Math. Program., Ser. A (2011) 127:399–424
DOI 10.1007/s10107-009-0292-7

FULL LENGTH PAPER

A first-order interior-point method for linearly constrained smooth optimization

Paul Tseng · Immanuel M. Bomze · Werner Schachinger

Received: 3 December 2007 / Accepted: 15 April 2009 / Published online: 3 June 2009

Abstract We propose a first-order interior-point method for linearly constrained smooth optimization that unifies and extends first-order affine-scaling method and replicator dynamics method for standard quadratic programming. Global convergence and, in the case of quadratic program, (sub)linear convergence rate and iterate convergence results are derived. Numerical experience on simplex constrained problems with 1000 variables is reported.

Keywords Linearly constrained optimization · Affine scaling · Replicator dynamics · Interior-point method · Global convergence · Sublinear convergence rate

Mathematics Subject Classification (2000) 49M37 · 65K05 · 90C20 · 90C25 · 90C26 · 90C30 · 90C51

1 Introduction

Consider a linearly constrained smooth optimization problem:

$$\max_{\mathbf{x}\in\Lambda} f(\mathbf{x}), \tag{1}$$

P. Tseng (deceased)
Department of Mathematics, University of Washington, Seattle, WA 98195, USA

I. M. Bomze (✉) · W. Schachinger
Department of Statistics and Decision Support Systems, University of Vienna,
Brünnerstr. 72, 1210 Vienna, Austria
e-mail: immanuel.bomze@univie.ac.at

W. Schachinger
e-mail: werner.schachinger@univie.ac.at

where $\Lambda = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}\}$ is polyhedral with $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. We assume without loss of generality that A has rank m. This problem has been well studied and many iterative methods have been proposed for its solution, such as active-set methods and interior-point methods; see, e.g., [1, Chapter 2], [10, Chapter 5], [6,9,11,19]. In what follows, $\text{ri}\,\Lambda = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{b},\ \mathbf{x} > \mathbf{0}\}$ denotes the relative interior of Λ and $\|\cdot\|$ denotes the Euclidean norm (2-norm). Bold letters without subscript denote vectors, with x_j denoting the jth component of a vector $\mathbf{x}$. For each index subset $\mathcal{J} \subseteq \{1,\ldots,n\}$, we denote by $\mathbf{x}_{\mathcal{J}}$ the vector composed of those components of $\mathbf{x} \in \mathbb{R}^n$ indexed by $j \in \mathcal{J}$.

An important special case of (1) is the standard quadratic program (StQP), where Λ is the unit simplex and f is homogeneous quadratic, i.e.,

$$A = \mathbf{e}^\top, \quad \mathbf{b} = 1, \quad f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x},$$

with $Q \in \mathbb{R}^{n\times n}$ symmetric and $\mathbf{e} \in \mathbb{R}^n$ a vector of ones. Applications of StQP include maximum clique, portfolio selection, graph isomorphism [2,23]. By adding a nonnegative multiple of $\mathbf{e}\mathbf{e}^\top$ to Q (which changes f by only a constant on Λ), we can assume that

$$q_{ii} > 0 \quad \text{and} \quad q_{ij} \geq 0 \quad \forall\, i, j, \tag{2}$$

so that $Q\mathbf{x} > \mathbf{0}$ componentwise for all $\mathbf{x} \in \Lambda$. In [2–4,23,24], a remarkably simple interior-point method called replicator dynamics (RD) was used for solving StQP:

$$\mathbf{x}^{k+1} = \frac{X^k\mathbf{g}^k}{(\mathbf{x}^k)^\top\mathbf{g}^k}, \quad k = 0, 1, \ldots, \quad \mathbf{x}^0 \in \text{ri}\,\Lambda, \tag{3}$$

where $\mathbf{g}^k = Q\mathbf{x}^k$, $X^k = \text{Diag}(\mathbf{x}^k)$. The Assumption (2) implies that $\mathbf{g}^k > 0$ and $\mathbf{x}^k \in \text{ri}\,\Lambda$ for all k. In fact, (2) is necessary and sufficient for $Q\mathbf{x} \geq \mathbf{0}$ and $\mathbf{x}^\top Q\mathbf{x} > 0$ for all $\mathbf{x} \in \Lambda$. This method is reminiscent of the power method for finding the largest (in magnitude) eigenvalue of a square matrix, with the unit Euclidean sphere replaced by the unit simplex. Starting the RD method in ri Λ is essential. If it is started at an $\mathbf{x}^0$ in the relative boundary of Λ, then all $\mathbf{x}^k$ will stay in the relative interior of the face of Λ that contains $\mathbf{x}^0$.

The RD method has a long history in mathematical biology, and it connects three different fields: optimization, evolutionary games, and qualitative analysis of dynamical systems; see [3] and references therein. It arises in population genetics under the name *selection equations* where it is used to model time evolution of haploid genotypes, with Q being the (symmetric) fitness matrix, and x_i^k representing the relative frequency of allele i in the population at time k (see, e.g., [20, Chapter III]). Since it also serves to model replicating entities in a much more general context, it is often called RD nowadays. The continuous-time version of the RD method is known to be a gradient system with respect to Shahshahani geometry; see [12]. This suggests that the method may be useful for local optimization. In fact, (3) has the remarkable property that, under Assumption (2), the generated sequence of iterates $\{\mathbf{x}^k\}$ converges

[15] (i.e., has a unique cluster point) and, given that we start in ri Λ, the limit $\bar{\mathbf{x}}$ is a first-order stationary point of (1). Additionally, the objective values $f(\mathbf{x}^k)$ increase with k, $\|\mathbf{x}^k - \bar{\mathbf{x}}\| = \mathcal{O}(1/\sqrt{k})$, and convergence rate is linear if and only if strict complementarity holds at $\bar{\mathbf{x}}$. This contrasts with other interior-point methods for solving (1), for which additional assumptions on Q are required to prove convergence of the generated iterates; see [6,11,19,27,29,31]. In [4], the RD method (3) was applied to solve medium-sized test problems from portfolio selection and was shown to be superior in performance to classical feasible ascent methods using exact line search, including Rosen's gradient projection method and Wolfe's reduced gradient method. A variant of the RD method that uses exact line search was also considered. Recently, some of the aforementioned results were extended to the case where Λ is a product of simplices, under the name of multi-standard quadratic program (MStQP); see [5].

However, in practice the RD method seems slow on large instances of both StQP and MStQP. Can this method be improved and extended to the general problem (1) while retaining its elegant simplicity? Denoting, as always in the sequel, $X^k = \mathrm{Diag}(\mathbf{x}^k)$, we can rewrite (3) as

$$\mathbf{x}^{k+1} - \mathbf{x}^k = \frac{X^k\big(\mathbf{g}^k - \big((\mathbf{x}^k)^\top \mathbf{g}^k\big)\mathbf{e}\big)}{(\mathbf{x}^k)^\top \mathbf{g}^k},$$

so we can interpret the numerator as the search direction and $1/(\mathbf{x}^k)^\top \mathbf{g}^k$ as the step-size; see [4, Eq. 15]. What if we do a line search along this search direction instead? This motivates the following interior-point method for solving the special case of (1) where Λ is the unit simplex (f need not be homogeneous quadratic):

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k, \quad \mathbf{d}^k = X^k \mathbf{r}\big(\mathbf{x}^k\big), \quad k = 0, 1, \ldots, \quad \mathbf{x}^0 \in \mathrm{ri}\,\Lambda, \tag{4}$$

with $0 < \alpha^k < -1/\min_j r_j(\mathbf{x}^k)$, where

$$\mathbf{r}(\mathbf{x}) = \nabla f(\mathbf{x}) - \mathbf{x}^\top \nabla f(\mathbf{x})\mathbf{e} = \big[I - \mathbf{e}\mathbf{x}^\top\big]\nabla f(\mathbf{x})$$

and r_j denotes the jth component of $\mathbf{r}$. The restriction on α^k ensures that $\mathbf{x}^k \in \mathrm{ri}\,\Lambda$ for all k. Note that, for $\mathbf{x} \in \Lambda$, we have $\mathbf{x}^\top \mathbf{r}(\mathbf{x}) = 0$ and hence $\min_j r_j(\mathbf{x}) < 0$ unless $\mathbf{x}$ is a stationary point of (1).

Among existing interior-point methods, the RD method (4) is most closely related to the first-order affine-scaling (AS) method of Dikin [7] for quadratic programming and extended by Gonzaga and Carlos for linearly constrained smooth convex minimization [11]. Their method, when specialized to (1), has the form

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k, \quad \mathbf{d}^k = \big(X^k\big)^2 \mathbf{r}\big(\mathbf{x}^k\big), \quad k = 0, 1, \ldots, \quad \mathbf{x}^0 \in \mathrm{ri}\,\Lambda, \tag{5}$$

where α^k is chosen by a limited maximization rule on $(0, -1/(\min_j d_j^k/x_j^k))$,

$$\mathbf{r}(\mathbf{x}) = \nabla f(\mathbf{x}) - \frac{\mathbf{x}^\top X \nabla f(\mathbf{x})}{\|\mathbf{x}\|^2}\mathbf{e} = \left[I - \frac{1}{\|\mathbf{x}\|^2}\mathbf{e}\mathbf{x}^\top X\right]\nabla f(\mathbf{x}),$$

and $X = \mathrm{Diag}(\mathbf{x})$. They showed that every cluster point of $\{\mathbf{x}^k\}$ is a stationary point of (1) when f is concave. The proof extends an idea of Dikin [8] and makes use of a constancy property of ∇f on each isocost line segment. Subsequently, Bonnans and Pola [6] and Monteiro and Wang [19] proposed first- and second-order AS trust-region methods based on generalizations of this search direction. In [6, Theorem 2.2], α^k is chosen by an Armijo-type rule and it is shown that every cluster point of the generated iterates is a stationary point provided a certain relaxed first-order optimality system has a unique solution or has isolated solutions, and an additional technical condition holds. In [19], the analysis in [11] is extended to show every cluster point of the generated iterates is a stationary point provided f is either concave or convex. In general, global convergence analysis for these kinds of interior-point methods, including (4), is nontrivial due to the search direction being componentwise proportional to the current iterate.

Upon comparing (4) with (5), we see that they differ mainly in that one scales its direction by X^k while the other scales by $(X^k)^2$. Also, $\mathbf{r}(\mathbf{x})$ is obtained by subtracting from $\nabla f(\mathbf{x})$ componentwise a weighted average of it components. In fact, the two methods (4) and (5) belong to a general class of first-order interior-point methods for the general problem (1) that has the form

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k, \quad \mathbf{d}^k = (X^k)^{2\gamma}\mathbf{r}(\mathbf{x}^k), \quad k = 0, 1, \ldots, \quad \mathbf{x}^0 \in \mathrm{ri}\,\Lambda, \tag{6}$$

where $\gamma > 0$, $0 < \alpha^k < -1/(\min_j d_j^k / x_j^k)$, and

$$\begin{aligned} \mathbf{r}(\mathbf{x}) &= \nabla f(\mathbf{x}) - A^\top \left(AX^{2\gamma}A^\top\right)^{-1} AX^{2\gamma}\nabla f(\mathbf{x}) \\ &= \left[I - A^\top\left(AX^{2\gamma}A^\top\right)^{-1}AX^{2\gamma}\right]\nabla f(\mathbf{x}). \end{aligned} \tag{7}$$

(here $X^{2\gamma}$ denotes X raised to the power 2γ, in contrast to X^k. The meaning of the exponent should be clear from the context). Thus, $\gamma = 1$ yields the first-order AS method while $\gamma = 1/2$ yields the RD method. Since A has rank m, $AX^{2\gamma}A^\top$ is positive definite and hence invertible whenever $\mathbf{x} > \mathbf{0}$.

We will discuss (inexact) line search strategies for choosing the stepsize α^k so as to achieve fast global convergence. We will show that if α^k is chosen by either an Armijo rule or a limited maximization rule, then every cluster point of $\{\mathbf{x}^k\}$ is a stationary point of (1) under a primal nondegeneracy assumption and additional assumptions such as f being concave or convex; see Theorem 1(c). Thus if f is concave, then every cluster point would be a global maximizer. In the special case where f is quadratic, we show that $\{f(\mathbf{x}^k)\}$ converges at a sublinear rate or, specifically, $\upsilon - f(\mathbf{x}^k) = \mathcal{O}(1/k^{1/\max\{\gamma, 2\gamma-1\}})$ where $\upsilon = \lim_{k\to\infty} f(\mathbf{x}^k)$; see Theorem 2. To our knowledge, this is the first rate of convergence result for a first-order interior-point method when the objective function f is nonlinear. Moreover, we extend the result in [15] to show that, for $\gamma < 1$, $\{\mathbf{x}^k\}$ converges sublinearly and, under primal nondegeneracy, its limit $\bar{\mathbf{x}}$ is a stationary point of (1). If in addition $\gamma \leq \frac{1}{2}$ and $\bar{\mathbf{x}}$ satisfies strict complementarity, then $\{f(\mathbf{x}^k)\}$ and $\{\mathbf{x}^k\}$ converge linearly. On the other hand, if $\frac{1}{2} \leq \gamma < 1$ and $\bar{\mathbf{x}}$ does not satisfy strict complementarity, then $\{\mathbf{x}^k\}$ cannot converge linearly. Why are we interested in a first-order method if its convergence rate can be sublinear? They have much simpler iterations compared to second-order interior-point methods [9,29,33] and hence may be suited for solving very large problems ($n \geq 10000$). The case of multiple simplex constraints is a good example. In this case, $AX^{2\gamma}A^\top$ has a block-diagonal structure corresponding to the simplices and $\mathbf{r}(\mathbf{x})$ decomposes accordingly. In general, if m is small or $AA^\top$ has a nice sparsity structure, then $\mathbf{r}(\mathbf{x})$ can be inexpensively computed from $\nabla f(\mathbf{x})$. Our analysis and numerical experience suggest that a value of $\gamma < 1$ is superior to values of $\gamma \geq 1$. Our main contributions are: a unified algorithmic framework, practical stepsize rules, a comprehensive global convergence analysis and, for quadratic f, convergence rate analysis, and implementation and testing of the method. Our results can be extended to handle upper bound constraint $\mathbf{x} \leq \mathbf{u}$ by working with $\mathrm{Diag}(\min\{\mathbf{x}, \mathbf{u} - \mathbf{x}\})$; see, e.g., [29]. For simplicity we do not consider this more general problem here.

