*Acta Mathematica Sinica, English Series*
Jan., 2019, Vol. 35, No. 1, pp. 1–16
Published online: December 15, 2017
https://doi.org/10.1007/s10114-017-7097-z
http://www.ActaMath.com

**Acta Mathematica Sinica, English Series**


# A Partial First-Order Affine-Scaling Method

**Ran GU** **Ya Xiang YUAN**
*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing* 100190*, P. R. China*
*E-mail*: *guran@lsec.cc.ac.cn* *yyx@lsec.cc.ac.cn*

**Abstract** We present a partial first-order affine-scaling method for solving smooth optimization with linear inequality constraints. At each iteration, the algorithm considers a subset of the constraints to reduce the complexity. We prove the global convergence of the algorithm for general smooth objective functions, and show it converges at sublinear rate when the objective function is quadratic. Numerical experiments indicate that our algorithm is efficient.

**Keywords** Linear inequality constraints, affine scaling, interior point

**MR(2010) Subject Classification** 65K05, 90C30, 90C51

## 1 Introduction

Affine scaling technique is widely used in interior point methods for solving linearly constrained optimization problems with the following form:

$$\min_{\theta\in\mathbb{R}^n} f(\theta),$$
$$\text{s.t. } A\theta = b, \quad \theta \geq 0,$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $f: \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. Without loss of generality, we assume that $A$ has rank $m$.

Dikin [9] first proposed a first-order affine-scaling (AS) method for quadratic programming. Proofs of Dikin's AS method was further developed by Adler et al. [1], Barnes [2], Monma and Morton [14], and Vanderbei et al. [19]. Gonzaga and Carlos [10] extended AS method to linearly constrained smooth convex minimization.

At the $t$-th iteration, given an interior point $\theta^t$ as the current iterate point, AS method decides the search direction by solving the following problem

$$\min_{d\in\mathbb{R}^n} \nabla f(\theta^t)^T d,$$
$$\text{s.t. } Ad = 0,$$
$$\|(\Theta^t)^{-1} d\| \leq \mu,$$

where $\Theta^t = \mathrm{Diag}(\theta^t)$ and $\mu < 1$ is a constant, and where $\|\cdot\|$ is the Euclidean norm $\|z\|_2 = \sqrt{z^T z}$. The objective function $\nabla f(\theta^t)^T d$ is a first-order approximation to $f(\theta^t + d) - f(\theta^t)$.

Received March 6, 2017, revised July 4, 2017, accepted August 31, 2017
This paper is partially supported by NSFC (Grant Nos. 11331012 and 11461161005)

$\|(\Theta^t)^{-1}d\| \leq \mu$ is an ellipsoid inside $\theta^t + d \geq 0$. The optimal solution $d^t$ has the following closed form solution

$$d^t = \mu \frac{\hat{d}^t}{\|\hat{d}^t\|}, \tag{1.1}$$

where

$$\hat{d}^t = -(\Theta^t)^2[I - A^T[A(\Theta^t)^2A^T]^{-1}A(\Theta^t)^2]\nabla f(\theta^t).$$

Then $\theta^{t+1} = \theta^t + \alpha^t d^t$, where $\alpha^t$ is chosen by a limited maximization rule on $(0, -1/(\min_j d_j^t/\theta_j^t))$ to ensure that $\theta^{t+1}$ is an interior point.

Bonnans and Pola [4] proposed a first-order AS trust region method, where stepsize $\alpha^t$ is chosen by an Armijo-type rule. They showed every cluster point of the generated iterates is a stationary point. Monteriro and Wang [15] proposed a second-order AS trust-region method based on a generalization of the search direction (1.1). More AS trust-region methods have been proposed (see [3, 21]).

Tseng et al. [18] extended the first order AS method by adding a parameter $\gamma > 0$. Their search direction is

$$d^t = -(\Theta^t)^{2\gamma}[I - A^T[A(\Theta^t)^{2\gamma}A^T]^{-1}A(\Theta^t)^{2\gamma}]\nabla f(\theta^t).$$

They showed that if $\alpha^t$ is chosen by Armijo rule, then every cluster point of $\{\theta^t\}$ is a stationary point under nondegeneracy assumption and additional assumptions such as $f$ being convex or concave. They discussed the value of $\gamma$ and suggested $\gamma < 1$.

In this paper, we consider the following problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^m} f(x), \\ &\text{s.t. } A^T x \leq b, \end{aligned} \tag{1.2}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^n$, $n \geq m$, and $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. Without loss of generality, we assume $A$ has rank $m$. Notice that the constraints are all linear inequalities. Problem (1.2) has been studied extensively, for example, see [7, 16, 22].

To use first-order AS method, we add a slack variable $s \geq 0$, such that $A^T x + s = b$. Then a subproblem for obtaining the search direction can be constructed as

$$\begin{aligned} &\min_{d\in\mathbb{R}^m, s\in\mathbb{R}^n} \nabla f(x^t)^T d, \\ &\text{s.t. } A^T d + s - s^t = 0, \\ &\|(S^t)^{-\gamma}(s - s^t)\| \leq \mu, \end{aligned} \tag{1.3}$$

where $s^t = b - A^T x^t$ and $S^t = \text{Diag}(s^t)$. Ignoring the length of the step-size, we see that the solution of Problem (1.3) is

$$d^t = -[A(S^t)^{-2\gamma}A^T]^{-1}\nabla f(x^t).$$

The complexity of computing $A(S^t)^{-2\gamma}A^T$ is $O(m^2n)$. When $n$ is much larger than $m$, the computation cost is very high. Notice that $A(S^t)^{-2\gamma}A^T = \sum_{j=1}^n (s_j^t)^{-2\gamma}A_jA_j^T$, where $A_j$ is the $j$-th column of $A$. Using the fact that smaller $s_j$ dominate the sum, we consider a partial sum of $(s_j^t)^{-2\gamma}A_jA_j^T$, $j \in I$, where $I$ is an index set. Intuitively, if $s_j^t$, $j \in I$ are much smaller than the others, then the partial sum losses a little. With this, the complexity can be reduced to $O(m^2|I|)$. We propose a partial first-order affine-scaling method, with the search direction defined by

$$d^t = -\Big[\sum_{j\in I^t}(s_j^t)^{-2\gamma}A_jA_j^T\Big]^{-1}\nabla f(x^t),$$

where $I^t$ is the index set corresponding to the $M$ smallest $s_j^t$, and $M$ is a constant such that $M \geq m$. Other ideas to reduce the calculation were provided by Dantzig and Ye [8] and Tseng [17].

In the rest of this paper, we discuss the property of our partial first-order affine-scaling method. We propose our algorithm in Section 2. Meanwhile we propose some assumptions to make our search direction meanful and also provide the Armijo-type rule. We show that if $\alpha^t$ is chosen by Armijo-type rule, then every cluster point of $x^t$ is a stationary point under some assumptions in Section 3. Then we show the sublinear convergence when $f$ is quadratic in Section 4. Numerical results are provided in Section 5. Final conclusion is given in Section 6.

**Preliminaries** $x^0$ is an initial interior point. $g^t = \nabla f(x^t)$. $\Lambda = \{x \in \mathbb{R}^m|\ A^Tx \leq b\}$. $\Lambda^0 = \{x \in \Lambda | f(x) \leq f(x^0)\}$. $s_I = (s_{i_1}, \ldots, s_{i_{|I|}})^T$, $A_I = (A_{i_1}, \ldots, A_{i_{|I|}})$, where $I = \{i_1, \ldots, i_{|I|}\}$. $S_I = \mathrm{Diag}(s_I)$. Thus the search direction can be rewritten as

$$d^t = -[A_{I^t}(S_{I^t}^t)^{-2\gamma}A_{I^t}^T]^{-1}g^t. \tag{1.4}$$

## 2 The Partial First-order Affine-scaling Method

Firstly, we provide the following nondegeneracy assumption.

**Assumption 2.1** For any $x \in \Lambda$, $J = \{j | s_j = b_j - A_j^T x = 0\}$, $A_J$ has rank $|J|$.

This implies that $|J| \leq M$.

To make $d^t$ well defined, we add the following assumption.

**Assumption 2.2** $A_{I^t}$ has rank $m$.

Under Assumption 2.2, $A_{I^t}(S_{I^t}^t)^{-2\gamma}A_{I^t}^T$ is invertible.

