INSTITUTE FOR SYSTEMS RESEARCH

TECHNICAL RESEARCH REPORT

A Simple, Quadratically Convergent Interior Point Algorithm for Linear Programming and Convex Quadratic Programming

by A.L. Tits and J.L. Zhou

T.R. 93-53

The Institute for Systems Research is supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

A Simple, Quadratically Convergent Interior Point Algorithm for Linear Programming and Convex Quadratic Programming[1]

André L. Tits
Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

Jian L. Zhou
Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

Abstract

An algorithm for linear programming (LP) and convex quadratic programming (CQP) is proposed, based on an interior point iteration introduced more than ten years ago by J. Herskovits for the solution of nonlinear programming problems. Herskovits' iteration can be simplified significantly in the LP/CQP case, and quadratic convergence from any initial point can be achieved. Interestingly the resulting algorithm is closely related to a popular scheme, proposed in 1989 by Kojima et al. independently of Herskovits' work.

Keywords: Linear programming, quadratic programming, global convergence, quadratic convergence.

1 Introduction

In 1982, J. Herskovits proposed an interior point approach to the solution of nonlinear constrained optimization problems, based on a novel idea [1–3] (whose rudiments can be found in [4]). Herskovits' algorithm roughly proceeds as follows. Given a strictly

[1]This research was supported in part by NSF's Engineering Research Centers Program No. NSFD-CDR-88-03012 and by NSF grant No. DMC-88-15996.

feasible point, a descent direction d_0 is obtained by solving a linear system in (x, λ) (primal-dual pair) part of which can be viewed as a quasi-Newton iteration for the solution x of $\nabla L(x, \lambda) = 0$, using a positive definite estimate of the Hessian of the Lagrangian function L. As this direction may not allow a large step to be taken within the feasible set, a modified search direction d is then computed by solving a linear system with identical matrix and perturbed (by a function of d_0) right hand side. In [3] global and local superlinear convergence of this algorithm is proved under certain conditions (see also [5]).

In [6] it is observed that Herskovits' iteration can be viewed as a quasi-Newton iteration in the pair (x, λ) on all the equalities in the KKT conditions of optimality, modified to preserve strict primal and dual feasibility. It had been pointed out by McCormick that the corresponding Jacobian matrix is nonsingular in the neighborhood of any strong minimizer and thus that the pure Newton iteration converges locally with a quadratic rate [7]; Polak and Tits had then proposed to force global convergence by combining this Newton iteration with a first order method [8]. Herskovits' observation that, under the condition of primal and dual feasibility and if a positive definite estimate of the Hessian of the Lagrangian is used, d_0 is a descent direction for the objective function, suggests a natural globalization of the corresponding quasi-Newton iteration. This is fully explored in [6], where an algorithm is proposed which, under mild assumptions, enjoys global and local superlinear convergence (in particular, Maratos-type effects are avoided).

While Herskovits' iteration was devised for the solution of general nonlinear programming problems, it can be applied, in a simplified form, to problems with linear constraints, in particular, to linear and quadratic programming problems. Applicability to linear programming was first pointed out by Herskovits [9,10]. When all constraints are linear, d_0 is appropriate as a search direction, removing the need to solve a perturbed linear system. Interestingly, in the linear programming case, the matrix in Herskovits' iteration is identical to that appearing in the path-following Newton iteration proposed independently by Kojima et al. [11]; the two right-hand sides become identical as well if Kojima et al.'s barrier parameter is set to zero, as was subsequently suggested by Monteiro et al. [12].

In this paper, we present and analyze a version of Herskovits' iteration tailored at linear programming (LP) and convex quadratic programming (CQP). It is directly inspired from the algorithm in [6], with the following modifications: (i) only one linear system of equations is solved at each iteration, (ii) the exact Hessian (possibly singular) is used, (iii) the iteration is adjusted to allow for quadratic convergence to take place. Under dual nondegeneracy assumption, we prove global convergence to the solution set (somewhat stronger than the result proved in [6] for the nonlinear case). Under second order sufficiency conditions with strict complementarity (in the LP case, this is identical to primal nondegeneracy), the local convergence rate is quadratic. Most proofs are given in the appendix.

2 Problem Definition and Algorithm Statement

Consider the problem

$$(P) \qquad \text{minimize } \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \ \text{ s.t. } \ Ax \leq b$$

where $H = H^T \geq 0$. If $H = 0$, (P) is a linear programming problem; otherwise it is a convex quadratic programming problem. Let $I = \{1, \cdots, m\}$, where m is the number of rows of A, and, for $i \in I$, let a_i be the ith row of A, let b_i be the ith entry of b, and let $g_i(x) = a_i x - b_i$. Also let $f(x) = \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle$. The feasible set S is given by

$$S = \{x \in \mathbb{R}^n : g_i(x) \leq 0 \ \ \forall i \in I\},$$

the strictly feasible S^0 set by

$$S^0 = \{x \in \mathbb{R}^n : g_i(x) < 0 \ \ \forall i \in I\},$$

and the solution set S^* by

$$S^* = \{x^* \in S : f(x^*) \leq f(x) \ \ \forall x \in S\}.$$

A point $x^* \in S$ is said to be *stationary* for (P) if there exists $\lambda^* \in \mathbb{R}^m$ such that

$$\begin{aligned} Hx^* + c + A^T\lambda^* &= 0 \\ \lambda_i^* g_i(x^*) &= 0 \quad \forall i \in I \end{aligned} \tag{1}$$

(in particular, all vertices of S are stationary). If furthermore $\lambda^* \geq 0$, then x^* is a *KKT point* for (P), i.e., since (P) is convex, $x^* \in S^*$.

