N}_E(\frac{3}{2})$ and fix an index j. Then by Lemma 2, $\|w\|_2^2 \leq 5n\mu$ and $\delta < 1$. So,

$$\frac{\|w\|_2^2}{x_j s_j} \leq \frac{5n\mu}{x_j s_j} \leq 5\mathbf{e}n.$$

Using Lemma 1, we find

$$\frac{x_j(\alpha)s_j(\alpha)}{(1-\alpha)\mu} \;\geq\; \frac{x_j s_j}{\mu}\left\{1 + \frac{\alpha}{1-\alpha}\left[\delta - \ln\left(\frac{x_j s_j}{\mu}\right)\right] - \frac{\alpha^2}{1-\alpha}5\mathbf{e}n\right\}, \tag{29}$$

and

$$\frac{x_j(\alpha)s_j(\alpha)}{(1-\alpha)\mu} \;\leq\; \frac{x_j s_j}{\mu}\left\{1 + \frac{\alpha}{1-\alpha}\left[\delta - \ln\left(\frac{x_j s_j}{\mu}\right)\right] + \frac{\alpha^2}{1-\alpha}5\mathbf{e}n\right\}. \tag{30}$$

Now, we can analyze a few cases:

Case 1: $\frac{x_j s_j}{\mu} \in [1/\mathbf{e}, 1/\sqrt{\mathbf{e}})$.

In this case,

$$\frac{1}{2} = -\ln(1/\sqrt{\mathbf{e}}) \leq \delta - \ln\left(\frac{x_j s_j}{\mu}\right) \leq 2.$$

The first inequality above uses the fact that $\delta \geq 0$ and the second inequality uses $\delta < 1$ (Lemma 3.2). Now, from inequality (29) we see that if $\alpha = \frac{1}{12\mathbf{e}n}$ then

$$\frac{x_j(\alpha)s_j(\alpha)}{(1-\alpha)\mu} \geq \frac{x_j s_j}{\mu} \geq \frac{1}{\mathbf{e}}.$$

Using $\delta - \ln\left(\frac{x_j s_j}{\mu}\right) \leq 2$, inequality (30) and $\alpha = \frac{1}{12\mathbf{e}n}$ we also find

$$\frac{x_j(\alpha)s_j(\alpha)}{(1-\alpha)\mu} \leq \mathbf{e}^2.$$

Case 2: $\frac{x_j s_j}{\mu} \in [1/\sqrt{\mathbf{e}}, 1)$.

In this case,

$$0 \leq \delta - \ln\left(\frac{x_j s_j}{\mu}\right) \leq \frac{3}{2}.$$

So, using inequality (29) for $\alpha = \frac{1}{12\mathbf{e}n}$ we have

$$\frac{x_j(\alpha)s_j(\alpha)}{(1-\alpha)\mu} \geq \frac{1}{\sqrt{\mathbf{e}}}\left(1 - \frac{\alpha^2}{1-\alpha}5\mathbf{e}n\right) \geq \frac{1}{\mathbf{e}}.$$

Also, using inequality (30) for $\alpha = \frac{1}{12\mathbf{e}n}$ we get

$$\frac{x_j(\alpha)s_j(\alpha)}{(1-\alpha)\mu} \leq 1 + \frac{3}{2}\frac{\alpha}{1-\alpha} + \frac{\alpha^2}{1-\alpha}5\mathbf{e}n \leq \mathbf{e}^2.$$

Estimations for the cases 3 ($\frac{x_j s_j}{\mu} \in [1, \mathbf{e})$), 4 ($\frac{x_j s_j}{\mu} \in [\mathbf{e}, \mathbf{e}^{3/2}]$), and 5 ($\frac{x_j s_j}{\mu} \in [\mathbf{e}^{3/2}, \mathbf{e}^2]$) are similar. Feasibility of the new iterates in the corresponding indices follows from the fact that $\alpha < 1$ and that $x(\alpha)$ and $s(\alpha)$ are continuous functions of α. □

Theorem 3.1 *Let $(x^0, s^0) \in \mathcal{N}_E(\frac{3}{2})$ such that $x^T s \leq 2^t$ be given. Then the interior-point algorithm that uses the above search directions and chooses at each iteration the largest possible step size to stay in the neighborhood $\mathcal{N}_E(\frac{3}{2})$ finds a solution $(x, s) \in \mathcal{N}_E(\frac{3}{2})$ such that $x^T s \leq 2^{-t}$ in $O(nt)$ iterations.*

Proof. It follows from Lemma 3.3 and Lemma 3.1 (b) that the duality gap decreases at least to the fraction $\left(1 - \frac{1}{12en}\right)^k$ in k iterations. □

Notice that, although the search directions are based on the potential function ψ, we do not prove polynomiality using this function. At each iteration, $n \ln \mu$ decreases by at least $\frac{1}{12e}$, but $n\delta$ might increase by n. However, since $0 \le \delta \le 1$ for all iterates, the potential function does decrease by a constant per iteration on average.