Therefore, our $d^t$ is well defined. Moreover, $d^t$ is a descent direction, due to the fact that

$$-(g^t)^Td^t = (g^t)^T[A_{I^t}(S_{I^t}^t)^{-2\gamma}A_{I^t}^T]^{-1}g^t > 0.$$

For general $f$, stepsize $\alpha^t$ is chosen by an Armijo-type rule: $\alpha^t$ is the largest $\alpha \in \{\alpha_0^t\beta^k\}_{k=0,1,\ldots}$ satisfying

$$f(x^t + \alpha d^t) \leq f(x^t) + \sigma_1\alpha(g^t)^Td^t, \tag{2.1}$$

where

$$\alpha_0^t = \sigma_2 \max\{\alpha | x^t + \alpha d^t \in \Lambda\}, \tag{2.2}$$

$0 < \beta, \sigma_1, \sigma_2 < 1$ are constants.

Since $d^t$ is a feasible decent direction at $x^t$, $\alpha^t$ is well defined.

In the following, we present our partial first-order affine-scaling method.

**Algorithm 2.3** *Partial first-order affine-scaling method.*

**Step 0** *Given* $x^0$, $s^0 = b - Ax^0$, $M \geq m$, $0 < \beta, \sigma_1, \sigma_2 < 1$, $t = 0$.

**Step 1** *Compute* $I^t = \{i_1, \ldots, i_M\}$, *such that* $s_{i_1}^t \leq s_{i_2}^t \leq \cdots \leq s_{i_M}^t \leq s_{i_{M+1}}^t \leq \cdots \leq s_{i_n}^t$.

**Step 2** *Compute* $d^t$ *by* (1.4).

**Step 3** *Compute* $\alpha_0^t$ *by* (2.2)*, set* $\alpha = \alpha_0^t$.

**Step 4** *Loop* $\alpha := \alpha\beta$ *until* (2.1) *is satisfied.*

**Step 5** $\alpha^t = \alpha$, $x^{t+1} = x^t + \alpha^t d^t$, $s^{t+1} = b - Ax^{t+1}$. $t := t+1$, go to Step 1.

From the descriptions of our algorithm, it is easy to see that the following lemma is true.

**Lemma 2.4** *Let* $\{x^t\}$ *be generated by Algorithm* 2.3*. Then* $x^t \in \mathrm{int}(\Lambda)$ *for all* $t$*. Moreover,* $\{f(x^t)\}$ *is decreasing.*

## 3 Global Convergence

In this section, we show the global convergence of Algorithm 2.3. Theorems 3.7 and 3.13 state the global convergence of our algorithm in different cases.

To prove the convergence, we need the following assumptions.

**Assumption 3.1** $\Lambda^0$ is bounded.

Since $\{x^t\} \in \Lambda^0$ and $\Lambda^0$ is bounded, $\{x^t\}$ has at least one cluster point. We define $X'$ as the set of all cluster points of $\{x^t\}$.

**Lemma 3.2** *Under Assumptions* 2.1, 2.2 *and* 3.1*, all the sequences* $\{x^t\}$*,* $\{s^t\}$*,* $\{g^t\}$ *and* $\{d^t\}$ *are bounded.*

*Proof* The boundedness of $\Lambda^0$ implies that $\{x^t\}$ is bounded and $\{s^t = b - Ax^t\}$ is bounded. Since $\{x^t\}$ is bounded and $\nabla f$ is continuous on $\Lambda^0$, it is obvious that $\{g^t\}$ is bounded.

We have $d^t = -[A_{I^t}(S_{I^t}^t)^{-2\gamma}A_{I^t}^T]^{-1}g^t$,

$$\begin{aligned}\|d^t\|^2 &\le \lambda_{\max}\{[A_{I^t}(S_{I^t}^t)^{-2\gamma}A_{I^t}^T]^{-2}\}\|g^t\|^2\\ &= \{\lambda_{\min}[A_{I^t}(S_{I^t}^t)^{-2\gamma}A_{I^t}^T]\}^{-2}\|g^t\|^2\\ &\le \{\lambda_{\min}[s_u^{-2\gamma}A_{I^t}A_{I^t}^T]\}^{-2}\|g^t\|^2\\ &= s_u^{4\gamma}\{\lambda_{\min}[A_{I^t}A_{I^t}^T]\}^{-2}\|g^t\|^2,\end{aligned}$$

where $s_u$ is the upper bound of $\{s^t\}$. Since the number of subsets $I^t$ is finite and $A_{I^t}$ has rank $m$, $\lambda_{\min}[A_{I^t}A_{I^t}^T]$ have a uniform lower bound which is larger than 0. Therefore, $\{d^t\}$ is bounded. □

In order to show the convergence, we introduce a new variable.

For any $x \in \mathrm{int}(\Lambda)$, we define $y \in \mathbb{R}^n$ by

$$\begin{aligned}y_I &= -(S_I)^{-2\gamma}A_I^T d = S_I^{-2\gamma}A_I^T(A_I S_I^{-2\gamma}A_I^T)^{-1}g,\\ y_{I^C} &= 0,\end{aligned}$$

where $s = b - A^T x$, $I$ is the index subset we choose, $g = \nabla f(x)$ and $d = (A_I S_I^{-2\gamma}A_I^T)^{-1}g$.

**Lemma 3.3** *Let* $\bar{x}$ *be any point on* $\Lambda$*,* $\bar{J} = \{j|\bar{s}_j = 0\}$*. Under Assumptions* 2.1, 2.2 *and* 3.1*, there exists* $\bar{\delta}$*, such that* $I(x) \supseteq \bar{J}$ *for all* $x \in B(\bar{x}, \bar{\delta}) \cap \Lambda$.

*Proof* Notice that $s$ is continuous. Combining $\bar{s}_j = 0$ for $j \in \bar{J}$ and $\bar{s}_j \ne 0$ for $j \in \bar{J}^C$, there exists a constant $\bar{\delta}$, such that $\max\{s_j|j \in \bar{J}\} < \min\{s_j|j \in \bar{J}^C\}$ for all $x \in B(\bar{x}, \bar{\delta}) \cap \Lambda$. Since $|\bar{J}| \le M \le |I(x)|$, we have $I(x) \supseteq \bar{J}$. □

**Lemma 3.4** *Let $\bar{x}$ be any point on $\Lambda$, $J=\{j|\bar{s}_j=0\}$, $J^C=\{j|\bar{s}_j>0\}$. Then under Assumptions 2.1, 2.2 and 3.1, there exists a constant $\bar{\delta}$ such that when $x\in B(\bar{x},\bar{\delta})\cap \mathrm{int}(\Lambda)$, $y$ is bounded. Moreover, let $\mathcal{T}$ be any subsequence such that $\{x^t\}_{t\in\mathcal{T}}$ converge to $\bar{x}$ and $I^t=I$ for all $t\in\mathcal{T}$, then*

$$\{y_J^t\}_{t\in\mathcal{T}}\to(A_J^T\bar{D}^{-1}A_J)^{-1}A_J^T\bar{D}^{-1}\bar{g},$$
$$\{d^t\}_{t\in\mathcal{T}}\to[\bar{D}^{-1}-\bar{D}^{-1}A_J(A_J^T\bar{D}^{-1}A_J)^{-1}A_J^T\bar{D}^{-1}]\bar{g},$$

*where $\bar{D}=A_{I\cap J^C}\bar{S}_{I\cap J^C}^{-2\gamma}A_{I\cap J^C}^T+A_JA_J^T$.*

*Proof* According to Lemma 3.3, there exists a constant $\bar{\delta}$ such that when $x\in B(\bar{x},\bar{\delta})\cap\mathrm{int}(\Lambda)$, $I\supseteq J$ . Let $W=S_J^{-2\gamma}-I$ and $D=A_{I\cap J^C}S_{I\cap J^C}^{-2\gamma}A_{I\cap J^C}^T+A_JA_J^T$. $D$ is bounded by

$$\min\{1,\max\{s_{I\cap J^C}\}^{-2\gamma}\}A_IA_I^T\preceq D\preceq\max\{1,\min\{s_{I\cap J^C}\}^{-2\gamma}\}A_IA_I^T.$$

Using Sherman–Morrison–Woodbury formula that

$$\begin{aligned}(A_IS_I^{-2\gamma}A_I^T)^{-1}&=(D+A_JWA_J^T)^{-1}\\&=D^{-1}-D^{-1}A_JW^{\frac{1}{2}}[I+W^{\frac{1}{2}}A_J^TD^{-1}A_JW^{\frac{1}{2}}]^{-1}W^{\frac{1}{2}}A_J^TD^{-1},\end{aligned}$$

it follows that

$$A_J(A_IS_I^{-2\gamma}A_I^T)^{-1}=W^{-1}(W^{-1}+A_J^TD^{-1}A_J)^{-1}A_J^TD^{-1}.$$

Then

$$y_J=S_J^{-2\gamma}A_J^T(A_IS_I^{-2\gamma}A_I^T)^{-1}g=S_J^{-2\gamma}W(W^{-1}+A_J^TD^{-1}A_J)^{-1}A_J^TD^{-1}g.$$

When $s_J$ goes to zero, $W^{-1}$ goes to zero and $S_J^{-2\gamma}W$ goes to identity. Therefore, $y_J$ is bounded. Moreover, when $I$ is fixed, $y_J$ goes to $(A_J^T\bar{D}^{-1}A_J)^{-1}A_J^T\bar{D}^{-1}\bar{g}$, and

$$d=(A_IS_I^{-2\gamma}A_I^T)^{-1}g\to[\bar{D}^{-1}-\bar{D}^{-1}A_J(A_J^T\bar{D}^{-1}A_J)^{-1}A_J^T\bar{D}^{-1}]\bar{g}.\qquad\square$$

**Lemma 3.5** *Under Assumptions 2.1, 2.2 and 3.1, when $\gamma\geq\frac{1}{2}$, $\inf\alpha_0^t>0$.*

*Proof* Let $\bar{x}$ be any cluster point of $\{x^t\}$. $J$ is defined by $J=\{j|\bar{s}_j=0\}$. Considering $x\in B(\bar{x},\bar{\delta})$, we have $I(x)$ is fixed.

