Generalization of the primal and dual affine scaling algorithms

F. G. M. CUNHA[†], A. W. M. PINTO[‡], P. R. OLIVEIRA[*§] and J. X. da C. NETO[¶]

[†]*Department of Mathematics, Ceará Federal Centre of Technological Education, Fortaleza, Brazil;* [‡]*Department of Mathematics, Federal University of Amazonas, Manaus, Brazil;* [§]*Department of Systems and Computing Engineering, Rio de Janeiro Federal University, Rio de Janeiro, Brazil;* [¶]*Department of Mathematics, Federal University of Piauí, Teresina, Brazil;*

(v1.0 finished 23 April 2005)

We obtain a new class of primal affine scaling algorithms for the linearly constrained convex programming. It is constructed through a family of metrics generated by $-r$ power, $r \geq 1$, of the diagonal iterate vector matrix. We prove the so-called weak convergence. It generalizes some known algorithms. Working in dual space, we generalize the dual affine scaling algorithm of Adler, Karmarkar, Resende and Veiga, similarly depending on a r-parameter and we give its global convergence proof for nondegenerate linear programs. With the purpose of observing the computational performance of the methods, we compare them with classical algorithms (when $r = 1$ or $r = 2$), implementing the proposed families and applying to some linear programs obtained from NETLIB library. In the case of primal family, we also apply it to some quadratic programming problems described in the Maros and Mészáros repository.

Keywords: Interior Point Algorithms; Affine Scaling Algorithms; Linear Convex Programming

1 Introduction

We are interested in *affine scaling potential algorithms.* If we restrict ourselves to primal interior point methods which is one class we are interested in, we mention the affine scaling algorithm for linear and quadratic cost functions [1] and the polynomial logarithmic barrier approach for convex self-concordant functions, as can be seen in [2]. As our algorithm can be easily particularized to the case without linear constraints, i.e., the minimization over the positive orthant, we mention the multiplicative method [3], analysed in [4].

[*]Corresponding author. Tel.: +55 21 2562-8675; Fax: +55 21 2562-8676; E-mail: poliveir@cos.ufrj.br

[†]E-mail: gevane@cos.ufrj.br; [‡]E-mail: awagner@fua.br; [¶]E-mail: jxavier@ufpi.br

In our approach, we let the positive orthant $\mathbb{R}^n_{++}$ as a Riemannian manifold. The knowledge of Riemannian geometry concepts is not necessary to understand this paper, but those interested can see: [5] for Riemannian geometry and [6,7] for that regarding specific connection with optimization. The connection between geometry and mathematical programming comes from [8]. Since then, some researches have been carried out (see, e.g. [6,7,9–16]). Most of the papers consider the so-called geodesic algorithms, where the usual straight line search of the cost function is substituted by the minimization along the geodesics which has, in general, a high computing cost. Our method follows the usual pattern — one line search is performed along the given direction. Following the same lines as this paper, we have proposed classes of algorithms for the nonlinear complementarity problem in [17] and for the minimization over the positive orthant within a proximal method set in [18].

The directions of our class, corresponding to $r = 1$ and $r = 2$ are, respectively, the classic multiplicative and affine scaling directions. The convergence result we got is the so-called *weak-convergence*, as obtained in [4] for the multiplicative method. Finally, let us mention the power affine scaling method [19] for linear programming, whose convergence results are also valid for the degenerate case.

Working in dual space we propose for linear programs, a generalization of the dual affine scaling algorithm of Adler, Karmarkar, Resende and Veiga [20], similarly depending on a r-parameter. This generalization is based on metrics defined on dual space. We give the global convergence proof of the algorithm for nondegenerate problems.

The paper is organized as follows. In section 2, we present the generalized primal affine scaling algorithm giving some background theory and its convergence. In section 3, we present the generalized dual affine scaling algorithm and give its global convergence proof for nondegenerate linear programs. Section 4 is dedicated to observing the computational performance of the methods and comparing them with classical algorithms (when $r = 1$ or $r = 2$). The paper ends reaching some conclusions and suggesting possible future researches.

2 Generalized Primal Affine Scaling Method

2.1 *Introduction*

We consider the resolution of

$$
\text{(P)} \quad \begin{array}{ll} \text{minimize}_{\boldsymbol{x}} & f(\boldsymbol{x}) \\ \text{subject to} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq 0, \end{array} \tag{1}
$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable and convex function, $\mathbf{A}$ is a $m \times n$ matrix and $\boldsymbol{b}$ a $\mathbb{R}^m$ vector.

Associated with $\mathbb{R}^n_{++}$, we define a metric through the diagonal matrix $\mathbf{G}(\boldsymbol{x}) = \mathbf{X}^{-r}$, for $r \geq 1$, $\mathbf{X}$ being the diagonal matrix whose nonnull elements are the $\boldsymbol{x} \in \mathbb{R}^n_{++}$ entries. We take $M = \left\{\boldsymbol{x} \in \mathbb{R}^n_{++} :\ \mathbf{A}\boldsymbol{x} = \boldsymbol{b}\right\}$ as a submanifold of $\left(\mathbb{R}^n_{++}, \mathbf{X}^{-r}\right)$, $r \geq 1$, with the induced metric.

As we show below, considering f defined on M, the opposite of its gradient is the affine scaling direction corresponding to each $r \geq 1$. We limit ourselves to presenting the theoretical results we need to construct the algorithm direction. Those interested in the detailed description of the concepts presented here can see [7].

We will now obtain the gradient of f restricted to M.

The first element we need is the tangent space to M, the usual null space of $\mathbf{A}$, given by

$$T_{\boldsymbol{x}}M = TM = \{\boldsymbol{d} \in \mathbb{R}^n :\ \mathbf{A}\boldsymbol{d} = 0\}\,, \tag{2}$$

for any $\boldsymbol{x}$ (it would be $\boldsymbol{x}$ dependent for nonlinear constraints).

In order, we need the expression of the projection operator onto TM. Denoting $\mathbf{I}$ as the identity $n \times n$ matrix, the projection onto TM, under any (positive definite matrix) metric $\mathbf{G}$, is

$$P_{\mathbf{G}}(\boldsymbol{x}) = \mathbf{I} - \mathbf{G}^{-1}(\boldsymbol{x})\mathbf{A}^T\left(\mathbf{A}\mathbf{G}^{-1}(\boldsymbol{x})\mathbf{A}^T\right)^{-1}\mathbf{A}. \tag{3}$$

In our case, with $\mathbf{G} = \mathbf{X}^{-r}$, this rewrites

$$P(\boldsymbol{x}) = \mathbf{I} - \mathbf{X}^r\mathbf{A}^T\left(\mathbf{A}\mathbf{X}^r\mathbf{A}^T\right)^{-1}\mathbf{A}. \tag{4}$$

Now, notice that the gradient of f restricted to $\left(\mathbb{R}^n_{++}, \mathbf{X}^{-r}\right)$ is

$$\nabla_{\mathbb{R}^n_{++}} f(\boldsymbol{x}) = \mathbf{G}^{-1}\nabla f(\boldsymbol{x}) = \mathbf{X}^r\nabla f(\boldsymbol{x}), \tag{5}$$

for any $\boldsymbol{x} \in \mathbb{R}^n_{++}$, where ∇ represents the usual (Euclidean) gradient.

Therefore, the gradient of f restricted to $(M, \mathbf{X}^{-r})$ is the projection of $\nabla_{\mathbb{R}^n_{++}} f(\boldsymbol{x})$ onto TM, i.e.,

$$\nabla_M f(\boldsymbol{x}) = P(\boldsymbol{x})\mathbf{X}^r\nabla f(\boldsymbol{x}).$$

It is easy to check that $\nabla_M f(\boldsymbol{x}) \in TM$.

For all $\boldsymbol{d} \in TM$ such that $\|\boldsymbol{d}\|_{\mathbf{G}} := \sqrt{\langle \boldsymbol{d}, \mathbf{G}\boldsymbol{d}\rangle} = 1$, using Cauchy–Schwartz inequality, we obtain

$$|\langle \boldsymbol{d}, \nabla_M f(\boldsymbol{x})\rangle_{\mathbf{G}}| \leq \|\boldsymbol{d}\|_{\mathbf{G}} \|\nabla_M f(\boldsymbol{x})\|_{\mathbf{G}} = \|\nabla_M f(\boldsymbol{x})\|_{\mathbf{G}}.$$

Consequently, $-\nabla_M f(\boldsymbol{x})/\|\nabla_M f(\boldsymbol{x})\|_{\mathbf{G}}$ is the steepest descent direction of f at $\boldsymbol{x}$.