After the initial writing of this paper, Takashi Tsuchiya informed us that the method (6), (7) had previously been studied by Saigal [26] for linear programming. It is shown in [26, Theorems 6 and 24] that $\{\mathbf{x}^k\}$ converges to a stationary point of (1) assuming f is linear, $\gamma > 1$, $\alpha^k = -\alpha/(\min_j d_j^k/x_j^k)$, and either (i) $0 < \alpha < 1$ and primal nondegeneracy (Assumption 1 in Sect. 2) or (ii) $0 < \alpha/(1-\alpha)^{2\gamma} < 2/(2\gamma - 1)$. It is remarked in [26, pp. 378, 415] that its analysis can be extended to $\frac{1}{2} < \gamma < 1$ though no detail is given. In contrast, we show convergence of $\{\mathbf{x}^k\}$ to a stationary point assuming f is quadratic, $\gamma < 1$, α^k is chosen by line search, and primal nondegeneracy; see Theorem 2(b).

2 Properties of search direction

The following lemma shows key feasible ascent properties of the search direction $\mathbf{d}^k$.

Lemma 1 *For any $\gamma > 0$ and $\mathbf{x} \in \mathrm{ri}\,\Lambda$, let $X = \mathrm{Diag}(\mathbf{x})$ and $\mathbf{d} = X^{2\gamma}\mathbf{r}(\mathbf{x})$, where $\mathbf{r}$ is given by (7). Then we have $A\mathbf{d} = \mathbf{0}$ and*

$$\nabla f(\mathbf{x})^\top \mathbf{d} = \|X^{-\gamma}\mathbf{d}\|^2 = \|X^\gamma \mathbf{r}(\mathbf{x})\|^2.$$

Moreover, $\mathbf{d}$ solves the following subproblem

$$\max_{\mathbf{u}\in\mathbb{R}^n} \left\{\nabla f(\mathbf{x})^\top \mathbf{u} \mid A\mathbf{u} = \mathbf{0},\ \|X^{-\gamma}\mathbf{u}\| \leq \|X^\gamma \mathbf{r}(\mathbf{x})\|\right\}. \tag{8}$$

Proof Let $\mathbf{g} = \nabla f(\mathbf{x})$ and denote by

$$P_\gamma = I - X^\gamma A^\top \left(AX^{2\gamma}A^\top\right)^{-1} AX^\gamma$$

the matrix of orthogonal projection onto the null space of AX^γ. Then $P_\gamma^\top = P_\gamma = (P_\gamma)^2$. We have from (7) that

$$X^{-\gamma}\mathbf{d} = X^\gamma \mathbf{r}(\mathbf{x}) = P_\gamma X^\gamma \mathbf{g},$$

so $X^{-\gamma}\mathbf{d}$ is in the null space of AX^{γ} and hence $A\mathbf{d} = (AX^{\gamma})(X^{-\gamma}\mathbf{d}) = \mathbf{0}$. Also,

$$\begin{aligned}\mathbf{g}^{\top}\mathbf{d} &= \left(X^{\gamma}\mathbf{g}\right)^{\top}\left(X^{-\gamma}\mathbf{d}\right)\\ &= \left(X^{\gamma}\mathbf{g}\right)^{\top} P_{\gamma}X^{\gamma}\mathbf{g}\\ &= \|P_{\gamma}X^{\gamma}\mathbf{g}\|^2\\ &= \|X^{-\gamma}\mathbf{d}\|^2,\end{aligned}$$

where the third equality uses $P_{\gamma} = P_{\gamma}^2$. The minimum-norm property of orthogonal projection implies that $\mathbf{d}$ solves the subproblem (8). □

We will make use of the following primal nondegeneracy assumption, which is standard in the analysis of AS methods, especially when the objective is nonquadratic; see [6,7,11,19].

Assumption 1 *For any $\mathbf{x} \in \Lambda$, the columns of A corresponding to $\{j \mid x_j \neq 0\}$ have rank m.*

Assumption 1 is satisfied when Λ is the unit simplex or a Cartesian product of simplices. The following result is well known; see [7,11].

Lemma 2 *Under Assumption 1, $AX^{2\gamma}A^{\top}$ is nonsingular for all $\mathbf{x} \in \Lambda$ and $\mathbf{r}$ is a continuous mapping on Λ, where X and $\mathbf{r}$ are given by (7).*

3 Stepsize rules

For general f, we propose to choose α^k by an Armijo-type rule [1, Sect. 2.2.1]: α^k is the largest $\alpha \in \{\alpha_0^k\beta^{\ell}\}_{\ell=0,1,\ldots}$ satisfying

$$f\left(\mathbf{x}^k + \alpha\mathbf{d}^k\right) \geq f\left(\mathbf{x}^k\right) + \sigma\alpha\left(\mathbf{g}^k\right)^{\top}\mathbf{d}^k, \tag{9}$$

where $\mathbf{g}^k = \nabla f\left(\mathbf{x}^k\right)$, $0 < \beta, \sigma < 1$ are constants and

$$0 < \alpha_0^k < \begin{cases} \infty & \text{if } \mathbf{d}^k \geq \mathbf{0};\\ \frac{-1}{\min_j d_j^k/x_j^k} & \text{else.}\end{cases} \tag{10}$$

Notice that if $\mathbf{d}^k = \mathbf{0}$, then (9) is satisfied by any $\alpha \geq 0$ and the Armijo rule yields $\alpha^k = \alpha_0^k$. Since $\mathbf{d}^k$ is a feasible ascent direction at $\mathbf{x}^k$ by Lemma 1 and $\alpha_0^k > 0$, we know that α^k is well defined and positive.

In the special case where f is a quadratic or cubic function, we can choose α^k by the limited maximization rule:

$$\alpha^k \in \arg\max_{0\leq\alpha\leq\alpha_0^k} f(\mathbf{x}^k + \alpha\mathbf{d}^k). \tag{11}$$

4 Global convergence

In this section we analyze the global convergence of the first-order interior-point method (6, 7). The proof uses ideas from [1, Sect. 1.2] and [11], [19, Appendix A]. As with RD and AS methods, the proof is complicated by the fact that the direction mapping $\mathbf{x} \mapsto X^{2\gamma}\mathbf{r}(\mathbf{x})$ is undefined on the relative boundary of Λ. Even when it is defined and continuous on the relative boundary of Λ, as is the case under Assumption 1, it may be zero at a non-stationary point.

Theorem 1 *Assume $\Lambda^0 = \{\mathbf{x} \in \Lambda \mid f(\mathbf{x}) \geq f(\mathbf{x}^0)\}$ is bounded. Let $\{\mathbf{x}^k\}$ be generated by the method (6), (7) with $\{\alpha^k\}$ chosen by the Armijo rule (9) and $\{\alpha_0^k\}$ satisfying (10). Then the following results hold with $\mathbf{g}^k = \nabla f(\mathbf{x}^k)$.*

(a) *$\mathbf{x}^k \in \operatorname{ri}\Lambda$ for all k, $\{f(\mathbf{x}^k)\}$ is nondecreasing, and $\{\mathbf{x}^k\}$, $\{\mathbf{d}^k\}$ are bounded.*
(b) *Assume $\inf_k \alpha_0^k > 0$. Then $\{(\mathbf{g}^k)^\top \mathbf{d}^k\} \to 0$, $\{(X^k)^\gamma \mathbf{r}(\mathbf{x}^k)\} \to 0$, and every cluster point $\bar{\mathbf{x}}$ of $\{\mathbf{x}^k\}$ satisfies*

$$\operatorname{Diag}(\bar{\mathbf{x}})\big(\nabla f(\bar{\mathbf{x}}) - A^\top \bar{\mathbf{p}}\big) = 0 \quad \text{for some } \bar{\mathbf{p}} \in \mathbb{R}^m. \tag{12}$$

If $\sup_k \alpha_0^k < \infty$, then iterate change goes to zero, i.e., $\{\mathbf{x}^{k+1} - \mathbf{x}^k\} \to \mathbf{0}$.
(c) *Suppose $\inf_k \alpha_0^k > 0$, $\sup_k \alpha_0^k < \infty$, Assumption 1 holds, and either (i) f is concave or convex or (ii) Λ_{cs} consists of isolated points or (iii) every $\mathbf{x} \in \Lambda_{\mathrm{cs}}$ satisfies strict complementarity (i.e., $x_j - r_j(\mathbf{x}) \neq 0$ for all j), where*

$$\Lambda_{\mathrm{cs}} = \Big\{\mathbf{x} \in \Lambda \mid \operatorname{Diag}(\mathbf{x})\mathbf{r}(\mathbf{x}) = \mathbf{0},\ f(\mathbf{x}) = \lim_{k\to\infty} f(\mathbf{x}^k)\Big\}.$$

Then every cluster point of $\{\mathbf{x}^k\}$ is a stationary point of (1). Under (ii), $\{\mathbf{x}^k\}$ converges.
(d) *If $\inf_k \alpha_0^k > 0$ and ∇f is Lipschitz continuous on Λ, then $\inf_k \alpha^k > 0$.*

Proof (a) Since $\mathbf{x}^0 \in \operatorname{ri}\Lambda$, by using (6) and Lemma 1 and an induction argument on k, we have that $\mathbf{x}^k \in \operatorname{ri}\Lambda$ and $\alpha^k > 0$ for all k. Also, α^k satisfies (9), so (6) implies

$$f\big(\mathbf{x}^{k+1}\big) - f\big(\mathbf{x}^k\big) \geq \sigma\alpha^k\big(\mathbf{g}^k\big)^\top \mathbf{d}^k = \sigma\alpha^k\|\big(X^k\big)^{-\gamma}\mathbf{d}^k\|^2 \quad \forall k, \tag{13}$$

where the equality uses Lemma 1. Thus $\{f(\mathbf{x}^k)\}$ is nondecreasing. Then $\mathbf{x}^k \in \Lambda^0$ for all k. Since Λ^0 is bounded, this implies $\{\mathbf{x}^k\}$ is bounded. Also, (6) implies

$$\|\mathbf{d}^k\| = \|\big(X^k\big)^{2\gamma}\mathbf{r}\big(\mathbf{x}^k\big)\| \leq \max_j \big(x_j^k\big)^\gamma \|(X^k)^\gamma \mathbf{r}\big(\mathbf{x}^k\big)\|$$

and (7) implies

$$\|\big(X^k\big)^\gamma \mathbf{r}\big(\mathbf{x}^k\big)\| = \|P_\gamma^k\big(X^k\big)^\gamma \mathbf{g}^k\| \leq \|\big(X^k\big)^\gamma \mathbf{g}^k\|,$$

where P_γ^k is the matrix of orthogonal projection onto the null space of $A(X^k)^\gamma$. Since $\{\mathbf{x}^k\}$ is bounded and ∇f is continuous so that $\{\mathbf{g}^k\}$ is bounded, this shows that $\{\mathbf{d}^k\}$ is bounded.