Since $A_J^Td=W^{-1}(W^{-1}+A_J^TD^{-1}A_J)^{-1}A_J^TD^{-1}g$, then $A_j^Td=O(s_j^{2\gamma})$ for $j\in J$. When $x$ is in a neighborhood of $\bar{x}$, $s_j/|A_j^Td|\geq C_Js_j^{1-2\gamma}$ for $j\in J$, where $C_J$ is a constant depending on $\bar{x}$ but not $x$. Because $d$ is bounded, $A_{J^C}^Td$ is bounded. It follows that $s_j/|A_j^Td|\geq C_{J^C}s_j$ for $j\in J^C$. The initial stepsize can be calculated by

$$\alpha_0=\sigma_2\min_{\{j:A_j^Td>0\}}s_j/|A_j^Td|.$$

Overall, there exists an $\bar{\epsilon}$ such that $\alpha_0\geq\bar{C}_0$ for all $x\in B(\bar{x},\bar{\epsilon})\cap\Lambda^0$, where $\bar{C}_0$ is a constant depending on $\bar{x}$ but not $x$.

For a proof by contradiction, we suppose that $\inf\alpha_0^t=0$. It follows that there exists a subsequence $\mathcal{T}$ such that $\alpha_0^t\to0$ for $t\in\mathcal{T}$. Since $\Lambda^0$ is bounded, there exists a subsequence $\mathcal{T}_1\subseteq\mathcal{T}$, $\{x^t\}_{t\in\mathcal{T}_1}\to\bar{x}\in X'$. When $t$ is large enough, $x^t$ is in $B(\bar{x},\bar{\epsilon})$ for $t\in\mathcal{T}_1$, then $\alpha_0^t\geq\bar{C}_0$. This contradicts to $\alpha_0^t\to0$ for $t\in\mathcal{T}_1$. $\square$

**Lemma 3.6** *Under Assumptions 2.1, 2.2 and 3.1, when $\gamma\geq\frac{1}{2}$, $(g^t)^Td^t\to0$.*

*Proof* Let $\bar{x}$ be a cluster point of $\{x^t\}$. Then $f(x^t) \to f(\bar{x})$, and $f(x^t) - f(x^{t+1}) \to 0$. According to the stepsize rule, we have

$$\alpha^t (g^t)^T d^t \to 0. \tag{3.1}$$

For a proof by contradiction, we suppose that $(g^t)^T d^t \nrightarrow 0$. This means that there exists a subsequence $\mathcal{T}$ and a constant $a$ such that

$$|(g^t)^T d^t| > a > 0 \tag{3.2}$$

for $t \in \mathcal{T}$. Due to the boundedness of $\{x^t\}_{t\in\mathcal{T}}$ and the finiteness of $I^t$, there exists a subsequence $\mathcal{T}_1 \in \mathcal{T}$ such that $\{x^t\}_{t\in\mathcal{T}_1}$ converges to $\bar{x}$ and $I^t = I$ for all $t \in \mathcal{T}_1$. By (3.1) and (3.2), we have $\{\alpha^t\}_{t\in\mathcal{T}_1} \to 0$. According to Lemma 3.5, $\alpha^t < \alpha_0^t$ when $t \in \mathcal{T}$ is large enough. From the stepsize rule, condition (2.1) is violated by $\alpha = \alpha^t/\beta$. We have

$$\frac{f(x^t + (\alpha^t/\beta)d^t) - f(x^t)}{\alpha^t/\beta} > \sigma_1 (g^t)^T d^t, \quad \forall t \in \mathcal{T}_1. \tag{3.3}$$

Since $g^t$ and $d^t$ have limits, we take the limit of the above inequality with $\{\alpha^t\}_{t\in\mathcal{T}_1}$, we obtain $\bar{g}^T\bar{d} \ge \sigma_1 \bar{g}^T\bar{d}$. Due to $0 < \sigma_1 < 1$, it follows that $\bar{g}^T\bar{d} \ge 0$. On the other hand, $(g^t)^T d^t < 0$, therefore $\bar{g}^T\bar{d} = 0$. This is a contrary to (3.2). □

**Theorem 3.7** *Let $\bar{x} \in X'$. Under Assumptions* 2.1, 2.2 *and* 3.1, *and $\gamma \ge \frac{1}{2}$ is satisfied, then there exists a unique $\bar{y}$ such that*

$$\begin{aligned} y^t &\to \bar{y}, \\ \bar{S}\bar{y} &= 0, \\ \bar{g} &= A\bar{y}. \end{aligned}$$

*Proof* From the fact

$$(g^t)^T d^t = (g^t)^T (A_{I^t}(S_{I^t}^t)^{-2\gamma} A_{I^t}^T)^{-1} g^t = \|(S_{I^t}^t)^{-\gamma} A_{I^t}^T (A_{I^t}(S_{I^t}^t)^{-2\gamma} A_{I^t}^T)^{-1} g^t\|^2 = \|(S^t)^\gamma y^t\|^2$$

and $(g^t)^T d^t \to 0$, it follows that

$$(S^t)^\gamma y^t \to 0.$$

For any subsequence $\mathcal{T}$ such that $I^t$ is fixed for all $t \in \mathcal{T}$, we have $\{y^t\}$ has limit. It yields $\bar{y}_{J^C} = 0$ from $\bar{s}_{J^C} \ne 0$. We take the limit $t \in \mathcal{T} \to \infty$ on $g^t = Ay^t$, then we have $\bar{g} = A\bar{y} = A_J \bar{y}_J$. Since $A_J$ has full column rank, $\bar{y}$ is unique. For different subsequences, we have $\bar{g}$ and $A_J$ are fixed, which means $\bar{y}$ are the same. □

**Lemma 3.8** *Under Assumptions* 2.1, 2.2 *and* 3.1, *if $\gamma \ge \frac{1}{2}$, then $\{x^{t+1} - x^t\} \to 0$.*

*Proof* Let $\bar{x} \in X'$, $\mathcal{T}$ be any subsequence such that $\{x^t\}_{t\in\mathcal{T}} \to \bar{x}$. We define $H = \{j | \bar{y}_j \ne 0\}$. So we have $\bar{s}_H = 0$, which implies $H \subseteq J$. Since $\alpha^t (g^t)^T d^t \to 0$ and $(g^t)^T d^t = \|(S^t)^\gamma y^t\|^2$, then $(\alpha^t)^{\frac{1}{2}} (S^t)^\gamma y^t \to 0$. Due to $\bar{y}_j \ne 0$ for $j \in H$, we have $\{(\alpha^t)^{\frac{1}{2}} (S_H^t)^\gamma\}_{t\in\mathcal{T}} \to 0$.

