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

TECHNICAL REPORT NO. 1170

July 1996

On Adjusting Parameters in Homotopy Methods for Linear Programming[1]

by

Michael J. Todd

[1]Research supported in part by NSF through grant DMS-9505155 and ONR through grant N00014-96-1-0050.

On Adjusting Parameters in Homotopy Methods for Linear Programming

Michael J. Todd *
School of Operations Research
and Industrial Engineering
Rhodes Hall
Cornell University
Ithaca, NY 14853

July 1996

Dedicated to Michael J. D. Powell on his 60th birthday.

Abstract

Several algorithms in optimization can be viewed as following a solution as a parameter or set of parameters is adjusted to a desired value. Examples include homotopy methods in complementarity problems and path-following (infeasible-) interior-point methods. If we have a metric in solution space that corresponds to the complexity of moving from one solution point to another, there is an induced metric in parameter space, which can be used to guide parameter-adjustment schemes. We investigate this viewpoint for feasible- and infeasible- interior-point methods for linear programming.

Key words: homotopy methods, path-following methods, interior-point algorithms, linear programming, Riemannian metrics

Running Header: Adjusting Parameters in Homotopy Methods

*Research supported in part by NSF through grant DMS-9505155 and ONR through grant N00014-96-1-0050.

1 Introduction

This paper is concerned with developing guidelines for optimal adjustment of parameters in homotopy or path-following algorithms in optimization, concentrating on interior-point methods for linear programming. The general idea of such algorithms is that, given a particular problem, we regard part of the data, and possibly some additional parameters, as a parameter vector. To obtain a solution corresponding to some fixed value of the parameters, we trace a path as the parameter vector is adjusted from an initial artificial value, for which the corresponding solution is known, to the desired value. The thesis of our work is that for some such algorithms, in particular for several interior-point methods for linear programming, there is a natural (Riemannian) metric on parameter space with two properties: firstly, the distance between two parameter vectors measures the complexity of obtaining the solution corresponding to the second given that corresponding to the first; and secondly, that this complexity can be attained by following the shortest path joining the two vectors in this metric.

While our focus here is on algorithms for linear programming, Mike Powell's work in optimization has been mainly concerned with nonlinear programming. However, the methods we consider use primarily ideas from nonlinear programming, particularly those of the classical barrier function approach to constrained optimization. In addition, Powell himself has made two very fine contributions to interior-point algorithms for linear programming: an analysis of Karmarkar's original algorithm [7] for discretizations of a simple two-variable semi-infinite programming problem [20, 19], and a general convergence proof [21] for Polyak's modified barrier method [18]. In the first, the concern was complexity: Powell showed that the number of iterations required could be close to the bound established by Karmarkar; in the second, the topic was proving convergence of a (primal) method that did not require an initial strictly feasible solution. Our interest here also revolves around these two themes of complexity and infeasible- interior-point methods, but now from a primal-dual viewpoint.

There is another connection to Powell's work, indeed to his pioneering paper with Fletcher [4] on Davidon's variable metric algorithm for unconstrained minimization [2]. These papers introduced the famous DFP update formula and more generally gave birth to the field of quasi-Newton methods. The idea of a variable metric is key in this and also in our work. For unconstrained minimization, the Hessian of the objective function or a positive definite approximation to it defines a local norm at the current solution, and steepest descent with respect to this metric permits good progress in decreasing the function. As the iterates move, the curvature of the function changes, and the appropriate metric to yield good behavior also changes.

In interior-point methods a variable metric is also fundamental, but here it is used to describe the *constraint set* K rather than the objective function. This is different from incorporating curvature from the constraint functions into the objective function via the Lagrangian as in sequential quadratic programming, for example, because curvature is induced even by linear constraints; the metric reflects the geometry of *all* the constraints, not just those estimated to be active at the solution. Just as in general relativity the presence of all (but especially nearby) masses curves space-time, so here the presence of all (but especially nearby) constraint boundaries determines the local geometry.

To give an example, for the nonnegative orthant $K := \Re^n_+$, at every point x of the positive orthant $\operatorname{int} K = \Re^n_{++}$ we can define a local norm by $\|v\|_x := \|X^{-1}v\|_2$, where here and below X denotes the diagonal matrix containing the components of x. First, this norm reflects in some sense the shape of K: $\{x + v : \|v\|_x < 1\} \subseteq \operatorname{int} K$. Second, it is the ellipsoidal norm defined by the Hessian of a barrier function for K, namely the logarithmic barrier function

$$\phi(x) := -\sum_{j=1}^{n} \ln(x^{(j)}), \tag{1.1}$$

where $x^{(j)}$ denotes the jth component of x: $\|v\|_x = (v^\top \phi''(x) v)^{1/2}$. Third, if we are close to a desired point in this norm, we can approximate it well by a single Newton step: if $\bar{x}$ solves $\min\{\mu^{-1}c^\top x + \phi(x) : Ax = b, x \in K\}$, and we have a point $x \in \operatorname{int} K$ with $\|\bar{x} - x\|_x \leq \delta$ for some absolute constant $\delta < 1$, then one Newton step from x will give a very good approximation to $\bar{x}$. (Results of this form have been proved in a very general framework by Nesterov and Nemirovskii [14].)

Actually, we shall use not only local norms, but also the induced Riemannian metrics (the distance between two points is the length of a shortest path, measured using the local norm, connecting them) in what follows. The resulting Riemannian metric in parameter space will have the two properties mentioned at the beginning.

This paper is organized as follows. In Section 2 we describe the homotopy methods of interest, first in an abstract framework and then for linear programming. Then Section 3 discusses metrics in solution space, first in general and then for linear programming, and then shows under a certain key hypothesis how the complexity of a path-following method can be bounded by a constant times the induced distance between two parameter vectors. In Section 4 we calculate the local norm in parameter space for linear programming, and Section 5 then finds some shortest paths. Some of these challenge the conventional wisdom that recommends approaching feasibility faster than (or at the same rate as) optimality in infeasible-interior-point methods. Finally, Section 6 sketches a justification for the crucial hypothesis concerning the metrics in the case of linear programming.

A forthcoming paper will show how the same ideas can be applied to pivoting algorithms and to interior-point methods for more general convex programming problems, and will contain proofs of some of the results that are omitted here.

Related ideas of how to adjust parameters in interior-point methods appear in the target-following work of Mizuno [11] and Jansen, Roos, Terlaky, and Vial [5, 6]. Of course, the key ideas of Nesterov and Nemirovskii [14] on how barrier functions induce metrics with computational significance were also an important catalyst. While this research was in progress, I became aware of related later work by Nesterov and Nemirovskii [15] on moving efficiently in a multi-parameter surface. I would also like to acknowledge very helpful conversations with Jim Renegar on this approach, with Clovis Gonzaga on infeasible-interior-point methods, and with Gongyun Zhao, who has had very similar ideas, on appropriate metrics.

Our notation is mostly standard. We will use upper case letters (like X, S, and W) to denote the diagonal matrices containing the components of the corresponding vectors (like x, s, and w), and e to denote the n-vector of ones, so that $Xe = x$, etc. Thus XSe denotes the vector whose components are the products of those of x and s, but we shall also use xs to denote the same vector, and similarly $x^{1/2}$, x^{-1}, and $\ln x$ to denote the vectors of square roots, reciprocals, and logarithms of the components of x, etc. Sequences will be indicated by subscripts, and components by superscripts inside parentheses, as in (1.1) above.