(b) Suppose $\inf_k \alpha_0^k > 0$. Let $\bar{\mathbf{x}}$ be any cluster point of $\{\mathbf{x}^k\}$. Since $\mathbf{x}^k \in \Lambda^0$ for all k and Λ^0 is closed, $\bar{\mathbf{x}} \in \Lambda^0$. Since f is continuous and, by (a), $\{f(\mathbf{x}^k)\}$ is nondecreasing, we have $\{f(\mathbf{x}^k)\} \uparrow f(\bar{\mathbf{x}})$ and hence $\{f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)\} \to 0$. Then (13) implies

$$\left\{\alpha^k \left(\mathbf{g}^k\right)^\top \mathbf{d}^k\right\} \to 0. \tag{14}$$

Consider any subsequence $\{\mathbf{x}^k\}_{k\in\mathcal{K}}$ ($\mathcal{K} \subseteq \{0, 1, \ldots\}$) converging to $\bar{\mathbf{x}}$. Let $\bar{\mathbf{g}} = \nabla f(\bar{\mathbf{x}})$. By further passing to a subsequence if necessary, we will assume that either (i) $\inf_{k\in\mathcal{K}} \alpha^k > 0$ or (ii) $\{\alpha^k\}_{k\in\mathcal{K}} \to 0$. In case (i), we have from (14) that $\{(\mathbf{g}^k)^\top \mathbf{d}^k\}_{k\in\mathcal{K}} \to 0$. In case (ii), we have from $\inf_k \alpha_0^k > 0$ that $\alpha^k < \alpha_0^k$ for all $k \in \mathcal{K}$ sufficiently large, implying that the ascent condition (9) is violated by $\alpha = \alpha^k/\beta$, i.e.,

$$\frac{f\left(\mathbf{x}^k + \frac{\alpha^k}{\beta}\mathbf{d}^k\right) - f\left(\mathbf{x}^k\right)}{\alpha^k/\beta} < \sigma \left(\mathbf{g}^k\right)^\top \mathbf{d}^k. \tag{15}$$

Since $\{\mathbf{d}^k\}$ is bounded, by further passing to a subsequence if necessary, we can assume that $\{\mathbf{d}^k\}_{k\in\mathcal{K}} \to$ some $\bar{\mathbf{d}}$. Since $\{\alpha^k\}_{k\in\mathcal{K}} \to 0$ and f is continuously differentiable, the above inequality yields in the limit that

$$\bar{\mathbf{g}}^\top \bar{\mathbf{d}} \le \sigma \bar{\mathbf{g}}^\top \bar{\mathbf{d}}.$$

Since $0 < \sigma < 1$, this implies $\bar{\mathbf{g}}^\top \bar{\mathbf{d}} \le 0$. Thus $\limsup_{k\in\mathcal{K}, k\to\infty} (\mathbf{g}^k)^\top \mathbf{d}^k \le 0$. Since $(\mathbf{g}^k)^\top \mathbf{d}^k \ge 0$ for all k, this implies $\{(\mathbf{g}^k)^\top \mathbf{d}^k\}_{k\in\mathcal{K}} \to 0$. Then Lemma 1 implies

$$\left\{\left(X^k\right)^\gamma \mathbf{r}\left(\mathbf{x}^k\right)\right\}_{k\in\mathcal{K}} \to 0,$$

and hence $\{\mathbf{r}_{\mathcal{J}^c}\left(\mathbf{x}^k\right)\}_{k\in\mathcal{K}} \to \mathbf{0}$, where $\mathcal{J}^c = \left\{j \mid \bar{x}_j > 0\right\}$. By (7), the system of linear equations in $\mathbf{p} \in \mathbb{R}^m$:

$$\left(\mathbf{g}^k - A^\top \mathbf{p}\right)_j = r_j\left(\mathbf{x}^k\right) \quad \forall j \in \mathcal{J}^c$$

has a solution. Let $\mathbf{p}^k$ be its least 2-norm solution. Since the coefficient matrix in this system does not change with k, it follows from $\{\mathbf{g}^k\}_{k\in\mathcal{K}} \to \nabla f(\bar{\mathbf{x}})$ and $\{\mathbf{r}_{\mathcal{J}^c}(\mathbf{x}^k)\}_{k\in\mathcal{K}} \to \mathbf{0}$ that $\{\mathbf{p}^k\}_{k\in\mathcal{K}}$ converges to some $\bar{\mathbf{p}}$ satisfying $(\nabla f(\bar{\mathbf{x}}) - A^\top \bar{\mathbf{p}})_j = 0$ for all $j \in \mathcal{J}^c$. Thus (12) holds.

Since $\left\{\mathbf{x}^k\right\}$ is bounded by (a), the above argument shows that $\{(\mathbf{g}^k)^\top \mathbf{d}^k\} \to 0$ as well as $\left\{\left(X^k\right)^\gamma \mathbf{r}\left(\mathbf{x}^k\right)\right\} \to \mathbf{0}$, and every cluster point $\bar{\mathbf{x}}$ satisfies (12).

By (13), for all k,

$$(\mathbf{g}^k)^\top \mathbf{d}^k = \|(X^k)^{-\gamma}\mathbf{d}^k\|^2 = \frac{1}{(\alpha^k)^2}\|(X^k)^{-\gamma}(\mathbf{x}^{k+1}-\mathbf{x}^k)\|^2$$
$$\geq \frac{1}{(\alpha_0^k)^2}\frac{\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2}{\max_j (x_j^k)^{2\gamma}}.$$

If $\sup_k \alpha_0^k < \infty$, then since $\{(\mathbf{g}^k)^\top \mathbf{d}^k\} \to 0$ and $\{\mathbf{x}^k\}$ is bounded by (a), this implies $\{\mathbf{x}^{k+1}-\mathbf{x}^k\} \to \mathbf{0}$.

(c) Suppose that $\inf_k \alpha_0^k > 0$, that $\sup_k \alpha_0^k < \infty$, and that Assumption 1 holds. Let $\bar{\mathbf{x}}$ be any cluster point of $\{\mathbf{x}^k\}$. Let $\bar{X} = \mathrm{Diag}(\bar{\mathbf{x}})$ and $\bar{\mathbf{r}} = \mathbf{r}(\bar{\mathbf{x}})$. We have from $\{(X^k)^\gamma \mathbf{r}(\mathbf{x}^k)\} \to \mathbf{0}$ in (b) that $\bar{X}\bar{\mathbf{r}} = \mathbf{0}$ or, equivalently,

$$\bar{X}(\bar{\mathbf{g}} - A^\top \bar{\mathbf{p}}) = \mathbf{0},$$

where $\bar{\mathbf{g}} = \nabla f(\bar{\mathbf{x}})$ and $\bar{\mathbf{p}} = (A\bar{X}^{2\gamma}A^\top)^{-1}A\bar{X}^{2\gamma}\bar{\mathbf{g}}$. Thus $\bar{\mathbf{x}}$ belongs to Λ_{cs}, and $\bar{\mathbf{x}}$ is a stationary point of (1) if and only if $\bar{\mathbf{r}} \leq \mathbf{0}$.

Suppose that f is concave or convex. We show below that $\bar{\mathbf{r}} \leq \mathbf{0}$. The argument is similar to one used by Gonzaga and Carlos [11]; also see [19, Sect. 3.3]. First, we have the key result that

$$\{\mathbf{r}(\mathbf{x}^k)\} \to \bar{\mathbf{r}}. \tag{16}$$

Its proof is given in Appendix A. If $\bar{\mathbf{r}} \not\leq \mathbf{0}$, then there would exist some $\bar{j} \in \{1,\ldots,n\}$ such that $\bar{r}_{\bar{j}} > 0$. Then $\bar{X}\bar{\mathbf{r}} = \mathbf{0}$ implies $\bar{x}_{\bar{j}} = 0$. By (16), there exists $\bar{k}$ such that $r_{\bar{j}}(\mathbf{x}^k) > 0$ for all $k \geq \bar{k}$, so that $d_{\bar{j}}^k > 0$ for all $k \geq \bar{k}$ and hence $x_{\bar{j}}^k > x_{\bar{j}}^{\bar{k}} > 0$ for all $k \geq \bar{k}$. This contradicts $\bar{x}_{\bar{j}} = 0$.

Suppose that, instead of f being concave or convex, Λ_{cs} consists of isolated points. Since $\{\mathbf{x}^{k+1}-\mathbf{x}^k\} \to \mathbf{0}$ so the set of cluster points of $\{\mathbf{x}^k\}$ is connected, (b) implies $\{\mathbf{x}^k\} \to \bar{\mathbf{x}}$. Then (16) holds and the same argument as above yields $\bar{\mathbf{r}} \leq \mathbf{0}$. Suppose instead that every $\mathbf{x} \in \Lambda_{\mathrm{cs}}$ satisfies strict complementarity. Let

$$\bar{\Lambda}_{\mathrm{cs}} = \left\{\mathbf{x} \in \Lambda_{\mathrm{cs}} \mid \mathbf{r}_{\bar{\mathcal{J}}_0}(\mathbf{x}) = \mathbf{0},\ \mathbf{r}_{\bar{\mathcal{J}}_+}(\mathbf{x}) > \mathbf{0},\ \mathbf{r}_{\bar{\mathcal{J}}_-}(\mathbf{x}) < \mathbf{0}\right\},$$

where $\bar{\mathcal{J}}_0 = \{j \mid \bar{r}_j = 0\}$, $\bar{\mathcal{J}}_+ = \{j \mid \bar{r}_j > 0\}$, $\bar{\mathcal{J}}_- = \{j \mid \bar{r}_j < 0\}$. Since every $\mathbf{x} \in \Lambda_{\mathrm{cs}}$ satisfies strict complementarity, $\bar{\Lambda}_{\mathrm{cs}}$ is isolated from the rest of Λ_{cs}, i.e., there exists a $\delta > 0$ such that $(\bar{\Lambda}_{\mathrm{cs}} + \delta B) \cap \Lambda_{\mathrm{cs}} = \bar{\Lambda}_{\mathrm{cs}}$, where B denotes the unit Euclidean ball centered at the origin. Since the set of cluster points of $\{\mathbf{x}^k\}$ is connected and, by (b), is contained in Λ_{cs}, this set must in fact be contained in $\bar{\Lambda}_{\mathrm{cs}}$. Hence for every $j \in \bar{\mathcal{J}}_+$ we have $r_j(\mathbf{x}^k) > 0$ for all k sufficiently large, so $d_j^k > 0$ for all k sufficiently large, implying $\liminf_k x_j^k > 0$, a contradiction to $\bar{x}_j = 0$. Thus $\bar{\mathcal{J}}_+ = \emptyset$, i.e., $\bar{\mathbf{r}} \leq \mathbf{0}$.

(d) Suppose that $\inf_k \alpha_0^k > 0$ and ∇f is Lipschitz continuous on Λ with Lipschitz constant $L \geq 0$. Then it is readily shown using the mean value theorem that

$$\nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) - \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|^2 \leq f(\mathbf{y}) - f(\mathbf{x}) \qquad \forall \mathbf{x}, \mathbf{y} \in \Lambda$$

(see [1, p. 667]). For each $k \in \{0, 1, \ldots\}$, either $\alpha^k = \alpha_0^k$ or else (9) is violated by $\alpha = \frac{\alpha^k}{\beta}$, i.e., (15) holds. In the second case, we apply the above inequality to $\mathbf{x} = \mathbf{x}^k$ and $\mathbf{y} = \mathbf{x}^k + \frac{\alpha^k}{\beta}\mathbf{d}^k$, so that using (15) we obtain

$$\frac{\alpha^k}{\beta}(\mathbf{g}^k)^\top\mathbf{d}^k - \frac{L}{2}\left(\frac{\alpha^k}{\beta}\right)^2\|\mathbf{d}^k\|^2 \leq f\left(\mathbf{x}^k + \frac{\alpha^k}{\beta}\mathbf{d}^k\right) - f\left(\mathbf{x}^k\right) < \frac{\alpha^k}{\beta}\sigma\left(\mathbf{g}^k\right)^\top\mathbf{d}^k.$$

Dividing both sides by $\frac{\alpha^k}{\beta}$ and rearranging terms yields

$$(1-\sigma)\left(\mathbf{g}^k\right)^\top\mathbf{d}^k \leq \frac{L}{2}\frac{\alpha^k}{\beta}\|\mathbf{d}^k\|^2.$$

By Lemma 1,

$$(1-\sigma)\frac{\|\mathbf{d}^k\|^2}{\max_j\left(x_j^k\right)^{2\gamma}} \leq (1-\sigma)\|\left(X^k\right)^{-\gamma}\mathbf{d}^k\|^2 \leq \frac{L}{2}\frac{\alpha^k}{\beta}\|\mathbf{d}^k\|^2.$$

Since $\alpha^k \neq \alpha_0^k$, we have $\mathbf{d}^k \neq \mathbf{0}$, so this yields $(1-\sigma)/\max_j\left(x_j^k\right)^{2\gamma} \leq L\frac{\alpha^k}{2\beta}$. Thus in both cases we have

$$\alpha^k \geq \min\left\{\alpha_0^k, \frac{2\beta(1-\sigma)}{L\max_j\left(x_j^k\right)^{2\gamma}}\right\}. \tag{17}$$

Since $\inf_k \alpha_0^k > 0$ and $\{\mathbf{x}^k\}$ is bounded, this shows that $\inf_k \alpha^k > 0$. □

The assumption in Theorem 1(c–d) of (10) and $\inf_k \alpha_0^k > 0$ is reasonable since, by (a), $\{\mathbf{d}^k\}$ is bounded, so the right-hand side of (10) is uniformly bounded away from zero. Theorem 1(d) will be used in the convergence rate analysis of the next section. Similar to the observation in [1, p. 45], Theorem 1(a–c) extend to the limited maximization rule (11) or any stepsize rule that yields a larger ascent than the Armijo rule at each iteration.

Corollary 1 *Theorem* 1(a–c) *still hold if, in the interior-point method* (6), (7), α^k *more generally satisfies*

$$0 < \alpha^k \leq \alpha_0^k, \qquad f\left(\mathbf{x}^k + \alpha^k\mathbf{d}^k\right) \geq f\left(\mathbf{x}^k + \alpha_{\text{armijo}}^k\mathbf{d}^k\right),$$

where α_{armijo}^k *is chosen by the Armijo rule* (9).

Proof Theorem 1(a) clearly holds. Theorem 1(b) holds since (13), (14), (15) in its proof still hold with α^k replaced by α^k_{armijo}. This yields $\{(\mathbf{g}^k)^\top \mathbf{d}^k\} \to 0$ and $\{\mathbf{d}^k\} \to \mathbf{0}$. The proof of Theorem 1(c) is modified accordingly. □

Theorem 1(c) under condition (i) is similar to [11, Sect. 3] and [19, Theorem 3.14] for the case of AS methods ($\gamma = 1$). Theorem 1(c) under condition (ii) is similar to [6, Theorem 2.2], [33, Theorem 3] for the case of AS methods. In particular, Assumption 1 is equivalent to (H3) in [6], and condition (ii) and Λ_{cs} are refinements of, respectively, (H1) and $(\text{OS})_I$ solutions in [6]. When f is quadratic, Λ_{cs} consists of isolated points if and only if it is a finite set.