Notice that

$$\begin{aligned} \lim_{\{t\in\mathcal{T}\}\to\infty} x^{t+1} - x^t &= \lim_{\{t\in\mathcal{T}\}\to\infty} \alpha^t d^t \\ &= \lim_{\{t\in\mathcal{T}\}\to\infty} \alpha^t (A_{I^t}(S_{I^t}^t)^{-2\gamma} A_{I^t}^T)^{-1} g^t \end{aligned}$$

$$= \lim_{\{t\in\mathcal{T}\}\to\infty} \alpha^t (A_{I^t}(S^t_{I^t})^{-2\gamma}A^T_{I^t})^{-1}A_H\bar{y}_H.$$

Since

$$A^T_H(A_{I^t}(S^t_{I^t})^{-2\gamma}A^T_{I^t})^{-1} = W_H^{-1}[D^{-1}A_J(W^{-1}+A_J^TD^{-1}A_J)^{-1}]^T_H,$$

then we have, for $t\in\mathcal{T}$,

$$\begin{aligned} x^{t+1}-x^t &\to \alpha^t[D^{-1}A_J(W^{-1}+A_J^TD^{-1}A_J)^{-1}]_H W_H^{-1} \\ &\to \alpha^t[\bar{D}^{-1}A_J(A_J^T\bar{D}^{-1}A_J)^{-1}]_H(S^t_H)^{2\gamma}\bar{y}_H. \end{aligned}$$

Notice that $D$ has both upper and lower bound. Thus

$$\{x^{t+1}-x^t\}_{t\in\mathcal{T}}\to 0.$$

Due to the arbitrariness of $\mathcal{T}$, it is easy to show that $\lim_{t\to\infty} x^{t+1}-x^t=0$. □

To prove the global convergence, we only need $\bar{y}\le 0$. In the rest of this section, we consider two cases. One is strict complementarity, the other is that the objective function is convex or concave. In either case, we show the global convergence.

**Theorem 3.9** *Under Assumptions* 2.1, 2.2 *and* 3.1*, if* $\gamma\ge\frac{1}{2}$ *and every* $x\in\Lambda_{CS}$ *satisfies strict complementarity* (*i.e.,* $s_j-y_j\neq 0$ *for all* $j$)*, where*

$$\Lambda_{CS}=\Big\{x\in\Lambda\cap X'|Sy=0,\ f(x)=\lim_{t\to\infty}f(x^t)\Big\}.$$

*Then every cluster point of* $\{x^t\}$ *is a stationary point of* (1.2)*.*

*Proof* Let $\bar{x}\in X'$ and

$$\bar{\Lambda}_{CS}=\{x\in\Lambda_{CS}|y_{\bar{J}_0}=0, y_{\bar{J}_+}>0, y_{\bar{J}_-}<0\},$$

where $\bar{J}_0=\{j|\bar{y}_j=0\}$, $\bar{J}_+=\{j|\bar{y}_j>0\}$, $\bar{J}_-=\{j|\bar{y}_j<0\}$.

For each $x\in\Lambda_{CS}$ satisfies strict complementarity, then there exists a $\delta$ such that $\bar{\Lambda}_{CS}+B(0,\delta)\cap\Lambda_{CS}=\bar{\Lambda}_{CS}$. By Lemma 3.8, we have $x^{t+1}-x^t\to 0$, hence it implies that $X'$ forms a continuum. Therefore $X'\subseteq\Lambda_{CS}$ and $X'\subseteq\bar{\Lambda}_{CS}$.

For each $j\in\bar{J}_+$, we have $y^t_{\bar{J}_+}>0$, when $t$ is large enough. Since $s^{t+1}_{J_+}-s^t_{J_+}=\alpha^t(S^t_{J_+})^{2\gamma}y^t_{J_+}$ $>0$, when $t$ is large enough. It follows that $\liminf_t s^t_{J_+}>0$, which contradicts with $\bar{s}_{J_+}=0$. Therefore, $J_+=\emptyset$, i.e., $\bar{y}\le 0$. □

In the following of this section, we analyze the convergence when $f$ is convex or concave.

**Lemma 3.10** *Let* $\bar{x}\in X'$*,* $H=\{j|\bar{y}_j\neq 0\}$*,* $\bar{\Lambda}=\{x\in\Lambda|s_H=0, f(x)=f(\bar{x})\}$*. If* $f$ *is convex or concave, then* $\bar{\Lambda}$ *is convex and* $\nabla f$ *is constant on* $\bar{\Lambda}$*.*

*Proof* Since $\bar{S}\bar{y}=0$, then $\bar{x}$ is the optimal solution of the following problem

$$\text{optimize}\{f(x)|A^T_Hx=b_H, A^T_{H^C}x\le b_{H^C}\}. \tag{3.4}$$

When $f$ is concave, "optimize" is "maximize". When $f$ is convex, "optimize" is "minimize". Therefore, $\bar{\Lambda}$ is the optimal solution set. Due to (3.4) is equivalent to a convex minimization problem, $\bar{\Lambda}$ is convex. Then according to [13], it follows that $\nabla f$ is constant on $\bar{\Lambda}$. □

**Lemma 3.11** *Under Assumptions* 2.1, 2.2 *and* 3.1*, if* $\gamma\ge\frac{1}{2}$ *and* $f$ *is convex or concave, then* $y(x)=\bar{y}$ *for all* $x\in\bar{\Lambda}\cap X'$*.*

*Proof* For any $\hat{x} \in \bar{\Lambda} \cap X'$, since $s_{\hat{J}} = 0$ and $H \subseteq \hat{J}$, we have $g = A_H \bar{y}_H = A_{\hat{J}} \bar{y}_{\hat{J}}$. It follows that $A_{\hat{J}} \hat{y}_{\hat{J}} = g = A_{\hat{J}} \bar{y}_{\hat{J}}$. Since $A_{\hat{J}}$ has full column rank, we have $\hat{y}_{\hat{J}} = \bar{y}_{\hat{J}}$. Furthermore, $\hat{y}_{\hat{J}^C} = 0 = \bar{y}_{\hat{J}^C}$. Therefore, $y(x) = \bar{y}$ for all $x \in \bar{\Lambda} \cap X'$. □

**Lemma 3.12** *Under Assumptions* 2.1, 2.2 *and* 3.1, *if* $\gamma \geq \frac{1}{2}$ *and* $f$ *is convex or concave, then* $X' = \bar{\Lambda} \cap X'$.

*Proof* Firstly, we show that there exists a constant $\delta > 0$ such that $|y_j| \geq \frac{1}{2}|\bar{y}_j|$ for all $x \in [\bar{\Lambda} + \delta B(0, \delta)] \cap X'$ and all $j \in H$.

If not, then there exists a subsequence $\mathcal{T}$ and an index $j$, such that $x^t \in (\Lambda \cap X') + B(0, \delta^t)$, $\delta^t \to 0$ and $|y_j| < \frac{1}{2}|\bar{y}_j|$. Due to the boundness of $\{x^t\}$, there exists a subsequence $\mathcal{T}_1$ of sequence $\mathcal{T}$, such that $x^t \to \hat{x} \in \Lambda \cap X'$. Since $y \to \hat{y} = \bar{y}$, it follows that $|y_j| \geq \frac{1}{2}|\bar{y}_j|$ for all $j \in H$. This is a contradiction.

For a proof by contradiction, we suppose that there exists $\hat{x} \in X'$ but $\hat{x} \notin \bar{\Lambda}$. By $\{x^{t+1} - x^t\} \to 0$, we have that $X'$ forms a continuum. Then there exists $\tilde{x} \in X'$ such that $\tilde{x} \in [\bar{\Lambda} + \delta B(0, \delta)] \cap X'$ but $\tilde{x} \notin \bar{\Lambda}$. It follows that there exists $j \in H$ such that $\tilde{s}_j > 0$. Therefore $\tilde{s}_j \tilde{y}_j \neq 0$ contradicts with $\tilde{x} \in X'$. □

Combining Lemmas 3.11 and 3.12, we claim $y^t \to \bar{y}$ for all $t$.

**Theorem 3.13** *Under Assumptions* 2.1, 2.2 *and* 3.1, *if* $\gamma \geq \frac{1}{2}$, $f$ *is convex or concave, then every cluster point of* $\{x^t\}$ *is a stationary point of* (1.2).

*Proof* We only need to prove $\bar{y} \leq 0$. Since $y^t \to \bar{y}$, we suppose $\bar{y} \nleq 0$, then there exists $j \in J$ and $T$ such that $y_j^t > 0$ for all $t > T$. Since $s_j^{t+1} - s_j^t = \alpha^t (S_j^t)^{2\gamma} y_j^t$, we have $s_j^t > s_j^T$ for all $t > T$. This contradicts with $\bar{s}_j = 0$. Therefore, $\bar{y} \leq 0$. □

We extend the proof in [18] to prove the convergence of our algorithm. However in our proof, we propose none assumptions on stepsize $\alpha_0^t$ and $\alpha^t$.

## 4 Sublinear Convergence When $f$ is Quadratic

In this section, we show $\{f(x^t)\}$ converges sublinearly when $f$ is quadratic. Moreover, we show $\{x^t\}$ converges sublinearly for $\gamma < 1$. If $\gamma = \frac{1}{2}$ and strict complementarity holds, then the convergence is linear.