2 Homotopy methods for mathematical programming

Here we describe formally the class of methods we are dealing with and show how interior-point algorithms for linear programming are included.

Solution vectors z lie in some subset Z of a Euclidean space, while parameter vectors λ lie in a subset Λ of another Euclidean space. (Perhaps it would be more accurate to say subsets of finite-dimensional real vector spaces, because we shall use metrics (Riemannian metrics) on these spaces, which may be very different from Euclidean metrics, in order to measure the complexity of homotopy algorithms.) The solution and the parameter are related implicitly by the equation

$$F(z, \lambda) = 0. \tag{2.1}$$

We assume that F is continuously differentiable and that for each $\lambda \in \Lambda$ there is a unique $z = z(\lambda) \in Z$ satisfying (2.1). We also assume that the partial derivative with respect to z, $F_z(z(\lambda), \lambda)$, is nonsingular for any $\lambda \in \Lambda$.

We suppose we know $z_0 := z(\lambda_0)$ (or a good approximation to it) for some $\lambda_0 \in \Lambda$, and we seek $z_1 := z(\lambda_1)$ for some $\lambda_1 \in \Lambda$. The basic idea of a homotopy or path-following method is to define a path $\gamma : [0,1] \to \Lambda$ from λ_0 to λ_1 (i.e., with $\gamma(i) = \lambda_i$ for $i = 0, 1$), and to trace or approximate the solutions $\zeta(t) := z(\gamma(t))$ as t goes from 0 to 1. The basic mechanics of the method determine how the tracing is to be performed (usually via Newton or Newton-like steps), whereas the key high-level question is how the path γ is to be chosen. Our criterion throughout will be the *complexity* of obtaining z_1, i.e., of tracing the path ζ. Simple examples show that the linear interpolation

$$\gamma(t) := (1 - t)\lambda_0 + t\lambda_1 \tag{2.2}$$

may not be best according to this criterion.

While the focus here is on short-step path-following methods, we suspect that the paths selected will also be appropriate for adaptive- or long-step methods, where the parameter is moved at each iteration as far along the path as a single Newton step can track accurately or where a long step along the path is attempted by taking a sequence of Newton steps.

We remark that a parallel theory can be developed for the case that $F(z, \lambda) = f(z) - \lambda$, with f piecewise-linear. In this case, pivots replace Newton steps at each iteration, and the metrics to be defined are not Riemannian but basically count the number of pieces of linearity encountered. Lemke's method [10] can be viewed as a homotopy method in this framework. For simplicity, we confine ourselves here to the continuously differentiable case. This still encompasses a great variety of computational algorithms, for example smooth continuation methods for zero-finding or fixed-point problems. Our motivation comes from interior-point, and in particular infeasible-interior-point methods, for linear programming and its extensions, originating with Karmarkar [7]. Here, due largely to the pioneering work of Nesterov and Nemirovskii [14], a great deal is known about the behavior of a Newton step at each iteration based on a fundamental metric defined by a barrier function. Our aim is to formalize the use of these metrics and use them to suggest "good" paths γ to be used in path-following methods.

Let us therefore describe the framework of path-following interior-point methods for linear programming. In fact, our development also applies to similar methods for more general conic problems as studied by Nesterov and Nemirovskii [14], where the nonnegative orthant is replaced by a closed convex solid pointed cone. If the cone has a *self-scaled* barrier in the sense of Nesterov and Todd [16, 17], then similar results can be established. Such problems include those of semidefinite programming, which have been much studied recently.

2.1 Interior-point methods for linear programming

Suppose we wish to find an optimal solution to the problem

$$(LP) \qquad \begin{array}{lrcl} \min_x & c_1^\top x & & \\ & Ax & = & b_1, \\ & x & \geq & 0, \end{array}$$

and its dual

$$(LD) \qquad \begin{array}{lrcrcl} \max_{y,s} & b_1^\top y & & & & \\ & A^\top y & + & s & = & c_1, \\ & & & s & \geq & 0, \end{array}$$

where $A \in \Re^{m\times n}$, $b_1 \in \Re^m$, $c_1 \in \Re^n$, $x, s \in \Re^n$, and $y \in \Re^m$. We always suppose A fixed and, without loss of generality, of full row rank. The right-hand-side vector and the cost vector may be fixed, but often will be part of the parameter vector. Our approach is via solutions to the system

$$\begin{array}{rcrcrcll} Ax & & & - & b & = & 0, & (x \geq 0) \\ A^\top y & + & s & - & c & = & 0, & (s \geq 0) \\ & & XSe & - & v^2 & = & 0 & \end{array} \tag{2.3}$$

for a sequence of values of b, c, and v. Recall that our notation here is that e always denotes a vector of ones in $\Re^n$, while X and S are diagonal matrices with $Xe = x$ and $Se = s$; v^2 denotes the vector whose components are the squares of those of $v \in \Re^n$ (the *target* vector). We also write xs for XSe when no confusion can result.

Note first that the optimality conditions for (LP) and (LD) are exactly (2.3) for $v = 0$. Interior-point methods consider solutions for vectors $v > 0$, in which case x and s are both positive (interior points of the cone $\Re^n_+$). We will therefore ignore the nonnegativity conditions in (2.3), and write the left-hand side of these equations as $F(z, \lambda)$, where always $z := (x, y, s) \in Z := \Re^n_{++} \times \Re^m \times \Re^n_{++}$.

Note that $F_z(z, \lambda)$ is then

$$\begin{bmatrix} 0 & A & 0 \\ A^\top & 0 & I \\ 0 & S & X \end{bmatrix},$$

which is nonsingular for any $x, s > 0$; it is also well-known that there is a unique solution $(x, y, s) \in Z$ for any b, c, and $v > 0$ for which $Ax = b$ and $A^\top y + s = c$ have solutions with $x, s > 0$.

There are several choices for Λ. Suppose first that (LP) and (LD) have *strictly feasible* solutions, with x and s positive. Then there is a very important path in the set of feasible solutions to (LP) and (LD), called the *central path*. It consists of all solutions to (2.3) with $b = b_1$, $c = c_1$, and $v^2 = \mu e$ for some positive μ. Suppose we have strictly feasible solutions x_0 and (y_0, s_0), with $x_0 s_0$ close to $\mu_0 e$ for some $\mu_0 > 0$. Then we can approximate a solution to (LP) and (LD) by following the central path; this corresponds to choosing $\lambda := \mu \in \Lambda := (0, \infty)$, fixing b and c as b_1 and c_1, and replacing v^2 by μe. In this case our parameter is one-dimensional, so the question of choosing the path γ is moot, but we will still be able to say something about the complexity of moving from $\mu = \mu_0$ to $\mu = \mu_1$ in terms of the length (suitably defined) of this path.

Recall that for any feasible solutions x and (y, s) to (LP) and (LD),

$$c^\top x - b^\top y = (A^\top y + s)^\top x - (Ax)^\top y = s^\top x \geq 0,$$

so the primal objective value is always at least the dual objective value. The difference is called the duality gap, and is zero at optimality. If $z = (x, y, s)$ solves (2.3) with $v^2 = \mu e$, then the duality gap is

$$s^\top x = e^\top XSe = e^\top(\mu e) = n\mu, \tag{2.4}$$

so finding a solution for a small value μ_1 guarantees a small duality gap. Hence central path-following algorithms (e.g., Monteiro-Adler [13] and Kojima-Mizuno-Yoshise [8, 9]) are included in our framework.

