CONVERGENCE OF INTERIOR POINT ALGORITHMS FOR THE MONOTONE LINEAR COMPLEMENTARITY PROBLEM

J. FRÉDÉRIC BONNANS AND CLOVIS C. GONZAGA

The literature on interior point algorithms shows impressive results related to the speed of convergence of the objective values, but very little is known about the convergence of the iterate sequences.

This paper studies the horizontal linear complementarity problem, and derives general convergence properties for algorithms based on Newton iterations. This problem provides a simple and general framework for most existing primal-dual interior point methods. The conclusion is that most of the published algorithms of this kind generate convergent sequences. In many cases (whenever the convergence is not too fast in a certain sense), the sequences converge to the analytic center of the optimal face.

1. **Introduction.** Among several different statements of the linear complementarity problem, we chose the following one, known as the monotone horizontal linear complementarity problem:

$$
\begin{array}{ll}
\text{minimize} & s^T x \\
\text{subject to} & Qx + Rs = b \\
& x, s \geq 0,
\end{array}
$$

(P)

where $b \in \mathbb{R}^m$, and $Q, R \in \mathbb{R}^{m \times n}$ are such that for any $u, v \in \mathbb{R}^n$,

$$
\text{if } Qu + Rv = 0 \quad \text{then } u^T v \geq 0.
$$

The problem will be formally studied in §2, where state two hypotheses: the existence of an interior feasible solution, and the existence of a strictly complementary optimal solution. As we shall see, this problem trivially includes the linear programming problem and the convex quadratic programming problem in their usual formulations, and thus provides a quite general framework for the study of algorithms. The algorithms studied in this paper are restricted to feasible interior point methods.

Rather than developing new methods, the scope of this paper is a unified study of the iterate convergence of methods for which an R-linear convergence of the objective values $x^T s$ is already guaranteed. We describe the usual variants of Newton iterations for the perturbed version of this problem (exact Newton steps, centering steps and affine-scaling steps) and concentrate our attention on methods based on combinations of these steps. We conclude that among the methods found in the literature, the following ones generate convergent iterate sequences:

• All the path following methods that work in a Euclidean norm neighborhood of the central path. Predictor-corrector algorithms (already studied by Gonzaga and

Tapia (1992) in the linear programming case) and linearly convergent methods generate iterates that converge linearly to the analytic center of the optimal face. One step superlinearly convergent methods that use only exact Newton iterates may converge to another optimal solution. In particular, all the known primal-dual algorithms with a complexity of $O(\sqrt{n} L)$ iterations generate convergent sequences.

• All methods for which superlinear convergence of the objective values has been proved. If the convergence in values is very fast (say, quadratic), the iterates usually converge to a noncentral optimal solution. Otherwise, the iterates converge linearly to the analytic center of the optimal face.

Among the methods that we have considered, the only ones that may generate nonconvergent sequences are the ones that do not follow the central path closely and do not achieve superlinear convergence.

Apart from this study of convergence, this paper has two other contributions that may be of interest:

• We formulate the problem carefully, and show that by a simple change in the order of the variables the properties of Newton steps become very clear. The least-squares problems associated with the systems, as well as the relationship between "large" and "small" variables become straightforward.

• We develop in §4 a general convergence theorem for nonlinear programming algorithms that may be useful in other contexts.

The algorithms to which these results apply cover the whole history of primal-dual interior point methods. The first algorithm was introduced by Kojima, Mizuno and Yoshise (1989), based on a work by Megiddo (1989). The same authors (1987) and independently Monteiro and Adler (1989), developed a low-complexity version of the algorithm. Kojima, Megiddo, Noma and Yoshise (1991) in their monograph make an extensive treatment of methods for linear complementarity problems that, like the Kojima-Mizuno-Yoshise algorithm, use exact Newton iterations followed or not by line searches. The largest possible step for a path following algorithm using complete Newton steps is described by McShane (1991) for linear programming and for linear complementarity (McShane (1994)). He proves superlinear convergence under the hypothesis that the iterate sequence converges: this hypothesis, as a consequence of our results, is naturally satisfied.

Another consequence that will be discussed ahead is that the extension of the large step primal path following methods developed by Gonzaga (1991) and by Roos and Vial (1989) to the primal-dual setting is always convergent to the central optimum.

Primal-dual algorithms for linear programming with superlinear convergence were developed by Zhang and Tapia (1991), without following the central path. Algorithms that follow the central path and achieve 2-step quadratic convergence are based on the Mizuno-Todd-Ye (1990) algorithm. This rate of convergence was first proved by Ye, Tapia and Zhang (1991), with the hypothesis of convergence of the sequence of iterates, and afterwards by Ye, Güler, Tapia and Zhang (1993) and by Mehrotra (1991) without this hypothesis.

The extension of these results to linear complementarity problems was done by Kojima, Kurita and Mizuno (1991), Kojima, Megiddo and Noma (1989), Ji, Potra, Tapia and Zhang (1991), Ji, Potra and Huang (1995), either using a nondegeneracy hypothesis or assuming that the sequences converge. The result without these hypotheses was found by Ye and Anstreicher (1993), who show convergence of order 2 for the objective values and low polynomial complexity.

The study of interior point algorithms is greatly simplified by the description of properties of the so-called w-weighted trajectories associated with the problems. These trajectories for the horizontal linear complementarity problem are described

by Monteiro and Tsuchiya (1992), extending results by Kojima, Mizuno and Noma (1990).

The convergence of the sequences of iterates for the Mizuno-Todd-Ye algorithm for linear programming was studied by Gonzaga and Tapia (1992). The present paper extends the results in this references to the horizontal linear complementarity problem, and studies algorithms that are not of the predictor-corrector type.

The paper is structured as follows: §2 describes the LCP and its main properties. Section 3 describes the class of algorithms under consideration and states the main results. Section 4 is a self-contained convergence study of nonlinear programming descent algorithms. Sections 5 and 6 contain the mathematical treatment of Newton methods, and the proofs of the results described in §3.

Conventions. Given a vector x, d, the corresponding upper case symbol denotes as usual the diagonal matrix X, D defined by the vector. The symbol e will represent the vector of all ones, with dimension given by the context.

We shall denote component-wise operations on vectors by the usual notations for real numbers. Thus, given two vectors u, v of the same dimension, $uv, u/v$, etc. will denote the vectors with components $u_i v_i, u_i/v_i$, etc. This notation is consistent as long as component-wise operations always have precedence in relation to matrix operations. Note that $uv \equiv Uv$ and if A is a matrix, then $Auv \equiv AUv$, but in general $Auv \neq (AU)v$.

We shall frequently use the $O(\cdot)$ and $\Omega(\cdot)$ notation to express the relationship between functions. Our most common usage will be associated with a sequence $\{x^k\}$ of vectors and a sequence $\{\mu_k\}$ of positive real numbers. In this case $x^k = O(\mu_k)$ means that there is a constant K (dependent on problem data) such that for every $k \in \mathbb{N}$, $\|x^k\| \leq \mu_k$. Similarly, if $x^k > 0$, $x^k = \Omega(\mu_k)$ means that $(x^k)^{-1} = O(1/\mu_k)$. Finally, $x^k \approx \mu_k$ means that $x^k = O(\mu_k)$ and $x^k = \Omega(\mu_k)$.

We use the same notations for a point x in a set parameterized by μ, say E_μ. We say that $x = O(\mu)$ (resp. $x = \Omega(\mu)$, $x \approx \mu$) whenever there is a constant K such that (for μ small enough) for all $x \in E_\mu$, $\|x\| \leq K\mu$ (resp. $x^{-1} = O(1/\mu)$, $x \approx \mu$). In particular, $x \approx 1$ in E means that there are constants $K_2 > K_1 > 0$, such that any $x \in E$ satisfies $K_1 \leq x_i \leq K_2$, $i = 1, \ldots, n$.

Given two vector functions x and y, $x \approx y$ means that $x_i \approx y_i$ for $i = 1, \ldots, n$.

2. The linear complementarity problem. The problem (P) can be stated in the following format, which will be used in the paper: Solve

$$xs = 0,$$

(P)
$$Qx + Rs = b,$$

$$x, s \geq 0,$$

where $b \in \mathbb{R}^n$, and $Q, R \in \mathbb{R}^{n \times n}$ are such that for any $u, v \in \mathbb{R}^n$,

$$\text{if } Qu + Rv = 0 \quad \text{then } u^T v \geq 0.$$

The feasible set for (P) is

$$F = \{(x, s) \in \mathbb{R}^{2n} | Qx + Rs = b, x, s \geq 0\},$$

and the set of interior solutions is

$$F^0 = \{(x, s) \in F \,|\, xs > 0\}.$$

We say that respectively x or s is feasible if there exists s or x such that $(x, s) \in F$.

REMARK. We shall prove below that $[Q\ R]$ has rank n under the monotonicity hypothesis. One can use $Q, R \in \mathbb{R}^{m \times n}$ with $m > n$, as long as $r([Q\ R]) = n$.

EXAMPLE: THE QUADRATIC PROGRAMMING PROBLEM. This format is quite general. For instance, the convex quadratic programming problem is*

$$
\begin{aligned}
\text{minimize} \quad & c^T x + \tfrac{1}{2} x^T H x \\
\text{subject to} \quad & Ax = b \\
& x \geq 0,
\end{aligned}
$$

where $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $H \in \mathbb{R}^{n \times n}$ is a positive semi-definite matrix. The necessary and sufficient optimality conditions for this problem are

$$x^T s = 0,$$

$$-Hx + A^T w + s = c,$$

$$Ax = b,$$

$$x, s \geq 0.$$

Let B be a matrix whose rows span the null space of A. Multiplying the second equation by B, one obtains the equivalent relation $-BHx + Bs = Bc$, so that (x, w, s) satisfies the first-order optimality system iff

$$x^T s = 0,$$

$$
\begin{bmatrix} A & 0 \\ -BH & B \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} b \\ Bc \end{bmatrix},
$$

$$x, s \geq 0.$$

Now let $u, v \in \mathbb{R}^n$ be such that $Au = 0$ and $-Hu + A^T w + v = 0$. Multiplying this equation by u^T, we obtain $u^T v = u^T H u \geq 0$, and conclude that the optimality conditions constitute a monotone linear complementarity problem. This is also trivially true for the linear programming problem, where $H = 0$.