Since f is convex, $\boldsymbol{x}^*$ is a solution for (1) if and only if there exist Lagrangian multipliers $\boldsymbol{y}^* \in \mathbb{R}^m$ and $\boldsymbol{s}^* \in \mathbb{R}^n_+$ such that

$$\begin{aligned} \mathbf{A}^T \boldsymbol{y}^* + \boldsymbol{s}^* &= \nabla f(\boldsymbol{x}^*) \\ \mathbf{A}\boldsymbol{x}^* &= \boldsymbol{b} \\ \boldsymbol{x}_i^* \boldsymbol{s}_i^* &= 0, \quad i = 1, ..., n \\ \boldsymbol{x}^* &\geq 0. \end{aligned} \tag{6}$$

From (6), some simple calculations lead to

$$\begin{aligned} &\boldsymbol{y}(\boldsymbol{x}^*) = \left(\mathbf{A}\mathbf{X}_*^r \mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{X}_*^r \nabla f(\boldsymbol{x}^*) \\ &\boldsymbol{s}(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*) - \mathbf{A}^T \boldsymbol{y}(\boldsymbol{x}^*) \\ &\mathbf{X}_*^r \boldsymbol{s}^* = 0. \end{aligned} \tag{7}$$

Therefore, it is natural to define the following functions on M:

$$\boldsymbol{y}(\boldsymbol{x}) \equiv \left(\mathbf{A}\mathbf{X}^r \mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{X}^r \nabla f(\boldsymbol{x}), \tag{8}$$

$$\boldsymbol{s}(\boldsymbol{x}) \equiv \nabla f(\boldsymbol{x}) - \mathbf{A}^T \boldsymbol{y}(\boldsymbol{x}) = P^T(\boldsymbol{x}) \nabla f(\boldsymbol{x}), \tag{9}$$

$$\boldsymbol{d}(\boldsymbol{x}) \equiv \mathbf{X}^r \boldsymbol{s}(\boldsymbol{x}) = P(\boldsymbol{x}) \mathbf{X}^r \nabla f(\boldsymbol{x}) = \nabla_M f(\boldsymbol{x}). \tag{10}$$

Under the nondegeneracy assumption, all above functions can be extended to the closure of M.

Now, we apply the gradient direction computed to define the generalized primal affine scaling algorithm (GPAS).

Algorithm GPAS:
initialization: Let $(\mathbf{A}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x}^0 \in M,$ *stopping criterion*, $\beta > 0$ and $\delta \in (0, 1))$
$k = 0$

while *stopping criterion* not satisfied

$$
\begin{aligned}
\boldsymbol{y}^k &= \left(\mathbf{A}\mathbf{X}_k^r\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{X}_k^r\nabla f(\boldsymbol{x}^k)\\
\boldsymbol{s}^k &= \nabla f(\boldsymbol{x}^k) - \mathbf{A}^T\boldsymbol{y}^k\\
\boldsymbol{d}^k &= \mathbf{X}_k^r\boldsymbol{s}^k\\
\mu^k &= \chi\left(\mathbf{X}_k^{-1}\boldsymbol{d}^k, 0\right)\\
\alpha^k &= \tfrac{\delta}{\beta+\mu^k}\\
t^k &= \arg\min_{t\in[0,\alpha^k]} f(\boldsymbol{x}^k - t\boldsymbol{d}^k)\\
\boldsymbol{x}^{k+1} &= \boldsymbol{x}^k - t^k\boldsymbol{d}^k\\
k &= k+1
\end{aligned}
$$

end while
end Algorithm GPAS
where $\chi(\boldsymbol{v})$ denotes the largest component of $\boldsymbol{v}$.

An immediate consequence is the fact that the algorithm is an interior point method.
Indeed, for $\boldsymbol{x}^k \in M$, $k \geq 0$, it holds that

$$
\left(\mathbf{X}_k^{-1}\boldsymbol{x}^{k+1}\right)_i = 1 - t^k(\boldsymbol{x}_i^k)^{-1}\boldsymbol{d}_i^k \geq 1, \quad \text{if} \;\; \boldsymbol{d}_i^k \leq 0,
$$

otherwise, we have

$$
t^k(\boldsymbol{x}_i^k)^{-1}\boldsymbol{d}_i^k \leq \frac{\delta\mu^k}{\beta+\mu^k} < 1.
$$

Therefore, $1-t^k(\boldsymbol{x}_i^k)^{-1}\boldsymbol{d}_i^k > 0$. Since $\boldsymbol{x}^0 > 0$, the assertion follows by induction.
To assure a large search interval $[0, \alpha^k]$, the *safety factors* δ must be near 1, and β near 0.

2.2 *Convergence of the Generalized Primal Affine Scaling Algorithm*

In this subsection we will study the convergence of the generalized primal affine scaling algorithm with the following assumptions:
Assumption 1: The matrix $\mathbf{A}$ has full rank m;
Assumption 2: The primal problem (P) has an interior feasible solution, i.e.,

$$
P^{++} = \{\boldsymbol{x} \in \mathbb{R}^n : \; \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \; \boldsymbol{x} > 0\} \neq \emptyset;
$$

Assumption 3: (P) is a nondegenerate problem;
Assumption 4: The starting level set, $L_{\boldsymbol{x}^0} = \left\{\boldsymbol{x} \in \overline{M} : f(\boldsymbol{x}) \leq f(\boldsymbol{x}^0)\right\}$, is

bounded, where $\overline{M}$ is the closure of M.

The following lemma shows some useful relations.

Lemma 2.1 *Let $\boldsymbol{x} \in M$ and $\boldsymbol{d} \in TM$ be given. Then, the following statements hold:*
(a) $\nabla f(\boldsymbol{x})^T \boldsymbol{d} = \boldsymbol{s}(\boldsymbol{x})^T \boldsymbol{d}$;
(b) $\nabla f(\boldsymbol{x})^T \boldsymbol{d}(\boldsymbol{x}) = \|\mathbf{X}^{r/2} \boldsymbol{s}(\boldsymbol{x})\|^2 = \|\mathbf{X}^{-r/2} \boldsymbol{d}(\boldsymbol{x})\|^2$,
where $\boldsymbol{s}(\boldsymbol{x})$ and $\boldsymbol{d}(\boldsymbol{x})$ are given, respectively, by (9) and (10).

Proof Using (9), we obtain

$$\begin{aligned} \boldsymbol{s}(\boldsymbol{x})^T \boldsymbol{d} &= \nabla f(\boldsymbol{x})^T \boldsymbol{d} - \left(\mathbf{A}^T \boldsymbol{y}(\boldsymbol{x})\right)^T \boldsymbol{d} \\ &= \nabla f(\boldsymbol{x})^T \boldsymbol{d} - \boldsymbol{y}(\boldsymbol{x})^T \mathbf{A} \boldsymbol{d} \\ &= \nabla f(\boldsymbol{x})^T \boldsymbol{d}, \end{aligned}$$

where the last equality follows from the fact that $\boldsymbol{d} \in TM$ implies $\mathbf{A}\boldsymbol{d} = 0$. So, part (a) follows.

We now prove part (b). Using part (a) above and (10), we have

$$\begin{aligned} \nabla f(\boldsymbol{x})^T \boldsymbol{d}(\boldsymbol{x}) &= \boldsymbol{s}(\boldsymbol{x})^T \boldsymbol{d}(\boldsymbol{x}) \\ &= \boldsymbol{s}(\boldsymbol{x})^T \mathbf{X}^r \boldsymbol{s}(\boldsymbol{x}) = \|\mathbf{X}^{r/2} \boldsymbol{s}(\boldsymbol{x})\|^2, \end{aligned}$$

which is the first equality of part (b). The second equality easily follows again from (10), by substituting $\boldsymbol{s}(\boldsymbol{x}) = \mathbf{X}^{-r} \boldsymbol{d}(\boldsymbol{x})$. □

Proposition 2.2 *The following statements hold for the primal affine scaling algorithm:*
(a) *The sequences $\{\boldsymbol{x}^k\}$ and $\{\boldsymbol{s}^k\}$ are bounded;*
(b) *For each $r \geq 1$, $\mathbf{X}_k \boldsymbol{s}^k \to 0$.*

Proof The boundedness of $\{\boldsymbol{x}^k\}$ follows as a consequence of the Assumption 4 and from the fact that $\boldsymbol{x}^k \in L_{\boldsymbol{x}^0}$. Due to Assumption 3, $\boldsymbol{s}(\boldsymbol{x})$ is continuous on the closure of M. Therefore, $\{\boldsymbol{s}^k\}$ is also bounded. This proves part (a).