Let (x, μ) be an estimate of a KKT pair (x^*, λ^*) for (P) and let us substitute for the left-hand side of (1) its first order expansion around (x, μ) evaluated at $(x + d, \lambda)$, i.e., consider the linear system of equations in (d, λ)

$$\begin{aligned} Hx + c + Hd + A^T\lambda &= 0 \\ \mu_i a_i d + g_i(x)\lambda_i &= 0 \quad \forall i \in I. \end{aligned} \tag{2}$$

It will be shown that, under mild assumptions, if $x \in S$ and $\mu_i > 0$ for all $i \in I$, then the solution d of (2), if it is nonzero, is a feasible direction which is also a direction of descent for f, a useful property when seeking global convergence to S^*. However, it is clear from (2) that if $g_i(x) = 0$ with $\mu_i > 0$, then $a_i d = 0$ and thus $g_i(x + td) = 0$ for all t, and g_i will be satisfied as an equality for all subsequent iterations, possibly preventing convergence to S^*. Therefore, *strict* primal feasibility must be maintained at all times. Note that another favorable effect of primal and dual feasibility is that it implies that $a_i d < 0$ whenever $\lambda_i < 0$ so that the iterate will tend to move away from stationary points that are not solution points.

A pure Newton iteration for the solution of (1) amounts to selecting $x^+ = x+d$ and $\mu^+ = \lambda$ as next iterates, where (d, λ) solves (2). Under appropriate nondegeneracy assumptions, this iteration yields a local Q-quadratic rate of convergence in (x, μ). However, even close to a solution of (1), this iteration may not preserve primal and dual feasibility. It turns out that these apparently contradictory requirements can be reconciled, as it can be shown that the quadratic rate of convergence of Newton's method is preserved if each component of the next iterate is merely "close enough" either to the corresponding component of the solution of the nonlinear system of equations being solved, or to the corresponding component of the next iterate given by the pure Newton iteration. In our context, given $C > 0$,

$$\begin{aligned} x^+ &= x + (1 - C\|d\|)d \\ \mu_i^+ &= \max\{\lambda_i, \|d\|^2\}, \quad i \in I \end{aligned} \tag{3}$$

meets both requirements close to the solution of (1). Updates closely related to these are used in Algorithm A below.

The following two assumptions are made throughout.

Assumption A1. $S^0 \neq \emptyset$.

Assumption A2. S^* is nonempty and bounded.

Clearly Assumption A2 implies that the nullspaces of H and A have a trivial intersection, i.e.,

$$\{d : Hd = 0\} \cap \{d : Ad = 0\} = \{0\} \ . \tag{4}$$

Algorithm A.

Parameters. $\beta \in (0, 1)$, $\overline{\mu} > 0$.

Data. $x^0 \in S^0$, $\mu_i^0 > 0 \; \forall i \in I$.

Step 0. Initialization. Set $k = 0$.

Step 1. Computation of a feasible descent direction d^k. Let (d^k, λ^k) solves the linear system in (d, λ)

$$Hd + A^T\lambda = -(Hx^k + c) \tag{5}$$

$$\mu_i^k a_i d + g_i(x^k)\lambda_i = 0 \quad \forall i \in I. \tag{6}$$

If $d^k = 0$, stop.

Step 2. Updates.

(i) Compute the largest feasible stepsize

$$\bar{t}^k = \begin{cases} \infty & \text{if } \; a_i d^k \leq 0 \;\; \forall i \in I, \\ \min\{(-g_i(x^k)/a_i d^k) : a_i d^k > 0, i \in I\} & \text{otherwise.} \end{cases} \tag{7}$$

Set

$$t^k = \min\{\max\{\beta\bar{t}^k,\ \bar{t}^k - \|d^k\|\},\ 1\}. \tag{8}$$

Set $x^{k+1} = x^k + t^k d^k$.

(ii) If $\lambda_i^k \le g_i(x^k)$ for some $i \in I$, set

$$\mu_i^{k+1} = \mu_i^0,\ \forall i \in I. \tag{9}$$

Otherwise, set

$$\mu_i^{k+1} = \min\{\max\{\lambda_i^k,\ \|d^k\|^2\},\ \overline{\mu}\},\ \forall i \in I. \tag{10}$$

(iii) Set $k = k + 1$. Go to *Step 1.*

□

In Step 2(i), $\bar{t}^k$ is the maximum step preserving feasibility ($x^k + \bar{t}^k d^k \in S$) and the term $\beta\bar{t}^k$ ensures that t^k is positive even when $\|d^k\|$ is large. It will be shown that, for some $C' > 0$, $\bar{t}^k \ge 1 - C'\|d^k\|$ so condition (3) does hold. In Step 2(ii), (9) is introduced to prevent oscillation between non-KKT stationary points and $\overline{\mu}$ in (10) to ensure boundedness of $\{\mu^k\}$ (these features are used in the proofs of Lemma 3.2 and Lemma A.2; neither may be essential in practice).

The following result, proved in the appendix, shows that Step 1 in Algorithm A is well defined.

Lemma 2.1 Let $x \in S$ and $\mu \in \mathbb{R}^m$ be such that $\{a_i : g_i(x) = 0\}$ is a linearly independent set and $\mu_i > 0$ for all $i \in I$. Then

$$M(x,\mu) = \begin{bmatrix} H & A^T \\ \operatorname{diag}(\mu_i)A & \operatorname{diag}(g_i(x)) \end{bmatrix}$$

is nonsingular. In particular $M(x,\mu)$ is nonsingular if $x \in S^0$ and $\mu_i > 0$ for all $i \in I$.
□

Corollary. Algorithm A generates a uniquely defined sequence.

Remark 1. Perhaps more typical than formulation (P) in the LP/QP literature (see, e.g., [11,12]) is the formulation

$$\text{minimize } \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \quad \text{s.t.} \quad Ax = b,\ x \ge 0. \tag{11}$$

(We chose (P) instead for sake of simpler exposition.) Adapted to this problem, linear system (5)-(6) becomes

$$\begin{aligned} Hd + A^T\psi + \lambda &= -(Hx^k + c) \\ Ad &= 0 \\ \mu_i^k d_i + x_i^k \lambda_i &= 0 \end{aligned} \tag{12}$$

(where ψ is the updated multiplier vector for the equality constraints) and Algorithm A can be transposed accordingly.