4 Effects of Different Parametrizations on Infeasible-Interior-Point Algorithms

In this section we investigate some algorithmic ramifications of different parametrizations of the KKT conditions. Let's consider again the KKT conditions in section 2. One of the ways we motivated the new search directions was by rewriting the last optimality condition as $\ln x_j + \ln s_j = \ln(\mu)$ for all j.

Indeed, our main point in suggesting the potential function ψ (and thus the corresponding directions) was the fact that its Hessian induced the desired primal-dual scaling in both primal and dual spaces. On the other hand, the motivation in terms of reparametrization of the KKT conditions suggests many other possibilities. Aiming for a more general treatment of different parametrizations, we can consider the set of all continuously differentiable strictly monotone functions f from $\mathbb{R}_+$ to $\mathbb{R}$. Recall that $f(V)$ denotes the diagonal matrix whose entries are $f(v_j)$; similarly $f'(V)$ denotes the diagonal matrix whose entries are $f'(v_j)$. Then the last optimality condition becomes:

$$f(Xs) \;=\; f(\mu e).$$

Considering the Jacobian (the monotonicity of f and the fact that A has full row rank ensure that the Jacobian has full rank), we see that the usual equality

$$Sdx + Xds = \gamma\mu e - Xs$$

is replaced by

$$Sf'(XS)dx \;+\; Xf'(XS)ds \;=\; f(\gamma\mu e) - f(Xs).$$

Since $f'(XS)$ is diagonal and positive definite, we can invert it, to get

$$Sdx + Xds \;=\; (f'(XS))^{-1}(f(\gamma\mu e) - f(Xs)). \tag{31}$$

Thus the scaling is always the same, however the right hand side depends on the parametrization.

This includes the directions of Jansen, Roos and Terlaky [7] and Nazareth [13] as special cases. For instance, we can take $f(v_j) = -1/v_j$ (the inverted path equations in Nazareth's note). Letting $\gamma \to 0$ yields the search direction proposed in [7].

Now we can (at least qualitatively) discuss what kind of functions f would be desirable for infeasible-interior-point algorithms, as the trade-off between achieving feasibility and complementarity seems to have important effects on the convergence of such algorithms (at least theoretically).

A family of functions could be $f(v_j) := v_j^\alpha$, for $\alpha > 0$, or $f(v_j) := -v_j^\alpha$, for $\alpha < 0$. In this case, the right-hand-side of (31) becomes

$$\frac{1}{\alpha}\left[(\gamma\mu)^\alpha (XS)^{1-\alpha} e - Xs\right].$$

So, for positive α, one component of the search direction (corresponding to the term $-Xs$) is the same (up to a positive scalar multiple) as the primal-dual affine-scaling direction, whereas the centering component is new: $(\gamma\mu)^\alpha (XS)^{1-\alpha} e$. (The new centering component coincides with the conventional one when (x, s) lies on the central path.)

The following fact is well-known:

Fact 4.1 *Suppose A has full row rank and $x > 0$, $s > 0$, and let D be given by (13). Then the following system of linear equations*

$$Adx = p \tag{32}$$

$$A^T dy + ds = q \tag{33}$$

$$Sdx + Xds = r \tag{34}$$

has the unique solution

$$\begin{aligned} dx &= S^{-1}(r - Xds) \\ dy &= \left(AD^2A^T\right)^{-1}\left[p - AS^{-1}(r - Xq)\right] \\ ds &= q - A^T dy. \end{aligned}$$

In particular, defining $\bar{A} := AD$, $\bar{q} := Dq$ and $\bar{r} := X^{-1/2}S^{-1/2}r$ we have

$$dx = D\left[P_{\bar{A}}\bar{r} - P_{\bar{A}}\bar{q} + \bar{A}^T(\bar{A}\bar{A}^T)^{-1}p\right] \tag{35}$$

$$ds = D^{-1}\left[(I - P_{\bar{A}})\bar{r} + P_{\bar{A}}\bar{q} - \bar{A}^T(\bar{A}\bar{A}^T)^{-1}p\right]. \tag{36}$$

Note that the equations (32) and (33) come from a system of linear equations, while equation (34) is derived from the quadratic equation $Xs = \mu e$. Intuitively, to have a balanced solution (search direction) of (32)–(34) the right-hand-side of (34) could be altered to compensate for the nonlinearity of the quadratic equation. In this respect, the parametrization $f(v_j) = v_j^\alpha$ for $\alpha = 1/2$ is an interesting choice, since it attempts to make the nonlinear (quadratic) equation $Xs = \mu e$ look more "linear." From the same point of view, the parametrization given by $f(v_j) = \ln(v_j)$ seems too aggressive (the duality gap may decrease considerably faster than the infeasibility).