Next suppose we again have strictly feasible solutions x_0 and (y_0, s_0), but $x_0 s_0$ is not close to a multiple of e. We could perform some "centering" steps to approach the central path and then proceed as above, but it may be more efficient to use a target-following method. Thus we again fix b and c as b_1 and c_1, but now use $\lambda := v \in \Lambda := \Re^n_{++}$ as our parameter vector. Note that the duality gap for the corresponding solution is now

$$s^\top x = e^\top XSe = e^\top v^2 = \|v\|_2^2, \tag{2.5}$$

so we want to adjust $\lambda = v$ from its initial value v_0 with $x_0 s_0 = v_0^2$ to some v_1 of small norm. Thus the methods of Mizuno [11] and Jansen-Roos-Terlaky-Vial [5, 6] are embraced.

If we do not have strictly feasible solutions, we can choose an arbitrary starting point $z_0 = (x_0, y_0, s_0) \in Z$ and set $b_0 = Ax_0$, $c_0 = A^\top y_0 + s_0$. There remain several choices for λ. If $x_0 s_0$ is close to $\mu_0 e$ for some $\mu_0 > 0$, we can choose $\lambda := (b, c, \mu) \in \Lambda := A(\Re^n_{++}) \times (A^\top(\Re^m) + \Re^n_{++}) \times \Re_{++}$, and follow solutions to (2.3), with v^2 replaced by μe, as λ moves from $\lambda_0 := (b_0, c_0, \mu_0)$ to $\lambda_1 := (b_1, c_1, \mu_1)$ for some suitably small μ_1. If not, we can choose $\lambda := (b, c, v)$ in the obvious Λ and proceed similarly. We can also replace b and c by $(1-\theta)b_0 + \theta b_1$ and $(1-\theta)c_0 + \theta c_1$, and use $\lambda := (\theta, \mu)$ or $\lambda := (\theta, v)$ as our parameter. Hence we include most *infeasible-interior-point* methods; see, e.g., Mizuno-Todd-Ye [12] and the references therein.

3 Metrics reflecting complexity

In this section we show how, given a metric d_Z on Z indicating the complexity of moving the solution vector, we can infer a metric d_Λ on Λ so that we can obtain (a good approximation to) $z(\lambda_1)$ from (a good approximation to) $z(\lambda_0)$ in $O(d_\Lambda(\lambda_0, \lambda_1))$ basic steps, and the corresponding path γ in Λ is the shortest path in Λ according to d_Λ.

We will suppose that Z is a differentiable manifold endowed with a Riemannian structure, whose distance reflects the complexity of moving between two different solution vectors. We then pull this metric back to Λ, which is also assumed to be a differentiable manifold. Then the distance in Λ from λ_0 to λ_1 will represent the "best" complexity of a path-following method to approximate $z(\lambda_1)$ given $z(\lambda_0)$, and the corresponding shortest path a desired path to follow in Λ to achieve this complexity. We illustrate the development using linear programming.

Let Z be a differentiable manifold equipped with a Riemannian structure (see, for example, Boothby [1]), so that for each $z \in Z$, there is an inner product $\langle \cdot, \cdot \rangle_z$ defined on the tangent space at z. Thus for every tangent vector $\dot{z}$ at z (we use $\dot{z}$ instead of dz to simplify the notation; it does not necessarily indicate the derivative of a path, but no confusion should result), we can define its z-norm by

$$\|\dot{z}\|_z := \langle \dot{z}, \dot{z} \rangle_z^{1/2}. \tag{3.1}$$

We call this the *local norm at z*. Then given a piecewise smooth path $\zeta : [0, 1] \to Z$, we define its length by

$$\ell_Z(\zeta) := \int_0^1 \|\dot{\zeta}(t)\|_{\zeta(t)} dt. \tag{3.2}$$

Finally, the distance between z_0 and z_1 in Z is defined to be

$$d_Z(z_0, z_1) := \inf_\zeta \ell_Z(\zeta), \tag{3.3}$$

where the infimum is taken over all piecewise smooth paths ζ from z_0 to z_1, i.e., with $\zeta(i) = z_i$, $i = 0, 1$. We also define a modified metric

$$d'_Z(z_0, z_1) := d_{z(\Lambda)}(z_0, z_1) := \inf_\zeta \ell_Z(\zeta)$$

for points $z_0 = z(\lambda_0)$ and $z_1 = z(\lambda_1)$, where now the infimum is taken over piecewise smooth paths $\zeta : [0, 1] \to z(\Lambda)$ from z_0 to z_1.

Given a point $\bar{z} = z(\lambda) \in Z$, we call $z \in Z$ an *η-approximation* to $\bar{z}$ if either

$$d_Z(z, \bar{z}) \leq \eta \quad \text{or} \quad \|z - \bar{z}\|_{\bar{z}} \leq \eta. \tag{3.4}$$

The second choice is often more convenient, and we shall use it in our application to linear programming, but it requires that $z - \bar{z}$ be a tangent vector at $\bar{z}$; this certainly holds if Z is an open set (as in linear programming), so that the tangent space at any point is the Euclidean space in which Z is embedded. Note that given only z and λ, it may be hard to recognize whether (3.4) holds; we will comment further on this in Section 6.

The key requirement we make on our metric is

Hypothesis 3.1 *There are constants $\epsilon > 0$ and $\eta \geq 0$ with the following property. Suppose $\bar{z} = z(\lambda)$ and $\bar{z}_+ = z(\lambda_+)$ for some $\lambda, \lambda_+ \in \Lambda$, and suppose z is an η-approximation to $\bar{z}$. Then, as long as*

$$d'_Z(\bar{z}, \bar{z}_+) \leq \epsilon, \tag{3.5}$$

the Newton step from z for the system $F(\cdot, \lambda_+) = 0$ is well-defined and yields an η-approximation z_+ to $\bar{z}_+$.

In other words, if we have a good approximation z to the solution vector corresponding to the parameter vector λ, and we adjust the parameter vector so that the solution vector moves a small amount (ϵ in the Riemannian metric), then we can recover a good approximation to the new solution vector by performing a single Newton step.

Verifying that Hypothesis 3.1 holds in a particular setting can be arduous. For now, we merely define the metric for the cases of linear programming, postponing any discussion of its validity until the final section.