The optimal face. We shall use several properties of problem (P), proved by Monteiro and Tsuchiya in (1992). The set of optimal solutions of (P) is a face of the polyhedron F, denoted $\mathscr{F}$.

The face is characterized by a partition $\{B, N, T\}$ of the index set, called optimal partition, such that $i \in T$ if $x_i = 0$ and $s_i = 0$ for all optimal solutions; $i \in B$ or $i \in N$ respectively if there exists an optimal solution (x, s) such that $x_i > 0$ or $s_i > 0$.

In this paper we shall study separately the behavior of the so-called large variables x_B, s_N and small variables x_N, s_B. A great simplification is obtained by the following assumptions. Assumption 1 is necessary for the construction of feasible interior

---

* The notation in the example is local, and shall not be used in the remainder of the paper.

algorithms. Assumption 2 is the strict complementarity hypothesis. Under Assumptions 1 and 2, the optimal face is bounded and has a relative interior $\mathscr{F}^0 \neq \varnothing$. Assumption 3 may look peculiar, but we explain below how it can be made without loss of generality.

ASSUMPTION 1. $F^0 \neq \varnothing$.

ASSUMPTION 2. $T = \varnothing$.

ASSUMPTION 3. $N = \varnothing$.

Ensuring assumption 3 by a reordering of the variables. Assuming that $T = \varnothing$, Assumption 3 means that x is the vector of large variables, s is the vector of small variables, so that for any optimal solution $x \geq 0$ and $s = 0$. This situation is easily obtained by a reordering of the variables, as we show now. The constraint $Qx + Rs = b$ can be rewritten as

$$[Q_B \quad R_N]\begin{bmatrix} x_B \\ s_N \end{bmatrix} + [R_B \quad Q_N]\begin{bmatrix} s_B \\ x_N \end{bmatrix} = b.$$

We can now rename the variables in the following sequence:

$$Q \leftarrow [Q_B \quad R_N], \qquad R \leftarrow [R_B \quad Q_N],$$

$$x \leftarrow \begin{bmatrix} x_B \\ s_N \end{bmatrix}, \qquad s \leftarrow \begin{bmatrix} s_B \\ x_N \end{bmatrix},$$

$$N \leftarrow \varnothing, \qquad B \leftarrow \{1, \ldots, n\}.$$

With this reordering, the optimal face is characterized simply by

$$\mathscr{F} = \{(x, s) \in \mathbb{R}^{2n} | s = 0, Qx = b, x \geq 0\}.$$

We note that monotonicity is not affected by reordering. In addition, the Newton directions as well as the neighborhood of the central path (defined below) are invariant with respect to this transformation.

This means that all algorithms based on the Newton step and the proximity of the central path are invariant with respect to permutation of variables. Of course the algorithms never use the knowledge of the optimal partition, which is unknown.

The standard form. Our treatment will be done for the horizontal form of LCP, i.e., problem (P). This form is obviously at least as general as the standard form. However, it is instructive to observe that a simple transformation reduces the horizontal form to the standard form, so that the two classes of problems are indeed equivalent. This has been recently shown (Güler (1992)), and we give here an elementary proof.

The standard form of linear complementarity problems is

(SP)
$$\begin{aligned} \text{minimize} \quad & s^T x \\ \text{subject to} \quad & s = Mx + q, \\ & x, s \geq 0. \end{aligned}$$

If the matrix R is $n \times n$ and invertible, then reduction of (P) to the standard form is trivial, simply take $M := -R^{-1}Q$ and $q := R^{-1}q$. We now prove that a reordering of variables can always be done that reduces to the case of an invertible R.

Let r be the rank of Q. By a convenient ordering of the rows and columns of Q, we can factorize it as a product $Q = LU$, with L invertible, the last $n - r$ lines of U being identically zero, and with an invertible block U_1 as follows:

$$Q = L \begin{pmatrix} U_1 & U_2 \\ 0 & 0 \end{pmatrix}.$$

Denote $D := L^{-1}R$. By u^1, u^2 we denote the components of u in the first r and $n - r$ indices, respectively, and similarly for v. The homogeneous relation $Qu + Rv = 0$ is equivalent to:

$$\begin{cases} U_1 u^1 + U_2 u^2 + D_{11} v^1 + D_{12} v^2 = 0, \\ \qquad\qquad\quad D_{21} v^1 + D_{22} v^2 = 0. \end{cases}$$

We now prove that D_{22} is invertible. Choose $v^1 = 0$, v^2 arbitrarily in $\mathcal{N}(D_{22})$, and $u^2 = -v^2$. Since U_1 is invertible, the above linear system has a solution, with u^2, v^1, v^2 as specified, and by monotonicity $0 \leq u^T v = -|v^2\|^2$. It follows that $v^2 = 0$, i.e., the square matrix D_{22} is invertible, as was to be shown, so that the linear system can be written as

$$\begin{pmatrix} u^1 \\ v^2 \end{pmatrix} = - \begin{pmatrix} U_1 & D_{12} \\ 0 & D_{22} \end{pmatrix}^{-1} \begin{pmatrix} U_2 u^2 + D_{11} v^1 \\ D_{21} v^1 \end{pmatrix}.$$

This shows in passing that the rank of $[Q \; R]$ is n. Defining the new s as (x^1, s^2), we obtain the desired result.

Later in this paper, we use this idea of reordering in order to identify x and s with the set of large and small variables, respectively. Then of course, the LCP is in general not in standard form.

Notation. Here we introduce the notation that will be used in the paper. This notation is quite standard, with exception of $d(x, s)$, which usually does not include the multiplication by μ. Given an interior solution (x, s), we define:

$$\mu(x, s) = s^T x / n,$$

$$w(x, s) = sx / \mu(x, s),$$

$$\delta(x, s) = \|w(x, s) - e\|,$$

$$d(x, s) = \sqrt{\mu(x, s) x / s}.$$

When no confusion can arise, we drop the reference to the variables, and continue to use other symbols in a consistent manner. For example $\bar{w} = w(\bar{x}, \bar{s})$ or $\mu_k = \mu(x^k, s^k)$.

The scalar μ and the vector w are the parameter and weight associated with the w-weighted trajectory that passes through (x, s) (see below). δ is the proximity measure of (x, s) in relation to the central path (see below), and d will be used as a scaling vector when solving Newton equations.

Using these definitions, we define two sets of interior points which are bounded away from the nonoptimal faces of F. Given $\epsilon > 0$ and $\mu_0 > 0$,

$$(1) \qquad F_\epsilon = \{(x, s) \in F \mid \mu(x, s) \le \mu_0, w(x, s) \ge \epsilon e\}.$$

Given $\alpha \in (0, 1)$,

$$(2) \qquad \mathcal{N}_\alpha = \{(x, s) \in F \mid \mu(x, s) \le \mu_0, \delta(x, s)\alpha\}.$$

F_ϵ is the set in which all published convergent feasible primal-dual algorithms operate (to our knowledge). The set $\mathcal{N}_\alpha$ is the Euclidean norm α-neighborhood of the central path, in which the path following algorithms to be studied in this paper operate.

Trajectories. Here we summarize some properties of the central path and the w-weighted trajectories for the problem (P). These trajectories and their properties are described in Monteiro and Tsuchiya (1992).

The solution (x, s) is called central if $xs = \mu e$ for $\mu > 0$. For the problem with the hypotheses above, there is a unique central point $(x(\mu), s(\mu))$ associated with each $\mu > 0$, and the map $\mu > 0 \mapsto (x(\mu), s(\mu))$ defines a smooth curve, known as the central trajectory, ending in an optimal solution. Note that $x(\mu)^T s(\mu) = n\mu$, and thus $\mu = \mu(x, s)$.

In the same vein, given a vector $w > 0$ such that $w^T e = n$, we can define a w-weighted trajectory $\mu > 0 \mapsto (x_w(\mu), s_w(\mu))$ composed by the solutions (x, s) such that $xs = \mu w$. Note that again $\mu = \mu(x, s)$, showing the consistency of our notational conventions.

Given $w > 0$ as above, the w-weighted barrier penalized function is defined as

$$\mu > 0, \qquad (x, s) \in F_0 \mapsto f_w(x, s, \mu) = x^T s - \mu \sum_{i=1}^{n} w_i \log x_i s_i.$$

The following facts are known.

- $(\bar{x}, \bar{s}) = (x_w(\mu), s_w(\mu))$ if and only if

$$(\bar{x}, \bar{s}) = \mathrm{argmin}\{f_w(x, s, \mu) \mid (x, s) \in F^0\}.$$

- The w-weighted trajectory ends at the optimal solution

$$(x_w^*, 0) = \lim_{\mu \to 0} (x_w(\mu), s_w(\mu)),$$

$$x_w^* = \mathrm{argmin}\left\{ -\sum_{i=1}^{n} w_i \log x_i \mid Qx = b, x > 0 \right\}.$$

In words, $(x_w^*, 0)$ is the w-weighted analytic center of the optimal set. In particular, the central path ends at $(x^*, 0)$, the analytic center of the optimal face.

- Given $\epsilon > 0$, the solutions $(x, s) \in F_\epsilon$ satisfy $x \approx 1$, $s \approx \mu(x, s)$. In particular, this is satisfied for all points in $\mathcal{N}_\alpha$, for a fixed $\alpha \in (0, 1)$.