Remark 2. As pointed out in the introduction, Monteiro et al. [12] have proposed an iteration for the LP case based on linear system (12) (with $H = 0$). The similarity with Algorithm A stops there, however, as the primal and dual update formulae in [12], when transposed to problem (P), are given by

$$\begin{aligned} x^{k+1} &= x^k + \alpha d^k \\ \mu^{k+1} &= \mu^k + \alpha(\lambda^k - \mu^k) \end{aligned}$$

(i.e., the same stepsize is used for primal and dual variables, and it is unchanged from iteration to iteration), with α a certain number less than $1/2$ (whereas our stepsize converges to 1). Under certain assumptions, Monteiro et al. prove polynomial convergence of their algorithm.

Remark 3. Monteiro et al. [12] point out that their iteration (in the LP case) can be thought of as an affine scaling iteration with scaling matrix (superscripts are left out for readability) $D = (\text{diag}(\mu_i)\text{diag}(x_i^{-1}))^{-1/2}$ (instead of $D = \text{diag}(x_i)$ in the scaling-steepest descent (SSD) method of Dikin's [13]; also see [14]). This analogy can still be made in the CQP case, with now $D = (H + \text{diag}(\mu_i)\text{diag}(x_i^{-1}))^{-1/2}$. Alternatively the SSD direction can be thought of as the direction d that Algorithm A (transposed to the framework of formulation (11)) would yield if $\text{diag}(\mu_i)$ were set to $\text{diag}(x_i^{-1}) - H\text{diag}(x_i)$, instead of being updated as in Step 2(ii). Clearly, letting $\mu^{k+1} \simeq \lambda^k$ as in Algorithm A accelerates convergence.

3 Convergence Analysis

First note that, if $d^k = 0$ for some k (i.e., if Algorithm A terminates at iteration k), then $\lambda^k = 0$ and $\nabla f(x^k) = Hx^k + c = 0$, thus x^k solves (P). In the sequel, it is assumed that Algorithm A generates an infinite sequence $\{x^k\}$.

3.1 Global Convergence

Given $x \in S$, we denote by $I(x)$ the index set of active constraints at x, i.e.

$$I(x) = \{i \in I : \; g_i(x) = 0\}.$$

Assumption A3. For all $x \in S$, $\{a_i : i \in I(x)\}$ is a linearly independent set.

We show that, under Assumptions A1-A3, the sequence $\{x^k\}$ converges to S^*, the set of solution points. First, at every iteration, the values of the objective function and of all constraint functions with negative multiplier estimates decrease.

Proposition 3.1. If $d^k \neq 0$, then

$$f(x^k + t^k d^k) = f(x^k) - t^k(1 - t^k/2)\langle d^k, Hd^k\rangle - t^k\langle \lambda^k, Ad^k\rangle < f(x^k) \tag{13}$$

and

$$g_i(x^k + t^k d^k) = g_i(x^k) + t^k a_i d^k < g_i(x^k) \quad \forall i \text{ s.t. } \lambda_i^k < 0. \tag{14}$$

Proof. See the appendix. □

Corollary. The sequence $\{x^k\}$ is bounded.

Proof. Assumption A2 implies that, given any $x^0 \in S$, the level set $\{x \in S : f(x) \leq f(x^0)\}$ is bounded. The claim then follows from the monotone decrease of $f(x^k)$. □

We first show that $\{x^k\}$ converges to stationary points of (P). The argument is a simplified version of that used in [6]. It is given here for ease of reference. The proofs of Lemmas 3.2 and 3.3 are given in the appendix.

Lemma 3.2. Let $x^* \in \mathbb{R}^n$ and suppose that K, an infinite index set, is such that $\{x^k\}$ converges to x^* on K. If $\{d^k\}$ converges to zero on K, then x^* is stationary and $\{\lambda^k\}$ converges to λ^* on K, where λ^* is the unique multiplier vector associated with x^*. □

Lemma 3.3. Let $x^* \notin S^*$ and suppose that K, an infinite index set, is such that $\{x^k\}$ converges to x^* on K. Then $\{d^k\}$ goes to zero on K. □

Proposition 3.4. $\{x^k\}$ converges to the set of stationary points of (P).

Proof. By contradiction. Suppose not. Then, since $\{x^k\}$ is bounded, there exists some infinite index set K and some x^* not stationary such that $x^k \to x^*$ as $k \to \infty$, $k \in K$. In view of Lemma 3.2, $\{d^k\}$ does not converge to zero on K. Thus there is $K' \subset K$ s.t. $\inf_{k \in K'} \|d^k\| > 0$. Since $x^k \to x^*$ as $k \to \infty$, $k \in K'$, this contradicts Lemma 3.3. Thus the claim holds. □

Now note that since f decreases at such iteration, if one limit point of $\{x^k\}$ is in S^*, then all of them are. Proceeding by contradiction, we assume that $\{x^k\}$ is bounded away from S^*.

Lemma 3.5. If $\{x^k\}$ is bounded away from S^*, then $\{d^k\} \to 0$.

Proof. By contradiction. Suppose there exists an infinite index set K such that $\inf_K \|d^k\| > 0$. Let $K' \subset K$, $x^* \in X$ be such that $x^k \to x^*$ as $k \to \infty$, $k \in K'$. Since $\{x^k\}$ is bounded away from S^*, it follows that $x^* \notin S^*$ which, in view of Lemma 3.3, leads to a contradiction. □

So far the proof has not made essential use of the structure of (P). In [6], however, convergence to KKT points could not be proven without the artificial assumption that stationary points are isolated. In the present context, no such assumption is needed. The following key lemma is proved in the appendix.

Lemma 3.6. Suppose $\{x^k\}$ is bounded away from S^*. Let x^* and x'^* be two limit points of $\{x^k\}$ and let λ^* and λ'^* be the associated multiplier vectors. Then $\lambda^* = \lambda'^*$. □

Theorem 3.7. $\{x^k\}$ converges to S^*.