3.1 Linear programming

Here $Z = \Re^n_{++} \times \Re^m \times \Re^n_{++}$, and at any $z \in Z$, the tangent space is just $\Re^n \times \Re^m \times \Re^n$. First we define primal and dual norms in both primal (x) and dual (s) spaces. We set

$$\|v\|_x := \|X^{-1}v\|_2, \quad \|u\|^*_x := \|Xu\|_2, \tag{3.6}$$

$$\|u\|_s := \|S^{-1}u\|_2, \quad \|v\|^*_s := \|Sv\|_2. \tag{3.7}$$

Note the simple property that, if $x \in \Re^n_{++}$ and $v \in \Re^n$ with $\|v\|_x < 1$, then $x \pm v \in \Re^n_{++}$, and similarly for s. To define the Riemannian structure at $z =$

$(x, y, s) \in Z$, we first compute $\mu := s^\top x/n$ and $w := x^{1/2}s^{-1/2}, t := s^{1/2}x^{-1/2}$, and then for any $\dot{z}_i = (\dot{x}_i, \dot{y}_i, \dot{s}_i)$ in the tangent space, $i = 1, 2$, we set

$$\langle \dot{z}_1, \dot{z}_2 \rangle_z := \mu^{-1}(\dot{x}_1^\top W^{-2}\dot{x}_2 + \dot{s}_1^\top T^{-2}\dot{s}_2), \tag{3.8}$$

so that

$$\|\dot{z}\|_z = \mu^{-1/2}(\|\dot{x}\|_w^2 + \|\dot{s}\|_w^{*2})^{1/2}, \tag{3.9}$$

where $\dot{z} =: (\dot{x}, \dot{y}, \dot{s})$ (this notation will be implicit from now on). This is not strictly a norm; it becomes one if we eliminate the free variables y, but they are convenient to retain.

To motivate the use of w here, let us note that the norm $\|\cdot\|_x$ and its dual $\|\cdot\|_x^*$ arising from the primal solution x are generally unrelated to the norm $\|\cdot\|_s$ and its dual $\|\cdot\|_s^*$ arising from the dual solution. To effect a compromise, we use the intermediate vector w. Let us recall the logarithmic barrier function

$$\phi(x) := -\sum_{j=1}^n \ln(x^{(j)}), \tag{3.10}$$

where $x^{(j)}$ denotes the jth component of x. Then $\|v\|_x = \langle \phi''(x)v, v\rangle^{1/2}$ and $\|u\|_x^* = \langle u, [\phi''(x)]^{-1}u\rangle^{1/2}$, and similarly for the norms associated with s, w, and t. Note that $\phi''(w)x = s$ (and similarly $\phi''(t)s = x$, with $\phi''(t) = [\phi''(w)]^{-1}$), so that the norms defined by w and t are dual and are symmetric between the primal and the dual. In interior-point terminology, methods based on the norm $\|\cdot\|_x$ are called primal-scaling methods, those using $\|\cdot\|_s$ are called dual-scaling, while those using $\|\cdot\|_w$ are called (symmetric) primal-dual-scaling methods.

In one important special case, these norms are all related. If z is *central*, so that $xs = \mu e$, then $w = \mu^{-1/2}x$ and (3.9) simplifies to

$$\|\dot{z}\|_z = (\|\dot{x}\|_x^2 + \|\dot{s}\|_s^2)^{1/2}. \tag{3.11}$$

Then $\|\dot{z}\|_z < 1$ implies that $\|\dot{x}\|_x$ and $\|\dot{s}\|_s$ are less than 1, so that $z \pm \dot{z} \in Z$. This does not hold in general, but if $xs \geq \theta^2\mu e$ for some $\theta \in (0, 1]$, it is easy to see that $\|\dot{z}\|_z < \theta$ implies that $z \pm \dot{z} \in Z$.

From this local norm, we define the metric as above.

3.2 The metric on Λ

Given a metric on Z, we pull it back to get a metric on Λ as follows. For any $\lambda \in \Lambda$ and $\dot{\lambda}$ in the tangent space to Λ at λ, we let $z = z(\lambda)$ and compute $\dot{z}$ from

$$F_z(z, \lambda)\dot{z} + F_\lambda(z, \lambda)\dot{\lambda} = 0 \tag{3.12}$$

(this uniquely defines $\dot{z}$ from our assumptions below (2.1)). Then we set

$$\|\dot{\lambda}\|_\lambda := \|\dot{z}\|_z; \tag{3.13}$$

we could similarly define $\langle \cdot, \cdot \rangle_\lambda$, but it is not needed. In this way we get a Riemannian structure and hence a metric on Λ.

Let $\gamma : [0, 1] \to \Lambda$ be a smooth path in Λ, and let $\zeta(t) = z(\gamma(t))$ define the corresponding path in Z. Then the implicit function theorem shows that $\dot{\zeta}(t)$ is the $\dot{z}$ corresponding to $\dot{\lambda} = \dot{\gamma}(t)$, and hence from (3.13) that

$$\|\dot{\gamma}(t)\|_{\gamma(t)} = \|\dot{\zeta}(t)\|_{\zeta(t)}.$$

It follows that

$$\ell_\Lambda(\gamma) = \ell_Z(\zeta), \tag{3.14}$$

and hence

$$d_\Lambda(\lambda_0, \lambda_1) = d'_Z(z(\lambda_0), z(\lambda_1)). \tag{3.15}$$

From this we obtain the result which justifies our interest in this metric.

Theorem 3.1 *Given Hypothesis 3.1, there is a path-following method which, from a good approximation to* $z_0 = z(\lambda_0)$*, obtains a good approximation to* $z_1 = z(\lambda_1)$*, and requires*

$$O(d_\Lambda(\lambda_0, \lambda_1)) \tag{3.16}$$

Newton steps.

Proof. Let γ be a path in Λ from λ_0 to λ_1 of length at most $2d_\Lambda(\lambda_0, \lambda_1)$, and let $\zeta(t) := z(\gamma(t))$, so that the length of ζ is also at most $2d_\Lambda(\lambda_0, \lambda_1)$. Now divide ζ into intervals $z_0 = \bar{z}_{(0)}, \bar{z}_{(1)}, \ldots, \bar{z}_{(k)} = z_1$, with $d'_Z(\bar{z}_{(i-1)}, \bar{z}_{(i)}) \leq \epsilon$ and $\bar{z}_{(i)} = z(\lambda_{(i)})$ for $1 \leq i \leq k$. Then $k \leq 2d_\Lambda(\lambda_0, \lambda_1)/\epsilon = O(d_\Lambda(\lambda_0, \lambda_1))$. We suppose we have an η-approximation $z_{(0)}$ to $\bar{z}_{(0)} = z_0$. In general, assuming that we have an η-approximation $z_{(i-1)}$ to $\bar{z}_{(i-1)}$, we can obtain according to Hypothesis 3.1 an η-approximation $z_{(i)}$ to $\bar{z}_{(i)}$ by taking a single Newton step for $F(\cdot, \lambda_{(i)}) = 0$ from $z_{(i-1)}$. Hence, by induction, in k Newton steps we will obtain an η-approximation $z_{(k)}$ to $\bar{z}_{(k)} = z_1$. $\square$

4 The metric on parameter space for linear programming

Let us first consider the general infeasible-interior-point target-following method, so that $\lambda = (b, c, v)$. The corresponding solution is $z = (x, y, s) = z(\lambda)$, at which the local norm is defined by (3.9). Let us denote

$$H = W^{-2} = X^{-1}S,$$

so that (3.9) can be written

$$\|\dot{z}\|_z = \mu^{-1/2}(\dot{x}^\top H\dot{x} + \dot{s}^\top H^{-1}\dot{s})^{1/2}. \tag{4.1}$$

Suppose we are given a displacement $\dot{\lambda} = (\dot{b}, \dot{c}, \dot{v})$ in parameter space. We then compute the corresponding displacement $\dot{z}$ via (3.12), which becomes

$$\begin{array}{rcccl} & A\dot{x} & & = & \dot{b} \\ A^\top \dot{y} & & + \quad \dot{s} & = & \dot{c} \\ & S\dot{x} & + \quad X\dot{s} & = & 2v\dot{v}, \end{array} \tag{4.2}$$

or equivalently

$$\begin{array}{rcccl} & A\dot{x} & & = & \dot{b} \\ A^\top \dot{y} & & + \quad \dot{s} & = & \dot{c} \\ & H\dot{x} & + \quad \dot{s} & = & 2W^{-1}\dot{v}, \end{array} \tag{4.3}$$

For later use, we find the solution $\dot{z}$ to (4.3) where $2W^{-1}\dot{v}$ is replaced with $\dot{g}$.