Let

$$f(x) = \frac{1}{2} x^T Q x + c^T x.$$

**Theorem 4.1** *Suppose Assumptions* 2.1, 2.2 *and* 3.1 *hold, and* $\gamma \geq \frac{1}{2}$. *Then the following results hold with* $\omega = 1/(\bar{\gamma} - 1)$ *and* $\bar{\gamma} = \max\{1 + \gamma, 2\gamma\}$.

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

$$0 \leq f(x^t) - v \leq C t^{-\omega}, \quad \forall t \geq 1. \tag{4.1}$$

(2) *Assume* $\frac{1}{2} \leq \gamma < 1$. *Then there exist* $\bar{x} \in \Lambda^0$ *and* $C' > 0$ (*depending on* $x^0$) *such that*

$$\|\bar{x} - x^t\| \leq C' t^{-\frac{1-\gamma}{2\gamma}}, \quad \forall t \geq 1.$$

$\bar{x}$ *is a stationary point.*

*Moreover, if* $\gamma = \frac{1}{2}$ *and* $\bar{s} - \bar{y} > 0$, *then* $\{f(x^t)\}$ *converges Q-linearly and* $\{\|\bar{x} - x^t\|\}$ *converges R-linearly.*

*Proof* Let $\eta^t = (S^t)^\gamma y^t$ and $v = \lim_t f(x^t)$. We have

$$f(x^t) - f(x^{t+1}) \geq \sigma_1 \alpha^t \|\eta^t\|^2 \tag{4.2}$$

and $\{f(x^t) - f(x^{t+1})\} \to 0$. By Lemma 3.6, we obtain $\{\eta^t\} \to 0$. For any index set $J \subseteq \{1, \ldots, n\}$, we define

$$\mathcal{T}_J = \{t | s_j \leq |\eta_j^t|^{\frac{1}{1+\gamma}} \ \forall j \in J, \ |y_j| < |\eta_j^t|^{\frac{1}{1+\gamma}}, \forall j \in J^C\}.$$

Since $(s_j^t)^\gamma |y_j^t| = |\eta_j^t|$, we have only one of $s_j \leq |\eta_j^t|^{\frac{1}{1+\gamma}}$ and $|y_j| < |\eta_j^t|^{\frac{1}{1+\gamma}}$ holds. The number of different $J$ is infinite, therefore there exists an index set $J$, such that $\mathcal{T}_J$ is infinite. Consider any infinite set $\mathcal{T}_J$, for each $t \in \mathcal{T}_J$, the following linear system in $(x, s, y)$

$$S_J = S_J^t, \quad Qx + c = Ay, \quad y_{J^C} = y_{J^C}^t, \quad s \geq 0, \quad A^T x + s = b$$

has a solution $(x^t, s^t, y^t)$. Let $\|\cdot\|_\nu$ denote the $\nu$-norm. Then we have

$$\|(s_J^t, y_{J^C}^t)\|_{1+\gamma}^{1+\gamma} \leq \|\eta^t\|_1, \quad \forall t \in \mathcal{T}_J.$$

Since $\{\eta^t\} \to 0$, it follows that $\{(s_J^t, y_{J^C}^t)\}_{t \in \mathcal{T}_J} \to 0$. Therefore, every cluster point of $\{x^t, s^t, y^t\}$ satisfies

$$s_J = 0, \quad Qx + c = Ay, \quad y_{J^C} = 0, \quad s \geq 0, \quad A^T x + s = b.$$

Denote the solution set of this linear system by $\Sigma_J$. Using an error bound of Hoffman [11], there exists $(\bar{x}^t, \bar{s}^t, \bar{y}^t) \in \Sigma_J$ such that

$$\|(\bar{x}^t, \bar{s}^t, \bar{y}^t) - (x^t, s^t, y^t)\| \leq C_1 \|(s_J^t, y_{J^C}^t)\|_{1+\gamma}, \quad \forall t \in \mathcal{T}_J,$$

where $C_1$ is a constant depending on $Q$, $A$ and $J$.

Firstly, we show $f$ is constant on $\Sigma_J$. Suppose $(x, s, y), (x', s', y') \in \Sigma_J$,

$$\begin{aligned} f(x') - f(x) &= \frac{1}{2}(x' - x)^T Q(x' - x) + (Qx + c)^T (x' - x) \\ &= \frac{1}{2}(x' - x)^T Q(x' - x) + y^T A^T (x' - x) \\ &= \frac{1}{2}(x' - x)^T Q(x' - x) + y^T (s - s') \\ &= \frac{1}{2}(x' - x)^T Q(x' - x). \end{aligned}$$

A symmetric argument yields $f(x) - f(x') = \frac{1}{2}(x' - x)^T Q(x' - x)$. It implies $f(x) = f(x')$.

For any $t \in \mathcal{T}_J$,

$$(Q\bar{x}^t + c)^T (x^t - \bar{x}^t) = (Q\bar{x}^t + c - A\bar{y}^t)^T (x^t - \bar{x}^t) + (\bar{y}^t)^T (\bar{s}^t - s^t) = (\bar{y}^t)^T (\bar{s}^t - s^t).$$

If $\gamma \leq 1$, then it follows that

$$\begin{aligned} |f(x^t) - f(\bar{x}^t)| &= \left| \frac{1}{2}(x^t - \bar{x}^t)^T Q(x^t - \bar{x}^t) + (Q\bar{x}^t + c)^T (x^t - \bar{x}^t) \right| \\ &= \left| \frac{1}{2}(x^t - \bar{x}^t)^T Q(x^t - \bar{x}^t) + (\bar{y}^t)^T (\bar{s}^t - s^t) \right| \\ &\leq \frac{1}{2} |(x^t - \bar{x}^t)^T Q(x^t - \bar{x}^t)| + \sum_{j \in J} |\bar{y}_j^t (\bar{s}_j^t - s_j^t)| \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2}|(x^t-\bar{x}^t)^TQ(x^t-\bar{x}^t)| + \sum_{j\in J}|\bar{y}_j^t s_j^t| \\
&\le C_2\|x^t-\bar{x}^t\|^2 + \sum_{j\in J}|(\bar{y}_j^t - y_j^t)s_j^t| + \sum_{j\in J}|y_j^t s_j^t| \\
&\le C_2\|x^t-\bar{x}^t\|^2 + \sum_{j\in J}|\bar{y}_j^t - y_j^t|s_j^t + \sum_{j\in J}|\eta_j^t|(s_j^t)^{1-\gamma} \\
&\le C_2C_1^2\|\eta^t\|_1^{\frac{2}{1+\gamma}} + \sum_{j\in J}C_1\|\eta^t\|_1^{\frac{1}{1+\gamma}}|\eta_j^t|^{\frac{1}{1+\gamma}} + \sum_{j\in J}|\eta_j^t|^{\frac{2}{1+\gamma}},
\end{aligned}$$

where $C_2$ depends on $Q$. The last inequality in true due to $\gamma \le 1$. We have

$$|f(x^t) - f(\bar{x}^t)| \le C_J\|\eta^t\|_1^{\frac{2}{1+\gamma}}, \quad \forall t \in \mathcal{T}_J, \tag{4.3}$$

where $C_J$ is a constant depending on $Q$, $A$, $J$, $x^0$. If $\gamma > 1$, then $|y_j^t s_j^t| = |\eta_j^t|^{\frac{1}{\gamma}}|y_j^t|^{1-\frac{1}{\gamma}}$. Using $\{y^t\}$ is bounded and $\frac{2}{1+\gamma} > \frac{1}{\gamma}$, we have

$$|f(x^t) - f(\bar{x}^t)| \le C_J\|\eta^t\|_1^{\frac{1}{\gamma}}, \quad \forall t \in \mathcal{T}_J. \tag{4.4}$$

Let $C_3 = \max\{C_J |\, |\mathcal{T}_J| = \infty\}$.

Since $f(x^t) \to \nu$ and $\{\eta^t\} \to 0$, it follows from (4.3) and (4.4) that $\{f(\bar{x}^t)\}_{t\in\mathcal{T}_J} \to \nu$. Since $\bar{x}^t \in \Sigma_J$ and $f$ is constant on $\Sigma_J$, we have $f(\bar{x}^t) = \nu$. Therefore, for all $t \in \mathcal{T}_J$,

$$\begin{aligned}
f(x^t) - \nu = f(x^t) - f(\bar{x}^t) &\le C_3\|\eta^t\|_1^{\min\{\frac{2}{1+\gamma},\frac{1}{\gamma}\}} \\
&= C_3\|\eta^t\|_1^{\frac{2}{\bar{\gamma}}} \\
&\le \kappa(f(x^t) - f(x^{t+1}))^{\frac{1}{\bar{\gamma}}}.
\end{aligned}$$

The last inequality is from (4.2) and $\inf_t \alpha^t > 0$. Denote $\Delta^t = f(x^t) - \nu$. Then

$$\Delta^{t+1} \le \Delta^t - \left(\frac{\Delta^t}{\kappa}\right)^{\bar{\gamma}} \tag{4.5}$$

holds for all $t \in \mathcal{T}_J$ and all $J$ such that $\mathcal{T}_J$ is infinite. It implies when $t$ is larger than some $T$, (4.5) holds.