The convergence of $\{\mathbf{x}^k\}$ for AS methods has been much studied. In the case of linear f, convergence has been shown for first-order AS methods; see [14,18,26,30,32] and references therein. In the cases of concave quadratic f or a more general class of quadratic f and box constraint, convergence has been shown for second-order AS methods [17,27,29,31]. For more general f, convergence has been shown for AS methods under Assumption 1 and condition (ii) in Theorem 1 [33, Theorem 3], [6, Theorem 2.2], and for second-order AS methods, assuming f is concave or convex and $\nabla^2 f$ has a constant null space property [28], [19, Theorem 4.12].

5 Sublinear convergence when f is quadratic

In this section we show that, in the special case where f is quadratic, $\{f(\mathbf{x}^k)\}$ generated by the first-order interior-point method (6), (7) converges sublinearly. The proof, which uses Lemma 1 and Theorem 1, adapts the linear convergence analysis of a second-order AS method with line search [29, Theorem 1]. To our knowledge, this is the first rate of convergence result for a first-order interior-point method when f is nonlinear, and it does not assume primal nondegeneracy (Assumption 1). Moreover, by adapting the proof of [15, Theorem 3.2], we show that the generated iterates $\{\mathbf{x}^k\}$ converge sublinearly for $\gamma < 1$ and, under primal nondegeneracy, the limit is stationary for (1). If in addition $\gamma \le \frac{1}{2}$ and strict complementarity holds, then convergence is linear. But if $\sup_k \alpha_0^k < \infty$, $\frac{1}{2} \le \gamma < 1$ and strict complementarity fails, then convergence cannot be linear. This suggests $\gamma < 1$ may be preferable to $\gamma \ge 1$, which is corroborated by the numerical results in Sect. 6.

Theorem 2 *Assume* $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{c}^\top \mathbf{x}$ *for some symmetric* $Q \in \mathbb{R}^{n\times n}$ *and* $\mathbf{c} \in \mathbb{R}^n$. *Assume* $\Lambda^0 = \{\mathbf{x} \in \Lambda \mid f(\mathbf{x}) \ge f(\mathbf{x}^0)\}$ *is bounded. Let* $\{\mathbf{x}^k\}$ *be generated by the method* (6), (7) *with* $\{\alpha^k\}$ *chosen by the Armijo rule* (9) *and* $\{\alpha_0^k\}$ *satisfying* (10) *and* $\inf_k \alpha_0^k > 0$. *Then the following results hold with* $\omega = 1/(\bar{\gamma} - 1)$ *and* $\bar{\gamma} = \max\{1 + \gamma, 2\gamma\}$.

(a) There exist $\upsilon \in \mathbb{R}$ *and* $C > 0$ *(depending on* $\mathbf{x}^0$*) such that*

$$0 \le \upsilon - f\left(\mathbf{x}^k\right) \le Ck^{-\omega} \quad \forall k \ge 1. \tag{18}$$

(b) *Assume* $\gamma < 1$. *Then there exist* $\bar{\mathbf{x}} \in \Lambda^0$ *and* $C' > 0$ *(depending on* $\mathbf{x}^0$*) such that*

$$\|\bar{\mathbf{x}} - \mathbf{x}^k\| \leq C' k^{-\frac{1-\gamma}{2\gamma}} \quad \forall k \geq 1.$$

Suppose Assumption 1 also holds. Then $\bar{\mathbf{x}}$ *is a stationary point of* (1)*. Moreover, if* $\gamma \leq \frac{1}{2}$ *and* $\bar{\mathbf{x}} - \mathbf{r}(\bar{\mathbf{x}}) > \mathbf{0}$*, then* $\{f(\mathbf{x}^k)\}$ *converges linearly in the quotient sense and* $\{\|\bar{\mathbf{x}} - \mathbf{x}^k\|\}$ *converges linearly in the root sense. If instead* $\sup_k \alpha_0^k < \infty$*,* $\gamma \geq \frac{1}{2}$ *and* $\bar{\mathbf{x}} - \mathbf{r}(\bar{\mathbf{x}}) \not> \mathbf{0}$*, then* $\{\|\bar{\mathbf{x}} - \mathbf{x}^k\|\}$ *cannot converge linearly.*

Proof We have from Theorem 1(a) and its proof that $\{f(\mathbf{x}^k)\}$ is nondecreasing and (13) holds or, equivalently,

$$f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \geq \sigma \alpha^k \|\boldsymbol{\eta}^k\|^2 \quad \forall k, \tag{19}$$

where we let

$$\boldsymbol{\eta}^k = \left(X^k\right)^\gamma \mathbf{r}^k \quad \text{with} \quad \mathbf{r}^k = \mathbf{r}\left(\mathbf{x}^k\right).$$

Thus $\{f(\mathbf{x}^k)\}$ converges to a limit υ and $\{f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)\} \to 0$. Since ∇f is Lipschitz continuous on Λ, Theorem 1(d) implies $\inf_k \alpha^k > 0$. Then (19) implies $\{\boldsymbol{\eta}^k\} \to \mathbf{0}$.

For any $\mathcal{J} \subseteq \{1, \ldots, n\}$, let

$$\mathcal{K}_{\mathcal{J}} = \left\{ k \in \{0, 1, \ldots\} \left| \begin{array}{l} x_j^k \leq |\eta_j^k|^{\frac{1}{1+\gamma}} \ \forall j \in \mathcal{J}, \\ |r_j^k| < |\eta_j^k|^{\frac{1}{1+\gamma}} \ \forall j \in \mathcal{J}^c \end{array} \right. \right\}, \tag{20}$$

where $\mathcal{J}^c = \{1, \ldots, n\} \setminus \mathcal{J}$. Since $(x_j^k)^\gamma |r_j^k| = |\eta_j^k|$ so that either $x_j^k \leq |\eta_j^k|^{\frac{1}{1+\gamma}}$ or else $|r_j^k| < |\eta_j^k|^{\frac{1}{1+\gamma}}$, it follows that each $k \in \{0, 1, \ldots\}$ belongs to exactly one set $\mathcal{K}_{\mathcal{J}}$ for some $\mathcal{J}$ (because evidently $\mathcal{J} \neq \mathcal{J}' \subseteq \{1, \ldots, n\}$ implies $\mathcal{K}_{\mathcal{J}} \cap \mathcal{K}_{\mathcal{J}'} = \emptyset$). Since the number of subsets $\mathcal{J}$ is finite, there is at least one $\mathcal{J}$ such that $\mathcal{K}_{\mathcal{J}}$ is infinite.

Consider any $\mathcal{J}$ such that $\mathcal{K}_{\mathcal{J}}$ is infinite. For each $k \in \mathcal{K}_{\mathcal{J}}$, consider the following linear system in $(\mathbf{x}, \mathbf{p})$:

$$\mathbf{x}_{\mathcal{J}} = \mathbf{x}_{\mathcal{J}}^k, \quad \mathbf{q}_j^\top \mathbf{x} - \mathbf{a}_j^\top \mathbf{p} = -c_j + r_j^k \ \ \forall j \in \mathcal{J}^c, \quad \mathbf{x} \geq \mathbf{0}, \quad A\mathbf{x} = \mathbf{b}. \tag{21}$$

This system has at least one solution, namely $(\mathbf{x}, \mathbf{p}) = (\mathbf{x}^k, \mathbf{p}^k)$ with $\mathbf{p}^k = (A(X^k)^{2\gamma} A^\top)^{-1} A(X^k)^{2\gamma} \mathbf{g}^k$. Here $\mathbf{q}_j$ and $\mathbf{a}_j$ denote the jth column of Q and A, respectively. Now, let $\|\cdot\|_\nu$ denote the ν-norm. (We drop the subscript ν for the Euclidean norm where $\nu = 2$.) By the definition (20),

$$\left(\| \left(\mathbf{x}_{\mathcal{J}}^k, \mathbf{r}_{\mathcal{J}^c}^k\right) \|_{1+\gamma}\right)^{1+\gamma} \leq \|\boldsymbol{\eta}^k\|_1 \quad \forall k \in \mathcal{K}_{\mathcal{J}}, \tag{22}$$

so $\{\boldsymbol{\eta}^k\} \to \mathbf{0}$ yields $\{(\mathbf{x}^k_{\mathcal{J}}, \mathbf{r}^k_{\mathcal{J}^c})\}_{k\in\mathcal{K}_{\mathcal{J}}} \to \mathbf{0}$. Thus, the right-hand side of (21) is uniformly bounded for $k \in \mathcal{K}_{\mathcal{J}}$, so an error bound of Hoffman [13] implies that (21) has a solution, say $(\mathbf{y}^k, \mathbf{t}^k)$, that is uniformly bounded for $k \in \mathcal{K}_{\mathcal{J}}$. Since $\{(\mathbf{x}^k_{\mathcal{J}}, \mathbf{r}^k_{\mathcal{J}^c})\}_{k\in\mathcal{K}_{\mathcal{J}}} \to 0$, any cluster point $(\mathbf{y}, \mathbf{t})$ of $\{(\mathbf{y}^k, \mathbf{t}^k)\}_{k\in\mathcal{K}_{\mathcal{J}}}$ satisfies

$$\mathbf{y}_{\mathcal{J}} = \mathbf{0}, \quad \mathbf{q}_j^\top \mathbf{y} - \mathbf{a}_j^\top \mathbf{t} = -c_j \ \ \forall j \in \mathcal{J}^c, \quad \mathbf{y} \geq \mathbf{0}, \quad A\mathbf{y} = \mathbf{b}. \tag{23}$$

Thus, this linear system has a solution. Let $\Sigma_{\mathcal{J}}$ denote the set of solutions for (23). Since $(\mathbf{x}^k, \mathbf{p}^k)$ is a solution of (21), Hoffman's error bound [13] implies there exists $(\bar{\mathbf{x}}^k, \bar{\mathbf{p}}^k) \in \Sigma_{\mathcal{J}}$ satisfying

$$\| \left(\bar{\mathbf{x}}^k, \bar{\mathbf{p}}^k\right) - \left(\mathbf{x}^k, \mathbf{p}^k\right) \| \leq C_1 \| \left(\mathbf{x}^k_{\mathcal{J}}, \mathbf{r}^k_{\mathcal{J}^c}\right) \|_{1+\gamma} \quad \forall k \in \mathcal{K}_{\mathcal{J}}, \tag{24}$$

where C_1 is a constant depending on γ, Q, A, and $\mathcal{J}$ only.

We claim that f is constant on each $\Sigma_{\mathcal{J}}$. If $(\mathbf{y}, \mathbf{t})$ and $(\mathbf{y}', \mathbf{t}')$ both belong to $\Sigma_{\mathcal{J}}$, then (23) yields

$$\begin{aligned} f(\mathbf{y}') - f(\mathbf{y}) &= \tfrac{1}{2}(\mathbf{y}' - \mathbf{y})^\top Q(\mathbf{y}' - \mathbf{y}) + (Q\mathbf{y} + \mathbf{c})^\top (\mathbf{y}' - \mathbf{y}) \\ &= \tfrac{1}{2}(\mathbf{y}' - \mathbf{y})^\top Q(\mathbf{y}' - \mathbf{y}) + (Q\mathbf{y} + \mathbf{c} - A^\top \mathbf{t})^\top (\mathbf{y}' - \mathbf{y}) \\ &= \tfrac{1}{2}(\mathbf{y}' - \mathbf{y})^\top Q(\mathbf{y}' - \mathbf{y}), \end{aligned}$$

where the second equality uses $A(\mathbf{y}' - \mathbf{y}) = \mathbf{0}$ and third equality uses $\mathbf{y}'_{\mathcal{J}} = \mathbf{y}_{\mathcal{J}}$ and $\mathbf{q}_j^\top \mathbf{y} - \mathbf{a}_j^\top \mathbf{t} = -c_j$ for all $j \in \mathcal{J}^c$. A symmetric argument yields

$$f(\mathbf{y}) - f(\mathbf{y}') = \tfrac{1}{2}(\mathbf{y} - \mathbf{y}')^\top Q(\mathbf{y} - \mathbf{y}').$$

Combining the above two equalities yields $f(\mathbf{y}') = f(\mathbf{y})$.