To prove part (b), first observe that since the sequences $\{\boldsymbol{x}^k\}$ and $\{\boldsymbol{s}^k\}$ are bounded, $\mathbf{X}^k \boldsymbol{s}^k \to 0$ if and only if $\mathbf{X}_k^{r/2} \boldsymbol{s}^k \to 0$ for all $r \geq 0$. Assume, on the contrary, that $\lim_{k\to\infty} \mathbf{X}_k^{r/2} \boldsymbol{s}^k \neq 0$. In this case, there exists some subsequence $\{\boldsymbol{x}^{k_j}\}$ and $\varepsilon > 0$ such that $\|\mathbf{X}_{k_j}^{r/2} \boldsymbol{s}^{k_j}\| \geq \varepsilon$. On the other hand, the boundedness of $\{\boldsymbol{x}^k\}$ allows, without loss of generality, to admit that $\boldsymbol{x}^{k_j} \to \overline{\boldsymbol{x}}$, for some $\overline{\boldsymbol{x}} \in M$. Additionally, the nondegeneracy assumption leads to $\boldsymbol{s}^{k_j} \to \boldsymbol{s}(\overline{\boldsymbol{x}}) = \overline{\boldsymbol{s}}$. At this point, we set

$$l = \sup_{\boldsymbol{x} \in L_{\boldsymbol{x}^0}} \left\{ \chi\left(\mathbf{X}^{-1} \boldsymbol{d}(\boldsymbol{x}), 0\right) \right\} = \sup_{\boldsymbol{x} \in L_{\boldsymbol{x}^0}} \left\{ \chi\left(\mathbf{X}^{r-1} \boldsymbol{s}(\boldsymbol{x}), 0\right) \right\}. \tag{11}$$

Clearly, l is well defined and finite, when $r \geq 1$. Then, the sequence $\{\alpha^k\}$ of the algorithm verifies

$$\frac{\delta}{\beta} \geq \alpha^k = \frac{\delta}{\beta + \mu^k} \geq \frac{\delta}{\beta + l}.$$

Thus, using the definition of the sequence of iterates, we get

$$f(\boldsymbol{x}^{k_j+1}) \leq f(\boldsymbol{x}^{k_j} - \alpha \boldsymbol{d}^{k_j}), \quad \text{for all } \alpha \in \left[0, \frac{\delta}{\beta + l}\right],$$

that is, through the first order Taylor expansion of f,

$$f(\boldsymbol{x}^{k_j+1}) - f(\boldsymbol{x}^{k_j}) \leq -\alpha \nabla f(\boldsymbol{x}^{k_j})^T \boldsymbol{d}^{k_j} + R(\boldsymbol{x}^{k_j}, \alpha \boldsymbol{d}^{k_j}). \tag{12}$$

Now, we utilize the boundedness of $L_{\boldsymbol{x}^0}$ and the monotonicity of $\{f(\boldsymbol{x}^k)\}$, to obtain $f(\boldsymbol{x}^{k_j+1}) - f(\boldsymbol{x}^{k_j}) \to 0$.
Due to part (b) of Lemma 2.1, $\nabla f(\boldsymbol{x}^{k_j})^T \boldsymbol{d}^{k_j} \to \|\overline{\mathbf{X}}^{r/2} \boldsymbol{s}(\overline{\boldsymbol{x}})\|^2 \geq \varepsilon$.
Taking the limit in (12), we obtain

$$0 \leq -\alpha \|\overline{\mathbf{X}}^{r/2} \boldsymbol{s}(\overline{\boldsymbol{x}})\|^2 + R(\overline{\boldsymbol{x}}, \alpha \overline{\boldsymbol{d}}), \text{ for all } \alpha \in \left[0, \frac{\delta}{\beta + l}\right],$$

where $\overline{\boldsymbol{d}}$ is the corresponding limit to $\{\boldsymbol{d}^{k_j}\}$.
Consequently,

$$\frac{R(\overline{\boldsymbol{x}}, \alpha \overline{\boldsymbol{d}})}{\alpha} \geq \|\overline{\mathbf{X}}^{r/2} \boldsymbol{s}(\overline{\boldsymbol{x}})\|^2 \geq \varepsilon,$$

which contradicts the differentiability of f at $\overline{\boldsymbol{x}}$. This proves the assertion. □

Corollary 2.3 *The direction generated by the algorithm verify* $\lim_{k\to\infty} \boldsymbol{d}^k = 0$.

Proof This is a direct consequence of the Proposition 2.2 and the application of equality (10), i.e., $\boldsymbol{d}^k = \mathbf{X}_k^r \boldsymbol{s}^k$. □

Corollary 2.4 *The iterates verify* $\lim_{k\to\infty}(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k) = 0$.

Proof From the algorithmic procedure, we have $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| = t^k \|\boldsymbol{d}^k\| \leq \alpha^k \|\boldsymbol{d}^k\| \leq (\delta/\beta)\|\boldsymbol{d}^k\|$, that goes to zero due to the last corollary. □

Proposition 2.2 shows that any accumulation point $\overline{\boldsymbol{x}}$ of the iteration sequence $\{\boldsymbol{x}^k\}$ verifies the complementarity condition $\overline{\mathbf{X}}\boldsymbol{s}(\overline{\boldsymbol{x}}) = 0$.

We will show that $\overline{\boldsymbol{s}} = \boldsymbol{s}(\overline{\boldsymbol{x}}) \geq 0$, so proving, by convexity of f, that any limit point of $\{\boldsymbol{x}^k\}$ is a solution of (1).

In that direction, fix a certain accumulation point $\overline{\boldsymbol{x}}$, set $\overline{\boldsymbol{s}} = \boldsymbol{s}(\overline{\boldsymbol{x}})$, and consider a partition $\{B, N\}$ of the set $\{1, 2, ..., n\}$, given by $N := \{i : \overline{\boldsymbol{s}}_i \neq 0\}$. In addition, if there exists another limit point $\widetilde{\boldsymbol{x}}$, then, since $\{f(\boldsymbol{x}^k)\}$ is a convergent sequence, $f(\overline{\boldsymbol{x}}) = f(\widetilde{\boldsymbol{x}})$.

Those remarks will induce us to define the following set:

$$\Omega := \left\{ \boldsymbol{x} \in \overline{M} : f(\boldsymbol{x}) = f(\overline{\boldsymbol{x}}) \text{ and } \boldsymbol{x}_N = 0 \right\}. \tag{13}$$

The first property of Ω is its compactness, as a consequence of the boundedness of $L_{\boldsymbol{x}^0}$. A second and obvious fact is that the accumulation point $\boldsymbol{x}$ such that $\boldsymbol{x}_N = 0$ is necessarily in Ω.

We dilate Ω, through the following definition:

$$\Omega_\delta := \left\{ \boldsymbol{x} \in \overline{M} : \|\boldsymbol{x} - \boldsymbol{y}\| \leq \delta, \text{ for some } \boldsymbol{y} \in \Omega \right\}. \tag{14}$$

A simple but essential property relating those sets is, that if $\boldsymbol{x} \in \overline{M}$ and $\boldsymbol{x} \notin \Omega_\delta$, then the point-set distance, $\boldsymbol{d}(\boldsymbol{x}, \Omega)$, satisfies $\boldsymbol{d}(\boldsymbol{x}, \Omega) > \delta$.