(1) Take $C \ge \max\{\kappa^{\frac{\bar{\gamma}}{\bar{\gamma}-1}}, (\frac{\kappa^{\bar{\gamma}}}{\bar{\gamma}-1})^{\frac{1}{\bar{\gamma}-1}}\}$, sufficiently large so that $f(x^t) - \nu \le Ct^{-\omega}$ holds for $t = 1, \dots, T$. When $\omega \le 1$, it follows from $C \ge \kappa^{\frac{\bar{\gamma}}{\bar{\gamma}-1}}$ that $(\frac{C}{\kappa})^{\bar{\gamma}} \ge C$. Then

$$\Delta^{t+1} \le \Delta^t - \left(\frac{\Delta^t}{\kappa}\right)^{\bar{\gamma}} \le \frac{C}{t^\omega} - \left(\frac{C}{\kappa t^\omega}\right)^{\bar{\gamma}} \le C\left(\frac{1}{t^\omega} - \frac{1}{t^{\omega\bar{\gamma}}}\right) \le \frac{C}{(t+1)^\omega}.$$

The last inequality is implied by $\omega\bar{\gamma} = 1 + \omega$ and $(1 - \frac{1}{t+1})^\omega \ge 1 - \frac{1}{t+1}$. When $\omega > 1$, since $C \ge (\frac{\kappa^{\bar{\gamma}}}{\bar{\gamma}-1})^{\frac{1}{\bar{\gamma}-1}}$ yields $(\frac{C}{\kappa})^{\bar{\gamma}} \ge \frac{C}{\bar{\gamma}-1}$, we have

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

because $(1 - \frac{1}{t+1})^\omega \ge 1 - \frac{\omega}{t+1}$.

(2) Assume $\gamma < 1$, then $\bar{\gamma} = 1 + \gamma < 2$. We have, for all $t > T$,

$$\begin{aligned}
\|s_{I^t}^{t+1} - s_{I^t}^t\| &= \alpha^t\|(S^t)^\gamma\eta^t\| \\
&\le \alpha^t\|s^t\|_\infty^\gamma\|\eta^t\|
\end{aligned}$$

$$
\begin{aligned}
&= \alpha^t \|s^t\|_\infty^\gamma \frac{\|\eta^t\|^2}{\|\eta^t\|} \\
&\le \|s^t\|_\infty^\gamma \sqrt{n} \frac{\alpha^t \|\eta^t\|^2}{\|\eta^t\|_1} \\
&\le \|s^t\|_\infty^\gamma \sqrt{n} \frac{f(x^t) - f(x^{t+1})}{\sigma_1 \|\eta^t\|_1} \\
&\le \|s^t\|_\infty^\gamma \sqrt{n} \frac{\Delta^t - \Delta^{t+1}}{\sigma_1} \left(\frac{C_3}{\Delta^t}\right)^{\frac{\bar{\gamma}}{2}}.
\end{aligned}
$$

On the other hand, $\|s_{I^t}^{t+1} - s_{I^t}^t\| = \|A_{I^t}^T(x^{t+1} - x^t)\| \ge C_A \|x^{t+1} - x^t\|$, where $C_A$ depends on $A$. This is because the number of different $I$ is finite, and for each $I$, $A_I$ has full row rank. Then

$$
\begin{aligned}
\|x^{t+1} - x^t\| &\le C_4 (\Delta^t - \Delta^{t+1})(\Delta^t)^{-\frac{\bar{\gamma}}{2}} \\
&\le C_4 \int_{\Delta^{t+1}}^{\Delta^t} t^{-\frac{\bar{\gamma}}{2}} dt \\
&= \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} [(\Delta^t)^{1-\frac{\bar{\gamma}}{2}} - (\Delta^{t+1})^{\frac{\bar{\gamma}}{2}}].
\end{aligned}
$$

For any $t_2 \ge t_1 \ge T$, we have

$$
\begin{aligned}
\sum_{t=t_1}^{t_2} \|x^{t+1} - x^t\| &\le \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} [(\Delta_1^t)^{1-\frac{\bar{\gamma}}{2}} - (\Delta^{t_2+1})^{\frac{\bar{\gamma}}{2}}] \\
&\le \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} (\Delta_1^t)^{1-\frac{\bar{\gamma}}{2}}.
\end{aligned}
$$

Since $\Delta^{t_1} \to 0$ as $t_1 \to \infty$, this shows that $\{x^t\}$ satisfies Cauchy's criterion for convergence. Thus $\{x^t\}$ has a unique cluster point $\bar{x}$. Using triangle inequality, we have

$$
\begin{aligned}
\|x^{t_2+1} - x^{t_1}\| &= \left\| \sum_{t=t_1}^{t_2} (x^{t+1} - x^t) \right\| \\
&\le \sum_{t=t_1}^{t_2} \|x^{t+1} - x^t\| \\
&\le \frac{C_4}{1 - \frac{\bar{\gamma}}{2}} (\Delta_1^t)^{1-\frac{\bar{\gamma}}{2}}.
\end{aligned}
$$

Let $t_2 \to \infty$, it follows from (4.1) that $\|\bar{x} - x^{t_1}\| = O((t_1^{-\omega})^{1-\frac{\bar{\gamma}}{2}}) = O(t_1^{-\frac{1-\gamma}{2\gamma}})$.

Moreover, suppose $\bar{s} - \bar{y} > 0$ and $\gamma = \frac{1}{2}$. We have $\{x^t\} \to \bar{x}$ and $\bar{y} \le 0$. Then there exist $\bar{J} \subseteq \{1, \ldots, n\}$ such that

$$
\bar{y}_{\bar{J}} < \bar{s}_{\bar{J}} = 0 \quad \text{and} \quad \bar{s}_{\bar{J}^C} > \bar{y}_{\bar{J}^C} = 0.
$$

Since $\eta^t = (S^t)^\gamma y^t$ and $(s^t, y^t) \to (\bar{s}, \bar{y})$ yield

$$
s_j^t = O(|\eta_j^t|^{\frac{1}{\gamma}}), \quad \forall j \in \bar{J}, \quad \text{and} \quad |y_j^t| = O(|\eta_j^t|), \quad \forall j \in \bar{J}^C,
$$

we have $\|(s_{\bar{J}}^t, y_{\bar{J}^C}^t)\|_1 \le \|\eta^t\|_1$ and $\|(\bar{x}^t, \bar{s}^t, \bar{y}^t) - (x^t, s^t, y^t)\| \le C_1' \|(s_{\bar{J}}^t, y_{\bar{J}^C}^t)\|_1$. Then it follows that

$$
|f(x^t) - f(\bar{x}^t)| \le C_2 (C_1')^2 \|\eta^t\|_1^2 + \sum_{j \in \bar{J}} C_1' \|\eta_j^t\|_1 |\eta_j^t|^2 + \sum_{j \in \bar{J}} |\eta_j^t|^2 \le C_3' \|\eta\|_1^2
$$

and

$$\Delta^t \leq C_3' \|\eta^t\|_1^2 \leq \kappa'(f(x^t) - f(x^{t+1})) = \kappa'(\Delta^t - \Delta^{t+1}).$$

Using this relationship, we have $\Delta^{t+1} \leq \frac{\kappa'-1}{\kappa'}\Delta^t$. It implies that $\{\Delta^t\} \to 0$ Q-linearly. In addition, $\|\bar{x} - x^{t_1}\| = O((\Delta^{t_1})^{\frac{1}{4}}) = O((1-\frac{1}{\kappa'})^{\frac{1}{4}t_1})$. This means that $\{\|\bar{x} - x^t\|\}$ converges to 0 R-linearly. □

## 5 Numerical Experiments

In this section, we test the practical behaviour of Algorithm 2.3. We test different problems containing convex quadratic programming and nonconvex programming. Our code is written in MATLAB R2012b, and all the experiments are performed on a Dell Optiplex 780 workstation with Intel Core Quad 2.83GHz CPU and 4GB of RAM.