For any $k \in \mathcal{K}_{\mathcal{J}}$, we have

$$\begin{aligned} (Q\bar{\mathbf{x}}^k + \mathbf{c})^\top (\mathbf{x}^k - \bar{\mathbf{x}}^k) &= (Q\bar{\mathbf{x}}^k + \mathbf{c} - A^\top \bar{\mathbf{p}}^k)^\top (\mathbf{x}^k - \bar{\mathbf{x}}^k) \\ &= \sum_{j\in\mathcal{J}} \left(\mathbf{q}_j^\top \bar{\mathbf{x}}^k + c_j - \mathbf{a}_j^\top \bar{\mathbf{p}}^k\right) x_j^k \\ &= \sum_{j\in\mathcal{J}} \left(\mathbf{q}_j^\top (\bar{\mathbf{x}}^k - \mathbf{x}^k) - \mathbf{a}_j^\top (\bar{\mathbf{p}}^k - \mathbf{p}^k) + r_j^k\right) x_j^k, \end{aligned}$$

where the first equality uses $A(\mathbf{x}^k - \bar{\mathbf{x}}^k) = \mathbf{0}$, and the second equality uses $(\bar{\mathbf{x}}^k, \bar{\mathbf{p}}^k) \in \Sigma_{\mathcal{J}}$. If $\gamma \leq 1$, then this together with

$$f(\mathbf{x}^k) - f(\bar{\mathbf{x}}^k) = \tfrac{1}{2}(\mathbf{x}^k - \bar{\mathbf{x}}^k)^\top Q(\mathbf{x}^k - \bar{\mathbf{x}}^k) + (Q\bar{\mathbf{x}}^k + \mathbf{c})^\top (\mathbf{x}^k - \bar{\mathbf{x}}^k)$$

and the definition of $\boldsymbol{\eta}^k = (X^k)^\gamma \mathbf{r}^k$ yields

$$\begin{aligned}
&|f(\mathbf{x}^k) - f(\bar{\mathbf{x}}^k)| \\
&\quad = \left| \tfrac{1}{2}(\mathbf{x}^k - \bar{\mathbf{x}}^k)^\top Q(\mathbf{x}^k - \bar{\mathbf{x}}^k) + \sum_{j\in\mathcal{J}} \left((\mathbf{q}_j^\top(\bar{\mathbf{x}}^k - \mathbf{x}^k) - \mathbf{a}_j^\top(\bar{\mathbf{p}}^k - \mathbf{p}^k))x_j^k + x_j^k r_j^k \right) \right| \\
&\quad \le \tfrac{1}{2}\left|(\mathbf{x}^k - \bar{\mathbf{x}}^k)^\top Q(\mathbf{x}^k - \bar{\mathbf{x}}^k)\right| \\
&\qquad + \sum_{j\in\mathcal{J}} \left(\left|(\mathbf{q}_j^\top(\bar{\mathbf{x}}^k - \mathbf{x}^k) - \mathbf{a}_j^\top(\bar{\mathbf{p}}^k - \mathbf{p}^k))x_j^k\right| + (x_j^k)^{1-\gamma}|\eta_j^k| \right) \\
&\quad \le C_2\|\mathbf{x}^k - \bar{\mathbf{x}}^k\|^2 + \sum_{j\in\mathcal{J}} \left(\|(\mathbf{q}_j, -\mathbf{a}_j)\|\, \|(\bar{\mathbf{x}}^k, \bar{\mathbf{p}}^k) - (\mathbf{x}^k, \mathbf{p}^k)\| x_j^k + (x_j^k)^{1-\gamma}|\eta_j^k| \right) \\
&\quad \le C_2 C_1^2 \|\boldsymbol{\eta}^k\|_1^{\frac{2}{1+\gamma}} + \sum_{j\in\mathcal{J}} \left(\|(\mathbf{q}_j, -\mathbf{a}_j)\|\, C_1\|\boldsymbol{\eta}^k\|_1^{\frac{1}{1+\gamma}} x_j^k + (x_j^k)^{1-\gamma}|\eta_j^k| \right) \\
&\quad \le C_2 C_1^2 \|\boldsymbol{\eta}^k\|_1^{\frac{2}{1+\gamma}} + C_1 \sum_{j\in\mathcal{J}} \left(\|(\mathbf{q}_j, \mathbf{a}_j)\|\, \|\boldsymbol{\eta}^k\|_1^{\frac{1}{1+\gamma}} |\eta_j^k|^{\frac{1}{1+\gamma}} + |\eta_j^k|^{\frac{2}{1+\gamma}} \right),
\end{aligned}$$

where C_2 is a constant depending on Q only; the first inequality uses $x_j^k|r_j^k| = (x_j^k)^{1-\gamma}|\eta_j^k|$; the third inequality uses (22) and (24); the last inequality uses (20), $\gamma \le 1$, and $k \in \mathcal{K}_\mathcal{J}$. It follows that

$$|f(\mathbf{x}^k) - f(\bar{\mathbf{x}}^k)| \le C_\mathcal{J}\|\boldsymbol{\eta}^k\|_1^{\frac{2}{1+\gamma}} \quad \forall k \in \mathcal{K}_\mathcal{J}, \tag{25}$$

where $C_\mathcal{J}$ is a constant depending on γ, Q, A, and $\mathcal{J}$ only. If $\gamma > 1$ instead, then by using $x_j^k|r_j^k| = |\eta_j^k|^{1/\gamma}|r_j^k|^{1-1/\gamma}$ and $\sup_k \|\mathbf{r}^k\|_\infty^{1-1/\gamma} < \infty$, we similarly obtain (25) but with "$\|\boldsymbol{\eta}^k\|_1^{\frac{2}{1+\gamma}}$" replaced by "$\|\boldsymbol{\eta}^k\|_1^{1/\gamma}$" (also using $\frac{2}{1+\gamma} > \frac{1}{\gamma}$). Let C_3 be the maximum of $C_\mathcal{J}$ over all $\mathcal{J}$ such that $\mathcal{K}_\mathcal{J}$ is infinite.

Since $\{f(\mathbf{x}^k)\} \uparrow \upsilon$ and $\{\boldsymbol{\eta}^k\} \to \mathbf{0}$, it follows from (25) that $\{f(\bar{\mathbf{x}}^k)\}_{k\in\mathcal{K}_\mathcal{J}} \to \upsilon$. Since $\bar{\mathbf{x}}^k \in \Sigma_\mathcal{J}$ for all $k \in \mathcal{K}_\mathcal{J}$ and f is constant on $\Sigma_\mathcal{J}$, this implies $f(\bar{\mathbf{x}}^k) = \upsilon$ for all $k \in \mathcal{K}_\mathcal{J}$. This, together with (25), the subsequent remark, and $C_\mathcal{J} \le C_3$, yields

$$\begin{aligned}
\upsilon - f(\mathbf{x}^k) &= f(\bar{\mathbf{x}}^k) - f(\mathbf{x}^k) \\
&\le C_3\|\boldsymbol{\eta}^k\|_1^{\min\left\{\frac{2}{1+\gamma}, \frac{1}{\gamma}\right\}} \\
&= C_3\|\boldsymbol{\eta}^k\|_1^{2/\bar{\gamma}} \\
&\le \kappa\left(f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \right)^{1/\bar{\gamma}}
\end{aligned} \tag{26}$$

for all $k \in \mathcal{K}_\mathcal{J}$ with some constant $\kappa > 0$, where the second inequality uses (19) and $\inf_k \alpha^k > 0$. The above inequality yields, upon letting $\Delta^k = \upsilon - f(\mathbf{x}^k)$ and rearranging terms,

$$\Delta^{k+1} \le \Delta^k - \left(\frac{\Delta^k}{\kappa}\right)^{\bar{\gamma}}. \tag{27}$$

This holds for all $k \in \mathcal{K}_{\mathcal{J}}$ and all $\mathcal{J}$ such that $\mathcal{K}_{\mathcal{J}}$ is infinite. Since each $k \in \{0, 1, \ldots\}$ belongs to $\mathcal{K}_{\mathcal{J}}$ for some $\mathcal{J}$, then (26) and (27) hold for all k sufficiently large, say $k \geq K$.

(a) Take $C \geq \kappa^{\frac{\bar{\gamma}}{\bar{\gamma}-1}}$ sufficiently large so that (18) holds for $k = 1, \ldots, K$. Then an induction argument shows that (18) holds for all $k \geq 1$. In particular, if (18) holds for some $k \geq 1$, then we have from $C \geq \kappa^{\frac{\bar{\gamma}}{\bar{\gamma}-1}}$ that $(C/\kappa)^{\bar{\gamma}} \geq C$ and hence (27) yields

$$\Delta^{k+1} \leq \Delta^k - \left(\frac{\Delta^k}{\kappa}\right)^{\bar{\gamma}} \leq \frac{C}{k^\omega} - \left(\frac{C}{\kappa k^\omega}\right)^{\bar{\gamma}} \leq C\left(\frac{1}{k^\omega} - \frac{1}{k^{\omega\bar{\gamma}}}\right) \leq \frac{C}{(k+1)^\omega},$$

where the last inequality holds since $\omega\bar{\gamma} = 1 + \omega$ and $(1 - \frac{1}{k+1})^\omega \geq 1 - \frac{1}{k+1}$ (using $\omega \leq 1$). This proves (18).

(b) Assume $\gamma < 1$. Then $\bar{\gamma} = 1 + \gamma < 2$. Using (6), we have for all $k \geq K$ that

$$\begin{aligned}
\|\mathbf{x}^{k+1} - \mathbf{x}^k\| &= \alpha^k \|(X^k)^\gamma \boldsymbol{\eta}^k\| \\
&\leq \alpha^k \|\mathbf{x}^k\|_\infty^\gamma \|\boldsymbol{\eta}^k\| \\
&= \alpha^k \|\mathbf{x}^k\|_\infty^\gamma \frac{\|\boldsymbol{\eta}^k\|^2}{\|\boldsymbol{\eta}^k\|} \\
&\leq \|\mathbf{x}^k\|_\infty^\gamma \sqrt{n} \frac{\alpha^k \|\boldsymbol{\eta}^k\|^2}{\|\boldsymbol{\eta}^k\|_1} \\
&\leq \|\mathbf{x}^k\|_\infty^\gamma \sqrt{n} \frac{\Delta^k - \Delta^{k+1}}{\sigma} \left(\frac{C_3}{\Delta^k}\right)^{\frac{\bar{\gamma}}{2}},
\end{aligned}$$

where the last inequality uses (19) and (26) which holds for all $k \geq K$ as explained above. Since $\mathbf{x}^k$ lies in the bounded set Λ^0 so that $\sup_k \|\mathbf{x}^k\|_\infty < \infty$, this implies

$$\begin{aligned}
\|\mathbf{x}^{k+1} - \mathbf{x}^k\| &\leq C_4(\Delta^k - \Delta^{k+1})(\Delta^k)^{-\frac{\bar{\gamma}}{2}} \\
&\leq C_4 \int_{\Delta^{k+1}}^{\Delta^k} t^{-\frac{\bar{\gamma}}{2}} dt \\
&= \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} \left((\Delta^k)^{1-\frac{\bar{\gamma}}{2}} - (\Delta^{k+1})^{1-\frac{\bar{\gamma}}{2}}\right)
\end{aligned}$$

for all $k \geq K$, where $C_4 > 0$ is some constant. Thus for any $k_2 \geq k_1 \geq K$, we have

$$\sum_{k=k_1}^{k_2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| \leq \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} \left((\Delta^{k_1})^{1-\frac{\bar{\gamma}}{2}} - (\Delta^{k_2+1})^{1-\frac{\bar{\gamma}}{2}}\right) \leq \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} (\Delta^{k_1})^{1-\frac{\bar{\gamma}}{2}}.$$

Since $\Delta^{k_1} \to 0$ as $k_1 \to \infty$, this shows that the sequence $\{\mathbf{x}^k\}$ satisfies Cauchy's criterion for convergence and hence has a unique cluster point $\bar{\mathbf{x}}$. Moreover, the triangle inequality yields

$$\|\mathbf{x}^{k_2+1} - \mathbf{x}^{k_1}\| = \left\| \sum_{k=k_1}^{k_2} (\mathbf{x}^{k+1} - \mathbf{x}^k) \right\| \leq \sum_{k=k_1}^{k_2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| \leq \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} (\Delta^{k_1})^{1-\frac{\bar{\gamma}}{2}},$$

so taking $k_2 \to \infty$ and then using (18) yields $\|\bar{\mathbf{x}} - \mathbf{x}^{k_1}\| = \mathcal{O}\left((k_1^{-\omega})^{1-\frac{\bar{\gamma}}{2}}\right)$. Moreover, $\bar{\gamma} = 1 + \gamma$ and $\omega = 1/\gamma$, so $\omega(1 - \frac{\bar{\gamma}}{2}) = \frac{1-\gamma}{2\gamma}$.
Assume also that Assumption 1 holds. Then $\{\mathbf{r}^k\}$ converges to $\bar{\mathbf{r}} = \mathbf{r}(\bar{\mathbf{x}})$ (since $\mathbf{r}$ is continuous by Lemma 2) and it readily follows from (6) and $\{\mathbf{x}^k\} \to \bar{\mathbf{x}}$ that $\bar{\mathbf{r}} \leq \mathbf{0}$. Suppose $\bar{\mathbf{x}} - \bar{\mathbf{r}} > 0$, i.e., there exists $\bar{\mathcal{J}} \subseteq \{1, \ldots, n\}$ such that

$$\bar{\mathbf{r}}_{\bar{\mathcal{J}}} < \bar{\mathbf{x}}_{\bar{\mathcal{J}}} = \mathbf{0} \quad \text{and} \quad \bar{\mathbf{x}}_{\bar{\mathcal{J}}^c} > \bar{\mathbf{r}}_{\bar{\mathcal{J}}^c} = \mathbf{0}.$$

Then $\{(\mathbf{x}^k, \mathbf{r}^k)\} \to (\bar{\mathbf{x}}, \bar{\mathbf{r}})$ and $\boldsymbol{\eta}^k = (X^k)^\gamma \mathbf{r}^k$ imply

$$x_j^k = \mathcal{O}(|\eta_j^k|^{1/\gamma}) \quad \forall j \in \bar{\mathcal{J}} \quad \text{and} \quad |r_j^k| = \mathcal{O}(|\eta_j^k|) \quad \forall j \in \bar{\mathcal{J}}^c.$$

Repeating the preceding argument yields (25) with "$\frac{2}{1+\gamma}$" replaced by "$\min\{2, \frac{1}{\gamma}\}$" and (27) with "$\bar{\gamma}$" replaced by "$\max\{1, 2\gamma\}$". The latter in turn implies that, for $\gamma \leq \frac{1}{2}$,

$$\Delta^{k+1} \leq (1 - 1/\kappa)\Delta^k$$

for all k sufficiently large, so that $\{\Delta^k\} \to 0$ linearly in the quotient sense and, by the above argument, $\{\|\mathbf{x}^k - \bar{\mathbf{x}}\|\} \to 0$ linearly in the root sense. Suppose instead $\sup_k \alpha_0^k < \infty$, as well as $\gamma \geq \frac{1}{2}$ and $\bar{r}_{\bar{j}} = \bar{x}_{\bar{j}} = 0$ for some $\bar{j}$. Then $\sup_k \alpha^k < \infty$ and (6) yields

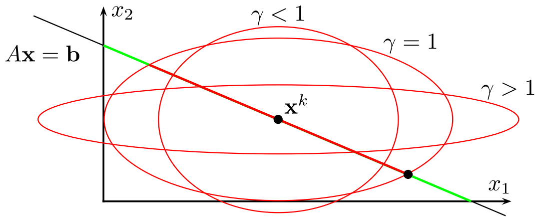
$$x_{\bar{j}}^{k+1}/x_{\bar{j}}^k = 1 + \alpha^k \left(x_{\bar{j}}^k\right)^{2\gamma-1} r_{\bar{j}}^k \to 1.$$

Thus $\{x_{\bar{j}}^k\}$ cannot converge linearly to 0 in the quotient or root sense, and hence neither can $\{\|\mathbf{x}^k - \bar{\mathbf{x}}\|\}$.