Proof. Proceeding again by contradiction, suppose that some limit point of $\{x^k\}$ is not in S^* and thus, since f takes on the same value at all limit points of $\{x^k\}$, that $\{x^k\}$ is bounded away from S^*. In view of Lemma 3.5, $\{d^k\} \to 0$. Let λ^* be the common multiplier vector associated with all limit points of $\{x^k\}$ (see Lemma 3.6). A simple contradiction argument shows that Lemma 3.2 then implies that $\{\lambda^k\} \to \lambda^*$. Since $\{x^k\}$ is bounded away from S^*, $\lambda^* \not\geq 0$. Let $i_0 \in I$ be such that $\lambda^*_{i_0} < 0$. Then $\lambda^k_{i_0} < 0$ for all k large enough. Proposition 3.1 and strict feasibility of $\{x^k\}$ then imply that, for k large enough,

$$0 > g_{i_0}(x^k) > g_{i_0}(x^{k+1}) > \dots$$

contradicting the fact that $\{g_{i_0}(x^k)\} \to 0$. □

3.2 Local Rate of Convergence

Let x^* be a limit point of $\{x^k\}$ and let λ^* be the corresponding KKT multiplier vector. We now assume that the second order sufficiency conditions of optimality with strict complimentary holds at x^*, i.e.,

Assumption A4. $\langle d, Hd \rangle > 0$ for all d such that $a_i d = 0 \ \forall i \in I(x^*)$,

Assumption A5. $\lambda^*_i > 0$ for all $i \in I(x^*)$.

Assumptions A4 and A5 ensure that x^* is the unique solution of (P) (in the LP case $(H = 0)$, under Assumption A3, they are equivalent to uniqueness of the solution, i.e., to primal nondegeneracy). Thus $\{x^k\} \to x^*$. The following result is a minor variation on a result pointed out by McCormick in 1971 [7]. It is related to Lemma 2.1 (but Assumptions A4 and A5 are not in force in that lemma). For the sake of completeness, a proof is given in the appendix.

Proposition 3.8. Let $\mu^*_i = \min(\lambda^*_i, \overline{\mu})$. Then $M(x^*, \mu^*)$ is nonsingular. □

This result was used in [6] to show that, in the general nonlinear case, if the Hessian of the Lagrangian is suitably approximated, $\{x^k\}$ converges to x^* two-step superlinearly. If the Hessian of the Lagrangian at (x^*, λ^*) is positive definite over the entire space, Q-superlinear convergence would follow. Both of these results obviously apply to the present case since the exact Hessian is used. Here however, as announced in the introduction, the pair $\{(x^k, \mu^k)\}$ also converges Q-quadratically to (x^*, λ^*). This is shown now. First a preliminary result, also derived in [6] (again, it is proved in the appendix for ease of reference).

Lemma 3.9. (i) $\{d^k\} \to 0$ and $\{\lambda^k\} \to \lambda^*$; (ii) for k large enough $\{i : \lambda^k_i > 0\} = I(x^*)$; (iii) if $\lambda^*_i \leq \overline{\mu} \ \forall i \in I$, then $\{\mu^k\} \to \lambda^*$. □

To prove Q-quadratic convergence of $\{(x^k, \mu^k)\}$, the following property of Newton's method will be used. Its proof is given in the appendix.

Proposition 3.10. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be twice continuously differentiable and let $z^* \in \mathbb{R}^n$ and $\rho > 0$ be such that $F(z^*) = 0$ and $\frac{\partial F}{\partial z}(z)$ is nonsingular whenever $z \in B(z^*, \rho) := \{z : \|z^* - z\| \leq \rho\}$. Let $d^N : B(z^*, \rho) \to \mathbb{R}^n$ be defined by $d^N(z) = -\left(\frac{\partial F}{\partial z}(z)\right)^{-1} F(z)$. Then given any $c_1 > 0$ there exists $c_2 > 0$ such that

$$\|z^+ - z^*\| \leq c_2 \|z - z^*\|^2 \quad \forall z \in B(z^*, \rho) \tag{15}$$

for every $z \in B(z^*, \rho)$ and $z^+ \in \mathbb{R}^n$ for which, for each $i \in \{1, \ldots, n\}$, either

(i) $|z_i^+ - z_i^*| \leq c_1 \|d^N(z)\|^2 \quad \forall z \in B(z^*, \rho)$

or

(ii) $|z_i^+ - (z_i + d_i^N(z))| \leq c_1 \|d^N(z)\|^2 \quad \forall z \in B(z^*, \rho)$.

□

Theorem 3.11. If $\lambda_i^* \leq \overline{\mu}\ \forall i \in I$, then $\{(x^k, \mu^k)\}$ converges to (x^*, λ^*) Q-quadratically.

Proof. With reference to Proposition 3.10, let $\rho > 0$ be such that $M(x, \mu)$ is non-singular for all $(x, \mu) \in B((x^*, \lambda^*), \rho)$ (in view of Proposition 3.8, such ρ exists). Since $\{(x^k, \mu^k)\} \to (x^*, \lambda^*)$ as $k \to \infty$, there exists k_0 such that $(x^k, \mu^k) \in B((x^*, \lambda^*), \rho)$ for all $k \geq k_0$. Now let us first consider $\{\mu^k\}$. For $i \in I(x^*)$, $\mu_i^{k+1} = \lambda_i^k$ for k large enough, so that condition (ii) in Proposition 3.10 holds for k large enough. For $i \notin I(x^*)$, for each k either again $\mu_i^{k+1} = \lambda_i^k$ or $\mu_i^{k+1} = \|d^k\|^2$. In the latter case, since $\lambda_i^* = 0$, condition (i) in Proposition 3.10 holds. Next, consider $\{x^k\}$. For $i \notin I(x^*)$,

$$\frac{|\mu_i^k|}{|\lambda_i^k|} = \frac{|g(x_i^k)|}{|a_i d^k|} \to \infty \quad \text{as} \quad k \to \infty.$$

Thus, if $I(x^*) = \emptyset$, then, in view of Step 2 (i) in Algorithm A, $t^k = 1$ for k large enough and otherwise

$$\bar{t}^k = \min\{\mu_i^k / \lambda_i^k : \ i \in I(x^*)\}$$

and

$$t^k = \min\{1, \frac{\mu_{i_k}^k}{\lambda_{i_k}^k} - \|d^k\|\} \tag{16}$$

for k large enough, for some $i_k \in I(x^*)$. In particular, t^k converges to 1. Thus, for k large enough and some $i_k \in I(x^*)$

$$\|x^{k+1} - (x^k + d^k)\| = |t^k - 1| \|d^k\|$$

$$\leq |\ \|d^k\| + \frac{\lambda_i^k - \mu^k}{\lambda_i^k} |\ \|d^k\|.$$

Since $\lambda_i^* > 0$ for all $i \in I(x^*)$, it follows that for some $C > 1$ and all k large enough

$$\begin{aligned} \|x^{k+1} - (x^k + d^k)\| &\leq (\|d^k\| + C\|\lambda^k - \mu^k\|)\|d^k\| \\ &\leq (1 + C)(\|d^k\| + \|\lambda^k - \mu^k\|)^2 \end{aligned}$$