LEMMA 2.5 *Ω, given by (13), is a convex set.*

Proof Let $\boldsymbol{u}$, $\boldsymbol{v} \in \Omega$ and $t \in [0, 1]$ be given. Then,

$$(t\boldsymbol{u} + (1-t)\boldsymbol{v})_N = t\boldsymbol{u}_N + (1-t)\boldsymbol{v}_N = 0. \tag{15}$$

By the convexity of f, we have

$$\begin{aligned} f(t\boldsymbol{u} + (1-t)\boldsymbol{v}) &\leq tf(\boldsymbol{u}) + (1-t)f(\boldsymbol{v}) \\ &= tf(\overline{\boldsymbol{x}}) + (1-t)f(\overline{\boldsymbol{x}}) = f(\overline{\boldsymbol{x}}). \end{aligned} \tag{16}$$

On the other hand, observe that $t\boldsymbol{u} + (1-t)\boldsymbol{v} - \overline{\boldsymbol{x}} \in TM$. Then, using Lemma 2.1, we obtain

$$\begin{aligned} \nabla f(\overline{\boldsymbol{x}})^T(t\boldsymbol{u} + (1-t)\boldsymbol{v} - \overline{\boldsymbol{x}}) &= \boldsymbol{s}(\overline{\boldsymbol{x}})^T(t\boldsymbol{u} + (1-t)\boldsymbol{v} - \overline{\boldsymbol{x}}) \\ &= \overline{\boldsymbol{s}}_N^T(t\boldsymbol{u} + (1-t)\boldsymbol{v} - \overline{\boldsymbol{x}})_N + \overline{\boldsymbol{s}}_B^T(t\boldsymbol{u} + (1-t)\boldsymbol{v} - \overline{\boldsymbol{x}})_B = 0, \end{aligned}$$

where the last equality follows due to $\overline{\boldsymbol{s}}_B = 0$ and $(t\boldsymbol{u} + (1-t)\boldsymbol{v} - \overline{\boldsymbol{x}})_N = 0$. By making use of the convexity of f again, we have

$$f(t\boldsymbol{u} + (1-t)\boldsymbol{v}) \geq f(\overline{\boldsymbol{x}}) + \nabla f(\overline{\boldsymbol{x}})^T(t\boldsymbol{u} + (1-t)\boldsymbol{v} - \overline{\boldsymbol{x}}) = f(\overline{\boldsymbol{x}}). \tag{17}$$

It follows, from (16) and (17), that $f(t\boldsymbol{u} + (1-t)\boldsymbol{v}) = f(\overline{\boldsymbol{x}})$. So, Ω is convex. $\square$

The following is a consequence of a result due to Mangasarian.

LEMMA 2.6 *Let f be a differentiable convex function on $\mathbb{R}^n$ and let $\Omega \subseteq \mathbb{R}^n$ be a convex set. Suppose that f is constant on Ω. Then, ∇f is also constant on Ω.*

Proof Since f is constant on Ω, f is constant on

$$Z = \left\{ \boldsymbol{z} : \boldsymbol{z} = \arg\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}) \right\} = \Omega.$$

Now, due to Lemma 1a of Mangasarian [22] for a differentiable function f, ∇f is also constant in Ω. $\square$

LEMMA 2.7 *Let $\boldsymbol{x} \in \Omega$ be given. Then, $\boldsymbol{s}(\boldsymbol{x}) = \overline{\boldsymbol{s}} = \boldsymbol{s}(\overline{\boldsymbol{x}})$.*

Proof Using (9) and the fact that in Ω, $\nabla f(\boldsymbol{x}) = \nabla f(\overline{\boldsymbol{x}})$, we obtain

$$\begin{aligned}
\boldsymbol{s}(\boldsymbol{x}) &= \left(\mathbf{I} - \mathbf{A}^T \left(\mathbf{A}\mathbf{X}^r\mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{X}^r\right) \nabla f(\boldsymbol{x}) \\
&= \left(\mathbf{I} - \mathbf{A}^T \left(\mathbf{A}\mathbf{X}^r\mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{X}^r\right) \nabla f(\overline{\boldsymbol{x}}) \\
&= \left(\mathbf{I} - \mathbf{A}^T \left(\mathbf{A}\mathbf{X}^r\mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{X}^r\right) \left(\overline{\boldsymbol{s}} + \mathbf{A}^T\overline{\boldsymbol{y}}\right) \\
&= \overline{\boldsymbol{s}} + \mathbf{A}^T\overline{\boldsymbol{y}} - \mathbf{A}^T \left(\mathbf{A}\mathbf{X}^r\mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{X}^r\overline{\boldsymbol{s}} - \mathbf{A}^T \left(\mathbf{A}\mathbf{X}^r\mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{X}^r\mathbf{A}^T\overline{\boldsymbol{y}} \\
&= \overline{\boldsymbol{s}},
\end{aligned}$$

where the last equality follows due to $\mathbf{X}^r\overline{\boldsymbol{s}} = 0$, which is a consequence of $\overline{\boldsymbol{s}}_B = 0$ and from the fact that $\boldsymbol{x} \in \Omega$ implies $\boldsymbol{x}_N = 0$. $\square$

The next proposition will be used to prove that $\overline{\boldsymbol{s}}_N > 0$. We need the parameter $\theta := \frac{1}{2}\min_{i \in N} |\overline{\boldsymbol{s}}_i|$.

PROPOSITION 2.8 *There exists $\delta > 0$ such that, for all $\boldsymbol{x} \in \Omega_\delta$, the following statements hold:*
(a) $\|\boldsymbol{s}(\boldsymbol{x}) - \overline{\boldsymbol{s}}\| \le \theta$*;*
(b) *If $\boldsymbol{x}_i > 0$, then $\boldsymbol{s}_i(\boldsymbol{x})\overline{\boldsymbol{s}}_i > 0$, and $\boldsymbol{d}_i(\boldsymbol{x})\overline{\boldsymbol{s}}_i > 0$, for all $i \in N$.*

Proof Due to Assumption 3, $\boldsymbol{s}(\boldsymbol{x})$ is uniformly continuous on the compact Ω_1, so implying the existence of some positive δ such that, for $\boldsymbol{x}$ and $\boldsymbol{y}$ in Ω_1 verifying $\|\boldsymbol{x} - \boldsymbol{y}\| \le \delta$, it holds that $\|\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{s}(\boldsymbol{y})\| \le \theta$. Thus, if $\boldsymbol{x} \in \Omega_\delta$, there exists some $\boldsymbol{y} \in \Omega$ that satisfies $\|\boldsymbol{x} - \boldsymbol{y}\| \le \delta$ and, consequently, $\|\boldsymbol{s}(\boldsymbol{x}) - \boldsymbol{s}(\boldsymbol{y})\| \le \theta$. However, $\boldsymbol{s}$ is constant on Ω, i.e., $\boldsymbol{s}(\boldsymbol{y}) = \boldsymbol{s}(\overline{\boldsymbol{x}}) = \overline{\boldsymbol{s}}$. Then, $\|\boldsymbol{s}(\boldsymbol{x}) - \overline{\boldsymbol{s}}\| \le 1/2 \min_{i \in N} |\overline{\boldsymbol{s}}_i|$. This proves part (a).

To prove part (b), observe that, if $i \in N$, the previous result allows us to write

$$|\boldsymbol{s}_i(\boldsymbol{x}) - \overline{\boldsymbol{s}}_i| \leq \|\boldsymbol{s}(\boldsymbol{x}) - \overline{\boldsymbol{s}}\| \leq \frac{1}{2} \min_{j \in N} |\overline{\boldsymbol{s}}_j| \leq \frac{1}{2} |\overline{\boldsymbol{s}}_i|.$$

This meaning that $\boldsymbol{s}_i(\boldsymbol{x})\overline{\boldsymbol{s}}_i > 0$, which is the first result to be achieved. Now, from (10), $\boldsymbol{d}_i(\boldsymbol{x}) = \boldsymbol{x}_i^r \boldsymbol{s}_i(\boldsymbol{x})$, so $\boldsymbol{d}_i(\boldsymbol{x})\overline{\boldsymbol{s}}_i > 0$, which finishes the proof. □

In order to obtain the positivity of $\overline{\boldsymbol{s}}_N$, we will show that the accumulation points of the sequence that are not in Ω, are outside Ω_δ too. After that, we will see that, if some entry of $\overline{\boldsymbol{s}}_N$ is negative, the sequence will possess a limit point outside Ω_δ.