3. Main results. This section describes a general model of primal-dual Newton methods and lists the main results to be shown in the subsequent sections.

The Newton step. Interior primal-dual algorithms take at each iteration a Newton step for solving the problem below, which corresponds to the search for a point in the central trajectory:

$$xs = \gamma \mu e,$$

(3)

$$Qx + Rs = b,$$

where $\gamma \in (0, 1)$ and, as usual $\mu = \mu(x, s)$. The Newton step for this problem is obtained by solving the system

$$su + xv = -xs + \gamma \mu e,$$

(4)

$$Qu + Rv = 0.$$

Under our hypotheses the Newton step always has a unique solution (u, v). The general algorithmic model below starts from a given initial point $(x^0, s^0) \in F_\epsilon$ and generates a sequence in F_ϵ. The generality is given by the choice of the aimed gap reduction γ and the steplength θ in each iteration.

ALGORITHM 3.1. *General Newton method. Data:* $\epsilon > 0$, $(x^0, s^0) \in F_\epsilon$.
$k := 0$
REPEAT
   $x := x^k$, $s := s^k$, $\mu := x^T s / n$.
   Choose $\gamma \equiv \gamma_k \in [0, 1]$ *and solve the Newton equations* (3).
   Choose $\theta \equiv \theta_k \in (0, 1]$ *such that* $(x + \theta u, s + \theta v) \in F_\epsilon$.
   $x^{k+1} := x + \theta u$, $s^{k+1} := s + \theta v$.
   $k := k + 1$.

We shall study the iterate convergence of a general class of algorithms based on this model. We shall not prove convergence of the objective values. Rather, we assume that the sequence (μ_k) converges R-linearly to zero, i.e., there exists a sequence $(\bar{\mu}_k)$ such that for $k \in \mathbb{N}$, $\mu_k \leq \bar{\mu}_k$ and $\bar{\mu}_{k+1} \leq \nu \bar{\mu}_k$, with $\nu < 1$.

The values of γ_k and θ_k at each iteration are chosen among the following three possibilities (of which the second and third are obviously special cases of the first):
   (i) Basic Newton step: $\gamma_k \in [0, 1]$, $\theta_k \in (0, 1]$.
   (ii) Affine-scaling step: $\gamma_k = 0$, $\theta_k \in (0, 1)$.
   (iii) Centering step: $\gamma_k = 1$, $\theta_k \in (0, 1]$ and $(x^k, s^k) \in \mathcal{N}_\alpha$ for $\alpha \leq 0.25$.

Convergence. Our main result is summarized in the following theorem, whose proof occupies the remainder of the paper.

THEOREM 3.2. *Consider the General Newton Method described above, and assume that* (μ_k) *converges R-linearly to zero. Choose* $\alpha \in (0, 0.25]$. *Then there exists* $\bar{\lambda} > 0$ *such that If* $\theta_k \gamma_k \leq \bar{\lambda}$ *for all k such that* $(x^k, s^k) \notin \mathcal{N}_\alpha$, *then*
   (i) *The sequence of iterates* (x^k, s^k) *converges to a point* $(\bar{x}, \bar{s}) \in \mathscr{F}^0$.
   (ii) *If* $\sum_{k=0}^{\infty} \theta_k \gamma_k = +\infty$ *then* $(\bar{x}, \bar{s}) = (x^*, s^*)$, *the analytic center of* $\mathscr{F}$. *If* $\sum_{k=0}^{\infty} \theta_k \gamma_k < +\infty$ *then possibly* $(\bar{x}, \bar{s}) \neq (x^*, s^*)$.

This theorem allows us to describe the convergence properties of several algorithms referenced in the introduction, according to the behavior of the sequences (θ_k) and (μ_k). In these comments when we refer to speed of convergence we mean the speed with which the sequence μ_k converges to zero. Actually there is a close connection between the speed of convergence of μ_k and the parameters θ_k and γ_k,

as we prove in Lemma 5.8 below that

$$\mu_{k+1} = (1 - \theta_k + \theta_k\gamma_k)\mu_k + \theta_k^2 O(\mu_k^2).$$

We deduce from Theorem 3.2 that:

(i) Algorithms for which $(x^k, s^k) \in \mathcal{N}_\alpha$ and $\theta_k\gamma_k$ is bounded away from zero in some subsequence: these algorithms generate sequences such that $x^k \to x^*$, because the series in the lemma is unbounded. This case includes the following methods:

—The predictor-corrector algorithm in $\mathcal{N}_\alpha$: this algorithm alternates affine-scaling steps and centering steps with $\gamma_k = 1$ and $\theta_k = 1$. The algorithm is 2-step quadratically convergent.

—Short steps algorithms: these methods use $\gamma_k > 0$ fixed, and θ_k bounded away from zero (either $\theta_k = 1$ or θ_k results from a line search). The algorithms are linearly convergent.

(ii) Algorithms for which $(x^k, s^k) \in F_e$, $\theta_k \to 1$, $\gamma_k \to 0$: this includes all Q-superlinearly convergent algorithms based on Newton steps and line searches. The sequences are convergent, but the limit point may not be the central optimum, depending on whether $\sum_{k=0}^{+\infty} \gamma_k = +\infty$ or not.

(iii) The longest step path following algorithm: $(x^k, s^k) \in \mathcal{N}_\alpha$, $\alpha \le 0.25$, $\theta_k = 1$, and $\gamma_k < 1$ such that $\delta(x^{k+1}, s^{k+1}) = \alpha$. From the theorem, the iterates converge. Using this fact and the results of McShane (1991), we conclude that the algorithm is superlinearly convergent. This means that $\gamma_k \to 0$ and we fall in case (ii) above.

(iv) Large steps path following algorithms: from (iii), it is clear that taking $\gamma_k < 1$ fixed, $\theta_k = 1$ and $(x^0, s^0) \in \mathcal{N}_\alpha$ sufficiently near the optimal face, the whole sequence will be in $\mathcal{N}_\alpha$. Then $x^k \to x^*$ as for short step algorithms. For an arbitrary initial point in $\mathcal{N}_\alpha$, each iteration can be constructed as in the primal large steps algorithm for linear programming developed by Gonzaga (1991) and by Roos and Vial (1989): use a sequence of damped Newton steps to obtain (x^{k+1}, s^{k+1}) such that $\|x^{k+1}s^{k+1}/\gamma_k\mu_k - e\| \le \alpha$. The argument above shows that for large values of k only one Newton step will be needed.

4. Descent algorithms for nonlinear programming. Let $\Gamma \subset \mathbb{R}^n$ be an open set, and $f: \Gamma \to \mathbb{R}$ a differentiable function with locally Lipschitz continuous gradient $x \in \Gamma \mapsto \nabla f(x)$.

We consider the unconstrained nonlinear programming problem

$$\text{minimize } f(x).$$
$$x \in \Gamma$$

We shall study general properties of descent algorithms for this problem. The format for the algorithms is the following, for a given $x^0 \in \mathbb{R}^n$:

ALGORITHM 4.1. *Model algorithm.*
$k := 0$
WHILE $\nabla f(x) \ne 0$ do
$\quad x^{k+1} := x^k + \lambda^k h^k.$
$\quad k := k + 1.$
END WHILE

where for each k, $h^k \in \mathbb{R}^n$ and $\lambda^k \in (0, +\infty)$ are such that $x^{k+1} \in \Gamma$. An algorithm will be called a "descent algorithm" if there exists $\Delta > 0$ such that for all $k \in \mathbb{N}$,
(i) $\|h^k\| \approx \|\Delta f(x^k)\|$,
(ii) $f(x^k + \lambda^k h^k) \le (x^k) - \Delta\lambda^k\|h^k\| \|\nabla f(x^k)\|.$

A sequence (x^k) in Γ will be said compact if it is contained in some compact set $\Gamma_0 \subset \Gamma$, and a point $x \in \Gamma$ is stationary if $\nabla f(x) = 0$.

The next lemma shows that there are two alternatives for a descent algorithm: either the steps are sufficiently large to reach the optimal set, and then all accumulation points are stationary, or the steps are short and the sequence converges to a nonstationary point. We are unaware of the existence of this result in the literature.

LEMMA 4.2.   *Assume that a descent algorithm as above generates an infinite compact sequence (x^k) in Γ. Then*
   (i) *Either $\sum_{k=0}^{\infty} \lambda^k < +\infty$ and (x^k) converges to a nonstationary point or*
   (ii) *$\sum_{k=0}^{\infty} \lambda^k = +\infty$ and all accumulation points of (x^k) are stationary.*

PROOF.   Consider sequences $(x^k), (h^k), (\lambda^k)$ generated by the algorithm, and assume that (x^k) is contained in a compact $\Gamma_0 \subset \Gamma$.

(i) If $\sum_{k=0}^{\infty} \lambda^k < +\infty$, then $\sum_{k=0}^{\infty} \lambda^k \|h^k\| < +\infty$, because $\|h^k\| \approx \|\nabla f(x^k)\|$ is bounded in Γ_0. Then (x^k) is a Cauchy sequence and thus $x^k \to \bar{x}$, with $\bar{x} \in \Gamma_0$. We must prove that $\bar{x}$ is not stationary.