Let us define

$$J := (AH^{-1}A^\top)^{-1}, \quad Q := A^\top JA, \quad P := H^{-1} - H^{-1}QH^{-1}. \tag{4.4}$$

Then it is easy to check that

$$\begin{array}{l} J, P \text{ and } Q \text{ are symmetric;} \\ AP = 0, \quad AH^{-1}Q = A; \\ PQ = 0, \quad QP = 0; \\ HP + QH^{-1} = H^{1/2}PH^{1/2} + H^{-1/2}QH^{-1/2} = I; \\ PHP = P, \quad QH^{-1}Q = Q. \end{array} \tag{4.5}$$

From these properties, it is straightforward to confirm that the solution to (4.3) (with its third right-hand side replaced by $\dot{g}$) is

$$\begin{array}{l} \dot{x} = P(\dot{g} - \dot{c}) + H^{-1}A^\top J\dot{b}, \\ \dot{y} = -JAH^{-1}(\dot{g} - \dot{c}) + J\dot{b}, \\ \dot{s} = HP\dot{c} + QH^{-1}\dot{g} - A^\top J\dot{b}. \end{array} \tag{4.6}$$

From this we calculate, using (4.5) again,

$$\begin{array}{rcl} \dot{x}^\top H\dot{x} + \dot{s}^\top H^{-1}\dot{s} &=& (\dot{g} - \dot{c})^\top P(\dot{g} - \dot{c}) + \dot{b}^\top J\dot{b} \\ &+& \dot{c}^\top P\dot{c} + \dot{g}^\top H^{-1}QH^{-1}\dot{g} + \dot{b}^\top J\dot{b} - 2\dot{b}^\top JAH^{-1}\dot{g}. \end{array} \tag{4.7}$$

We simplify this expression using the new variable

$$\dot{p} := \dot{c} - QH^{-1}\dot{c} + A^\top J\dot{b}. \tag{4.8}$$

Note that $AH^{-1}\dot{p} = \dot{b}$. Moreover, if we replace $\dot{c}$ in (4.3) by $\dot{p}$ (which differs by a vector in the range of $A^\top$), the only change in the solution occurs in $\dot{y}$. Thus the quantity in (4.7) remains unchanged. Let us therefore replace $\dot{b}$ by $AH^{-1}\dot{p}$ and $\dot{c}$ by $\dot{p}$ in (4.7), to get

$$\begin{array}{rcl} \dot{x}^\top H\dot{x} + \dot{s}^\top H^{-1}\dot{s} &=& (\dot{g} - \dot{p})^\top P(\dot{g} - \dot{p}) + \dot{p}^\top H^{-1}QH^{-1}\dot{p} \\ && +\dot{p}^\top P\dot{p} + \dot{g}^\top H^{-1}QH^{-1}\dot{g} \\ && +\dot{p}^\top H^{-1}QH^{-1}\dot{p} - 2\dot{p}^\top H^{-1}QH^{-1}\dot{g} \\ &=& (\dot{g} - \dot{p})^\top(P + H^{-1}QH^{-1})(\dot{g} - \dot{p}) + \dot{p}^\top(P + H^{-1}QH^{-1})\dot{p} \\ &=& (\dot{g} - \dot{p})^\top H^{-1}(\dot{g} - \dot{p}) + \dot{p}^\top H^{-1}\dot{p}. \end{array}$$

We conclude that

$$\begin{array}{rcl} \|\dot{z}\|_z &=& \mu^{-1/2}(\dot{x}^\top H\dot{x} + \dot{s}^\top H^{-1}\dot{s})^{1/2} \\ &=& \mu^{-1/2}((\dot{g} - \dot{p})^\top H^{-1}(\dot{g} - \dot{p}) + \dot{p}^\top H^{-1}\dot{p})^{1/2}. \end{array} \tag{4.9}$$

The expression (4.9) is useful in the general case of conic programming, but for linear programming it can be further simplified. Note that $\mu := s^\top x/n = e^\top XSe/n = \|v\|_2^2/n$. Also, $H^{-1/2}\dot{g} = 2\dot{v}$. Thus we obtain

$$\|\dot{z}\|_z = \frac{n^{1/2}}{\|v\|_2}(\|2\dot{v} - \dot{q}\|_2^2 + \|\dot{q}\|_2^2)^{1/2}, \tag{4.10}$$

where $\dot{q} := H^{-1/2}\dot{p}$.

We have therefore proved

Theorem 4.1 *For the case of linear programming with parameter space $\Lambda = \{(b, c, v)\}$, the local norm (3.13) defining the metric on Λ is given by*

$$\|(\dot{b}, \dot{c}, \dot{v})\|_{(b,c,v)} = \frac{n^{1/2}}{\|v\|_2}(\|2\dot{v} - \dot{q}\|_2^2 + \|\dot{q}\|_2^2)^{1/2}, \tag{4.11}$$

where $z = (x, y, s) := z(b, c, v)$ and

$$\dot{q} := (I - H^{-1/2}A^\top(AH^{-1}A^\top)^{-1}AH^{-1/2})H^{-1/2}\dot{c} + H^{-1/2}A^\top(AH^{-1}A^\top)^{-1}\dot{b},$$

with $H := X^{-1}S$. □

Let us consider various special cases.

First we suppose we have feasible solutions so that we are addressing feasible-interior-point target-following methods. Then b and c remain fixed as b_1 and c_1, and the natural parameter space is $\Lambda = \{v\} := \Re^n_{++}$. We can identify this with $\{(b_1, c_1, v)\} = \{b_1\} \times \{c_1\} \times \Re^n_{++}$, a subset of the parameter space considered above. To obtain the corresponding metric, we only need to set $\dot{b}$ and $\dot{c}$ to zero in (4.11). Thus $\dot{q}$ is zero and we find

$$\|\dot{v}\|_v = 2n^{1/2}\frac{\|\dot{v}\|_2}{\|v\|_2}. \tag{4.12}$$

We will determine shortest path geodesics corresponding to the metric corresponding to (4.12) in the next section.

Second, let us return to the infeasible case but suppose that we are concerned with infeasible-interior-point central-path-following methods. Then the natural parameter space is $\Lambda := \{(b, c, \mu)\}$ but we can identify this with the subset $\{(b, c, \mu^{1/2}e)\}$ of our space $\{(b, c, v)\}$ above. Corresponding to a displacement $\dot{\lambda} = (\dot{b}, \dot{c}, \dot{\mu})$ in Λ we have the displacement $(\dot{b}, \dot{c}, \dot{v})$ where $\dot{v} := \frac{1}{2}\mu^{-1/2}\dot{\mu}e$. We thus obtain the appropriate metric from the local norm given by

$$\|(\dot{b}, \dot{c}, \dot{\mu})\|_{(b,c,\mu)} = \mu^{-1/2}(\|\mu^{-1/2}\dot{\mu}e - \dot{q}\|_2^2 + \|\dot{q}\|_2^2)^{1/2}, \tag{4.13}$$

where $\dot{q}$ is defined below (4.11).