Firstly, we consider LASSO problem that

$$\min_{\theta \in \mathbb{R}^n} \frac{1}{2}\|\tilde{A}\theta - \tilde{b}\|^2 + \lambda\|\theta\|_1,$$

where $\tilde{A} \in \mathbb{R}^{m\times n}$, $\tilde{b} \in \mathbb{R}^m$. The dual problem of LASSO is that

$$\begin{aligned} &\min_{x\in\mathbb{R}^m} \frac{1}{2}x^Tx - \frac{\tilde{b}}{\lambda}^T x, \\ &\text{s.t. } \tilde{A}^T x \leq (1,\ldots,1)^T, \\ &\quad -\tilde{A}^T x \leq (1,\ldots,1)^T, \end{aligned}$$

which has the same form of Problem (1.2). It has $m$ variables and $2n$ constraints. Moreover, it is a convex quadratic programming.

In our experiments, $\tilde{A}$ and $\tilde{b}$ are generated by synthetic method. Firstly we generate a sparse variable $x$ and a random matrix $\tilde{A}$ with gaussian distribution. Then we set $\tilde{b}$ to be $\tilde{A}x$ with a noise (1%). In our implementation, we use three different values of $\gamma$ ($\gamma = 0.6, 0.8, 1$). For each case we use five different $M$ ($M = 2n, n, 2m, 1.6m, 1.2m$). When $M = 2n$, the method is the classic Affine-Scaling Method. We use an exact-type rule

$$\alpha^t = \min\left(\frac{-(g^t)^T d^t}{\|d^t\|^2}, \alpha_0^t\right)$$

to exchange the Armijio-type rule. The new $\alpha^t$ satisfies (2.1) with $\sigma_1 = \frac{1}{2}$, and $\inf_t \alpha^t > 0$. Thus, the convergence can be guaranteed. In addition, we set $\sigma_2 = 0.9$ and take 0 to be the initial point. We terminate the method when both $\|y_+^t\|$ and $\|S^t y^t\|$ are less than a tolerance $\text{tol} = 10^{-2}$, where $(y_+^t)_j = \max\{y_j^t, 0\}$.

Table 1 reports the number of iterations (iter), cpu time (in seconds), final objective value, final $\|y_+^t\|$, and $\|S^t y^t\|$ (complementary). We see from Table 1 that it has better performance at $\gamma = 0.6, 0.8$ rather than $\gamma = 1$. When $\gamma = 0.5$, it leads to numerical error that the sequence $\{x^t\}$ goes to boundary rapidly. For each case, it saves about half of the time, when $M$ is down from $2n$ to $n$. For the most cases, $M = 2m$ yields less cpu time than $M = n$. But when $M$ reduces to a number close to $m$, the number of iterations grows rapidly. When $M = m$, the algorithm is numerically unstable and does not converge in the most cases. This states that the appropriate choice of $M$ makes the algorithm efficient.

| Problem | | | $\gamma$ | $M$ | iter | cpu | obj | $\lVert y_+^t \rVert$ | com. |
|---|---|---|---|---|---|---|---|---|---|
| $m$ | $n$ | $\lambda$ | | | | | | | |
| 500 | 2000 | 1e − 3 | 0.6 | 4000 | 61 | 8.162 | −47186.2 | 9.53e − 4 | 9.60e − 3 |
| | | | | 2000 | 52 | 4.240 | −47186.2 | 1.45e − 3 | 6.83e − 3 |
| | | | | 1000 | 47 | 2.775 | −47186.2 | 2.12e − 3 | 9.00e − 3 |
| | | | | 800 | 57 | 3.259 | −47186.2 | 1.52e − 3 | 9.19e − 3 |
| | | | | 600 | 66 | 3.499 | −47186.2 | 9.73e − 4 | 8.91e − 3 |
| | | | 0.8 | 4000 | 64 | 8.564 | −47186.2 | 2.66e − 3 | 9.98e − 3 |
| | | | | 2000 | 43 | 3.360 | −47186.2 | 2.09e − 3 | 8.53e − 3 |
| | | | | 1000 | 78 | 4.836 | −47186.2 | 1.45e − 3 | 9.81e − 3 |
| | | | | 800 | 60 | 3.329 | −47186.2 | 1.55e − 3 | 9.25e − 3 |
| | | | | 600 | 136 | 7.343 | −47186.2 | 1.34e − 3 | 9.94e − 3 |
| | | | 1 | 4000 | 86 | 10.766 | −47186.1 | 5.11e − 3 | 9.35e − 3 |
| | | | | 2000 | 86 | 6.725 | −47186.1 | 5.36e − 3 | 9.95e − 3 |
| | | | | 1000 | 130 | 7.904 | −47186.1 | 3.88e − 3 | 9.74e − 3 |
| | | | | 800 | 132 | 7.037 | −47186.1 | 4.16e − 3 | 9.89e − 3 |
| | | | | 600 | 411 | 21.102 | −47186.2 | 2.49e − 3 | 9.95e − 3 |
| 500 | 4000 | 1e − 3 | 0.6 | 8000 | 107 | 26.928 | −46993.9 | 1.10e − 4 | 3.23e − 3 |
| | | | | 4000 | 88 | 13.421 | −46993.9 | 1.50e − 4 | 7.92e − 3 |
| | | | | 1000 | 53 | 4.208 | −46993.9 | 1.40e − 4 | 3.95e − 3 |
| | | | | 800 | 47 | 3.503 | −46993.9 | 1.55e − 4 | 4.66e − 3 |
| | | | | 600 | 45 | 2.913 | −46993.9 | 0.99e − 4 | 4.88e − 3 |
| | | | 0.8 | 8000 | 52 | 12.244 | −46993.9 | 1.73e − 4 | 8.25e − 3 |
| | | | | 4000 | 45 | 6.284 | −46993.9 | 1.79e − 4 | 9.99e − 3 |
| | | | | 1000 | 36 | 2.642 | −46993.9 | 1.19e − 4 | 8.49e − 3 |
| | | | | 800 | 34 | 2.310 | −46993.9 | 0.96e − 4 | 7.28e − 3 |
| | | | | 600 | 45 | 2.739 | −46993.9 | 1.02e − 4 | 7.98e − 3 |
| | | | 1 | 8000 | 98 | 24.196 | −46993.9 | 2.23e − 4 | 9.70e − 3 |
| | | | | 4000 | 69 | 10.332 | −46993.9 | 3.12e − 4 | 9.75e − 3 |
| | | | | 1000 | 95 | 8.026 | −46993.9 | 2.18e − 4 | 9.92e − 3 |
| | | | | 800 | 116 | 9.514 | −46993.9 | 2.20e − 4 | 9.78e − 3 |
| | | | | 600 | 127 | 9.518 | −46993.9 | 2.08e − 4 | 9.07e − 3 |
| 250 | 1000 | 1e − 3 | 0.6 | 2000 | 53 | 1.628 | −24884.8 | 1.74e − 3 | 6.70e − 3 |
| | | | | 1000 | 46 | 0.966 | −24884.8 | 1.24e − 3 | 6.38e − 3 |

continued on next page

continued from previous page

| Problem $m$ | Problem $n$ | Problem $\lambda$ | $\gamma$ | $M$ | iter | cpu | obj | $\|y_+^t\|$ | com. |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 500 | 35 | 0.562 | −24884.8 | 2.52e − 3 | 8.45e − 3 |
| | | | | 400 | 41 | 0.618 | −24884.8 | 1.91e − 3 | 9.34e − 3 |
| | | | | 300 | 56 | 0.792 | −24884.8 | 2.28e − 3 | 8.93e − 3 |
| | | | 0.8 | 2000 | 38 | 1.086 | −24884.8 | 2.42e − 3 | 9.29e − 3 |
| | | | | 1000 | 32 | 0.606 | −24884.8 | 2.61e − 3 | 9.29e − 3 |
| | | | | 500 | 37 | 0.579 | −24884.8 | 2.72e − 3 | 8.56e − 3 |
| | | | | 400 | 34 | 0.506 | −24884.8 | 2.92e − 3 | 9.45e − 3 |
| | | | | 300 | 126 | 1.812 | −24884.8 | 1.98e − 3 | 9.67e − 3 |
| | | | 1 | 2000 | 61 | 1.811 | −24884.7 | 6.97e − 3 | 9.93e − 3 |
| | | | | 1000 | 63 | 1.259 | −24884.7 | 6.45e − 3 | 9.28e − 3 |
| | | | | 500 | 204 | 3.349 | −24884.8 | 3.59e − 3 | 9.94e − 3 |
| | | | | 400 | 142 | 2.126 | −24884.7 | 4.71e − 3 | 9.96e − 3 |
| | | | | 300 | 469 | 6.703 | −24884.8 | 2.58e − 3 | 9.92e − 3 |
| 250 | 2000 | 1e − 3 | 0.6 | 4000 | 79 | 4.548 | −24854.9 | 1.46e − 3 | 6.64e − 3 |
| | | | | 2000 | 67 | 2.433 | −24854.9 | 9.77e − 4 | 5.06e − 3 |
| | | | | 500 | 43 | 0.902 | −24854.9 | 6.54e − 4 | 4.81e − 3 |
| | | | | 400 | 38 | 0.754 | −24854.9 | 7.14e − 4 | 5.72e − 3 |
| | | | | 300 | 42 | 0.778 | −24854.9 | 5.90e − 4 | 9.31e − 3 |
| | | | 0.8 | 4000 | 39 | 2.048 | −24854.9 | 4.53e − 4 | 9.26e − 3 |
| | | | | 2000 | 35 | 1.160 | −24854.9 | 5.21e − 4 | 6.01e − 3 |
| | | | | 500 | 29 | 0.564 | −24854.9 | 7.28e − 4 | 9.61e − 3 |
| | | | | 400 | 39 | 0.762 | −24854.9 | 6.48e − 4 | 9.57e − 3 |
| | | | | 300 | 47 | 0.824 | −24854.9 | 4.82e − 4 | 9.17e − 3 |
| | | | 1 | 4000 | 118 | 6.733 | −24854.9 | 6.95e − 4 | 9.69e − 3 |
| | | | | 2000 | 115 | 4.173 | −24854.9 | 7.32e − 4 | 9.91e − 3 |
| | | | | 500 | 64 | 1.323 | −24854.9 | 1.00e − 3 | 8.59e − 3 |
| | | | | 400 | 52 | 1.030 | −24854.9 | 1.41e − 3 | 8.55e − 3 |
| | | | | 300 | 79 | 1.447 | −24854.9 | 1.13e − 3 | 7.29e − 3 |