Let $\hat{\Gamma} = \{x \in \Gamma_0 \mid \nabla f(x) = 0\}$ be the set of stationary points in Γ_0. If $\hat{\Gamma} = \emptyset$, then there is nothing to be proved. Otherwise, let L be a global Lipschitz constant for ∇f in Γ_0. Since $\|h^k\| \approx \|\nabla f(x^k)\|$, $\|h^k\| \leq K\|\nabla f(x^k)\|$, where $K > 0$ is a constant. It follows that

$$\|x^{k+1} - x^k\| \leq \lambda^k \|h^k\| \leq \lambda^k K \|\nabla f(x^k)\|.$$

After an iteration,

$$\|\nabla f(x^{k+1})\| \geq \|\nabla f(x^k)\| - L\|x^{k+1} - x^k\|,$$

$$\geq (1 - KL\lambda^k)\|\nabla f(x^k)\|.$$

Since $\lambda^k \to 0$, for k_0 sufficiently large $1 - KL\lambda^k > 0$ and $\log(1 - KL\lambda^k) \geq -2KL\lambda^k$. Thus for $k > k_0$,

$$\|\nabla f(x^k)\| \geq \|\nabla f(x^{k_0})\| \prod_{j=k_0}^{k-1} (1 - KL\lambda^j),$$

and

$$\log\|\nabla f(x^k)\| \geq \log\|\nabla f(x^{k_0})\| + \sum_{j=k_0}^{k-1} \log(1 - KL\lambda^j) \geq \log\|\nabla f(x^{k_0})\| - 2KL \sum_{j=k_0}^{k-1} \lambda^j.$$

It follows that $\log\|\nabla f(x^k)\|$ has a lower bound, and thus $|\nabla f(x^k)\|$ is bounded away from zero, establishing (i).

(ii) Assume now that $\sum_{k=0}^{\infty} \lambda^k = +\infty$. For an arbitrary $\epsilon > 0$, define

$$\Gamma_\epsilon = \{x \in \Gamma_0 \mid \|\nabla f(x)\| > \epsilon\}.$$

By definition of descent algorithm,

$$f(x^k) \leq f(x^0) - \Delta \sum_{k=0}^{\infty} \lambda^k \|\nabla f(x^k)\| \|h^k\|.$$

Since f is bounded in Γ_0 and $\|h^k\| \approx \|\nabla f(x^k)\|$, we conclude that

$$(5) \qquad \sum_{k=0}^{\infty} \lambda^k \|\nabla f(x^k)\|^2 < +\infty.$$

Now assume by contradiction that (x^k) has an accumulation point $\bar{x} \in \Gamma_\epsilon$. Then the set $\mathcal{K} = \{k \in \mathbb{N} \mid x^k \in \Gamma_\epsilon\}$ is infinite.

From (5), with $\|\nabla f(x^k)\| \geq \epsilon$ for $k \in \mathcal{K}$,

$$(6) \qquad \sum_{k \in \mathcal{K}} \lambda^k < +\infty,$$

and we conclude that the set of indices $\overline{\mathcal{K}} = \mathbb{N} - \mathcal{K}$ is also infinite. For each $k \in \mathcal{K}$, let $k' > k$ be the first index in $\overline{\mathcal{K}}$. Then

$$(7) \qquad \|x^{k'} - x^k\| \leq \sum_{j=k}^{k'-1} \lambda^j \|h^j\| = O\left(\sum_{j=k}^{k'-1} \lambda^j\right).$$

Using (6), we see that as $k \in \mathcal{K}$ grows, $\|x^{k'} - x^k\| \to 0$, and thus the subsequences $(x^k)_{k \in \mathcal{K}}$ and $(x^{k'})_{k \in \mathcal{K}}$ have the same accumulation points. But all accumulation points of $(x^k)_{k \in \mathcal{K}}$ satisfy $\|\nabla f(x)\| \leq \epsilon$. This implies $\bar{x} \notin \Gamma_\epsilon$, contradicting the hypothesis and completing the proof.

Perturbed descent algorithms. Consider the model Algorithm 4.1 and a sequence of "perturbations" (η^k) in $\mathbb{R}^n$. A perturbed algorithm is an algorithm that follows the model, but has the step given by

$$x^{k+1} := x^k + \lambda^k h^k + \eta^k.$$

The next lemma shows that if the perturbations are summable in the sense that

$$\sum_{k=0}^{\infty} \|\eta^k\| < \infty,$$

then the capacity of a descent algorithm to find stationary points is not affected. But now the first part of the lemma above is not guaranteed, since "clever" perturbations might drive the sequence to a stationary point.

LEMMA 4.3. *Consider a descent algorithm with perturbed steps, with perturbations* (η^k) *such that* $\sum_{k=0}^{\infty} \|\eta^k\| < \infty$, *and assume that it generates a compact sequence of iterates. Then*

(i) *Either* $\sum_{k=0}^{\infty} \lambda^k < +\infty$ *and* (x^k) *converges to a possibly nonstationary point or*

(ii) $\sum_{k=0}^{\infty} \lambda^k = +\infty$ *and all accumulation points of* (x^k) *are stationary.*

PROOF. The proof follows that of Lemma 4.2 step by step. We shall only indicate the changes in the proof. Part (i) is trivial, because the resulting sequence is a Cauchy sequence. To prove (ii), define Γ_ϵ as before and note that using the Lipschitz condition for ∇f in the compact Γ_0, there exists a global Lipschitz constant L_f for f

in Γ_0 and hence

$$f(x^{k+1}) = f(x^k + \lambda^k h^k + \eta^k)$$

$$\leq (x^k + \lambda^k h^k) + L_f \|\eta^k\|$$

$$\leq f(x^k) - \Delta \lambda^k \|\nabla f(x^k)\| \|h^k\| + L_f \|\eta^k\|.$$

Since f is bounded in Γ_0,

$$\sum_{k=0}^{\infty} \left(\Delta \lambda^k \|\nabla f(x^k)\| \|h^k\| + L_f \|\eta^k\| \right) < +\infty.$$

As $\|h^k\| \approx \|\nabla f(x^k)\|$, we conclude that $\sum_{k=0}^{\infty} \lambda^k \|\nabla f(x^k)\|^2 < +\infty$. The rest of the proof follows directly the proof of Lemma 4.2, with expression (7) replaced by

$$\|x^{k'} - x^k\| \leq \sum_{j=k}^{k'-1} \left(\lambda^j \|h^j\| + \|\eta^j\| \right) = O\left(\sum_{j=k}^{k'-1} \lambda^j \right) + \sum_{j=k}^{k'-1} \|\eta^j\|.$$

An interesting situation is the one in which $\lambda^k \to 0$. In this case, the set of accumulation points form a continuum, and if the series of (λ^k) is infinite all these points are stationary.

Algorithmic maps. A descent direction for a differentiable function at a point is a vector that makes an angle of less than 90 degrees with the negative gradient direction. Descent algorithms can be obtained by means of maps that associate with each point a set of uniform descent directions, as follows.

Let $x \in \Gamma \to H(x) \subset \mathbb{R}^n$ be a point to set map. We say that $H(\cdot)$ is a map of descent directions in a set $\Gamma_0 \subset \Gamma$ if there exists a constant $\Delta > 0$ such that for any $x \in \Gamma_0$ and $h \in H(x)$:

(i) $\|h\| \approx \|\nabla f(x)\|$,
(ii) $\nabla f(x)^T h \leq -\Delta \|\nabla f(x)\| \|h\|$.

Given a map of descent directions defined in a set $\Gamma_0 \subset \Gamma$, algorithms can be constructed as in Algorithm 4.1. A complete algorithm with perturbations (η^k) will be:

Algorithm 4.4. *Perturbed algorithm based on a map of descent directions: given* $x^0 \in \Gamma_0$.
  $k := 0$
  WHILE $\nabla f(x) \neq 0$ do
    choose $h^k \in H(x^k)$.
    $x^{k+1} := x^k + \lambda^k h^k + \eta^k \in \Gamma_0$.
    $k := k + 1$.
  END WHILE

In each iteration k, $\lambda^k \in (0, +\infty)$ must be such that $x^{k+1} \in \Gamma$. The lemma below will show that if the λ^k are small and Γ_0 is compact, the resulting algorithm will be a descent algorithm, hence with the convergence properties described above. Two situations in which Γ_0 is easily constructed are the following: in the first case, which is frequent in the literature, the level set $\{x \in \Gamma \mid f(x) \leq f(x^0)\}$ is compact, and λ^k is given by a line search that guarantees $f(x^{k+1}) \leq f(x^k)$. The second case is our direct

interest, where the primal-dual algorithms are responsible for generating iterates in the set F_ϵ defined in §2.

LEMMA 4.5. *Let $H(\cdot)$ by a descent map defined as above in a compact $\Gamma_0 \subset \Gamma$. There exists $\lambda > 0$ such that for any $x \in \Gamma_0$, $h \in H(x)$ and $\lambda \in (0, \bar{\lambda}]$:*

$$f(x + \lambda h) \leq f(x) - \frac{\Delta}{2} \lambda \|\nabla f(x)\| \|h\|.$$

PROOF. We begin by using the Lipschitz condition to show that there exists $\lambda^1 > 0$ such that for any $x \in \Gamma_0$, $h \in H(x)$ and $\lambda \in (0, \lambda^1)$:

$$(8) \qquad \|\nabla f(x + \lambda h) - \nabla f(x)\| \leq \lambda L \|h\|.$$

Since Γ_0 is compact and Γ is open, there exists $\epsilon > 0$ such that $\Gamma_1 = \Gamma_0 + B_\epsilon \subset \Gamma$, where $B_\epsilon = \{y \in \mathbb{R}^n \mid \|y\| \leq \epsilon\}$. So ∇f is locally Lipschitz continuous in the compact Γ_1, and consequently satisfies a global Lipschitz condition in Γ_1, with a constant L. This means that for $\|\lambda h\| \leq \epsilon$, $x \in \Gamma_0$ we have $x + \lambda h \in \Gamma_1$ and (8) is satisfied. We must still specify the value of λ^1: since $\|h\| \approx \|\nabla f(x)\|$ and $\nabla f(x)$ is bounded in Γ_0, there exists $K_1 > 0$ such that for $x \in \Gamma_0$, $\|h\| \leq K_1$. Choosing $\lambda^1 = \epsilon/K_1$, the result is complete.