Now, we try to make the above vague arguments more concrete: In an infeasible-interior-point algorithm, p is $b - Ax$, the primal infeasibility, and q is the dual infeasibility at the current iterate. Since A has full row rank, $b = A\tilde{x}$ for some $\tilde{x}$ (in fact, for any feasible $\tilde{x}$, although our choice does not have to be feasible), so we can write $p = Al$. Let us write $\bar{l} := D^{-1}l$ so that $p = \bar{A}\bar{l}$. Then equations (35) and (36) can be rewritten as

$$dx = D\left[P_{\bar{A}}\bar{r} - P_{\bar{A}}\bar{q} + (I - P_{\bar{A}})\bar{l}\right] \tag{37}$$

$$ds = D^{-1}\left[(I - P_{\bar{A}})\bar{r} + P_{\bar{A}}\bar{q} - (I - P_{\bar{A}})\bar{l}\right]. \tag{38}$$

It is clear from (32)–(33) that a step α in the direction given by dx and ds will multiply the current primal and dual infeasibility by the factor $1 - \alpha$ (if $\alpha = 1$ does not violate the nonnegativity constraints then a feasible solution can be reached in just one step). So, if we would like to achieve complementarity and feasibility in a balanced manner, then one of the conditions we might want to ensure is to have

$$(x + \alpha dx)^T(s + \alpha ds) \approx (1 - \alpha)n\mu;$$

recall $n\mu := x^T s$. Alternatively, to ensure that infeasibility is reduced at least as fast as complementarity, we would like

$$(x + \alpha dx)^T(s + \alpha ds) \geq (1 - \alpha)n\mu.$$

Constraints similar to these can be found in the analyses of Kojima, Megiddo, and Mizuno [9], Mizuno [10], Potra [16], and Zhang [19] for primal-dual methods and Freund [4] for a primal method.

Defining $x(\alpha)$ and $s(\alpha)$ as before and denoting the projection of r onto the null space (row space) of $\bar{A}$ by r_p (r_q) etc., we have from (37)–(38)

$$x(\alpha)^T s(\alpha) = x^T s + \alpha(e^T r) + \alpha^2 \left[\bar{q}_p^T(\bar{r}_p - \bar{q}_p) + \bar{l}_q^T(\bar{r}_q - \bar{l}_q)\right]. \quad (39)$$

So, we want

$$e^T r = -x^T s \quad (40)$$

(for balance) or

$$e^T r \geq -x^T s$$

(for decrease of complementarity no faster than of infeasibility) and

$$\left[\bar{q}_p^T(\bar{r}_p - \bar{q}_p) + \bar{l}_q^T(\bar{r}_q - \bar{l}_q)\right] \approx 0.$$

Note that, using the reparametrization with f,

$$r = [f'(XS)]^{-1}\left[f(\gamma\mu)e - f(Xs)\right].$$

So, $e^T r = -x^T s$ if and only if

$$\sum_{j=1}^{n}\left[\frac{f(\gamma\mu) - f(x_j s_j)}{f'(x_j s_j)}\right] = -x^T s$$

(and similarly for $e^T r \geq -x^T s$). This can be viewed as an equation (or inequality) for γ. For example, if $f(v_j) := v_j^\alpha$ for $\alpha > 0$ or $f(v_j) := v_j^\alpha$ for $\alpha > 0$, we find the following: for $\alpha = 1$ — f is the identity — γ should be equal to (or at least) 0. Otherwise, γ should be equal to (or at least)

$$\left[\frac{1}{1-\alpha}\left(\frac{\sum_{j=1}^{n}(x_j s_j/\mu)^{1-\alpha}}{n}\right)\right]^{-1/\alpha}.$$

Using the concavity of $\phi(\tau) := \tau^\beta$ if $\beta \in (0,1)$ or its convexity if $\beta < 0$, we find that for $\alpha < 1$, γ must be at least $(1-\alpha)^{1/\alpha}$ (and more if the current iterate is not centered). This value is 1/4 for $\alpha = 1/2$ (trying to make the quadratic equation look linear) and 1/2 for $\alpha = -1$ (Nazareth's choice in [13]). For $f(v_j) := \ln(v_j)$ it is straightforward to see that γ should be equal to (or at least) $e^{\delta-1}$. This corresponds to equation (20) and shows that, if the current iterate is not well-centered, a large value of γ must be chosen so that complementarity doesn't decrease faster than infeasibility.