Thus condition (ii) of Proposition 3.10 holds. The claim then follows from Lemma 3.9 and Proposition 3.10. □

4 Discussion

A simple algorithm, inspired from an iteration due to Herskovits, has been proposed for the solution of linear programming and convex quadratic programming problems. Under nondegeneracy assumptions, convergence is global, with a local Q-quadratic rate. The importance of a fast rate of convergence, even for problems that can be solved in finitely many iterations (such as LP and CQP problems) has been stressed by several authors (e.g., [15,16] and references therein) and recently an interior point method has been shown to be both polynomial-time and Q-quadratically convergent [17]. The algorithm proposed here however has the advantage of greater simplicity (no barrier parameter to be iteratively adjusted).

5 Appendix: Proof of Some Results

Proof of Lemma 2.1. Let (d, λ) be such that $M(x, \mu)\begin{pmatrix} d \\ \lambda \end{pmatrix} = 0$. Thus

$$Hd + A^T\lambda = 0 \tag{17}$$

$$\text{diag }(\mu_i)Ad + \text{ diag }(g_i(x))\lambda = 0. \tag{18}$$

Taking the inner product of both sides of (17) by d yields

$$\langle d, Hd\rangle + \langle d, A^T\lambda\rangle = 0. \tag{19}$$

Since $\mu_i > 0$ for all $i \in I$, left multiplying both sides of (18) by $\text{diag}(\mu_i^{-1})$ and taking the inner product with λ yields

$$\langle\lambda, Ad\rangle + \langle\lambda, \text{diag }(g_i(x)/\mu_i)\lambda\rangle = 0. \tag{20}$$

From (19) and (20) we get

$$\langle d, Hd\rangle = \langle\lambda, \text{diag }(g_i(x)/\mu_i)\lambda\rangle,$$

in which the left hand side is nonnegative and the right hand side nonpositive. Thus $Hd = 0$ and $\text{diag}(g_i(x))\lambda = 0$. In view of (18) the latter implies that $Ad = 0$. It then follows from (4) that $d = 0$ and, from (17), $A^T\lambda = 0$. This together with Assumption A3 and the fact that $\text{diag}(g_i(x))\lambda = 0$ implies that $\lambda = 0$.

Proof of Proposition 3.1. The proof will make use of the following lemma.

Lemma A.1. Let $x \in S$, $\mu \in \mathbb{R}^m$ such that $\mu_i > 0$ for all $i \in I$, and let (d, λ) satisfy

$$\text{diag }(\mu_i)Ad + \text{diag }(g_i(x))\lambda = 0 \tag{21}$$

Then (i) $\langle \lambda, Ad \rangle \geq 0$ and (ii) $\langle \lambda, Ad \rangle = 0$ only if $Ad = 0$.

Proof. Left multiplying both sides of (21) by diag (μ_i^{-1}) and taking the inner product with λ yields

$$\langle \lambda, Ad \rangle + \langle \lambda, \text{diag } (g_i(x)/\mu_i)\lambda \rangle = 0.$$

Thus $\langle \lambda, Ad \rangle \geq 0$. Moreover, if $\langle \lambda, Ad \rangle = 0$ then $\lambda_i = 0$ for every i for which $g_i(x) \neq 0$. Together with (21), this implies that $Ad = 0$.

Proof of Proposition 3.1. Since f is quadratic, it follows from (5) that

$$\begin{aligned} f(x^k + t^k d^k) &= f(x^k) + t^k \langle \nabla f(x^k), d \rangle + \frac{1}{2}(t^k)^2 \langle d^k, Hd^k \rangle \\ &= f(x^k) - t^k \left(\langle d^k, Hd^k \rangle + \langle d^k, A^T \lambda^k \rangle \right) + \frac{1}{2}(t^k)^2 \langle d^k, Hd^k \rangle \\ &= f(x^k) - t^k(1 - t^k/2) \langle d^k, Hd^k \rangle - t^k \langle \lambda^k, Ad^k \rangle. \end{aligned}$$

Next, since $H \geq 0$ and since, by construction, $t_k \in (0, 1]$, it follows from Lemma A.1(i) that

$$f(x^k) - t^k(1 - t^k/2)\langle d^k, Hd^k \rangle - t^k \langle \lambda^k, Ad^k \rangle \leq f(x^k),$$

with equality holding only if $\langle d^k, Hd^k \rangle = \langle \lambda^k, Ad^k \rangle = 0$. In view of Lemma A.1(ii) and of (4) and since $d^k \neq 0$, this cannot occur, thus (13) is proved. Next, since g is linear,

$$g_i(x^k + t^k d^k) = g_i(x^k) + t^k a_i d^k \quad i = 1, \dots, m.$$

Since $\mu_i^k > 0$ for all $i \in I$, it follows from (6) that $a_i d^k < 0$ whenever $\lambda_i^k < 0$, proving (14).

Proof of Lemma 3.2. Suppose $\{d^k\} \to 0$ as $k \to \infty$, $k \in K$. Since $\{\mu^k\}$ is bounded, it follows from (6) that for all i for which $g_i(x^*) < 0$, $\{\lambda_i^k\} \to 0$ as $k \to \infty$, $k \in K$. Since in view of (5), $\{A^T \lambda^k\}$ converges on K, it follows from Assumption A3 that $\{\lambda^k\}$ converges on K, say to λ^*. Taking limits in (5)-(6) then yields

$$Hx^* + c + A^T \lambda^* = 0$$

$$\lambda_i^* g_i(x^*) = 0, \ i = 1, \dots, m,$$

implying that x^* is stationary, with multiplier vector λ^*.