Now we prove the lemma. Consider a point $x \in \Gamma_0$, $\lambda \in (0, \lambda^1)$, ad $h \in H(x)$. Then

$$f(x + \lambda h) = f(x) + \lambda \int_0^1 h^T \nabla f(x + \sigma \lambda h) \, d\sigma,$$

$$= f(x) + \lambda h^T \nabla f(x) + \lambda \int_0^1 h^T [\nabla f(x + \sigma \lambda h) - \nabla f(x)] \, d\sigma,$$

$$\leq f(x) - \lambda \Delta \|\nabla f(x)\| \|h\| + \frac{\lambda^2}{2} L \|h\|^2.$$

As $H(\cdot)$ is a map of descent direction, there exists $K_2 > 0$ such that $\|h\| \leq K_2 \|\nabla f(x)\|$, for all $x \in \Gamma_0$, $h \in H(x)$. Choosing $\bar{\lambda} = \min\{\lambda^1, \Delta/K_2 L\}$,

$$\frac{\lambda^2}{2} L \|h\|^2 \leq \lambda \frac{\Delta}{2} \|\nabla f(x)\| \|h\|,$$

for $\lambda \in (0, \bar{\lambda})$. The substitution of this into the expression above leads to the desired result, completing the proof.

5. Study of the Newton step. In this section we study general aspects of the Newton step associated with the problem (P). We begin by studying properties of feasible directions.

Properties of feasible directions. Consider the equality constraint in (P). A pair $(u, v) \in \mathbb{R}^{2n}$ is a feasible direction if and only if $Qu + Rv = 0$. Feasible directions can also be defined separately for x and s, as follows:

$$\mathcal{U} = \{u \in \mathbb{R}^n \mid Qu + Rv = 0 \text{ for some } v \in \mathbb{R}^n\},$$

$$\mathcal{V} = \{v \in \mathbb{R}^n \mid Qu + Rv = 0 \text{ for some } u \in \mathbb{R}^n\}.$$

The next lemma resumes the nice geometrical properties of feasible directions under

the monotonicity hypothesis, and is a particular case of Monteiro and Tsuchiya (1992, Lemma 3.3). We give the proof in order to make the paper self-contained.

LEMMA 5.1. $\mathscr{U} \subset \mathscr{R}(R^T)$ and $\mathscr{V} \subset \mathscr{R}(Q^T)$.

PROOF. Consider $u \in \mathscr{U}$. By definition, for some $v \in \mathscr{V}$, $Qu + Rv = 0$. Given $v' \in \mathscr{N}(R)$, and $\lambda \in \mathbb{R}$,

$$Qu + R(v + \lambda v') = 0,$$

and from the monotonicity assumption,

$$u^T(v + \lambda v') \geq 0.$$

This implies in $u^T v' = 0$. Since v' is arbitrary in $\mathscr{N}(R)$, $u \in \mathscr{N}(R)^\perp = \mathscr{R}(R^T)$. The second inclusion in the lemma is shown similarly, completing the proof.

Orthogonal projections. Now we summarize some properties of orthogonal projections into an affine space.

Given the affine space defined by $Ax = q$, where $A \in \mathbb{R}^{m \times n}$, $q \in \mathbb{R}^m$, we define the projection operators $P_{A,q}$ and P_A by

$$x \mapsto P_{A,q}x = \operatorname{argmin}\{\|w - x\| \mid Aw = q\},$$

and $P_A = P_{A,0}$. Since P_A is a linear operator in $\mathbb{R}^n$, the same notation will be used for its matrix representation.

LEMMA 5.2. *For any* $x \in \mathbb{R}^n$ *and* $q \in \mathbb{R}^m$, $P_{A,q}x = P_A x + P_{A,q}0$.

PROOF. This follows directly from the linearity of the mapping $(x, q) \mapsto P_{A,q}x$. The linearity of this map is proven by examining the optimality conditions that characterize the projection, as follows:

Consider $w^1 = P_{A,q^1}x^1$ and $w^2 = P_{A,q^2}x^2$. The optimality conditions for the projections are $Aw^i = q^i$, $x^i - w^i \perp \mathscr{N}(A)$, $i = 1, 2$. It follows that for $\lambda \in \mathbb{R}$, $A(w^1 + \lambda w^2) = q^1 + \lambda q^2$ and $x^1 + \lambda x^2 - (w^1 + \lambda w^2) \perp \mathscr{N}(A)$. This means that $w^1 + \lambda w^2 = P_{A,q^1+\lambda q^2}(x^1 + \lambda x^2)$, completing the proof.

The next lemma describes properties of the orders of magnitude of projections.

LEMMA 5.3. *Let* $\mathscr{D} \subset \mathbb{R}^n$ *be such that* $d \approx 1$ *whenever* $d \in \mathscr{D}$.[1] *Then for* $y \in \mathbb{R}^n$, $q \in \mathscr{R}(A)$, $d \in \mathscr{D}$,

$$P_{AD,q}0 = O(\|q\|),$$

$$P_{AD,q}y = O(\|q\|) + O(\|y\|),$$

$$\|DP_{AD}Dy\| \approx \|P_A y\|.$$

PROOF. Let A^+ be a right inverse of A, i.e., A^+ is an $(n \times m)$ matrix such that $AA^+q = q$ for all $q \in \mathscr{R}(A)$. Then $\hat{z} = D^{-1}A^+q$ is feasible for the problem

$$\text{minimize}\{\|z - y\| \mid ADz = q\}.$$

---

[1] I.e., there are constants $K_2 > K_1 > 0$, such that any $d \in \mathscr{D}$ satisfies $K_1 \leq d_i \leq K_2$, $i = 1, \ldots, n$.

It follows that

$$\|P_{AD,q}y - y\| \leq \|\hat{z} - y\| \leq \|D^{-1}A^+q\| + \|y\|.$$

The first two results follow from this inequality.

As a consequence of these results, for any $z \in \mathbb{R}^n$, $DP_{AD}Dz = O(\|z\|)$. Taking $z = P_A y = y - A^T w$ for some $w \in \mathbb{R}^m$,

$$P_{AD}Dz = P_{AD}Dy - P_{AD}DA^T w = P_{AD}Dy.$$

It follows that $DP_{AD}Dy = O(\|P_Q y\|)$. By the same process, with a change of variables $\bar{y} = Dy$, we get $P_A y = O(\|DP_{AD}Dy\|)$, completing the proof.

Shifted scalings.   The following lemma, shown in Gonzaga and Tapia (1992), will be useful ahead. We reproduce the proof here for completeness.

LEMMA 5.4.   *Let $g \in \mathbb{R}^n$ be such that $\|g - e\|_\infty \leq \alpha$, where $\alpha \in (0,1)$. Set $G = \text{diag}(g)$, and consider the projections $\hat{h} = P_A \, \rho$, $h = gP_{AG}g \, \rho$. Then*

$$\|h - \hat{h}\| \leq \alpha(1 + \alpha)\frac{2 - \alpha}{1 - \alpha}\|\hat{h}\|.$$

In particular, if $\alpha \in (0, 0.25)$:

$$\|h - \hat{h}\| \leq 3\alpha\|\hat{h}\|.$$

PROOF.   Since $\rho = \hat{h} + A^T w$ for some $w \in \mathbb{R}^m$,

$$g\rho = g\hat{h} + (AG)^T w$$

and thus

$$P_{AG}g \, \rho = P_{AG}g\hat{h}.$$

It follows that

$$g^{-1}h = P_{AG}g\hat{h}.$$

On the other hand, by definition of projection,

$$g\hat{h} = P_{AG}g\hat{h} + y,$$

where $y \in \mathscr{R}(GA^T)$. Merging the last expressions,

$$g\hat{h} = g^{-1}h + y,$$

where $g^{-1}h \in \mathscr{N}(AG)$ and $y \in \mathscr{R}(GA^T)$. Subtracting $g^{-1}\hat{h} \in \mathscr{N}(AG)$ from both sides,

$$(g^{-1} - g)\hat{h} = g^{-1}(h - \hat{h}) + y,$$

and from the orthogonality of the right-hand side terms,

$$\|(g^{-1} - g)\hat{h}\| \geq \|g^{-1}(h - \hat{h})\|.$$

Now use the following facts: $\|(h - \hat{h})\| \leq \|g\|_\infty \|g^{-1}(h - \hat{h})\|$ and $\|(g^{-1} - g)\hat{h}\| \leq \|(g^{-1} - g)\|_\infty \|\hat{h}\|$. Combining these three expressions leads to

$$\|h - \hat{h}\| \leq \|g\|_\infty \|g^{-1} - g\|_\infty \|\hat{h}\|.$$

Finally, $\|g\|_\infty \|g^{-1} - g\|_\infty \leq (1 + \alpha)(1/(1 - \alpha) + (1 - \alpha))$. The inequality for $\alpha \in (0, 0.25)$ follows trivially, completing the proof.

The scaled equations. The right-hand side of the first equation in (4) will vary according to the algorithm. At this point we shall study general results for the Newton step with a generic right-hand side, namely

$$(9) \qquad\qquad su + xv = f, \qquad Qu + Rv = 0,$$

where $f \in \mathbb{R}^n$. Our scope now is to characterize the direction u in relation to v. We begin by scaling the equations, to obtain a simpler situation. The scaling will use the scaling vector $d = \sqrt{\mu x/s}$ defined above.

The scaled equations are obtained from (9) by multiplying the first equation by $d/\mu x$, which results in

$$d\frac{s}{\mu x}u + \frac{d}{\mu}v = \frac{df}{\mu x},$$

and then using the definition of d,

$$(10) \qquad\qquad d^{-1}u + d\frac{v}{\mu} = \frac{d}{x}\frac{f}{\mu}, \qquad Qu + Rv = 0.$$

Defining now the scaled variables

$$(11) \qquad\qquad \bar{u} = d^{-1}u, \qquad \bar{v} = dv/\mu,$$

we obtain the scaled equations

$$(12) \qquad\qquad \bar{u} + \bar{v} = \frac{df}{\mu x}, \qquad QD\bar{u} = -\mu RD^{-1}\bar{v},$$

and $\bar{u}^T\bar{v} \geq 0$ whenever $\bar{u}, \bar{v}$ satisfy the second relation in (12).