Finally, we combine these two cases to consider feasible-interior-point central-path-following methods. Either by replacing v and $\dot{v}$ in (4.12) as in the previous paragraph, or by setting $\dot{q} = 0$ in (4.13), we find that for $\Lambda = \{\mu\} := \Re_{++}$, the appropriate metric comes from the local norm

$$\|\dot{\mu}\|_\mu = n^{1/2}\mu^{-1}|\dot{\mu}|. \tag{4.14}$$

Again, the next section gives the corresponding distances and shortest path geodesics for this case.

We note that only in the feasible cases are the local norms given in closed form in terms of the parameters alone, as in (4.12) and (4.14). In the more general cases, the local norm also involves $\dot{q}$, which is a function of the parameter changes $\dot{b}$ and $\dot{c}$, but uses projections depending on the corresponding solution vector z. It is for this reason that determining the corresponding distances and shortest path geodesics is in general intractable for the infeasible case, and we can only obtain insight from very special cases.

5 Shortest paths and examples

This section calculates some shortest path geodesics for the metrics derived in the previous section.

We start with the simplest case: feasible-interior-point central-path-following methods, where $\Lambda = \{\mu\} := \Re_{++}$. In this case the metric is given by the local norm

$$\|\dot{\mu}\|_\mu = n^{1/2}\mu^{-1}|\dot{\mu}|$$

by (4.14). Consider the mapping $\mu \to n^{1/2}\ln\mu$ from Λ to $\Re$. We see that this is an isometry between Λ with the metric given above and $\Re$ with the usual Euclidean metric. It follows that the shortest path geodesic between μ_0 and $\mu_1 < \mu_0$ in Λ is just the segment $[\mu_1, \mu_0]$ and, less trivially, that

$$d_\Lambda(\mu_0, \mu_1) = n^{1/2}|\ln(\mu_0/\mu_1)|. \tag{5.1}$$

Thus as long as Hypothesis 3.1 is true, Theorem 3.1 shows that

$$O(n^{1/2}|\ln(\mu_0/\mu_1)|) \tag{5.2}$$

iterations are sufficient to move from an approximate center corresponding to the parameter μ_0 to that corresponding to μ_1. This agrees with the complexity bounds in the usual analyses; see Monteiro-Adler [13] and Kojima-Mizuno-Yoshise [9].

Now we turn to feasible target-following methods. Thus $\Lambda = \{v\} := \Re^n_{++}$, and the metric comes from the local norm given by

$$\|\dot{v}\|_v = 2n^{1/2}\frac{\|\dot{v}\|_2}{\|v\|_2} \tag{5.3}$$

according to (4.12). Again we seek an isometry to a subset of Euclidean space. Consider the mapping $v \to (\rho, u) := (\ln \|v\|, v/\|v\|_2) \in \Re \times S^n_{++}$, where $S^n_{++} := \{u \in \Re^n_{++} : \|u\|_2 = 1\}$. We find $v = \exp(\rho)u$, so $\dot{v} = \exp(\rho)[\dot{\rho}u + \dot{u}]$. Since $u^\top u \equiv 1$ implies $u^\top \dot{u} = 0$, we have $\|\dot{v}\|_2 = \exp(\rho)\|(\dot{\rho}, \dot{u})\|_2$, so

$$\|\dot{v}\|_v = 2n^{1/2}\|(\dot{\rho}, \dot{u})\|_2. \tag{5.4}$$

Shortest path geodesics then correspond in (ρ, u)-space to moving at uniform speed from ρ_0 to ρ_1 and at uniform speed along the great circle from u_0 to u_1. We find

$$d_\Lambda(v_0, v_1) = 2n^{1/2}\left[(\ln \frac{\|v_0\|_2}{\|v_1\|_2})^2 + (\arccos[\frac{v_0^\top v_1}{\|v_0\|_2\|v_1\|_2}])^2\right]^{1/2}. \tag{5.5}$$

For example, the shortest path geodesic from $v_0 \in \Lambda$ to some point v_1 with $\|v_1\| = \epsilon$ is the straight line segment from v_0 to $v_1 := \epsilon v_0/\|v_0\|_2$. This corresponds to following a weighted path, see, e.g., Ding and Li [3].

Let us observe that these geodesics differ from the paths recommended in Jansen-Roos-Terlaky-Vial [5, 6]; their paths always become more centered, while $v_1 = \epsilon v_0/\|v_0\|_2$ implies that ours maintain the same degree of centrality. The main reason for this discrepancy is that Hypothesis 3.1 does not hold generally in this case. It is necessary to restrict Λ to triples (b, c, v) where $v > \theta\mu^{1/2}e = \theta\|v\|_2 n^{-1/2}e$ for some fixed $\theta \in (0, 1]$, and then the hypothesis holds with η and ϵ depending strongly on θ.

Jansen et al. define a distance $\delta(v, \bar{v})$ which, for infinitesimally close points, corresponds to the local norm

$$\|\dot{v}\|'_v := \frac{\|\dot{v}\|}{\min(v)}, \tag{5.6}$$

where $\min(v) := \min(v^{(j)})$, the smallest component of v. The same norm is implicit in the second neighborhood used by Mizuno [11]. This differs by at most a multiplicative constant from $\|\dot{v}\|_v$ in (5.3) as long as $v \geq \theta\mu^{1/2}e$, but the constant depends strongly on θ. Section 2 of [11] and Section 3 of each of [5, 6] show the appropriateness of the measure (5.6); roughly, if v is moved by a small (respectively moderate) distance according to this measure, a single Newton step (bounded number of steps) will yield a good approximation.

The local norm (5.6) "corresponds" to the local norm in solution space that differs from (3.9) in that $\mu^{-1/2}$ is replaced by $(\min(v))^{-1}$, where $v :=$

$x^{1/2}s^{1/2}$. The reason for the quotes above, and the reason we did not use this local norm, is that it is not smooth, but only piecewise smooth, so that we do not obtain Riemannian metrics. However, it is quite possible, based on the results of Jansen et al., that this is a more appropriate, if less smooth, metric for the target-following case. Note that the metrics coincide on the central path, corresponding to choosing $\theta = 1$ so that $\min(v) = \mu^{1/2}$.

Finally, we consider an infeasible case. Here we only address a particular instance.