Proof of Lemma 3.3. Let $J^k = \{i \in I : \lambda_i^k \leq g_i(x^k)\}$. The proof will make use of the following result.

Lemma A.2. Let K be an infinite index set such that

$$\inf\{\|d^{k-1}\| : \ k \in K, \ J^{k-1} = \emptyset\} > 0.$$

Then $\{d^k\} \to 0$ as $k \to \infty$, $k \in K$.

Proof. In view of (9) and (10), for all $i \in I$, μ_i^k is bounded away from zero on K. Proceeding by contradiction, assume that, for some infinite index set $K' \subset K$,

$\inf_{k\in K'} \|d^k\| > 0$. Since $\{x^k\}$ and $\{\mu^k\}$ are bounded, we may assume, without loss of generality, that for some x^* and μ^*, with $\mu_i^* > 0$ for all i,

$$\begin{aligned} \{x^k\} &\to x^* \quad \text{as} \quad k \to \infty, \quad k \in K' \\ \{\mu^k\} &\to \mu^* \quad \text{as} \quad k \to \infty, \quad k \in K'. \end{aligned}$$

Since in view of Lemma 2.1 and Assumption A3, $M(x^*, \mu^*)$ is nonsingular, it follows that, for some d^* and λ^*, with $d^* \neq 0$ (since $\inf_{k\in K'} \|d^k\| > 0$),

$$\begin{aligned} \{d^k\} &\to d^* \quad \text{as} \quad k \to \infty, \quad k \in K', \\ \{\lambda^k\} &\to \lambda^* \quad \text{as} \quad k \to \infty, \quad k \in K'. \end{aligned}$$

From (6) and (7) it follows that

$$\bar{t}^k = \begin{cases} \infty & \text{if } \lambda_i \leq 0 \ \ \forall i \in I, \\ \min\{(\mu_i^k/\lambda_i^k) : \lambda_i^k > 0, \ i \in I\} & \text{otherwise.} \end{cases}$$

Since, on K', $\{\lambda^k\}$ is bounded and, for each $i \in I$, $\{\mu_i^k\}$ is bounded away from zero, it follows that $\bar{t}^k$ is bounded away from zero on K', and so is t^k (Step 2 (i) in Algorithm A). On the other hand, in view of Lemma A.1 (i) and Proposition 3.1, it follows that

$$\frac{d}{dt} f(x^k + td^k) = -\Big((1-t)\langle d^k, Hd^k\rangle + \langle \lambda^k, Ad^k\rangle\Big) \leq 0 \quad \forall t \in [0,1].$$

Since $t^k \in (0,1]$ for all k, it then follows from Proposition 3.1 that there exists $\underline{t} > 0$ such that

$$f(x^k + t^k d^k) \leq f(x^k) - (\underline{t} - \frac{1}{2}\underline{t}^2)\langle d^k, Hd^k\rangle - \underline{t}\langle \lambda^k, Ad^k\rangle \quad \forall k \in K'.$$

Now, in view of (4), either (or both) $\langle d^*, Hd^*\rangle > 0$ or $Ad^* \neq 0$. In the latter case, in view of Lemma A.1, since $\mu_i^* > 0$ for all $i \in I$, $\langle \lambda^*, Ad^*\rangle > 0$, so $\langle \lambda^k, Ad^k\rangle > 0$ for k large enough, $k \in K'$. Since, in view of Proposition 3.1, $f(x^k)$ is monotonic nonincreasing, it follows that $f(x^k) \to -\infty$ as $x^k \to x^*$, contradicting continuity of f.

Proof of Lemma 3.3. Let us again proceed by contradiction, i.e., suppose $\{d^k\}$ does not converge to zero as $k \to \infty$, $k \in K$. In view of Lemma A.2, there exists an infinite index set $K' \subset K$ such that $J^{k-1} = \emptyset$ for all $k \in K'$ and $\inf_{k\in K'} \|d^{k-1}\| = 0$. Without loss of generality, assume that $\{d^{k-1}\} \to 0$ as $k \to \infty$, $k \in K'$. Since $\{x^k\} \to x^*$ as $k \to \infty$, $k \in K$ and $\|x^k - x^{k-1}\| = \|t^{k-1}d^{k-1}\| \leq \|d^{k-1}\|$, it follows that $\{x^{k-1}\} \to x^*$ as $k \to \infty$, $k \in K'$ which implies, in view of Lemma 3.2, that x^* is stationary and $\{\lambda^{k-1}\} \to \lambda^*$ as $k \to \infty$, $k \in K'$, where λ^* is the corresponding multiplier vector. Since $J^{k-1} = \emptyset$ for all $k \in K'$, it follows that $\lambda_i^* \geq 0$ for all $i \in I(x^*)$, thus $x^* \in S^*$, a contradiction.

Proof of Lemma 3.6. Let L be the set of limit points of $\{x^k\}$ (in view of Proposition 3.4, all of these are stationary points of (P)). We first prove three auxiliary lemmas.

Lemma A.3. Let $x, x' \in L$ be such that $I(x) = I(x')$. Then $H(x - x') = 0$.

Proof. Let $I^* := I(x)(= I(x'))$. Then for some λ, λ',

$$Hx + c + \sum_{i \in I^*} \lambda_i a_i^T = 0,$$

$$Hx' + c + \sum_{i \in I^*} \lambda_i' a_i^T = 0,$$

which implies that, for all $\alpha \in (0,1)$,

$$Hx_\alpha + c + \sum_{i \in J} \lambda_{\alpha,i} a_i^T = 0, \tag{22}$$

with $x_\alpha = (1-\alpha)x + \alpha x'$ and $\lambda_\alpha = (1-\alpha)\lambda + \alpha\lambda'$. Now $I^* = I(x) = I(x')$ implies that $a_i(x' - x) = 0$ for all $i \in I^*$ which, together with (22), implies that

$$\langle Hx_\alpha + c, x' - x\rangle = 0 \quad \forall \alpha \in (0,1).$$

Since $x_\alpha = x + \alpha(x' - x)$, we get, for all $\alpha \in (0,1)$,

$$\langle Hx + \alpha H(x' - x) + c, (x' - x)\rangle = 0,$$

$$\langle Hx + c, x' - x\rangle + \alpha\langle x' - x, H(x' - x)\rangle = 0.$$

Thus $\langle x' - x, H(x' - x)\rangle = 0$. Since $H \geq 0$, the claim follows.