The least squares problem. Lemma 5.1 will be used to relate large and small variables by means of a least squares problem. We start by a simple situation, to which we will reduce after a scaling of the problem:

$$(13) \qquad\qquad u + v = c, \qquad Qu + Rv = 0,$$

where $c \in \mathbb{R}^n$ is a given vector.

LEMMA 5.5. *Consider the system* (13) *and assume that v is known. Let $\bar{v} \in \mathcal{V}$ be chosen arbitrarily. Then u solves the problem*

$$\text{minimize}_{\tau \in \mathbb{R}^n} \qquad \|\tau - c + \bar{v}\|$$

$$\text{subject to} \qquad Q\tau = -Rv.$$

PROOF. The optimality conditions for the least squares problem above are $Qu = -Rv$ and $u - c - \bar{v} \in \mathcal{R}(Q^T)$. The first one is trivially satisfied. To check the second one, substitute $c - u = v$ to obtain the condition $v + \bar{v} \in \mathcal{R}(Q^T)$. This follows from Lemma 5.1, completing the proof.

We are ready to apply Lemma 5.5 to (12). Let $\bar{v} \in \mathcal{V}$ be an arbitrary feasible direction for the small variables, and let $d\bar{v}/\mu$ be the corresponding direction for the scaled problem. Assume that v is given. Then $\bar{u}$ solves the problem

(14)
$$\text{minimize}_{\tau \in \mathbf{R}^n} \quad \left\| \tau - \frac{df}{\mu x} + \frac{d\bar{v}}{\mu} \right\|$$

$$\text{subject to} \qquad QD\tau = -Rv.$$

Using orthogonal projections and Lemma 5.2, we obtain from (14):

(15)
$$\bar{u} = P_{QD}\left(\frac{df}{\mu x} - \frac{d\bar{v}}{\mu} \right) + P_{QD, Rv}0,$$

$$u = d\bar{u}.$$

Orders of magnitude for the general Newton step.

LEMMA 5.6. *Consider the general Newton step as above, starting at a point $(x, s) \in F_\varepsilon$. Then*:

$$x \approx 1, \quad s \approx \mu, \quad d \approx 1, \quad w \approx 1;$$

$$\bar{u} = O\left(\frac{f}{\mu}\right), \qquad \bar{v} = O\left(\frac{f}{\mu}\right),$$

$$\bar{u} = O(f/\mu), \quad v = O(f/\mu),$$

$$u = O(f/\mu), \quad v = O(f).$$

PROOF. The magnitudes of x, s, d and w follow immediately from the discussion in §2.

From the scaled equations (12), taking norms,

$$\|\bar{u}\|^2 + \|\bar{v}\|^2 + 2\bar{u}^T\bar{v} = \left\| \frac{df}{\mu x} \right\|^2.$$

By the monotonicity hypothesis, $\bar{u}^T\bar{v} \geq 0$, and it follows that

$$\|\bar{u}\| \leq \left\| \frac{df}{\mu x} \right\|, \qquad \|\bar{v}\| \leq n \left\| \frac{df}{\mu x} \right\|.$$

Since $d \approx 1$ and $x \approx 1$, this implies the desired magnitudes for $\bar{u}$ and $\bar{v}$. Finally, $u = d\bar{u} = O(\bar{u})$, and $v = d^{-1}\mu\bar{v} = O(\mu\bar{v})$, because $d \approx 1$. This shows the two last equalities, completing the proof.

Algorithms using the exact Newton step. The exact Newton step from a given interior solution (x, s), $\gamma \in (0, 1)$ is the solution of the system (4). We shall begin by specializing (15) for this system.

LEMMA 5.7.   *Let $\tilde{s}$ be an arbitrary feasible vector of small variables, and assume that v is known. Then u is given by*

(16)
$$\bar{u} = P_{QD}\, d\left(-\frac{\tilde{s}}{\mu} + \gamma x^{-1}\right) + P_{QD,\,Rv}0,$$

$$u = d\bar{u}.$$

In particular, taking $\tilde{s} = 0$,

(17)
$$\bar{u} = \gamma P_{QD}\, dx^{-1} + P_{QD,\,Rv}0,$$

$$u = d\bar{u}.$$

PROOF.   Let $\tilde{s}$ be an arbitrary feasible vector of small variables (note that by definition $\tilde{s} = 0$ is feasible). The direction $\bar{v} = \tilde{s} - s$ is feasible. Substituting $f = -xs + \gamma\mu e$ from (4), the first equality in (15) becomes

$$\bar{u} = P_{QD}\frac{d}{\mu}(-s + \bar{v} + \gamma\mu x^{-1}) + P_{QD,\,Rv}0.$$

The desired expressions are obtained by substituting $\tilde{s} = s + \bar{v}$.

Orders of magnitude for the exact Newton step.

LEMMA 5.8.   *Consider an exact Newton step from a point $(x, s) \in \Gamma_e$. Then*

$$\bar{u} = O(\gamma) + O(\mu), \qquad u = O(\gamma) + O(\mu);$$

$$\bar{v} = O(1), \qquad v = O(\mu);$$

$$u^T v = O(\mu^2);$$

$$\mu(x + \theta u, s + \theta v) = (1 - \theta + \theta\gamma)\mu + \theta^2 O(\mu^2).$$

PROOF.   For the exact Newton step, $f = -xs + \gamma\mu e = O(\mu)$ by Lemma 5.6. From the same lemma it follows that $\bar{v} = O(1)$ and $v = O(\mu)$. To prove the results for the large variables, we shall use the Lemma 5.3 and expression (17) to obtain $\bar{u} = \gamma O(dx^{-1}) + O(Rv)$. This immediately implies the desired results, since $dx^{-1} \approx 1$ and $Rv = O(v) = O(\mu)$. Let us now turn to the analysis of $u^T v$. From (11) and Lemma 5.7 we get

$$u^T v = \mu\bar{u}^T\bar{v} = \mu\left(\gamma P_{QD}\, dx^{-1} + O(\mu)\right)^T \bar{v}.$$

By Lemma 5.1 adapted to the scaled problem (12) we find that $\bar{v} \in \mathscr{R}(DQ^T) = \mathscr{N}(QD)^\perp$ so that $\bar{v}^T P_{QD}\, dx^{-1} = 0$. As $\bar{v} = O(1)$, the relation $u^T v = O(\mu^2)$ follows. We now prove the last relation. By (4),

$$(x + \theta u)(s + \theta v) = xs + \theta(xv + su) + \theta^2 uv,$$

$$= (1 - \theta)xs + \theta\gamma\mu e + \theta^2 uv,$$

so that

$$\mu(x + \theta u, s + \theta v) = (1 - \theta + \theta\gamma)\mu + \theta^2 u^T v/n.$$

Using $u^T v = O(\mu^2)$, the result follows.

6. Reduction to the optimal face. Consider a sequence (x^k, s^k) such that $\mu_k \to 0$. As k grows, (x^k, s^k) approaches the optimal face and $s^k \to 0$. Let the optimal face for the large variables be defined by

$$\mathscr{X} = \{x \in \mathbb{R}^n \mid Qx = b; x \geq 0\}.$$

Denote its relative interior by $\overset{\circ}{\mathscr{X}}$; i.e.,

$$\mathscr{X} = \{x \in \mathbb{R}^n \mid Qx = b; x > 0\}.$$

To $x \in \mathbb{R}^n$ we associate its projection onto $\mathscr{X}$:

$$y(x) = \operatorname{argmin}\{\|x - y\| \mid y \in \mathscr{X}\}.$$

The next lemma shows, in particular, that the sequences (x^k) and (y^k) have the same limit points. The rest of this section is devoted to the analysis of the asymptotic behaviour of $\{y^k\}$.

LEMMA 6.1. *Consider the map* $(x, s) \in F \mapsto y(x) \in \mathscr{X}$ *defined above. Assuming* μ_0 *(in the definition of F_ε) small enough, for all* $(x, s) \in F_\varepsilon$, *one has* $y(x) \approx 1$ *and* $x - y(x) = O(\mu)$.

PROOF. Given $(x, s) \in F_\varepsilon$, let $\hat{y} = P_{Q,b}x$. Since $Qx = b - Rs$, we can use the linearity of the map $(x, q) \mapsto P_{Q,q}x$ to deduce

$$x = P_{Q,b-Rs}x = P_{Q,b}x - P_{Q,Rs}0 = \hat{y} + P_{Q,Rs}0.$$

Using Lemma 5.3 and the fact that $s = O(\mu)$, we conclude that $x = \hat{y} + O(\mu)$. Finally, since $x \approx 1$ in F_ε, for μ sufficiently small, $\hat{y} \geq x/2$ and thus $\hat{y} \approx 1$.

In this case, $\hat{y} > 0$ and thus $\hat{y} \in \mathscr{X}$. If follows from the definition of projection that $y(x) = \hat{y}$, completing the proof.

Without loss of generality, we shall assume for the remaining of the paper that the set F_ε is defined with $\mu_0 = \tilde{\mu}$, given by the lemma above. So the corresponding projections in $\mathscr{X}$ will be in the set

(18)
$$\mathscr{X}_\varepsilon := \{y(x) \mid (x, s) \in F_\varepsilon\} \subset \overset{\circ}{\mathscr{X}}.$$

Since $y \approx 1$ in $\mathscr{X}_\varepsilon$, this set is bounded away from the boundary of $\mathscr{X}$, and thus its closure satisfies

(19)
$$\operatorname{cl}\mathscr{X}_\varepsilon \subset \overset{\circ}{\mathscr{X}}.$$

We observe that, by (17) and Lemma 5.8, the increment on $\{x^k\}$ satisfies

(20)
$$x^{k+1} = x^k + \theta_k\gamma_k d^k P_{QD^k}d^k(x^k)^{-1} + O(\mu_k).$$

We now derive a similar relation for $\{y^k\}$.