Lemma 2.9 *Let $\widetilde{\boldsymbol{x}}$ a limit point of $\{\boldsymbol{x}^k\}$. Then, $\widetilde{\boldsymbol{x}} \in \Omega$ or $\widetilde{\boldsymbol{x}} \notin \Omega_\delta$.*

Proof Assume that $\widetilde{\boldsymbol{x}} \in \Omega_\delta$, but $\widetilde{\boldsymbol{x}} \notin \Omega$ and set $\widetilde{\boldsymbol{s}} = \boldsymbol{s}(\widetilde{\boldsymbol{x}})$. The convergence of $\{f(\boldsymbol{x}^k)\}$ leads to $f(\widetilde{\boldsymbol{x}}) = f(\overline{\boldsymbol{x}})$, where $\overline{\boldsymbol{x}}$ is the accumulation point we have fixed. Then, since $\widetilde{\boldsymbol{x}} \notin \Omega$, there exists some $i \in N$ such that $\widetilde{\boldsymbol{x}}_i > 0$. Hence, by Proposition 2.8, $\widetilde{\boldsymbol{s}}_i\overline{\boldsymbol{s}}_i > 0$, and then $\widetilde{\boldsymbol{s}}_i \neq 0$. Consequently, we have $\widetilde{\boldsymbol{x}}_i\widetilde{\boldsymbol{s}}_i \neq 0$, which contradicts the complementarity condition. □

Lemma 2.10 *Suppose that $\overline{\boldsymbol{s}}_N = \boldsymbol{s}_N(\overline{\boldsymbol{x}})$ has a negative component. Then, $\{\boldsymbol{x}^k\}$ has an accumulation point $\widetilde{\boldsymbol{x}}$ with $\widetilde{\boldsymbol{x}} \notin \Omega_\delta$.*

Proof By Lemma 2.9, if an accumulation point is in Ω_δ, it is perforce in Ω, which is contained in the relative interior of Ω_δ. In this case, the whole sequence would be contained in Ω_δ, from some iteration k_0 on. Assume that there exists $i \in N$ such that $\overline{\boldsymbol{s}}_i < 0$ and all accumulation points of $\{\boldsymbol{x}^k\}$ are in Ω_δ, thereby in Ω. Then, for $k \geq k_0$, $\boldsymbol{x}^k \in \Omega_\delta$ and, due to Proposition 2.8, $\boldsymbol{d}_i^k < 0$. In this case, it holds that

$$\boldsymbol{x}_i^{k+1} = \boldsymbol{x}_i^k - t^k \boldsymbol{d}_i^k \geq \boldsymbol{x}_i^k \geq \boldsymbol{x}_i^{k_0} > 0,$$

from which we get the nondecreasing property of $\{\boldsymbol{x}_i^k\}$. However, this contradicts $\liminf \boldsymbol{x}_i^k = 0$. □

Finally, we can show the positivity of $\overline{\boldsymbol{s}}_N$.

Lemma 2.11 *Let $\overline{\boldsymbol{x}}$ be the above fixed accumulation point of the algorithm iterations. It holds that $\overline{\boldsymbol{s}}_N > 0$, where $\overline{\boldsymbol{s}} = \boldsymbol{s}(\overline{\boldsymbol{x}})$.*

Proof Assume that there exists some $i \in N$ such that $\overline{\boldsymbol{s}}_i < 0$. From Lemma 2.10 we know that, if there exists some $i \in N$ such that $\overline{\boldsymbol{s}}_i < 0$, the sequence $\{\boldsymbol{x}^k\}$ has an accumulation point outside Ω_δ. On the other hand, by definition of Ω, $\overline{\boldsymbol{x}}$ is an accumulation point in Ω, then in Ω_δ. So, using the closure property of Ω_δ, we conclude immediately that there exists a subsequence $\{\boldsymbol{x}^{k_j}\}$ such that $\boldsymbol{x}^{k_j} \in \Omega_\delta$ and $\boldsymbol{x}^{k_j+1} \notin \Omega_\delta$. To simplify notations, we can suppose that those sequences are convergent. Now, from Lemma 2.9, it follows that for some $\boldsymbol{u} \in \Omega$ and $\boldsymbol{v} \notin \mathrm{ri}(\Omega_\delta)$, $\boldsymbol{x}^{k_j} \to \boldsymbol{u}$ and $\boldsymbol{x}^{k_j+1} \to \boldsymbol{v}$. This is equivalent to $\lim_{k\to\infty} \|\boldsymbol{x}^{k_j+1} - \boldsymbol{x}^{k_j}\| \geq \delta$, contradicting Corollary 2.4. Then, the lemma is proved. □

Now, we present the main theorem which is an immediate consequence of the previous results.

THEOREM 2.12 *Let $\{\boldsymbol{x}^k\}$ be the sequence generated by the generalized primal affine scaling algorithm and assume that Assumptions 1–4 hold. Then, the following statements hold:*
(a) *The sequence $\{\boldsymbol{x}^k\}$ is bounded;*
(b) $\lim_{k\to\infty}(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k) = 0$*;*
(c) *Any accumulation point of $\{\boldsymbol{x}^k\}$ solves (1).*

Proof Part (a) follows from part (a) of Proposition 2.2. Part (b) is the Corollary 2.4. Finally, part (c) follows from part (b) of Proposition 2.2 and from Lemma 2.11. □

3 Generalized Dual Affine Scaling Method

3.1 *Introduction*

In this section, we generalize, by variable metric, the dual affine scaling algorithm of Adler, Karmarkar, Resende and Veiga [20] for linear programming.

The dual affine scaling algorithm has proved reliable in practice in linear programming problems [20, 23–25], large-scale network flow problems [26] and large-scale assignment problems [27, 28]. Notwithstanding its success in practice, no polynomial time proof has been found so far for the primal or dual variants of the algorithm.

The dual affine scaling algorithm, like the primal one, also consists of three key parts, namely, starting with an interior dual feasible solution, moving towards a better interior solution, and stopping at an optimal dual solution.

Let $\boldsymbol{c}$ and $\boldsymbol{x}$ be $\mathbb{R}^n$ vectors, $\boldsymbol{b}$ and $\boldsymbol{y}$ be $\mathbb{R}^m$ vectors and $\mathbf{A}$ a $m \times n$ matrix, we will consider the linear programming problem

$$
\begin{array}{lll}
\text{(P)} & \text{minimize}_{\boldsymbol{x}} & \boldsymbol{c}^T\boldsymbol{x} \\
 & \text{subject to} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b} \\
 & & \boldsymbol{x} \geq 0
\end{array}
\tag{18}
$$

and its dual

$$\text{(D)} \quad \begin{array}{ll} \text{maximize}_{\boldsymbol{y}} & \boldsymbol{b}^T\boldsymbol{y} \\ \text{subject to} & \mathbf{A}^T\boldsymbol{y} \leq \boldsymbol{c} \end{array} \tag{19}$$

Introducing slack variables to the formulation of (D), we have

$$\text{(D)} \quad \begin{array}{ll} \text{maximize}_{(\boldsymbol{y},\boldsymbol{s})} & \boldsymbol{b}^T\boldsymbol{y} \\ \text{subject to} & \mathbf{A}^T\boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c} \\ & \boldsymbol{s} \geq 0 \end{array} \tag{20}$$

where $\boldsymbol{s}$ is the n-vector of slack variables.

The dual affine scaling algorithm solves the linear programming problem (18), indirectly, by solving its dual (19). The algorithm starts with an initial solution

$$\boldsymbol{y}^0 \in \{\boldsymbol{y} | \boldsymbol{c} - \mathbf{A}^T\boldsymbol{y} > 0\}$$

and proceeds to generate a sequence of feasible interior points $\{\boldsymbol{y}^1, \boldsymbol{y}^2, ..., \boldsymbol{y}^k, ...\}$, i.e., $\boldsymbol{c} - \mathbf{A}^T\boldsymbol{y}^k > 0$, such that

$$\boldsymbol{b}^T\boldsymbol{y}^{k+1} > \boldsymbol{b}^T\boldsymbol{y}^k,$$

until a *stopping criterion* be satisfied.