Lemma A.4. If $\{x^k\}$ is bounded away from S^*, then, for all $x, x' \in L$, $H(x' - x) = 0$.

Proof. Since there are only finitely many possible combinations of binding constraints, in view of Lemma A.3, L is a finite union of "parallel" affine sets of the form $L \cap (x + \mathcal{N}(H))$, with $x \in L$, where $\mathcal{N}(H) = \{d : \ Hd = 0\}$. In view of Lemma 3.3, all these affine sets must be identical, which proves the claim.

Lemma A.5. If $\{x^k\}$ is bounded away from S^*, then L is connected.

Proof. Suppose not. Then there exists $E, F \subset \mathbb{R}^n$, both nonempty, such that $L = E \cup F$, $\overline{E} \cap F = \emptyset$, $E \cap \overline{F} = \emptyset$. Since L is compact E and F must be compact. Thus $\delta := \min\limits_{x \in E, x' \in F} \|x - x'\| > 0$. A simple contradiction argument using the fact that $\{x^k\}$ is bounded shows that, for k large enough, $\min_{x \in L} \|x^k - x\| \leq \delta/3$, i.e., either $\min_{x \in E} \|x^k - x\| \leq \delta/3$ or $\min_{x \in F} \|x^k - x\| \leq \delta/3$. Moreover, since both E and F are nonempty (i.e., contain limit points of $\{x^k\}$), each of these situations occurs infinitely many times. Thus $K := \{k : \min_{x \in E} \|x^k - x\| \leq \delta/3, \min_{x \in F} \|x^{k+1} - x\| \leq \delta/3\}$ is an infinite index set and $\|d^k\| \geq \delta/3 > 0$ for all $k \in K$. On the other hand since $\{x^k\}_{k \in K}$ is bounded and bounded away from S^*, it has some limit point $x^* \notin S^*$. In view of Lemma 3.3, this is a contradiction.

Proof of Lemma 3.6. Given any $x \in L$, let $\lambda(x)$ be the multiplier vector associated with x and let $J(x)$ be the index set of "binding" constraints at x, i.e.,

$$J(x) = \{i \in I : \lambda_i(x) \neq 0\}.$$

We first show that, if $x, x' \in L$ are such that $J(x) = J(x')$, then $\lambda(x) = \lambda(x')$. Indeed, in view of (1), it follows from Lemma A.4 that

$$\sum_{j \in J(x)} \lambda_j(x) a_j^T = \sum_{j \in J(x)} \lambda_j(x') a_j^T,$$

and the claim follows from linear independence of $\{a_j^T : j \in I(x)\}$ (Assumption A3) and from the fact that $J(x) \subset I(x)$. To conclude the proof, we show that, for any $x, x' \in L$, $J(x) = J(x')$. Let $\tilde{x} \in L$ be arbitrary, let $J = J(\tilde{x})$, and let $E := \{x \in L : J(x) = J(\tilde{x})\}$ and $F := \{x \in L : J(x) \neq J(\tilde{x})\}$. Proceeding by contradiction, suppose that F is nonempty. Since L is connected (Lemma A.5), $\overline{E} \cap F \neq \emptyset$. Let $\hat{x} \in \overline{E} \cap F$ and let $\{y^\ell\} \to \hat{x}$, $\{y^\ell\} \subset E$. Since $J(y^\ell) = J$ for all ℓ, it follows from the first part of this proof that $\lambda(y^\ell) = \lambda$ for all ℓ for some λ, i.e,

$$Hy^\ell + c + A^T\lambda = 0 \quad \forall \ell.$$

Thus

$$H\hat{x} + c + A^T\lambda = 0.$$

Also, for all ℓ, $g_j(y^\ell) = 0$ for all j such that $\lambda_j \neq 0$, so that $g_j(\hat{x}) = 0$ for all j such that $\lambda_j \neq 0$. It follows that $\lambda(\hat{x}) = \lambda$ and thus $J(\hat{x}) = J(\tilde{x})$, contradicting that fact that $\hat{x} \in F$.

Proof of Proposition 3.8. Let (d, λ) be such that $M(x^*, \mu^*) \begin{pmatrix} d \\ \lambda \end{pmatrix} = 0$. Thus

$$Hd + A^T\lambda = 0 \tag{23}$$

$$\mathrm{diag}(\mu_i^*)Ad + \mathrm{diag}(g_i(x^*))\lambda = 0. \tag{24}$$

In view of Assumption A5, (24) implies that

$$a_i d = 0 \quad \forall i \in I(x^*) \tag{25}$$

and, since $\mu_i^* = \lambda_i^* = 0$ when $i \notin I(x^*)$,

$$\lambda_i = 0 \quad \forall i \notin I(x^*). \tag{26}$$

It follows that

$$\langle d, A^T\lambda \rangle = \langle Ad, \lambda \rangle = 0.$$

Taking the inner product of both sides of (23) with d thus yields

$$\langle d, Hd \rangle = 0,$$

i.e., since $H \geq 0$,

$$Hd = 0.$$

In view of (25) and Assumption A4, it follows that $d = 0$. Finally, it follows from (23), (26) and Assumption A3 that $\lambda = 0$.