LEMMA 6.2. *Consider a sequence* (x^k, s^k) *in* F_ϵ *constructed by an algorithm as described above, and let* $y^k = y(x^k)$. *Then* $x^k - y^k = O(\mu_k)$, *the sequences* x^k *and* y^k *have the same accumulation points and*

$$(21) \qquad\qquad y^{k+1} = y^k + \theta_k \gamma_k d^k P_{QD^k} d^k (y^k)^{-1} + O(\mu_k).$$

PROOF. By Lemma 6.1, $x^k - y^k = O(\mu_k)$ and consequently the sequences (x^k) and (y^k) have the same accumulation points. Since $x^k \approx 1$, $(x^k)^{-1} - (y^k)^{-1} = O(\mu_k)$, and thus $d^k P_{QD^k} d^k (x^k)^{-1} = d^k P_{QD^k} d^k (y^k)^{-1} + O(\mu_k)$. It follows from (20), Lemma 5.3 and the linearity of projections that

$$y^{k+1} = x^{k+1} + O(\mu_{k+1})$$

$$= y^k + O(\mu_k) + \theta_k \gamma_k d^k P_{QD^k} d^k (y^k)^{-1} + O(\mu_k) + O(\mu_{k+1}).$$

By Lemma 5.8 we know that $\mu_{k+1} = O(\mu_k)$. This completes the proof.

At this point we may clarify our strategy: the scope is studying the convergence of (x^k, s^k). Since $s^k \to 0$, it is enough to study (x^k), but this sequence has the same asymptotic properties as (y^k), which is apparent from a comparison of the expressions (20) and (21).

We will show now that (21) can be interpreted as a perturbed descent algorithm for the logarithmic barrier function defined in $\mathcal{X}$.

The barrier function in $\mathcal{X}$. The logarithmic barrier function is defined by

$$x \in \mathbb{R}^n, \qquad x > 0 \mapsto p(x) = -\sum_{i=1}^{n} \log x_i.$$

The properties of this function are well known; see for instance Gonzaga (1992). The analytic center of $\mathcal{X}$ is

$$x^* = \operatorname{argmin}\{ p(y) \mid y \in \overset{\circ}{\mathcal{X}} \}.$$

The gradient of the barrier function restricted to $\overset{\circ}{\mathcal{X}}$ is

$$g(y) = P_Q \nabla p(y) = -P_Q y^{-1}.$$

The Newton direction for minimizing the barrier function in $\mathcal{X}$ is given by

$$\hat{u}(y) = y P_{QY} y y^{-1} = y P_{QY} e.$$

The primal proximity of a point $y \in \overset{\circ}{\mathcal{X}}$ to x^* is measured by the norm of the primal scaled Newton centering step in $\mathcal{X}$,

$$\delta(y) := \| P_{QY} e \|.$$

Using an orthonormal basis of $\mathcal{N}(Q)$, we may reduce the minimization of the barrier function over $\mathcal{N}(Q)$ to an unconstrained problem. The gradient of the latter is precisely $g(y)$, and the Newton direction is invariant. This allows us to apply the analysis of §4. After a Newton step, the barrier function decreases if the point is sufficiently close to x^*. It appears that we can control the (primal) proximity of $y(x)$

in the optimal face, since it is related to the (primal-dual) proximity $\delta(x, s)$ by the following lemma:

LEMMA 6.3. *Given $(x, s) \in F_\epsilon$ and $y = y(x)$, $\delta(y) \le \delta(x, s) + O(\mu)$.*

PROOF. (i) $\delta(y) = \|P_{QY}e\| = \|P_{QX}e\| + O(\mu)$.
(ii) $\|P_{QX}e\| \le \|e - w\| = \delta(x, s)$. By definition of projection,

$$\|P_{QY}e\| = \min\{\|e + YQ^Tz\| \mid z \in \mathbb{R}^m\} = \min\{\|e - z\| \mid z \in \mathscr{R}(YQ^T)\}.$$

But $s \in \mathscr{R}(Q^T)$ by Lemma 5.1, and thus $ys/\mu \in \mathscr{R}(YQ^T)$. Hence, from the relation above,

$$\|P_{QY}e\| \le \|e - ys/\mu\|$$

$$\le \|e - xs/\mu\| + \|(x - y)s/\mu\|.$$

But $(x - y) = O(\mu)$ by Lemma 6.2, and $s = O(\mu)$ by Lemma 5.6. Hence the last term is bounded by $O(\mu)$, completing the proof.

We are now ready to state conditions allowing us to apply Lemma 4.3.

LEMMA 6.4. *Consider the algorithmic map $L(\cdot)$ that associates with each $y \in \mathscr{X}_\epsilon$ the set*

$$L(y) = \{l = dP_{QD}\, dy^{-1} \mid d = d(x, s), (x, s) \in F_\epsilon\}.$$

Then L is a map of descent directions, i.e., for any $y \in \mathscr{X}_\epsilon$, $l \in L(y)$,

$$(22) \qquad \|l\| \approx \|g(y)\|$$

and there exists $\Delta_1 > 0$ such that

$$(23) \qquad g(y)^T l \le -\Delta_1 \|g(y)\| \|l\|.$$

In addition, there exists $\bar{\lambda}$ such that, if one of the two following conditions holds at each step k of Algorithm 3.1:
  (i) $\theta_k \gamma_k < \bar{\lambda}$,
  (ii) $(x^k, s^k) \in \mathscr{N}_\alpha$, with $\alpha \le 0.25$, and $\theta_k \gamma_k \in (0, 1]$,
then (21) is a descent algorithm with perturbation, i.e., there exists $\Delta > 0$ such that

$$p(y^{k+1}) \le p(y^k) - \theta_k \gamma_k \Delta \|l^k\| \|g(y^k)\|,$$

where $l^k = d^k P_{QD^k}\, d^k(y^k)^{-1}$.

PROOF. (a) We first check that L is a map of descent directions. In F_ϵ, we have $d \approx 1$. From Lemma 5.3, $\|dP_{QD}\, dy^{-1}\| \approx \|P_Q y^{-1}\| = \| - g(y)\|$, proving (22).
Given d, by definition of projection,

$$d^{-1}l - dy^{-1} \in \mathscr{R}(DQ^T).$$

Hence $d^{-2}l - y^{-1} \in \mathscr{R}(Q^T)$. Similarly, $g(y) + y^{-1} \in \mathscr{R}(Q^T)$, and it follows that

$$d^{-2}l + g(y) \in \mathscr{R}(Q^T).$$

But $l \in \mathcal{N}(Q)$, and multiplying the expression above by l^T,

$$-l^T D^{-2} l = g(y)^T l.$$

Since $d \approx 1$ and $\|l\| \approx \|g(y)\|$, there exists $\Delta_1 > 0$ such that $l^T D^{-2} l \geq \Delta_1 \|l\| \|g(y)\|$. This fact and the expression above establish (23).

(b) The condition of decrease is, in case (i), an immediate consequence of Lemma 4.5. It remains to study case (ii). We may drop the indices k. As $p(y)$ is a convex mapping it suffices to establish Lemma 6.4 when $\theta_k \gamma_k = 1$. The method of proof consists in comparing l and the Newton direction $\hat{l}$ for the minimization of $p(\cdot)$ in $\mathcal{Z}_e$. We have

$$\hat{l} = y P_{QY} y y^{-1}, \qquad l = d P_{QD} dy^{-1}.$$

To use Lemma 5.4 on shifted scalings, let us define

$$\hat{h} = y^{-1} \hat{l} = P_{QY} e, \qquad h = y^{-1} l = dy^{-1} P_{QD} dy^{-1},$$

and analyze the values of $\|dy^{-1} - e\|_\infty$. Since $d = \sqrt{\mu x/s} = x/\sqrt{w}$ and $\|x^{-1} - y^{-1}\| = O(\mu)$,

$$\|dy^{-1} - e\|_\infty = \|dx^{-1} - e\|_\infty + O(\mu) = \left\| \frac{1}{\sqrt{w}} - e \right\|_\infty + O(\mu).$$

But $(x, s) \in \mathcal{N}_\alpha$ means that $\|w - e\| \leq 0.25$ and consequently $w_i \geq 0.75$, and $|1/\sqrt{w_1} - 1| \leq 0.1548$. It follows that for μ sufficiently small, $\|dy^{-1} - e\|_\infty \leq 0.155$. From Lemma 5.4, we conclude that $h = \hat{h} + \rho$, with $\|\rho\| \leq 0.39\|\hat{h}\|$. We can now compute the variation of $p(\cdot)$ along l:

$$p(y + l) - p(y) = p(e + y^{-1} l) - p(e) = p(e + y^{-1} l) = p(e + h).$$

The quadratic approximation of the barrier function gives us the following property, shown for instance in Gonzaga (1989): if $h \in \mathcal{N}(Q)$ and $\|h\| < 1$ then

$$p(e + h) \leq -e^T h + \frac{\|h\|^2}{2} + \frac{1}{3} \frac{\|h\|^3}{1 - \|h\|}.$$

We know that $\|\hat{h}\| = \delta(y) \leq \delta(x, s) + O(\mu) \leq 0.25 + O(\mu)$, so that $\|h\| \leq \|\hat{h}\| + \|\rho\| \leq 1.4\|\hat{h}\| \leq 0.35 + O(\mu)$. In particular, for k large enough $\|h\| < 1$ and the formula above applies.