The coefficient of α^2 is zero if, for example,

$$\bar{q}_p^T(\bar{r}_p - \bar{q}_p) = \bar{l}_q^T(\bar{r}_q - \bar{l}_q) = 0. \tag{41}$$

This specifies the projections of $\bar{r}$ along $\bar{q}_p$ and $\bar{l}_q$, but it is otherwise free. One possibility is to use a *variable parametrization*, i.e., to vary the parametrization of equality $Xs = \mu e$ from iteration to iteration so that (at least approximately) (40) and (41) hold.

To conclude, we note that the basic idea of reparametrizing the condition $Xs = \mu e$ can be applied to other related problems, such as linear complementarity problems or semidefinite programming problems. For linear complementarity problems, the techniques above can be used directly. For semidefinite programming problems, the variables X and S are $n \times n$ symmetric positive semidefinite matrices. We could for example reparametrize the usual equation $XS = \mu I$ as

$$Q^T(XS)\text{diag}[(f(\lambda(XS))]Q(XS) \;=\; \text{diag}[(f(\mu e)],$$

where $Q^T(XS)\text{diag}[\lambda(XS)]Q(XS)$ is the spectral decomposition of XS. Even though this seems acceptable "in principle", it may not be as practical (as say in linear programming problems or linear complementarity problems).

References

[1] D. A. Bayer and J. C. Lagarias, "The nonlinear geometry of linear programming: I. Affine and projective scaling trajectories, II. Legendre transform coordinates and central trajectories," *Transactions of the American Mathematical Society*, vol. 314, pp. 499–581, 1989.

[2] I. I. Dikin, Iterative solution of problems of linear and quadratic programming, *Soviet Mathematics Doklady*, vol. 8, pp. 674–675, 1967.

[3] S. Erlander, "Entropy in linear programs," *Mathematical Programming*, vol. 21, pp. 137–151, 1981.

[4] R. M. Freund, "A potential reduction algorithm with user-specified phase I-phase II balance for solving linear program from an infeasible warm start," *SIAM Journal on Optimization*, vol. 5, pp. 247–268, 1995.

[5] C. C. Gonzaga, "Path following methods for linear programming," *SIAM Review*, vol. 34, pp. 167–227, 1992.

[6] 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," *Mathematics of Operations Research*, vol. 20, pp. 415–440, 1995.

[7] B. Jansen, C. Ross and T. Terlaky, "A polynomial primal-dual Dikin-type algorithm for linear programming," Report No. 93–36, Faculty of Technical Mathematics and Informatics, Delft University of Technology, The Netherlands, 1993.

[8] N. Karmarkar, "A new polynomial time algorithm for linear programming," *Combinatorica*, vol. 4, pp. 373–395, 1984.

[9] M. Kojima, N. Megiddo, and S. Mizuno, "Theoretical convergence of large-step primal-dual interior point algorithms for linear programming," *Mathematical Programming*, vol. 59, pp. 1–21, 1993.

[10] S. Mizuno, "Polynomiality of infeasible–interior–point algorithms for linear programming," *Mathematical Programming*, vol. 67, pp. 109–119, 1994.

[11] S. Mizuno, M.J. Todd, and Y. Ye, "On adaptive-step primal-dual interior-point algorithms for linear programming," *Mathematics of Operations Research*, vol. 18, pp.964–981, 1993.

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

[13] J. L. Nazareth, "A reformulation of the central path equations and its algorithmic implications," manuscript, Washington State University, Pullman, WA, 1993.

[14] Y. E. Nesterov and A. S. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Publications, SIAM, Philadelphia, 1994.

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

[16] F. A. Potra, "A quadratically convergent predictor–corrector method for solving linear programs from infeasible starting points," *Mathematical Programming*, vol. 67, pp. 383–406, 1994.

[17] M. J. Todd, "Scaling, shifting and weighting in interior-point methods," *Computational Optimization and Applications*, vol. 3, pp. 305–315, 1994.

[18] M. J. Todd, "Potential reduction methods in mathematical programming," *Technical Report 1112*, School of OR and IE, Cornell University, Ithaca, NY, U.S.A. (January 1995).

[19] Y. Zhang, "On the convergence of a class of infeasible interior–point methods for the horizontal linear complementarity problem," *SIAM Journal on Optimization*, vol. 4, pp. 208–227, 1994.