As usual, our affine variant of Karmarkar's algorithm consists of a scaling operation applied to (D), with a subsequent search procedure that determines the next iterate. At iteration k, this linear transformation is given by

$$\hat{\boldsymbol{s}} = \mathbf{S}_k^{-r/2}\boldsymbol{s},$$

where

$$\mathbf{S}_k = diag(\boldsymbol{s}_1^k, ..., \boldsymbol{s}_m^k).$$

In terms of the scaled slack variables, problem (19) is

$$(\hat{\text{D}}) \quad \begin{array}{ll} \text{maximize}_{(\boldsymbol{y},\boldsymbol{s})} & \boldsymbol{b}^T\boldsymbol{y} \\ \text{subject to} & \mathbf{A}^T\boldsymbol{y} + \mathbf{S}_k^{r/2}\hat{\boldsymbol{s}} = \boldsymbol{c} \\ & \boldsymbol{s} \geq 0 \end{array} \tag{21}$$

Under the full-rank assumption for $\mathbf{A}^T$, there is a one-to-one relationship between the feasible solutions set for D,

$$Y = \left\{ \boldsymbol{y} \in \mathbb{R}^m | \mathbf{A}^T \boldsymbol{y} \leq \boldsymbol{c} \right\},$$

and the feasible scaled slacks set in $\hat{\mathrm{D}}$,

$$\hat{S} = \left\{ \hat{\boldsymbol{s}} \in \mathbb{R}^n | \exists \boldsymbol{y} \in Y, \mathbf{A}^T \boldsymbol{y} + \mathbf{S}_k^{r/2} \hat{\boldsymbol{s}} = \boldsymbol{c} \right\},$$

given by

$$\hat{\boldsymbol{s}}(\boldsymbol{y}) = \mathbf{S}_k^{-r/2} (\boldsymbol{c} - \mathbf{A}^T \boldsymbol{y}) \tag{22}$$

and

$$\boldsymbol{y}(\hat{\boldsymbol{s}}) = \left(\mathbf{A}\mathbf{S}_k^{-r} \mathbf{A}^T \right)^{-1} \mathbf{A}\mathbf{S}_k^{-r/2} \left(\mathbf{S}_k^{-r/2} \boldsymbol{c} - \hat{\boldsymbol{s}} \right), \tag{23}$$

as observed in [29].

Similarly, there is also a corresponding one-to-one relationship linking feasible directions $\boldsymbol{dy}$ in Y and $\boldsymbol{d\hat{s}}$ in $\hat{S}$, with

$$\boldsymbol{d\hat{s}} = -\mathbf{S}_k^{-r/2} \mathbf{A}^T \boldsymbol{dy} \tag{24}$$

and

$$\boldsymbol{dy} = -\left(\mathbf{A}\mathbf{S}_k^{-r} \mathbf{A}^T \right)^{-1} \mathbf{A}\mathbf{S}_k^{-r/2} \boldsymbol{d\hat{s}}. \tag{25}$$

From (24), the feasible direction in $\hat{S}$ lies on the range space of $\mathbf{S}_k^{-r/2} \mathbf{A}^T$.

In line with other presentations of affine variants of Karmarkar's algorithm, we use the projected objective function gradient with respect to the scaled variables as the search direction chosen at each iteration.

Using (21) and (23), the gradient of the objective function with respect to $\hat{\boldsymbol{s}}$ is

$$\nabla_{\hat{\boldsymbol{s}}} \boldsymbol{b}(\boldsymbol{y}(\hat{\boldsymbol{s}})) = (\nabla_{\hat{\boldsymbol{s}}} \boldsymbol{y}(\hat{\boldsymbol{s}}))^T \nabla_{\boldsymbol{y}} \boldsymbol{b}(\boldsymbol{y}) = -\mathbf{S}_k^{-r/2} \mathbf{A}^T \left(\mathbf{A}\mathbf{S}_k^{-r} \mathbf{A}^T \right)^{-1} \boldsymbol{b},$$

which lies in the range space of $\mathbf{S}_k^{-r/2} \mathbf{A}^T$. Consequently, a projection is unnecessary and the search direction in $\hat{S}$ is given by

$$\boldsymbol{d\hat{s}} = -\mathbf{S}_k^{-r/2} \mathbf{A}^T \left(\mathbf{A}\mathbf{S}_k^{-r} \mathbf{A}^T \right)^{-1} \boldsymbol{b}. \tag{26}$$

From (25) and (26), we compute the corresponding feasible direction in Y,

$$\boldsymbol{dy} = \left(\mathbf{A}\mathbf{S}_k^{-r}\mathbf{A}^T\right)^{-1}\boldsymbol{b}. \tag{27}$$

The corresponding feasible direction for the unscaled slacks is obtained by applying the inverse affine transformation to $\boldsymbol{d\hat{s}}$,

$$\boldsymbol{ds} = -\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-r}\mathbf{A}^T\right)^{-1}\boldsymbol{b}. \tag{28}$$

Unboundedness is detected if $\boldsymbol{ds} \geq 0$, in the case of $\boldsymbol{b} \neq 0$. On the contrary, we compute the next iterate by taking the maximum feasible step in the direction $\boldsymbol{ds}$, and retracting back to an interior point according to a *safety factor* γ, $0 < \gamma < 1$, i.e.,

$$\boldsymbol{y}^{k+1} = \boldsymbol{y}^k + \alpha\boldsymbol{dy}, \tag{29}$$

where

$$\alpha = \gamma \times \min\left\{-\boldsymbol{s}_i^k/(\boldsymbol{ds})_i \mid (\boldsymbol{ds})_i < 0,\ i = 1, ..., n\right\}. \tag{30}$$

At each iteration, an attempt at primal solution is computed by

$$\boldsymbol{x}^k = \mathbf{S}_k^{-r}\mathbf{A}^T\left(\mathbf{A}\mathbf{S}_k^{-r}\mathbf{A}^T\right)^{-1}\boldsymbol{b}. \tag{31}$$

It is easy to check $\mathbf{A}\boldsymbol{x}^k = \boldsymbol{b}$, but $\boldsymbol{x}^k$ may not necessarily be positive.

Even if $\boldsymbol{dy}$ is not computed exactly in (27), we can still obtain a pair of feasible directions by computing

$$\boldsymbol{ds} = -\mathbf{A}^T\boldsymbol{dy}. \tag{32}$$

Generalized dual affine scaling algorithm (GDAS) can be described in pseudo-code below.

Algorithm GDAS:
initialization: Let ($\mathbf{A}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{y}^0 \in Y$, *stopping criterion* and $\gamma \in (0, 1)$)
$k = 0$

while *stopping criterion* not satisfied

$$
\begin{aligned}
\boldsymbol{s}^k &= \boldsymbol{c} - \mathbf{A}^T \boldsymbol{y}^k \\
\boldsymbol{dy}^k &= \left(\mathbf{A}\mathbf{S}_k^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b} \\
\boldsymbol{ds}^k &= -\mathbf{A}^T \boldsymbol{dy}^k \\
\textbf{if } \boldsymbol{ds}^k &\geq 0 \rightarrow \textbf{stop} \\
\alpha_k &= \gamma \times \min\left\{-\boldsymbol{s}_i^k/(\boldsymbol{ds}^k)_i | (\boldsymbol{ds}^k)_i < 0,\ i = 1, ..., n\right\} \\
\boldsymbol{y}^{k+1} &= \boldsymbol{y}^k + \alpha \boldsymbol{dy}^k \\
k &= k + 1
\end{aligned}
$$

end while
end Algorithm GDAS

3.2 *Convergence of the Generalized Dual Affine Scaling Algorithm*

We will now prove the convergence of the generalized dual affine scaling algorithm with the following assumptions:
Assumption 1: The matrix $\mathbf{A}$ has full rank m;
Assumption 2: $\boldsymbol{b} \neq 0$;
Assumption 3: The dual problem (D) has an interior feasible solution, i.e.,

$$D^{++} = \left\{(\boldsymbol{y}, \boldsymbol{s}) \in \mathbb{R}^m \times \mathbb{R}^n :\ \mathbf{A}^T\boldsymbol{y} + \boldsymbol{s} = \boldsymbol{c},\ \boldsymbol{s} > 0\right\} \neq \emptyset;$$

Assumption 4: The primal problem (P) has a feasible solution.

We now introduce some functions which are used in the description and in the analysis of the dual affine scaling algorithm. For every $\boldsymbol{s} \in \mathbb{R}^n_{++}$, let

$$\boldsymbol{x}(\boldsymbol{s}) \equiv \mathbf{S}^{-r}\mathbf{A}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b}, \tag{33}$$

$$\boldsymbol{dy}(\boldsymbol{s}) \equiv \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b}, \tag{34}$$

$$\boldsymbol{ds}(\boldsymbol{s}) \equiv -\mathbf{A}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b} = -\mathbf{A}^T\boldsymbol{dy}(\boldsymbol{s}) = -\mathbf{S}^r\boldsymbol{x}(\boldsymbol{s}), \tag{35}$$

where $\mathbf{S} \equiv diag(\boldsymbol{s})$.