Setting $h = \hat{h} + \rho$, $e^T h = e^T P_{QY} e = \|\hat{h}\|^2$,

$$p(e + h) \leq -\|\hat{h}\|^2 - e^T \rho + \frac{\|\hat{h}\|^2 + \|\rho\|^2 + 2\hat{h}^T \rho}{2} + \frac{1}{3} \frac{\|h\|^3}{1 - \|h\|}.$$

Let us check that $\rho \in \mathcal{N}(QY)$. As $\hat{h} \in \mathcal{N}(QY)$ we have

$$QY_\rho = QY(h - \hat{h}) = QYh = QY dy^{-1} P_{QD} dy^{-1} = QD P_{QD} dy^{-1} = 0.$$

Using $\hat{h} = P_{QY}e$, it follows that $\hat{h}^T\rho = e^T\rho$. We deduce

$$p(e + h) \leq -\frac{\|\hat{l}\|^2}{2} + \frac{\|\rho\|^2}{2} + \frac{1}{3}\frac{\|h\|^3}{1 - \|h\|}$$

$$\leq \left(-\frac{1}{2} + \frac{0.16}{2} + \frac{1}{3}\frac{\|h\|}{1 - \|h\|}\frac{\|h\|^2}{\|\hat{h}\|^2}\right)\|\hat{h}\|^2.$$

Using again $\|\hat{h}\| \leq 0.25 + O(\mu)$ and $\|h\| \leq 1.4\|\hat{h}\| \leq 0.35 + O(\mu)$, we obtain for μ sufficiently small,

$$p(e + h) \leq -0.06\|h\|^2.$$

From (22) and $y \approx 1$ we deduce that $\|h\| \geq \Delta\|g\|$ for some positive Δ. With the above relation we get the conclusion.

Proof of main result. We now prove Theorem 3.2.

PROOF. Let $\{(x^k, s^k)\}$ be a sequence generated by Algorithm 3.1. Under our hypotheses, $\mu_k \to 0$, and consequently $s^k \to 0$. It remains to study the convergence of (x^k).

For k sufficiently large, we have $\mu_k < \bar{\mu}$ as in Lemma 6.1, and then $y^k = y(x^k) \in \mathscr{X}_e^{\circ} \subset \mathscr{X}^{\circ}$ as in (18). It also follows that (y^k) and (x^k) have the same limit points.

By Lemma 6.4, there exists $\bar{\lambda} > 0$ such that under the conditions of the main theorem, (21) defines a descent algorithm with perturbations for the barrier function in the optimal face. The perturbations are $\eta_k := O(\mu_k)$, which satisfy $\sum_{k=0}^{\infty}\|\eta^k\| < \infty$ as a consequence of the R-linear convergence of (μ_k). The sequence (y^k) is compact because it is in cl $\mathscr{X}_e^{\circ} \subset \mathscr{X}^{\circ}$ by (19).

Hence we can apply Lemma 4.3. The analytic center of the primal optimal face x^* is the unique stationary point of the barrier function $p(\cdot)$. We conclude that under the hypotheses of the main theorem, either $\sum_{k=0}^{\infty}\theta_k\gamma_k = +\infty$ and $y^k \to x^*$, or $\sum_{k=0}^{\infty}\theta_k\gamma_k < +\infty$ and (y^k) converges to a point $\bar{x} \in \mathscr{X}$, possibly $\bar{x} \neq x^*$. Since (y^k) and (x^k) have the same limit points, this completes the proof.

Concluding remarks. We essentially proved that algorithms that either operate in the neighborhood $\mathscr{N}_\alpha$ or use moderate centering, generate convergent sequences of iterates.

In all cases, centering results in directions of descent for the barrier function in the optimal face. Note that the primal-dual direction results in a centering direction in $\mathscr{X}$ that generally differs from the Newton centering direction by an amount determined by the vector w. As a result, one can only expect linear convergence of the iterates, even when they converge to the central optimum.

If the convergence is fast, i.e., the series of centering steplengths is bounded, then it is expected that $x^k \to \bar{x} \neq x^*$, due to Lemma 4.2. Even if the series is unbounded but $\theta_k\gamma_k$ becomes very small, one should expect a slow convergence of the iterates to the central optimum, while the gap converges to zero very fast.

Algorithms that generate nonconvergent sequences of iterates must be methods that operate in large neighborhoods of the central path, and take medium sized steps ($\theta_k\gamma_k$ bounded away from zero in some subsequence). This includes algorithms that may be very efficient in practice (see Gonzaga 1991), such as predictor-corrector algorithms based on other neighborhoods of the central path.

References

Gonzaga, C. C. (1989). Conical projection algorithms for linear programming, *Math. Programming* **43** 151–173.

——— (1991). Large steps path-following methods for linear programming. Part I: Barrier function method, *SIAM J. Optim.* **1** 268–279.

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

———, R. A. Tapia (1992). *On the convergence of the Mizuno-Todd-Ye algorithm to the analytic center of the solution set*, Technical Report, TR-92-36, Dept. of Mathematical Sciences, Rice University, Houston, TX 77251, USA.

Güler, O. (1993). *Generalized linear complementarity problems and interior point algorithms for their solutions.* (Manuscript).

Ji, J., F. Potra, S. Huang (1995). A predictor-corrector method for linear complementarity problems with polynomial complexity and superlinear convergence, *J. Optim. Theory Appl.* **84** (to appear).

———, ———, R. A. Tapia, Y. Zhang (1991). *An interior-point method with polynomial complexity and superlinear convergence for linear complementarity problems*, Technical Report, TR-91-23, Dept. of Mathematical Sciences, Rice University, Houston, TX 77251, USA.

Kojima, M., Y. Kurita, S. Mizuno (1991). *Large-step interior point algorithms for linear complementarity problems.* Research Reports on Information Sciences, Ser. B: Operations Research, B-243, Dept. of Information Sciences, Tokyo Institute of Technology, Oh-Okayama, Meguro-ku, Tokyo 152, Japan.

———, N. Megiddo, T. Noma, (1989). *Homotopy continuation methods for complementarity problems.* Research Report, RJ 6638 (63949), IBM Almaden Research Center, San Jose, CA 95120-6099, USA.

———, ———, ———, A. Yoshise (1991). A unified approach to interior point algorithms for linear complementarity problems, Lecture Notes in Computer Science, 538, Springer Verlag, Berlin, Germany.

———, S. Mizuno, T. Noma (1990). Limiting behavior of trajectories by a continuation method for monotone complementarity problems, *Math. Oper. Res.* **15** 662–675.

———, ———, A. Yoshise, (1989a). A primal-dual interior point algorithm for linear programming, N. Megiddo, ed, *Progress in Mathematical Programming: Interior Point and Related Methods*, Springer Verlag, New York, 29–47.

———, ———, ——— (1989b). A polynomial-time algorithm for a class of linear complementarity problems, *Math. Programming* **44** 1–26.

McShane, K. A. (1991). *A superlinearly convergent $O(\sqrt{n}\,L)$ iteration primal-dual linear programming algorithm*, Manuscript, U.S. Government, 2537 Villanova Drive, Vienna, Virginia, USA.

——— (1994). A superlinearly convergent $O(\sqrt{n}\,L)$ iteration interior-point algorithms for linear programming and the monotone linear complementarity problem, *SIAM J. Optim* **4** 247–261.

Megiddo, N. (1989). Pathways to the optimal set in linear programming, N. Megiddo, ed, *Progress in Mathematical Programming: Interior Point and Related Methods*, Springer Verlag, New York, 131–158. Identical version in: *Proceedings of the 6th Mathematical Programming Symposium of Japan*, Nagoya, Japan, 1–35, 1986.

Mehrotra, S. (1991). *Quadratic convergence in a primal-dual method*, Technical Report, 91-15, Dept. of Industrial Engineering and Management Science, Northwestern University, Evanston, IL 60208, USA.

Mizuno, S., M. J. Todd, Y. Ye (1993). *On adaptive-step primal-dual interior-point algorithms for linear programming*, *Math. Oper. Res.* **18** 945–963.

Monteiro, R. D. C., I. Adler (1989). Interior path following primal-dual algorithms: Part I: Linear programming, *Math. Programming* **44** 27–41.

———, T. Tsuchiya (1992). *Limiting behavior of the derivatives of certain trajectories associated with a monotone horizontal linear complementarity problem*, Working Paper, 92-28, December, Dept. of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721, USA.

Roos, C., J. P. Vial (1989). Long steps with the logarithmic penalty barrier function in linear programming, *Economic Decision-Making: Games, Economics and Optimization*, dedicated to J. H. Dreze, J. Gabszevwicz and J. F. Richard and L. Wolsey, Elsevier Science Publisher B.V., Amsterdam, The Netherlands, 433–441.

Ye, Y., K. M. Anstreicher, (1993). On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for the linear complementary problem, *Math. Programming* **62** 537–551.

___, O. Güler, R. A. Tapia, Y. Zhang (1993). A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming, *Math. Programming* **59** 151–162.

___, R. A. Tapia, Y. Zhang (1991). *A superlinearly convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming*, Technical Report, TR-91-22, Dept. of Mathematical Sciences, Rice University, Houston, TX 77251, USA.

Zhang, Y., R. A. Tapia (1991). *A polynomial-time and superlinearly convergent interior point algorithm for linear programming*, Technical Report, March, Dept. of Mathematical Sciences, Rice University, Houston, TX 77251, USA, *SIAM J. Optim.* (to appear).

J. F. Bonnans: INRIA, B.P. 105, 78153 Rocquencourt, France; e-mail: Frederic.Bonnans@inria.fr

C. C. Gonzaga: Department of Mathematics, Federal University of Santa Catarina, Cx Postal 476, 88010-970 Florianópolis, SC, Brazil