□

As γ decreases, ω increases while the proof of Theorem 2(a) suggests that C increases (since $\kappa^{\frac{\bar{\gamma}}{\bar{\gamma}-1}} \to \infty$ as $\bar{\gamma} \to 1$). This cautions against taking γ too small. Theorem 2(b) shows $\gamma = \frac{1}{2}$ is a good choice in theory. However, the numerical results in Sect. 6 suggest the resulting method may be prone to roundoff errors. The convergence of $\{\mathbf{x}^k\}$ for $\gamma = 1$ remains an open question. Since the proof of Theorem 2(a) is adapted

Fig. 1 The ellipsoid associated with $\mathbf{d}^k$, centered at $\mathbf{x}^k$ $(n = 2)$



from that of [29, Theorem 1], why can we prove only sublinear convergence and not linear convergence? This is because the amount of ascent $f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)$ is only guaranteed to be in the order of $\|(X^k)^\gamma \mathbf{r}^k\|^2$ (see (19)) instead of $\|X^k \mathbf{r}^k\|$ (see [29, Lemma 1]). To prove linear convergence, we would need to establish not a constant lower bound on α^k, as in Theorem 1(d), but a lower bound in the order of $1/\|(X^k)^\gamma \mathbf{r}^k\|$. This may be the price we pay for the simpler iterations of a first-order interior-point method. The linear convergence results in [19, Lemma 4.11] and [29, Theorem 1] hold for second-order AS methods only.

When $\gamma < 1$, the ellipsoid associated with $\mathbf{d}^k$ (see (8)) tends to be rounder (since $(x_j)^\gamma < x_j$ when $x_j > 1$ and $(x_j)^\gamma > x_j$ when $x_j < 1$). This may give some intuition for the better convergence behavior of $\{\mathbf{x}^k\}$ (Fig. 1).

6 Numerical experience

In order to better understand its practical performance, we have implemented in Matlab the first-order interior-point method (6), (7), with α^k chosen by the Armijo-type rule (9), to solve (1) with simplex constraint ($A = \mathbf{e}^\top$, $\mathbf{b} = 1$) and large n. In this section, we describe our implementation and report our numerical experience on test problems with objective functions f from Moré et al. [21], negated for maximization.

In our implementation, we use six different values of γ ($\gamma = 0.5, 0.8, 0.9, 1, 1.1, 1.2$). We use the standard setting of $\beta = .5$, $\sigma = .1$ for the Armijo rule (9), and we choose

$$\alpha_0^k = \min\left\{\alpha_{\text{feas}}^k, \max\left\{10^{-5}, \frac{\alpha^{k-1}}{\beta^2}\right\}\right\}, \quad \alpha_{\text{feas}}^k = \frac{0.95}{-\min_j d_j^k / x_j^k},$$

with $\alpha^{-1} = \infty$. Since $\mathbf{e}^\top \mathbf{d}^k = 0$, we have $\alpha_{\text{feas}}^k > 0$ whenever $\mathbf{d}^k \neq 0$. Thus $\{\alpha_0^k\}$ satisfies (10) and $\inf_k \alpha_0^k > 0$. Moreover, if α^{k-1} is small, then so is α_0^k. This can save on function evaluations since often $\alpha^k \approx \alpha^{k-1}$. In our experience, this choice of α_0^k yields much better performance than the choice $\alpha_0^k = 0.99/\|X^k \mathbf{r}(\mathbf{x}^k)\|$ used in [6, Algorithm 1].

For f, we choose 9 test functions with $n = 1000$ from the set of nonlinear least square functions used by Moré et al. [21] and negate them. These functions, listed in Table 1 with the numbering from [21, pp. 26–28] shown in parentheses, are chosen for their diverse characteristics: convex or nonconvex, sparse or dense Hessian, well-conditioned or ill-conditioned Hessian, and are grouped accordingly. The first three

Table 1 Behavior of first-order interior-point method with inexact line search on 9 test functions from [21], with $\mathbf{x}^0 = \mathbf{e}/n$ and $n = 1000$

Problem	γ	iter	nf	cpu	obj	resid	step	sc
ER (21)	0.5	6	7	0.03	498.002	3.4×10^{-7}	0.47	0.001
	0.8	5	6	0.01	498.002	5.2×10^{-7}	7.9×10^{3}	0.001
	0.9	5	6	0.01	498.002	6.8×10^{-7}	4.4×10^{4}	0.001
	1	7	8	0.03	498.002	2.6×10^{-7}	3.5×10^{6}	0.001
	1.1	8	9	0.02	498.002	8.4×10^{-7}	5.6×10^{7}	0.001
	1.2	10	11	0.03	498.002	3.7×10^{-7}	3.5×10^{9}	0.001
DBV (28)	0.5	117	352	0.4	4.5×10^{-8}	9.8×10^{-5}	59.3	0.0003
	0.8	99	299	0.42	4.9×10^{-8}	9.8×10^{-5}	1.8×10^{3}	0.0003
	0.9	107	323	0.45	4.7×10^{-8}	9.7×10^{-5}	7.4×10^{3}	0.0003
	1	146	440	0.52	4.5×10^{-8}	9.8×10^{-5}	2.9×10^{4}	0.0003
	1.1	191	575	0.82	4.3×10^{-8}	9.9×10^{-5}	1.1×10^{5}	0.0003
	1.2	240	722	1.02	4.0×10^{-8}	9.9×10^{-5}	4.7×10^{5}	0.0003
BT (30)	0.5	3,893	3,894	4.25	999.030	9.9×10^{-4}	0.15	0.007
	0.8	9,146	27,409	20.68	999.031	9.9×10^{-4}	0.50	0.006
	0.9	38,124	114,346	87.05	999.033	9.9×10^{-4}	0.69	0.006
	1	17,559	52,665	23.31	999.055	9.9×10^{-4}	0.45	0.01
	1.1	83,209	249,222	191.06	999.077	9.9×10^{-4}	0.65	0.01
	1.2	527,757	1.58×10^{6}	1,192.55	999.081	9.9×10^{-4}	1.14	0.01
TRIG (26)	0.5	41	121	0.11	1.1×10^{-6}	9.0×10^{-7}	623.6	0.0006
	0.8	39	115	0.14	9.5×10^{-7}	9.9×10^{-7}	3.6×10^{4}	0.0006
	0.9	73	214	0.24	9.0×10^{-7}	9.8×10^{-7}	8.6×10^{4}	0.0006
	1	95	271	0.22	9.7×10^{-7}	8.3×10^{-7}	5.3×10^{5}	0.0006
	1.1	52	143	0.17	8.8×10^{-7}	8.1×10^{-7}	2.3×10^{6}	0.0006
	1.2	86	223	0.27	1.1×10^{-6}	9.9×10^{-7}	6.4×10^{6}	0.0006
BAL (27)	0.5	6[a]	63	0.02	9.98998×10^{8}	1.9×10^{-5}	6.5×10^{-21}	0.001
	0.8	8[a]	84	0.02	9.98998×10^{8}	1.4×10^{-6}	6.0×10^{-21}	0.001
	0.9	6	7	0.01	9.98998×10^{8}	6.1×10^{-7}	113.648	0.001
	1	7	8	0.03	9.98998×10^{8}	3.6×10^{-8}	1947.55	0.001
	1.1	9[a]	94	0.03	9.98998×10^{8}	1.0×10^{-6}	6.4×10^{-21}	0.001
	1.2	8	9	0.02	9.98998×10^{8}	3.7×10^{-7}	123,466	0.001
EPS (22)	0.5	120	313	0.43	1.3×10^{-6}	9.9×10^{-4}	39.23	0.0001
	0.8	424	1,269	1.89	1.3×10^{-6}	9.9×10^{-4}	1503.22	0.0002
	0.9	644	1,929	2.85	2.4×10^{-6}	9.9×10^{-4}	5984.43	0.0002
	1	987	2,958	3.52	3.9×10^{-6}	9.9×10^{-4}	23824.5	0.0003
	1.1	870	2,608	3.81	5.9×10^{-6}	9.9×10^{-4}	47423.4	0.0003
	1.2	1,963	5,887	8.76	8.0×10^{-6}	9.4×10^{-4}	188.796	0.0004
VD (25)	0.5	17,503[a]	17,505	18.78	6.22504×10^{22}	2.9×10^{10}	9.5×10^{-22}	-2×10^{10}

Table 1 continued

	0.8	22	74	0.04	6.22504×10^{22}	2.4×10^{-9}	9.5×10^{-14}	1
	0.9	19	43	0.03	6.22504×10^{22}	1.6×10^{-8}	3.7×10^{-19}	1
	1	19	46	0.04	6.22504×10^{22}	2.6×10^{-8}	6.3×10^{-19}	1
	1.1	20	86	0.04	6.22504×10^{22}	8.8×10^{-8}	2.1×10^{-20}	1
	1.2	18	50	0.05	6.22504×10^{22}	5.7×10^{-9}	6.5×10^{-19}	0.9
`LR1` (33)	0.5	30,595[a]	30,624	50.12	3.32834×10^{8}	1.5×10^{-4}	1.4×10^{-12}	0.9
	0.8	20	21	0.04	3.32834×10^{8}	9.5×10^{-7}	2.0×10^{-4}	0.9
	0.9	19	20	0.04	3.32834×10^{8}	9.5×10^{-7}	2.3×10^{-3}	0.9
	1	19	20	0.03	3.32839×10^{8}	3.5×10^{-7}	0.031	1
	1.1	19	20	0.03	3.32911×10^{8}	5.9×10^{-7}	0.16	0.9
	1.2	20	21	0.03	3.33481×10^{8}	9.9×10^{-7}	0.77	0.9
`LR1Z` (34)	0.5	3,454[a]	3,538	3.96	251.125	145.94	8.8×10^{-21}	−7.9
	0.8	21[a]	88	0.05	251.125	119.46	9.9×10^{-21}	−6.5
	0.9	34[a]	129	0.07	251.125	3.7×10^{-6}	8.8×10^{-21}	0.01
	1	25[a]	102	0.05	251.125	170.84	8.8×10^{-21}	−9.3
	1.1	30[a]	119	0.06	251.125	5.8×10^{-7}	7.35	0.008
	1.2	65[a]	224	0.13	251.125	3.7×10^{-7}	41.94	0.01

[a] Quit due to roundoff error causing Armijo ascent condition not met when $\alpha < 10^{-20}$

functions `ER`, `DBV`, `BT` are nonconvex, with sparse Hessian. The next two functions `TRIG`, `BAL` are nonconvex, with dense Hessian. The sixth function `EPS` is convex, with sparse Hessian. The seventh function `VD` is strongly convex with dense Hessian. The last two functions `LR1`, `LR1Z` are convex quadratic with dense Hessian of rank 1. The functions `ER` and `EPS` have block-diagonal Hessians, and `VD`, `LR1`, `LR1Z` have ill-conditioned Hessians. Upon negation, these convex functions become concave functions. The functions and gradients are coded in Matlab using vector operations.

Since the starting points given in [21] may not satisfy the simplex constraint, we use the starting point

$$\mathbf{x}^0 = \mathbf{e}/n$$

for the first-order interior-point method. We terminate the method when the residual $\|\min\{\mathbf{x}^k, -\mathbf{r}^k\}\|$ is below a tolerance $tol > 0$. Note that $\mathbf{r}^k$ depends on γ. We set $tol = 10^{-6}$ for all problems except `DBV`, `BT`, `EPS`, and `LR1Z`. For `DBV`, `BT` and `EPS`, this is too tight and $tol = 10^{-4}$, $tol = 10^{-3}$ and $tol = 10^{-3}$ are used instead. For `LR1Z`, this is too loose and $tol = 10^{-7}$ is used instead. Roundoff error in Matlab occasionally causes this termination criterion never to be met, in which case we quit. In particular, we quit at iteration k if the Armijo ascent condition (9) is still not met when α falls below 10^{-20}.

Table 1 reports the number of iterations (iter), number of f-evaluations (nf), cpu time (in seconds), final objective value (for the original minimization problems), final residual (resid), final stepsize (step), and strict complementarity measure

$\min_j(x_j - r_j(x))$ (sc). All runs are performed on an HP DL360 workstation, running Red Hat Linux 3.5 and Matlab (Version 7.0). We see from Table 1 that iter and nf vary with γ but the best performance is obtained at $\gamma = .8$, with lower nf and iter on average. The ratio of nf/iter is typically below 4, suggesting that the Armijo rule typically uses few f-evaluations before accepting a stepsize. For a simple first-order method, the number of iterations looks quite reasonable on all problems except BT (whose Hessian is tridiagonal), for which interestingly $\gamma = .5$ yields the best performance. However, the method seems prone to roundoff error when $\gamma = .5$. The final stepsize tends to increase with γ. This is consistent with the lower bound (17), which increases with γ (since $\max_j x_j^k < 1$). When $\gamma > 1$, a smaller residual may be needed to achieve the same accuracy in the final objective value; see LR1 and LR1Z.