Table 1 Behavior of partial first-order affine-scaling method to LASSO

We now compare the numerical results of our method with fmincon Interior Point Algorithm on nonconvex problems. The Matlab function fmincon is to solve nonlinear programming. More extensive description can be found in [5, 6, 20].

For $f$, we choose 5 test functions from the set of nonlinear least square functions used by Moré et al. [12]. Among the 5 nonconvex functions, ER, DBV, BT are with sparse Hessian, and TRIG, BAL are with dense Hassion. We use the default starting point $x^0$ given in [12]. For constraints, we generate $A$ with gaussian distribution, then the primary constraints are $A^T x \leq \mu(1,\ldots,1)^T$, where $\mu > 0$ is given. After that, we translate the center of the feasible region from the origin to $x^0$, which means $A^T(x - x^0) \leq \mu(1,\ldots,1)^T$. We terminate both methods when the first order optimality measure is below a tolerance (tol $> 0$) times $\|\nabla f(x^0)\|_\infty$. We set tol $= 1\text{e}-3$ except BAL. For BAL, we set tol $= 1\text{e}-10$. For Algorithm 2.3, we set $\gamma = 0.8$ and $M = 2m$, since such selections may bring good performance due to the experience from Table 1.

The results are shown in Table 2. We see that Algorithm 2.3 costs less time than fmincon in the most cases. Especially, for BT and BAL, Algorithm 2.3 performs much better. For DEV, Algorithm 2.3 costs more time than fmincon, but with a better objective function value.

| Problem | $m$ | $n$ | $\mu$ | Algorithm 1 | | | fmincon | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | obj | iter | cpu | obj | iter | cpu |
| ER | 250 | 2000 | 1 | $3.436\text{e}-2$ | 1457 | 14.96 | $3.601\text{e}-2$ | 81 | 66.89 |
| DEV | 250 | 2000 | $1\text{e}-4$ | $7.971\text{e}-8$ | 351 | 10.77 | $5.184\text{e}-7$ | 19 | 9.66 |
| BT | 250 | 2000 | 1 | $2.885\text{e}-1$ | 432 | 4.20 | $3.640\text{e}-1$ | 51 | 43.52 |
| TRIG | 250 | 2000 | $1\text{e}-3$ | $1.311\text{e}-4$ | 187 | 7.14 | $2.218\text{e}-4$ | 16 | 8.36 |
| BAL | 250 | 2000 | $1\text{e}-1$ | $2.947\text{e}+6$ | 30 | 0.74 | $2.947\text{e}+6$ | 27 | 21.21 |

Table 2 Comparing Algorithm 2.3 with fmincon to nonconvex problems

## 6 Conclusion

In this paper, we have proposed a partial first-order affine-scaling method for solving smoothing programming with inequality constraints. The method chooses a subset of constraints in consideration at each iteration. We have proved the global convergence and showed the local convergence rate under some assumptions. Numerical experiments show that the method can solve Problem (1.2) efficiently, and a appropriate $M$ can improve the efficiency of affine-scaling method.

## References


[1] Adler, I., Resende, G. C., Veiga, G., et al.: An implementation of karmarkar's algorithm for linear programming. *Math. Programming*, **44**(1), 297–335 (1989)

[2] Barnes, E. R.: A variation on karmarkar's algorithm for solving linear programming problems. *Math. Programming*, **36**(2), 174–182 (1986)

[3] Bellavia, S., Macconi, M., Morini, B.: An affine scaling trust-region approach to bound-constrained nonlinear systems. *Appl. Numer. Math.*, **44**(3), 257–280 (2003)

[4] Bonnans, J. F., Pola, C.: A trust region interior point algorithm for linearly constrained optimization. *SIAM J. Optimiz.*, **7**(3), 717–731 (1997)

[5] Byrd, R. H., Gilbert, J. C., Nocedal, J.: A trust region method based on interior point techniques for nonlinear programming. *Math. Programming*, **89**(1), 149–185 (2000)

[6] Byrd, R. H., Hribar, M. E., Nocedal, J.: An interior point algorithm for large-scale nonlinear programming. *SIAM J. Optimiz.*, **9**(4), 877–900 (1999)

[7] Coleman, T. F., Li, Y.: A trust region and affine scaling interior point method for nonconvex minimization with linear inequality constraints. *Math. Programming*, **88**(1), 1–31 (2000)

[8] Dantzig, G. B., Ye, Y.: A build-up interior-point method for linear programming: Affine scaling form. Tech. Rep., Department of Management Science, University of Iowa, Iowa City IA, 1991

[9] Dikin, I. I.: Iterative solution of problems of linear and quadratic programming. **8**(8), 747–748 (1967)

[10] Gonzaga, C. C., Carlos, L. A.: A primal affine-scaling algorithm for constrained convex programs. Report ES-238/90, COPPE - Federal University of Rio de Janeiro, 1990. Revised in September, 2002

[11] Hoffman, A. J.: On approximate solutions of systems of linear inequalities. *J. Research National Bureau Standards*, **49**(4), 263–265 (2015)

[12] Garbow, J. J., B. S., Hillstrom, K. E.: Algorithm 566: Fortran subroutines for testing unconstrained optimization software [c5], [e4]. *ACM Trans. Math. Software*, **7**(1), 136–140 (1981)

[13] Mangasarian, O. L.: A simple characterization of solution sets of convex programs. *Operations Research Letters*, **7**(1), 21–26 (1988)

[14] Monma, C. L., Morton, A. J.: Computational experience with a dual affine variant of karmarkar's method for linear programming. *Operations Research Letters*, **6**(6), 261–267 (1987)

[15] Monteiro, R. D. C., Wang, Y.: Trust region affine scaling algorithms for linearly constrained convex and concave programs. *Math. Programming*, **80**(3), 283–313 (1998)

[16] Tits, A., Absil, P. A., Woessner, W. P.: Constraint reduction for linear programs with many inequality constraints. *SIAM J. Optimiz.*, **17**(1), 119–146 (2006)

[17] Tseng, P.: Partial affine-scaling for linearly constrained minimization. *Math. Operations Research*, **20**(3), 678–694 (1995)

[18] Tseng, P., Bomze, I. M., Schachinger, W.: A first-order interior-point method for linearly constrained smooth optimization. *Math. Programming*, **127**(2), 399–424 (2011)

[19] Vanderbei, R. J., Meketon, M. S., Freedman, B. A.: A modification of karmarkar's linear programming algorithm. *Algorithmica*, **1**(1), 395–407 (1986)

[20] Waltz, R. A., Morales, J. L., Nocedal, J., et al.: An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Math. Programming*, **107**(3), 391–408 (2006)

[21] Wang, X., Yuan, Y. X.: A trust region method based on a new affine scaling technique for simple bounded optimization. *Optimization Methods & Software*, **28**(4), 1–18 (2013)

[22] Winternitz, L. M. B.: Primal-dual interior-point algorithms for linear programs with many inequality constraints. Dissertations & Theses-Gradworks, University of Maryland, College Park, Md., 2010