$\boldsymbol{x}(\boldsymbol{s})$ is called *primal estimate* associated with the point $\boldsymbol{s}$ and the pair $(\boldsymbol{dy}(\boldsymbol{s}), \boldsymbol{ds}(\boldsymbol{s}))$ is referred to as the pair of *dual affine scaling directions* associated with $\boldsymbol{s}$.

It is easy to note that Assumption 1–4 implies that the inverse of $\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T$ exists for every $\boldsymbol{s} > 0$ and $\boldsymbol{ds}$ must have at least one negative component. Thus, the expression which determines $\boldsymbol{y}^{k+1}$ in the dual affine scaling algorithm is well defined.

The following result provides characterizations of the dual affine scaling direction $\boldsymbol{dy}(\boldsymbol{s})$ and the primal estimate $\boldsymbol{x}(\boldsymbol{s})$ as optimal solutions of certain quadratic programming problems.

PROPOSITION 3.1 *The following statements hold.*

(a) *For every* $\boldsymbol{s} > 0$, $\boldsymbol{dy}(\boldsymbol{s}) = \sqrt{\boldsymbol{b}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b}}\ \boldsymbol{u}(\boldsymbol{s})$, *where* $\boldsymbol{u}(\boldsymbol{s})$ *is the unique optimal solution of the following Quadratic Programming problem*

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{u}} & \boldsymbol{b}^T\boldsymbol{u} \\ \text{subject to} & \boldsymbol{u}^T\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\boldsymbol{u} \leq 1, \end{array} \tag{36}$$

(b) *For every* $\boldsymbol{s} > 0$, $\boldsymbol{x}(\boldsymbol{s})$ *is the unique optimal solution of the following Quadratic Programming problem*

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{x}} & \frac{1}{2}\|\mathbf{S}^{r/2}\boldsymbol{x}\|^2 \\ \text{subject to} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \end{array} \tag{37}$$

where $\mathbf{S} \equiv diag(\boldsymbol{s})$.

Proof Since the solution of (36) will lie on the boundary of the ellipsoid, using the Lagrangian theory, this solution is

$$\begin{aligned} \boldsymbol{u}(\boldsymbol{s}) &= \tfrac{1}{\lambda(\boldsymbol{s})}\left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1}\boldsymbol{b} \\ \lambda(\boldsymbol{s}) &= \sqrt{\boldsymbol{b}^T\left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1}\boldsymbol{b}} \end{aligned}$$

and, thus, for each $\boldsymbol{s}$,

$$\boldsymbol{dy}(\boldsymbol{s}) = \lambda(\boldsymbol{s})\boldsymbol{u}(\boldsymbol{s}).$$

So, part (a) follows.

To prove part (b), by the Lagrange multipliers theorem, if $\boldsymbol{x}^*(\boldsymbol{s})$ is solution of (37), there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$\mathbf{S}^r\boldsymbol{x}^* + \mathbf{A}^T\boldsymbol{\lambda}^* = 0 \tag{38}$$

$$\mathbf{A}\boldsymbol{x}^* = \boldsymbol{b} \tag{39}$$

Multiplying (38) by $\mathbf{A}\mathbf{S}^{-r}$ and using (39), we obtain

$$\boldsymbol{\lambda}^* = -\left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1}\boldsymbol{b}$$

Now, substituting $\boldsymbol{\lambda}^*$ in (38) multiplied by $\mathbf{S}^{-r}$, we obtain

$$\boldsymbol{x}^* = \mathbf{S}^{-r}\mathbf{A}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b}$$

Therefore, part (b) also follows. □

LEMMA 3.2

$$\|\mathbf{S}^{-r/2}\boldsymbol{ds}(\boldsymbol{s})\|^2 = \|\mathbf{S}^{r/2}\boldsymbol{x}(\boldsymbol{s})\|^2 = \boldsymbol{b}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b} > 0, \ \ \forall \boldsymbol{s} > 0, \tag{40}$$

where $\mathbf{S} \equiv diag(\boldsymbol{s})$.

Proof The first equality is an immediate consequence of the fact that $\boldsymbol{d}(\boldsymbol{s}) = -\mathbf{S}^r\boldsymbol{x}(\boldsymbol{s})$, given in (35). The second equality may be obtained, also using (35), from the following simple calculation

$$\begin{aligned}\|\mathbf{S}^{-r/2}\boldsymbol{ds}(\boldsymbol{s})\|^2 &= \| - \mathbf{S}^{-r/2}\mathbf{A}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b}\|^2 \\ &= \boldsymbol{b}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \mathbf{A}\mathbf{S}^{-r/2}\mathbf{S}^{-r/2}\mathbf{A}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b} \\ &= \boldsymbol{b}^T \left(\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T\right)^{-1} b.\end{aligned}$$

The inequality in (40) follows from Assumption 1, which implies that $\mathbf{A}\mathbf{S}^{-r}\mathbf{A}^T$ is positive definite, and from Assumption 2. □

THEOREM 3.3 $\boldsymbol{s}^k > 0$ *for all* k*, and the sequence* $\{\boldsymbol{b}^T\boldsymbol{y}^k\}$ *is monotone increasing, and converges. Also* $\sum_{i=0}^{\infty} \boldsymbol{b}^T(\boldsymbol{y}^{k+1} - \boldsymbol{y}^k) < \infty$.

Proof The first part follows by induction. Since $\boldsymbol{s}^0 > 0$, it is sufficient to prove that $\boldsymbol{s}^k > 0$ implies that $\boldsymbol{s}^{k+1} > 0$. From the algorithm procedure and (35), we have

$$\boldsymbol{s}^{k+1} = \boldsymbol{c} - \mathbf{A}^T\boldsymbol{y}^{k+1} = \boldsymbol{c} - \mathbf{A}^T(\boldsymbol{y}^k + \alpha_k\boldsymbol{dy}^k) = \boldsymbol{s}^k - \alpha_k\mathbf{A}^T\boldsymbol{dy}^k = \boldsymbol{s}^k + \alpha_k\boldsymbol{ds}^k.$$

Assume that $\boldsymbol{s}^k > 0$. If $\boldsymbol{ds}_i^k \geq 0$, then $\boldsymbol{s}_i^k + \alpha_k\boldsymbol{ds}_i^k > 0$, $\forall\alpha_k \geq 0$. Otherwise, if $\boldsymbol{ds}_i^k < 0$, then $\boldsymbol{s}_i^k + \alpha_k\boldsymbol{ds}_i^k > 0$ if and only if $\alpha_k < -\boldsymbol{s}_i^k/\boldsymbol{ds}_i^k$, which is satisfied by the choice of α_k.

To obtain the second part, we use (34) and Lemma 3.2 to have

$$\begin{aligned}\boldsymbol{b}^T\boldsymbol{y}^{k+1} &= \boldsymbol{b}^T\boldsymbol{y}^k + \frac{\gamma}{\chi\left[-\mathbf{S}_k^{-1}\boldsymbol{ds}^k\right]}\boldsymbol{b}^T \left(\mathbf{A}\mathbf{S}_k^{-r}\mathbf{A}^T\right)^{-1} \boldsymbol{b} \\ &\geq \boldsymbol{b}^T\boldsymbol{y}^k + \gamma\frac{\|\mathbf{S}_k^{-r/2}\boldsymbol{ds}^k\|^2}{\|\mathbf{S}_k^{-1}\boldsymbol{ds}^k\|}.\end{aligned} \tag{41}$$

From (40), the additional term in the above formula is positive. This implies that $\boldsymbol{b}^T\boldsymbol{y}^{k+1} > \boldsymbol{b}^T\boldsymbol{y}^k$. Therefore, $\{\boldsymbol{b}^T\boldsymbol{y}^k\}$ is monotone increasing. Due to Assumption 4, this sequence is limited and, therefore, converges, say to $\boldsymbol{b}^*$.