For comparison, we ran MINOS (Version 5.5.1), a well-known Fortran implementation of an active-set method for constrained smooth optimization [22], on the same workstation to solve the same 9 test problems. MINOS was compiled using the Gnu F-77 compiler (Version 3.2.57). Default settings and initialization were used. Table 2 reports the number of iterations (iter), number of f-evaluations (nf), cpu time (in seconds), and final objective value (negated) for MINOS. Comparing Tables 1 and 2, we see that the first-order interior-point method with $\gamma = .9$ has better performance (lower nf, cpu, and obj) on 4 of the problems (ER, TRIG, BAL, VD) while MINOS has better performance on 3 of the problems (BT, LR1, LR1Z). On DBV, the first-order interior-point method has higher nf and cpu but a much lower obj than MINOS, which may suggest multiple local minima. On EPS, the reverse is true, and obj is near zero for both. Thus the first-order interior-point method might complement existing methods by being more efficient at solving certain classes of problems.

The sc values in Table 1 suggest that strict complementarity holds for all problems. To see how the method performs on a degenerate problem, we test it on $f(\mathbf{x}) = -(\mathbf{e}^\top \mathbf{x})^2 - \sum_{j=1}^{n-1} x_j^2$, for which (1) has a unique stationary point $\bar{\mathbf{x}} = (0, \ldots, 0, 1)^\top$ that violates strict complementarity at all except one component. On this problem, $\gamma = 0.5$ yields the best performance, e.g., when $tol = 10^{-6}$, it terminates with $iter = 7$, $nf = 8$, $cpu = 0.01$, and $sc = 2 \times 10^{-8}$.

7 Monotonely transformed methods

In [4,23–25] an interesting exponential variant of (4) is studied for solving the special case of (1) with homogeneous quadratic f and simplex constraints. We extend this variant to allow for general f and general monotone transformation as follows. Suppose Λ is the unit simplex. Consider the method (6) but, instead of (7), we use

$$\mathbf{r}(\mathbf{x}) = \psi(\nabla f(\mathbf{x})) - \frac{\mathbf{e}^\top X^{2\gamma} \psi(\nabla f(\mathbf{x}))}{\mathbf{e}^\top X^{2\gamma} \mathbf{e}} \mathbf{e} \quad \text{with } X = \mathrm{Diag}(\mathbf{x}), \tag{28}$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is any strictly increasing function, and $\psi(\mathbf{y})$ means applying ψ to $\mathbf{y} \in \mathbb{R}^n$ componentwise. Taking $\psi(\tau) = e^\tau$ yields the exponential variant when $\gamma = 1/2$. Taking $\psi(\tau) = \tau$ yields the RD method (4) when $\gamma = 1/2$ and the first-order AS method (5) when $\gamma = 1$. To avoid numerical overflow, we can replace the

Table 2 Behavior of MINOS on 9 test functions from [21], with $\mathbf{x}^0$ chosen by default initialization procedure and $n = 1000$

Problem	iter	nf	cpu	obj
`ER` (21)	1,049	2,105	2.49	498.00
`DBV` (28)	10	48	0.01	5.96×10^{-2}
`BT` (30)	11	24	0.00	999.03
`TRIG` (26)	2,023	4,051	27.23	1.3×10^{-6}
`BAL` (27)	2	32	1.9	9.9899×10^{8}
`EPS` (22)	3,199	6,534	10.08	3.0×10^{-7}
`VD` (25)	0[a]	6	0.01	6.2749×10^{22}
`LR1` (33)	0	5	0.00	3.3283×10^{8}
`LR1Z` (34)	2	8	0.00	251.12

[a] MINOS exits due to problem being badly scaled

exponential by a polynomial when k exceeds some threshold. This monotone transformation, which is related to payoff monotonic game dynamics [24, Eq. (8)], seems to apply only in the case of simplex constraints.

Lemma 3 *Suppose Λ is the unit simplex, i.e., $A = \mathbf{e}^\top$, $\mathbf{b} = 1$. For any $\mathbf{x} \in \Lambda$, let X and $\mathbf{r}(\mathbf{x})$ be given by (28) with ψ any strictly increasing function. Then the following results hold.*

(a) $X\mathbf{r}(\mathbf{x}) = \mathbf{0}$ *if and only if* $X(\nabla f(\mathbf{x}) - p(\mathbf{x})\mathbf{e}) = \mathbf{0}$ *and* $\mathbf{r}(\mathbf{x}) \le \mathbf{0}$ *if and only if* $\nabla f(\mathbf{x}) - p(\mathbf{x})\mathbf{e} \le \mathbf{0}$, *where*

$$p(\mathbf{x}) = \psi^{-1}\left(\frac{\mathbf{e}^\top X^{2\gamma}\psi(\nabla f(\mathbf{x}))}{\mathbf{e}^\top X^{2\gamma}\mathbf{e}}\right).$$

(b) $\mathbf{d} = X^{2\gamma}\mathbf{r}(\mathbf{x})$ *satisfies* $\mathbf{e}^\top\mathbf{d} = 0$ *and* $\nabla f(\mathbf{x})^\top\mathbf{d} \ge 0$. *Moreover,* $\nabla f(\mathbf{x})^\top\mathbf{d} = 0$ *if and only if* $X\mathbf{r}(\mathbf{x}) = \mathbf{0}$.

Proof (a) This follows from (28) and the strictly increasing property of ψ.
(b) For any $\mathbf{x} \in \Lambda$, the direction $\mathbf{d} = X^{2\gamma}\mathbf{r}(\mathbf{x})$ satisfies $\mathbf{e}^\top\mathbf{d} = 0$ and

$$\begin{aligned}\mathbf{g}^\top\mathbf{d} &= \mathbf{g}^\top X^{2\gamma}\left(\psi(\mathbf{g}) - \frac{\mathbf{e}^\top X^{2\gamma}\psi(\mathbf{g})}{\mathbf{e}^\top X^{2\gamma}\mathbf{e}}\mathbf{e}\right)\\ &= (\mathbf{g} - \rho\mathbf{e})^\top X^{2\gamma}\psi(\mathbf{g})\\ &= (\mathbf{g} - \rho\mathbf{e})^\top X^{2\gamma}(\psi(\mathbf{g}) - \psi(\rho)\mathbf{e})\\ &\ge 0,\end{aligned}$$

where we let $\mathbf{g} = \nabla f(\mathbf{x})$ and $\rho = \frac{\mathbf{g}^\top X^{2\gamma}\mathbf{e}}{\mathbf{e}^\top X^{2\gamma}\mathbf{e}}$. We used monotonicity of ψ, which implies $(y-z)(\psi(y)-\psi(z)) \ge 0$ for any two $y, z \in \mathbb{R}$. Moreover, the inequality is strict unless $g_j = \rho$ for all j with $x_j \ne 0$. In this case, $X\psi(\mathbf{g}) = \psi(\rho)\mathbf{x}$ and

$X^{2\gamma}\psi(\mathbf{g}) = \psi(\rho)X^{2\gamma}\mathbf{e}$, so that

$$X\mathbf{r}(\mathbf{x}) = X\left(\psi(\mathbf{g}) - \psi(\rho)\frac{\mathbf{e}^\top X^{2\gamma}\mathbf{e}}{\mathbf{e}^\top X^{2\gamma}\mathbf{e}}\mathbf{e}\right) = \psi(\rho)\mathbf{x} - \psi(\rho)\mathbf{x} = \mathbf{0}.$$

The converse implication $X\mathbf{r}(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{g}^\top\mathbf{d} = 0$ follows similarly.

□

Remark 1 In the case of $\psi(\tau) = \exp(\theta\tau)$ with $\theta > 0$, we have

$$\nabla f(\mathbf{x})^\top\mathbf{d} \geq 4\left\|X^\gamma\left(\exp\left(\tfrac{\theta}{2}\mathbf{g}\right) - \exp\left(\tfrac{\theta}{2}\rho\right)\mathbf{e}\right)\right\|^2$$

(cf. Lemma 1). This follows from the inequality $(y-z)(\exp(y)-\exp(z)) \geq 4\big(\exp(\frac{y}{2}) - \exp(\frac{z}{2})\big)^2$, which holds for any $y, z \in \mathbb{R}$.

Extension of the above analysis to the case of multiple simplices is straightforward.

Lemma 4 *Let Λ be a Cartesian product of m unit simplices of dimensions $n_1, \ldots, n_m$ (so that $\sum_{i=1}^m n_i = n$), corresponding to $\mathbf{b} \in \mathbb{R}^m$ being a vector of ones and $A \in \{0,1\}^{m\times n}$ satisfying $A^\top\mathbf{b} = \mathbf{e}$, $A\mathbf{e} = (n_1, \ldots, n_m)^\top$. For any $\mathbf{x} \in \Lambda$, let $\mathbf{g} = \nabla f(\mathbf{x})$ and*

$$\mathbf{r}(\mathbf{x}) = [I - A^\top(AX^{2\gamma}A^\top)^{-1}AX^{2\gamma}]\psi(\mathbf{g}) \quad \text{with} \quad X = \operatorname{Diag}(\mathbf{x}), \tag{29}$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is any strictly increasing function. Then the following results hold.

(a) $X\mathbf{r}(\mathbf{x}) = \mathbf{0}$ *if and only if* $X(\mathbf{g} - A^\top\mathbf{p}(\mathbf{x})) = \mathbf{0}$ *and* $\mathbf{r}(\mathbf{x}) \leq \mathbf{0}$ *if and only if* $\mathbf{g} - A^\top\mathbf{p}(\mathbf{x}) \leq \mathbf{0}$*, where*

$$\mathbf{p}(\mathbf{x}) = \psi^{-1}\big(\big(AX^{2\gamma}A^\top\big)^{-1}AX^{2\gamma}\psi(\mathbf{g})\big). \tag{30}$$

(b) $\mathbf{d} = X^{2\gamma}\mathbf{r}(\mathbf{x})$ *satisfies* $A\mathbf{d} = \mathbf{0}$ *and* $\mathbf{g}^\top\mathbf{d} \geq 0$*. Moreover,* $\mathbf{g}^\top\mathbf{d} = 0$ *if and only if* $X\mathbf{r}(\mathbf{x}) = \mathbf{0}$.

Proof The proof of (a) is as in Lemma 3 above. For the proof of (b), we analogously let $\mathbf{s} = (AX^{2\gamma}A^\top)^{-1}AX^{2\gamma}\mathbf{g}$. Furthermore we observe that $A^\top\psi(\mathbf{s}) = \psi(A^\top\mathbf{s})$ by the properties of A, and obtain

$$\begin{aligned}
\mathbf{g}^\top\mathbf{d} &= \mathbf{g}^\top X^{2\gamma}\left(\psi(\mathbf{g}) - A^\top\left(AX^{2\gamma}A^\top\right)^{-1}AX^{2\gamma}\psi(\mathbf{g})\right)\\
&= \left(\mathbf{g} - A^\top\mathbf{s}\right)^\top X^{2\gamma}\psi(\mathbf{g})\\
&= \left(\mathbf{g} - A^\top\mathbf{s}\right)^\top X^{2\gamma}\left(\psi(\mathbf{g}) - A^\top\psi(\mathbf{s})\right)\\
&= \left(\mathbf{g} - A^\top\mathbf{s}\right)^\top X^{2\gamma}\left(\psi(\mathbf{g}) - \psi(A^\top\mathbf{s})\right)\\
&\geq 0,
\end{aligned}$$

where the inequality is justified as in the proof of Lemma 3. Again, the inequality is strict only if $X(\psi(\mathbf{g}) - A^\top\psi(\mathbf{s})) = \mathbf{0}$, which can be used to show that $X\mathbf{r}(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{g}^\top\mathbf{d} = 0$. □

Lemma 4 remains true if different ψ functions are used for different simplices. Lemmas 3 and 4 generalize [4, Theorem 3] for the special case of $\gamma = \frac{1}{2}$ and $\psi(\tau) = \exp(\theta\tau)$ with $\theta > 0$. By Lemma 4(b), $\mathbf{d} = X^{2\gamma}\mathbf{r}(\mathbf{x})$ is a feasible ascent direction at every $\mathbf{x} \in \operatorname{ri}\Lambda$. Using Lemma 4, we can extend Theorem 1(a–c) and Corollary 1 to the monotonely transformed method, assuming furthermore that ψ is continuous.

Theorem 3 *Assume Λ is a Cartesian product of m unit simplices, with A and $\mathbf{b}$ given as in Lemma 4. Then Theorem 1(a–b) still holds. Moreover, Theorem 1(c) under condition (ii) or (iii) still holds if $\mathbf{r}$ is instead given by (29) with ψ any continuous strictly increasing function. This remains true when the Armijo rule is replaced by any stepsize rule that yields a larger ascent as described in Corollary 1.*

Proof It is readily verified that Assumption 1 holds so that, by Lemma 2 and continuity of ψ, the mapping $\mathbf{r}$ given by (29) is continuous on Λ. It is then readily seen from its proof that Theorem 1(a) still holds.