Proof of Lemma 3.9. We first prove by contradiction that $\{d^k\} \to 0$. Thus suppose that there exists an infinite index set K such that $\inf_K \|d^k\| > 0$. In view of Lemma A.2, there exists an infinite index set $K' \subset K$ such that $\{d^{k-1}\} \to 0$ as $k \to \infty$, $k \in K'$ and, for all $k \in K'$, $\lambda_i^{k-1} > g_i(x^{k-1})$ for all $i \in I$. It then follows from Lemma 3.2 that $\{\lambda^{k-1}\} \to \lambda^*$ as $k \to \infty$, $k \in K'$ and from the update rule for μ^k in Step 2 (ii) of Algotihm A that, for all $i \in I$, $\mu_i^k \to \mu_i^* = \min(\lambda_i^*, \overline{\mu})$ as $k \to \infty$, $k \in K'$. In view of Proposition 3.8, $M(x^*, \mu^*)$ is nonsingular and thus, since $x^* \in S^*, \{d^k\} \to 0$ as $k \to \infty$, $k \in K'$, a contradiction. Thus $\{d^k\} \to 0$. It now follows from Lemma 3.2 that $\{\lambda^k\} \to \lambda^*$ and, in view of Assumption A5, $J^k = \emptyset$ for k large enough. Lemma 3.2 and the update rule for μ^k again implies that $\{\mu_i^k\} \to \min(\lambda_i^*, \overline{\mu})$ for all $i \in I$.

Proof of Proposition 3.10. First, let $i \in \{1, \ldots, n\}$ be such that (i) holds. Since $\left(\frac{\partial F}{\partial z}(z)\right)^{-1}$ is bounded in $B(z^*, \rho)$ and F is Lipschitz continuous in the same ball, there exists $c_2 > 0$ such that, for all $z \in B(z^*, \rho)$

$$|z_i^+ - z_i^*| \leq c_1 \|\frac{\partial F}{\partial z}(z)^{-1}\|^2 \|F(z) - F(z^*)\|^2 \leq c_2 \|z - z^*\|^2.$$

Next, suppose (ii) holds. Then

$$\begin{aligned} |z_i^+ - z_i^*| &\leq |z_i^+ - (z + d^N(z))_i| + |z_i^* - (z + d^N(z))_i| \\ &\leq c_1 \|d^N(z)\|^2 + \|z^* - (z + d^N(z))\| \\ &\leq c_1 \|d^N(z)\|^2 + \|\frac{\partial F}{\partial z}(z)^{-1}\| \|F(z) + \frac{\partial F}{\partial z}(z)(z^* - z)\|. \end{aligned}$$

The first term in the right hand side is as in (i). Boundedness of $\frac{\partial F}{\partial z}(z)^{-1}$ in $B(z^*, \rho)$ and regularity of F thus again imply that the claim holds.

References

[1] J. Herskovits, "A Two-Stage Feasible Direction Algorithm for Nonlinear Constrained Optimization," paper TuA.1.1, Symposium Volume of 11th International Symposium on Mathematical Programming, Bonn, West Germany, August 1982.

[2] J. Herskovits, "A Two-Stage Feasible Direction Algorithm Including Variable Metric Techniques," paper FA.1.3, Symposium Volume of 11th International Symposium on Mathematical Programming, Bonn, West Germany, August 1982.

[3] J. Herskovits, *Développement d'une Méthode Numérique pour l'Optimization Non-Linéaire*, Ph.D. Thesis, Université Paris IX - Dauphine, Paris, January 1982.

[4] S. Segenreich, N. Zouain & J. Herskovits, "An Optimality Criteria Method Based on Slack Variables Concept for Large Structural Optimization," in *Proceedings of the Symposium on Applications of Computer Methods in Engineering*, Los Angeles, California, 1977, 563–572.

[5] J. Herskovits, "A Two-Stage Feasible Directions Algorithm for Nonlinear Constrained Optimization," *Math. Programming* 36 (1986), 19–38.

[6] E.R. Panier, A.L. Tits & J.N. Herskovits, "A QP-Free, Globally Convergent, Locally Superlinearly Convergent Algorithm for Inequality Constrained Optimization," *SIAM J. Control Optim.* 26 (1988), 788–811.

[7] G.P. McCormick, "Penalty Function Versus Non-Penalty Function Methods for Constrained Nonlinear Programming Problems," *Math. Programming* 1 (1971), 217–238.

[8] E. Polak & A.L. Tits, "On Globally Stabilized Quasi-Newton Methods for Inequality Constrained Optimization Problems," in *Proceedings of the 10th IFIP Conference on System Modeling and Optimization — New York, NY, August-September 1981*, R.F. Drenick & F. Kozin, eds., Lecture Notes in Control and Information Sciences #38 , Springer-Verlag, New York–Heidelberg–Berlin, 1982, 539–547.

[9] J. Herskovits, "A New Interior Point Method for Linear Programming," paper 113, Final Program of SIAM Conference on Optimization, Houston, Texas, May 1987.

[10] J. Herskovits, "A General Approach for Interior Point Methods in Mathematical Programming. Part I – Linear Programming," paper WE2B4, Program and Abstracts of 14th International Symposium on Mathematical Programming, Tokyo, Japan, August-September 1988.

[11] M. Kojima, S. Mizuno & A. Yoshise, "A Primal-Dual Interior Point Method for Linear Programming.," in *Progress in Mathematical Programming, Interior-Point and Related Methods*, N. Megiddo, ed., Springer-Verlag, New York, 1989, 29–47.

[12] R.D.C. Monteiro, I. Adler & M.G.C. Resende, "A Polynomial-Time Primal-Dual Affine Scaling Algorithm for Linear and Convex Quadratic Programming and Its Power Series Extension," *Mathematics of Operations Research* 15 (1990), 191–214.

[13] I.I. Dikin, "Iterative Solution of Problems of Linear and Quadratic Programming," *Soviet Math. Dokl.* 8 (1967), 674–675.

[14] C.C. Gonzaga, "Path-Following Methods for Linear Programming," *SIAM Rev.* 34 (1992), 167–224.

[15] Y. Zhang, R.A. Tapia & J.E. Dennis, "On the Superlinear and Quadratic Convergence of Primal-Dual Interior Point Linear Programming Algorithms," *SIAM J. on Optimization* 2 (1992), 304–324.

[16] Y. Ye, "On the Q-Order of Convergence of Interior-Point Algorithms for Linear Programming," The College of Business Administration, The Univerity of Iowa, Preprint, Iowa City, Iowa, 1992.

[17] Y. Ye, O. Güler, R.A. Tapia & Y. Zhang, "A Quadratically Convergent $O(\sqrt{n}L)$-Iteration Algorithm for Linear Programming," The College of Business Administration, The Univerity of Iowa, Working Paper No 91-14, Iowa City, Iowa, August 1991.