Example 5.1 Let $A = [I, 0] \in \Re^{m\times n}$, and similarly partition the vectors $c = (c_f; c_u)$, $x = (x_f; x_u)$, and $s = (s_f; s_u)$, so that our problems are

$$(LP) \qquad \begin{array}{rccccl} \min & c_f^\top x_f & + & c_u^\top x_u & & \\ & x_f & & & = & b \\ & x_f, & & x_u & \geq & 0, \end{array}$$

and

$$(LD) \qquad \begin{array}{rccccl} \max & b^\top y & & & & \\ & y & + & s_f & = & c_f \\ & & & s_u & = & c_u \\ & & & s_f, \; s_u & \geq & 0. \end{array}$$

(For the primal, subscript "f" denotes fixed, while "u" denotes unconstrained.) We will just consider central-path-following methods, so that our parameter space is $\Lambda = \{(b, c, \mu)\} := \Re^m_{++} \times (\Re^m \times \Re^{n-m}_{++}) \times \Re_{++}$. We find, for $\lambda = (b, c, \mu) \in \Lambda$, $z(\lambda) = (x, y, s)$ where

$$x_f = b, \; x_u = \mu c_u^{-1}, \; y = c_f - \mu b^{-1}, \; s_f = \mu b^{-1}, \; s_u = c_u.$$

Thus $H = \mathrm{Diag}(\mu b^{-2}, \mu^{-1} c_u^2)$, $J = \mathrm{Diag}(\mu b^{-2})$, $Q = \mathrm{Diag}(\mu b^{-2}, 0)$, and $P = \mathrm{Diag}(0, \mu c_u^{-2})$. Thus

$$\dot{p} = \begin{pmatrix} \dot{c}_f \\ \dot{c}_u \end{pmatrix} - \begin{pmatrix} \dot{c}_f \\ 0 \end{pmatrix} + \begin{pmatrix} \mu b^{-2}\dot{b} \\ 0 \end{pmatrix} = \begin{pmatrix} \mu b^{-2}\dot{b} \\ \dot{c}_u \end{pmatrix}$$

and $\dot{q} = \begin{pmatrix} \mu^{1/2} b^{-1}\dot{b} \\ \mu^{1/2} c_u^{-1}\dot{c}_u \end{pmatrix}$. Now we use (4.13) to obtain

$$\|(\dot{b}, \dot{c}, \dot{\mu})\|_{(b,c,\mu)} = \left(\left\| \begin{pmatrix} \mu^{-1}\dot{\mu} e_f - b^{-1}\dot{b} \\ \mu^{-1}\dot{\mu} e_u - c_u^{-1}\dot{c}_u \end{pmatrix} \right\|_2^2 + \left\| \begin{pmatrix} b^{-1}\dot{b} \\ c_u^{-1}\dot{c}_u \end{pmatrix} \right\|_2^2 \right)^{1/2}.$$

(Note again that this is not strictly a metric, since c_f is not involved, as we should expect; c_f only affects the free variable y. It *is* a metric when we restrict to the subvector (b, c_u, μ).)

Now consider the map $(b, c, \mu) \to (\ln b, \ln c_u, \ln \mu)$ (componentwise). Then our metric above is induced by a fixed ellipsoidal metric in the log-space. Hence shortest path geodesics in Λ correspond to straight lines in $(\ln b, \ln c_u, \ln \mu)$-space, and

$$d_\Lambda((b_0, c_0, \mu_0), (b_1, c_1, \mu_1)) = \left\| \begin{pmatrix} \ln(\mu_0/\mu_1)e_f - \ln(b_0 b_1^{-1}) \\ \ln(\mu_0/\mu_1)e_u - \ln(c_{0u} c_{1u}^{-1}) \\ \ln(b_0 b_1^{-1}) \\ \ln(c_{0u} c_{1u}^{-1}) \end{pmatrix} \right\|_2 .$$

We remark that these geodesics are quite different from straight lines in (b, c, μ)-space, as used by most infeasible-interior-point methods. One consequence is that we can obtain better complexity bounds: if

$$2^{-L} e_f \le b_i \le 2^L e_f \text{ and } 2^{-L} e_u \le c_{iu} \le 2^L e_u$$

for $i = 0, 1$, then

$$d_\Lambda((b_0, c_0, \mu_0), (b_1, c_1, \mu_1)) = O(\sqrt{n}[\ln(\mu_0/\mu_1) + L]),$$

and the number of iterations of a central-path following method is of the same order, whereas many infeasible-interior-point methods replace the $\sqrt{n}$ with n^α for $\alpha \ge 1$. We hasten to add that this is just one instance, which is trivial to solve directly; but it may suggest the value of trying to obtain geodesics for more general problems.

Finally, consider the extremely trivial case with $m = n = 1$, so that b is one-dimensional and c_u disappears. Then shortest path geodesics are straight lines in $(\ln b, \ln \mu)$ space. If b moves through a smaller multiplicative range than μ, these geodesics have the property that feasibility is attained at a slower rate than optimality, in strong contradiction to the conventional wisdom that the reverse should be the case.

6 Justification of the metric

Here we will discuss Hypothesis 3.1 for linear programming and also consider two other aspects of our approach: how to recognize good approximations and whether good approximations yield acceptable solutions to the original problems. Almost all proofs will be omitted, and we concentrate on central-path-following methods.

We will use the second condition for η-approximations, i.e.,

$$\|z - \bar{z}\|_{\bar{z}} \leq \eta \tag{6.1}$$

from (3.4), in confirming Hypothesis 3.1. This raises the natural question: given only $z \in Z$ and $\lambda \in \Lambda$, how can we check (6.1) for $\bar{z} = z(\lambda)$? In fact, for the case $\lambda = \mu$ or $\lambda = (b, c, \mu)$ this is straightforward. Suppose $Ax = b$, $A^\top y + s = c$, and $\mu = s^\top x/n$. Then for $\bar{z} = z(\lambda)$ it can be shown that

$$\|\frac{xs}{\mu} - e\|_2 \leq \delta \leq .1 \text{ implies } \begin{cases} \|\bar{x} - x\|_x \leq 3\delta, \ \|\bar{s} - s\|_s \leq 3\delta \\ \|x - \bar{x}\|_{\bar{x}} \leq 9\delta/2, \ \|s - \bar{s}\|_{\bar{s}} \leq 9\delta/2 \\ \|z - \bar{z}\|_{\bar{z}} \leq 7\delta. \end{cases} \tag{6.2}$$

Note that the condition on the left-hand side can easily be checked, and that if it holds for $\delta \leq \min\{\eta/7, .1\}$ we know that (6.1) holds.

Next we address whether having a good approximation z to an acceptable solution $\bar{z}$ suffices. Since we are using a Newton step at each iteration, linear constraints will be satisfied exactly, so that if $A\bar{x} = b_1$ and $A^\top \bar{y} + \bar{s} = c_1$, the same will be true for x and (y, s). Also, (6.1) will assure that $x > 0$ and $s > 0$ for sufficiently small η, so feasibility is assured. The only remaining concern is the duality gap $s^\top x$. We find

$$\begin{aligned} s^\top x &= [\bar{s} + (s - \bar{s})]^\top [\bar{x} + (x - \bar{x})] \\ &\leq \bar{s}^\top \bar{x} + \|s - \bar{s}\|_{\bar{w}}^* \|\bar{x}\|_{\bar{w}} + \|\bar{s}\|_{\bar{w}}^* \|x - \bar{x}\|_{\bar{w}} + \|s - \bar{s}\|_{\bar{w}}^* \|x - \bar{x}\|_{\bar{w}}, \end{aligned}$$

where $\bar{w} := \bar{x}^{1/2}\bar{s}^{-1/2}$. But (6.1) ensures that $\|x - \bar{x}\|_{\bar{w}}$ and $\|s - \bar{s}\|_{\bar{w}}^*$ are at most $\mu^{1/2}\eta$, where $\mu := \bar{s}^\top \bar{x}/n$, and it is not hard to show that $\|\bar{x}\|_{\bar{w}} = \|\bar{s}\|_{\bar{w}}^* = \mu^{1/2}n^{1/2}$, so that

$$s^\top x \leq 1.3\bar{s}^\top \bar{x} \tag{6.3}$$

as long as $\eta \leq .1$.