The proof of Theorem 1(b) yields that $\{(\mathbf{g}^k)^\top\mathbf{d}^k\}_{k\in\mathcal{K}} \to 0$, where $\{\mathbf{x}^k\}_{k\in\mathcal{K}}$ ($\mathcal{K} \subseteq \{0, 1, \ldots\}$) is any subsequence of $\{\mathbf{x}^k\}$ converging to some $\bar{\mathbf{x}}$. Since $\mathbf{r}$ is continuous on Λ, this yields in the limit that $\bar{\mathbf{g}}^\top\bar{X}^{2\gamma}\mathbf{r}(\bar{\mathbf{x}}) = 0$, where $\bar{\mathbf{g}} = \nabla f(\bar{\mathbf{x}})$ and $\bar{X} = \operatorname{Diag}(\bar{\mathbf{x}})$. By Lemma 4(b), $\bar{X}\mathbf{r}(\bar{\mathbf{x}}) = \mathbf{0}$. Assuming $\sup_k \alpha_0^k < \infty$, we prove that $\{\mathbf{x}^{k+1} - \mathbf{x}^k\} \to 0$ as follows: Since $\{\mathbf{x}^k\}$ is bounded and every cluster point $\bar{\mathbf{x}}$ satisfies $\bar{X}^\gamma\mathbf{r}(\bar{\mathbf{x}}) = \mathbf{0}$, the continuity of $\mathbf{r}$ implies $\{\mathbf{d}^k\} = \{(X^k)^{2\gamma}\mathbf{r}(\mathbf{x}^k)\} \to \mathbf{0}$. Since $\sup_k \alpha_0^k < \infty$ so that $\sup_k \alpha^k < \infty$, we obtain $\{\mathbf{x}^{k+1} - \mathbf{x}^k\} \to \mathbf{0}$.

The proof of Theorem 1(c) under conditions (ii) or (iii) still holds. The proof shows that every cluster point $\bar{\mathbf{x}}$ of $\{\mathbf{x}^k\}$ satisfies $\mathbf{r}(\bar{\mathbf{x}}) \leq \mathbf{0}$ as well as $\bar{X}\mathbf{r}(\bar{\mathbf{x}}) = \mathbf{0}$. By Lemma 4(a), $\bar{\mathbf{x}}$ satisfies $\bar{X}(\bar{\mathbf{g}} - A^\top\mathbf{p}(\bar{\mathbf{x}})) = \mathbf{0}$ and $\bar{\mathbf{g}} - A^\top\mathbf{p}(\bar{\mathbf{x}}) \leq \mathbf{0}$, where $\mathbf{p}$ is given by (30). Thus $\bar{\mathbf{x}}$ is a stationary point of (1). □

It is not known if Theorem 1(c) under condition (i) or Theorem 2 can be extended to the monotonely transformed method.

Appendix A

In this section we assume f is concave or convex and prove (16) following the line of analysis in [11] and [19, Appendix A]. Let $\bar{\mathcal{J}} = \{j \mid \bar{r}_j = 0\}$, $\bar{\mathcal{J}}^c = \{1, \ldots, n\}\backslash\bar{\mathcal{J}}$, and

$$\bar{\Lambda} = \{\mathbf{x} \in \Lambda \mid \mathbf{x}_{\bar{\mathcal{J}}^c} = 0,\ f(\mathbf{x}) = f(\bar{\mathbf{x}})\}.$$

Lemma 5 *$\bar{\Lambda}$ is convex.*

Proof Since $\bar{X}\bar{\mathbf{r}} = \mathbf{0}$, the point $\bar{\mathbf{x}}$ satisfies the Karush-Kuhn-Tucker (KKT) optimality condition for the restricted problem

$$\text{optimize}\left\{f(\mathbf{x}) \mid A\mathbf{x} = \mathbf{b}\,,\ \mathbf{x}_{\bar{\mathcal{J}}} \geq \mathbf{0}, \mathbf{x}_{\bar{\mathcal{J}}^c} = \mathbf{0}\right\},$$

where "optimize" means "maximize" when f is concave and means "minimize" when f is convex. Thus $\bar{\mathbf{x}}$ is an optimal solution of this restricted problem and $\bar{\Lambda}$ is its optimal solution set. Since this problem is equivalent to a convex minimization problem, its optimal solution set is convex. □

Lemma 6 *If f is constant on a convex subset of $\mathbb{R}^n$, then ∇f is constant on this subset.*

Proof See, e.g., [16]. □

Using Lemmas 5 and 6, we have the following lemma.

Lemma 7 $\mathbf{r}(\mathbf{x}) = \bar{\mathbf{r}}$ *for all* $\mathbf{x} \in \bar{\Lambda}$.

Proof For any $\mathbf{x} \in \bar{\Lambda}$, Lemmas 5 and 6 imply $\nabla f(\mathbf{x}) = \bar{\mathbf{g}} = \bar{\mathbf{r}} + A^\top \bar{\mathbf{p}}$. Thus (7) yields

$$\begin{aligned}\mathbf{r}(\mathbf{x}) &= [I - A^\top (AX^{2\gamma} A^\top)^{-1} AX^{2\gamma}](\bar{\mathbf{r}} + A^\top \bar{\mathbf{p}}) \\ &= [I - A^\top (AX^{2\gamma} A^\top)^{-1} AX^{2\gamma}]\bar{\mathbf{r}} = \bar{\mathbf{r}},\end{aligned}$$

where the last equality uses $\bar{\mathbf{r}}_{\bar{\mathcal{J}}} = \mathbf{0}$ and $\mathbf{x}_{\bar{\mathcal{J}}^c} = \mathbf{0}$, so that $X^{2\gamma}\bar{\mathbf{r}} = \mathbf{0}$. □

Using Theorem 1(b) and Lemma 7, we have the following lemma.

Lemma 8 *Every cluster point of* $\{\mathbf{x}^k\}$ *is in* $\bar{\Lambda}$.

Proof We argue by contradiction. Suppose there exists a cluster point $\hat{\mathbf{x}}$ of $\{\mathbf{x}^k\}$ that is not in $\bar{\Lambda}$. Since $\hat{\mathbf{x}} \in \Lambda$ and $f(\hat{\mathbf{x}}) = f(\bar{\mathbf{x}})$, this implies $\hat{x}_{\hat{j}} > 0$ for some $\hat{j} \in \bar{\mathcal{J}}$.

Since $f(\bar{\mathbf{x}}) \geq f(\mathbf{x}^0)$, the set $\bar{\Lambda}$ lies inside the bounded set Λ^0, so $\bar{\Lambda}$ is compact. Then, by Lemma 2, $\mathbf{r}$ is uniformly continuous over $\bar{\Lambda}$. Lemma 7 implies that, for all $\delta > 0$ sufficiently small, we have

$$|r_j(\mathbf{x})| \geq |\bar{r}_j|/2 \quad \forall j \in \bar{\mathcal{J}}^c,\ \forall \mathbf{x} \in \bar{\Lambda} + \delta B, \tag{31}$$

where B denotes the unit Euclidean ball centered at the origin. Take δ small enough so that $\delta < \hat{x}_{\hat{j}}$. Then $\hat{\mathbf{x}} \notin \bar{\Lambda} + \delta B$ (since $|\hat{x}_{\hat{j}} - x_{\hat{j}}| = \hat{x}_{\hat{j}} > \delta$ for all $\mathbf{x} \in \bar{\Lambda}$). By Theorem 1(c), $\{\mathbf{x}^{k+1} - \mathbf{x}^k\} \to 0$, so the set of cluster points of $\{\mathbf{x}^k\}$ is connected. Since there exists a cluster point in $\bar{\Lambda}$ and another not in $\bar{\Lambda} + \delta B$, there must exist a cluster point $\tilde{\mathbf{x}}$ in $\bar{\Lambda} + \delta B$ but not in $\bar{\Lambda}$. Since $\tilde{\mathbf{x}} \in \Lambda$ and $f(\tilde{\mathbf{x}}) = f(\bar{\mathbf{x}})$, the latter implies $\tilde{\mathbf{x}}_{\bar{\mathcal{J}}^c} \neq \mathbf{0}$. Since $\tilde{\mathbf{x}}$ is in $\bar{\Lambda} + \delta B$, (31) implies $|r_j(\tilde{\mathbf{x}})| \geq |\bar{r}_j|/2$ for all $j \in \bar{\mathcal{J}}^c$. Thus $\tilde{X}r(\tilde{\mathbf{x}}) \neq \mathbf{0}$, where $\tilde{X} = \mathrm{Diag}(\tilde{\mathbf{x}})$, a contradiction of Theorem 1(b). □

Since $\bar{\Lambda}$ is compact and $\mathbf{r}$ is a continuous mapping on Λ, Lemmas 7 and 8 imply $\{\mathbf{r}(\mathbf{x}^k)\} \to \bar{\mathbf{r}}$. This proves (16).

Acknowledgment The research of the first author is supported by the National Science Foundation, Grant No. DMS-0511283.

References

1. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont (1999)
2. Bomze, I.M.: On standard quadratic optimization problems. J. Global Optim. **13**, 369–387 (1998)
3. Bomze, I.M.: Regularity vs. degeneracy in dynamics games and optimization: a unified approach to different aspects. SIAM Rev. **44**, 394–414 (2002)
4. Bomze, I.M.: Portfolio selection via replicator dynamics and projection of indefinite estimated covariances. Dyn. Cont. Discrete Impuls. Syst. Ser. B **12**, 527–564 (2005)
5. Bomze, I. M., Schachinger, W.: Multi-standard quadratic optimization problems. Comput. Optim. Appl. doi:10.1007/s10589-009-9243-8 (2009)
6. Bonnans, J.F., Pola, C.: A trust region interior point algorithm for linearly constrained optimization. SIAM J. Optim. **7**, 717–731 (1997)
7. Dikin, I.I.: Iterative solution of problems of linear and quadratic programming. Soviet Math. Dokl. **8**, 674–675 (1967)
8. Dikin, I.I.: Letter to the editor. Math. Program. **41**, 393–394 (1988)
9. Forsgren, A., Gill, P.E., Wright, M.H.: Interior methods for nonlinear optimization. SIAM Rev. **44**, 525–597 (2002)
10. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization. Academic Press, New York (1981)
11. Gonzaga, C.C., Carlos, L.A.: A primal affine scaling algorithm for linearly constrained convex programs. Tech. Report ES-238/90, Department of Systems Engineering and Computer Science, COPPE Federal University of Rio de Janeiro, Rio de Janeiro, December 1990
12. Hofbauer, J., Sigmund, K.: Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge (1998)
13. Hoffman, A.J.: On approximate solutions of systems of linear inequalities. J. Res. Natl. Bur. Stand. **49**, 263–265 (1952)
14. Luo, Z.-Q., Tseng, P.: On the convergence of the affine-scaling algorithm. Math. Program. **56**, 301–319 (1992)
15. Lyubich, Y., Maistrowskii, G.D., Ol'khovskii, Yu.G.: Selection-induced convergence to equilibrium in a single-locus autosomal population. Probl. Inf. Transm. **16**, 66–75 (1980)
16. Mangasarian, O.L.: A simple characterization of solution sets of convex programs. Oper. Res. Lett. **7**, 21–26 (1988)
17. Monteiro, R.D.C., Tsuchiya, T.: Global convergence of the affine scaling algorithm for convex quadratic programming. SIAM J. Optim. **8**, 26–58 (1998)
18. Monteiro, R.D.C., Tsuchiya, T., Wang, Y.: A simplified global convergence proof of the affine scaling algorithm. Ann. Oper. Res. **46/47**, 443–482 (1993)
19. Monteiro, R.D.C., Wang, Y.: Trust region affine scaling algorithms for linearly constrained convex and concave programs. Math. Program. **80**, 283–313 (1998)
20. Moran, P.A.P.: The statistical Processes of Evolutionary Theory. Clarendon Press, Oxford (1962)
21. Moré, J.J., Garbow, B.S., Hillstrom, K.E.: Testing unconstrained optimization software. ACM Trans. Math. Softw. **7**, 17–41 (1981)
22. Murtagh, B.A., Saunders, M.A.: MINOS 5.5 user's guide. Report SOL 83-20R, Department of Operations Research, Stanford University, Stanford (Revised July 1998)
23. Pelillo, M.: Replicator equations, maximal cliques, and graph isomorphism. Neural Comput. **11**, 1933–1955 (1999)
24. Pelillo, M.: Matching free trees, maximal cliques, and monotone game dynamics. IEEE Trans. Pattern Anal. Machine Intell. **24**, 1535–1541 (2002)
25. Pelillo, M., Siddiqi, K., Zucker, S.W.: Matching hierarchical structures using association graphs. IEEE Trans. Pattern Anal. Mach. Intell. **21**, 1105–1120 (1999)
26. Saigal, R.: The primal power affine scaling method. Ann. Oper. Res. **62**, 375–417 (1996)
27. Sun, J.: A convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions. Math. Program. **60**, 69–79 (1993)
28. Sun, J.: A convergence analysis for a convex version of Dikin's algorithm. Ann. Oper. Res. **62**, 357–374 (1996)
29. Tseng, P.: Convergence properties of Dikin's affine scaling algorithm for nonconvex quadratic minimization. J. Global Optim. **30**, 285–300 (2004)
30. Tsuchiya, T.: Global convergence of the affine scaling methods for degenerate linear programming problems. Math. Program. **52**, 377–404 (1991)

31. Tsuchiya, T.: Global convergence of the affine scaling algorithm for primal degenerate strictly convex quadratic programming problems. Ann. Oper. Res. **46/47**, 509–539 (1993)
32. Tsuchiya, T., Muramatsu, M.: Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems. SIAM J. Optim. **5**, 525–551 (1995)
33. Ye, Y.: On affine scaling algorithms for nonconvex quadratic programming. Math. Program. **56**, 285–300 (1992)