The fundamental reason that Hypothesis 3.1 holds is the quadratic convergence of Newton's method, but we need explicit constants and also many applications of the fact that norms evaluated at neighboring points are close. The basic property here is that

$$\|x_+ - x\|_x \leq \delta < 1 \text{ implies } \|v\|_{x_+} \leq (1 - \delta)^{-1}\|v\|_x \tag{6.4}$$

for any $x, x_+ \in \Re^n_{++}$ and $v \in \Re^n$. This is trivial to show directly, since

$$\|v\|_{x_+} = \|(X_+^{-1}X)(X^{-1}v)\|_2 \leq \|X_+^{-1}X\|_2 \|X^{-1}v\|_2.$$

Note that from (6.4) the second line of implications in (6.2) follows from the first, since $(1-3\delta)^{-1}3\delta \leq (.7)^{-1}3\delta \leq 9\delta/2$. (The third line then follows directly from the definition (3.11).)

Finally, let us state two results which are key ingredients in establishing Hypothesis 3.1.

Lemma 6.1 *Let $x, x_+ \in \Re^n_{++}$ with $\delta := \|x_+ - x\|_x < 1$. Then*

$$\| - x_+^{-1} + x^{-1} - X^{-2}(x_+ - x)\|_x^* \leq \frac{\delta^2}{1-\delta}. \tag{6.5}$$

□

Note that the norm on the left-hand side can also be written in terms of the logarithmic barrier function ϕ defined in (3.10): it becomes

$$\|\phi'(x_+) - \phi'(x) - \phi''(x)(x_+ - x)\|_x^*.$$

This makes it clear that the lemma is bounding the error in the first-order Taylor approximation to $\phi'(x_+)$.

The next result refers to the points appearing in Hypothesis 3.1, and assumes a central-path-following method, so that $\lambda = \mu$ or $\lambda = (b, c, \mu)$.

Lemma 6.2 *If $z, \bar{z}_+$, and z_+ are as in Hypothesis 3.1, with $\bar{z}_+ = z(\mu_+)$ or $z(b, c, \mu_+)$, and $H = W^{-2}$ with $w = x^{1/2}s^{-1/2}$, then*

$$\|x_+ - \bar{x}_+\|_w^2 + \|s_+ - \bar{s}_+\|_w^{*2} = \|\mu_+(-\bar{x}_+^{-1} + x^{-1}) - H(\bar{x}_+ - x)\|_w^{*2}. \tag{6.6}$$

□

Note that the left-hand side of (6.6) is closely related to $\|z_+ - \bar{z}_+\|_{\bar{z}_+}$ (except that the norm is wrong), while its right-hand side is closely related to the quantity bounded in (6.5) (except that the norm is wrong and H/μ_+ replaces X^{-2}). Putting all these pieces together enables one to prove

Theorem 6.1 *For (feasible or infeasible) central-path-following methods (with $\lambda = \mu$ or $\lambda = (b, c, \mu)$), Hypothesis 3.1 holds with $\epsilon = \eta = .04$.* □

We briefly mention, as we have hinted above, that in the case of target-following methods it is necessary to restrict v so that $v \geq \theta \|v\|_2 n^{-1/2} e$, where $\theta \in (0,1]$ is a constant. Thus $\Lambda = \{v\} := \{v \in \Re^n_{++} : v \geq \theta \|v\|_2 n^{-1/2} e\}$ or $\Lambda = \{(b,c,v)\} := A(\Re^n_{++}) \times (A^\top(\Re^m) + \Re^n_{++}) \times \{v \in \Re^n_{++} : v \geq \theta \|v\| n^{-1/2} e\}$. With this restriction, Hypothesis 3.1 holds with $\epsilon = \eta = .04\theta$; note the unpleasant dependence on θ.

References

[1] W. M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry.* Academic Press, New York, 1986.

[2] W. C. Davidon. Variable metric methods for minimization. Report ANL-5990, Argonne National Laboratories, Argonne, IL, 1959, reprinted in *SIAM Journal on Optimization*, 1:1–17, 1991..

[3] J. Ding and T. Y. Li. An algorithm based on weighted logarithmic barrier functions for linear complementarity problems. *Arabian Journal for Science and Engineering*, 15(4):769–685, 1990.

[4] R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. *Computer Journal*, 6:163–168, 1963.

[5] B. Jansen, C. Roos, T. Terlaky, and J. P. Vial. Primal-dual target-following algorithms for linear programming. Technical Report 93–107, Faculty of Technical Mathematics and Informatics, TU Delft, NL–2600 GA Delft, The Netherlands, November 1993.

[6] B. Jansen, C. Roos, T. Terlaky, and J. P. Vial. Long-step primal-dual target-following algorithms for linear programming. *Zeitschrift für Operations Research — Mathematical Methods of Operations Research*, 44(1), 1996.

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

[8] M. Kojima, S. Mizuno, and A. Yoshise. A primal–dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in*

Mathematical Programming : Interior Point and Related Methods, pages 29–47. Springer Verlag, New York, 1989.

[9] M. Kojima, S. Mizuno, and A. Yoshise. A polynomial–time algorithm for a class of linear complementarity problems. *Mathematical Programming*, 44:1–26, 1989.

[10] C. E. Lemke. Bimatrix equilibrium points and mathematical programming. *Management Science*, 11:681–689, 1965.

[11] S. Mizuno. A new polynomial time method for a linear complementarity problem. *Mathematical Programming*, 56:31–43, 1992.

[12] S. Mizuno, M. J. Todd, and Y. Ye. A surface of analytic centers and primal–dual infeasible–interior–point algorithms for linear programming. *Mathematics of Operations Research*, 20:135–162, 1995.

[13] R. D. C. Monteiro and I. Adler. Interior path following primal–dual algorithms : Part I : Linear programming. *Mathematical Programming*, 44:27–41, 1989.

[14] Yu. E. Nesterov and A. S. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming : Theory and Algorithms*. SIAM Publications. SIAM, Philadelphia, USA, 1993.

[15] Yu. E. Nesterov and A. S. Nemirovskii. Multi-parameter surfaces of analytic centers and long-step surface-following interior point methods. Research Report 3/95, Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel.

[16] Yu. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. Technical Report No. 1091, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA, 1994. To appear in *Mathematics of Operations Research*.

[17] Yu. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. Technical Report No. 1125, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853-3801, USA, 1995.

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

[19] M. J. D. Powell. The complexity of Karmarkar's algorithm for linear programming. In D. F. Griffiths and G. A. Watson, editors, *Numerical Analysis 1991*, volume 260 of *Pitman Research Notes in Mathematics*, pages 142–163. Longman, Burnt Hill, UK, 1992.

[20] M. J. D. Powell. On the number of iterations of Karmarkar's algorithm for linear programming. *Mathematical Programming*, 62:153–197, 1993.

[21] M. J. D. Powell. Some convergence properties of the modified log barrier method for linear programming. *SIAM Journal on Optimization*, 5:695–739, 1995.