Also, $\sum_{i=0}^{\infty} \boldsymbol{b}^T(\boldsymbol{y}^{k+1} - \boldsymbol{y}^k) = \boldsymbol{b}^* - \boldsymbol{b}^T\boldsymbol{y}^0$, where $\boldsymbol{b}^T\boldsymbol{y}^k \to \boldsymbol{b}^*$, and this concludes the proof. □

The following technical lemmas will be used in the proof of the Proposition 3.6.

LEMMA 3.4 *Given a vector $\boldsymbol{c} \in \mathbb{R}^k$ with $\boldsymbol{c} \neq 0$, a matrix $\mathbf{Q} \in \mathbb{R}^{k\times l}$ of rank k and a diagonal matrix $\mathbf{D} \in \mathbb{R}^{l\times l}$ with positive diagonal entries. Then, there exists a constant $p(\mathbf{Q}, \boldsymbol{c}) > 0$, independent of $\mathbf{D}$, such that if $\boldsymbol{x}^*$ solves*

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{x}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{x}^T\mathbf{QDQ}^T\boldsymbol{x} \leq 1, \end{array} \tag{42}$$

then

$$|\boldsymbol{c}^T\boldsymbol{x}^*| \geq p(\mathbf{Q}, \boldsymbol{c})\|\boldsymbol{x}^*\|. \tag{43}$$

Proof See the proof in [19]. □

LEMMA 3.5 *Let $\mathbf{H} \in \mathbb{R}^{l\times p}$ and $\boldsymbol{h} \in \mathbb{R}^l$ be given such that $\boldsymbol{h} \in Im(\mathbf{H})$. Then, there exists a constant $M > 0$ such that for every diagonal matrix $\mathbf{D} > 0$, the (unique) optimal solution $\bar{\boldsymbol{w}} = \bar{\boldsymbol{w}}(\mathbf{D}) \in \mathbb{R}^p$ of the problem*

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{w}} & \|\mathbf{D}\boldsymbol{w}\| \\ \text{subject to} & \mathbf{H}\boldsymbol{w} = \boldsymbol{h} \end{array} \tag{44}$$

satisfies $\|\bar{\boldsymbol{w}}\| \leq M$.

Proof The proof is based on Hoffman's Lemma and can be found in [30]. □

From now on, we denote the sequence of the primal estimates $\{\boldsymbol{x}(\boldsymbol{s}^k)\}$ simply by $\{\boldsymbol{x}^k\}$.

PROPOSITION 3.6 *Let $\{\boldsymbol{b}^T\boldsymbol{y}^k\}$ be bounded above, and converge to $\boldsymbol{b}^*$. Then, the following statements hold for the generalized dual affine scaling algorithm:*
(a) *The sequence $\{(\boldsymbol{y}^k, \boldsymbol{s}^k)\}$ converges.*
(b) *There is a $\rho > 0$ such that for every $k = 1, 2, ...$*

$$\frac{\boldsymbol{b}^* - \boldsymbol{b}^T\boldsymbol{y}^k}{\|\boldsymbol{y}^* - \boldsymbol{y}^k\|} \geq \rho.$$

(c) *The sequence* $\{\boldsymbol{x}^k\}$ *is bounded;*
(d) *For each* $r \geq 1$, $\mathbf{X}_k \boldsymbol{s}^k \to 0$.

Proof From Proposition 3.1, $\boldsymbol{dy}^k = \lambda_k \boldsymbol{u}^k$, where $\boldsymbol{u}^k$ solves

$$\begin{array}{ll} \text{maximize}_{\boldsymbol{u}} & \boldsymbol{b}^T \boldsymbol{u} \\ \text{subject to} & \boldsymbol{u}^T \mathbf{A} \mathbf{S}_k^{-r} \mathbf{A}^T \boldsymbol{u} \leq 1. \end{array} \tag{45}$$

From Lemma 3.4, $\exists q(\mathbf{A}, \boldsymbol{b}) > 0$ such that $|\boldsymbol{b}^T \boldsymbol{u}^k| \geq q(\mathbf{A}, \boldsymbol{b}) \|\boldsymbol{u}^k\|$.
Then,

$$|\boldsymbol{b}^T \boldsymbol{dy}^k| = \lambda_k |\boldsymbol{b}^T \boldsymbol{u}^k| \geq \lambda_k q(\mathbf{A}, \boldsymbol{b}) \|\boldsymbol{u}^k\| = q(\mathbf{A}, \boldsymbol{b}) \|\boldsymbol{dy}^k\|.$$

From Theorem 3.3, it follows, observing that $\boldsymbol{b}^T \boldsymbol{dy}^k > 0$,

$$\begin{aligned} \infty > \sum_{i=0}^{\infty} \boldsymbol{b}^T (\boldsymbol{y}^{k+1} - \boldsymbol{y}^k) &= \sum_{i=0}^{\infty} \alpha_k \boldsymbol{b}^T \boldsymbol{dy}^k \\ \geq \sum_{i=0}^{\infty} \alpha_k q(\mathbf{A}, \boldsymbol{b}) \|\boldsymbol{dy}^k\| &= q(\mathbf{A}, \boldsymbol{b}) \sum_{i=0}^{\infty} \|\boldsymbol{y}^{k+1} - \boldsymbol{y}^k\| \end{aligned}$$

Thus, $\{\boldsymbol{y}^k\}$ is a Cauchy sequence and, therefore, converges. Since for each k, $\boldsymbol{s}^k = \boldsymbol{c} - \mathbf{A}^T \boldsymbol{y}^k$, the sequence $\{\boldsymbol{s}^k\}$ also converges. This proves part (a).

To prove part (b), using the triangular inequality, we obtain

$$\begin{aligned} \boldsymbol{b}^* - \boldsymbol{b}^T \boldsymbol{y}^k &= \sum_{j=0}^{\infty} \boldsymbol{b}^T (\boldsymbol{y}^{k+1+j} - \boldsymbol{y}^{k+j}) \geq q(\mathbf{A}, \boldsymbol{b}) \sum_{i=0}^{\infty} \|\boldsymbol{y}^{k+1+j} - \boldsymbol{y}^{k+j}\| \\ &\geq q(\mathbf{A}, \boldsymbol{b}) \|\boldsymbol{y}^* - \boldsymbol{y}^k\|. \end{aligned}$$

Part (c) follows from part (b) of Proposition 3.1 and Lemma 3.5.

Finally, to prove part (d), since $\{\boldsymbol{b}^T \boldsymbol{y}^k\}$ converges and using (41), we have

$$\gamma \frac{\|\mathbf{S}_k^{-r/2} \boldsymbol{ds}^k\|^2}{\|\mathbf{S}_k^{-1} \boldsymbol{ds}^k\|} = \gamma \frac{\|\mathbf{S}_k^{r/2} \boldsymbol{x}^k\|^2}{\|\mathbf{S}_k^{r-1} \boldsymbol{x}^k\|} \to 0, \text{ with } 0 < \gamma < 1.$$

Since $\{\boldsymbol{s}^k\}$ converges and $\{\boldsymbol{x}^k\}$ is bounded, when $r \geq 1$, the denominator of the above expression is bounded. Thus, $\mathbf{S}_k^{r/2} \boldsymbol{x}^k \to 0$. But this is only possible if $\mathbf{S}_k \boldsymbol{x}^k \to 0$. □

We now show, with the additional assumption of dual nondegeneration, that the sequence $\{\boldsymbol{x}^k\}$ also converges, and that the limit points of the primal and dual sequences are optimal for their respective problems.

THEOREM 3.7 *Assume that Assumptions 1–4 hold and that the dual is nondegenerated. Then, there exist vectors $\boldsymbol{x}^*$, $\boldsymbol{y}^*$ and $\boldsymbol{s}^*$ such that*
(a) $\boldsymbol{x}^k \to \boldsymbol{x}^*$,
(b) $\boldsymbol{y}^k \to \boldsymbol{y}^*$,
(c) $\boldsymbol{s}^k \to \boldsymbol{s}^*$,
where $\boldsymbol{x}^*$ is an optimal solution of primal problem and $(\boldsymbol{y}^*, \boldsymbol{s}^*)$ is an optimal solution of the dual problem.

Proof Using part (a) of the Proposition 3.6, let $(\boldsymbol{y}^*, \boldsymbol{s}^*)$ be a cluster point of the sequence $(\boldsymbol{y}^k, \boldsymbol{s}^k)$. It is clear that this limit point belongs to the boundary of the linear programming problem polyhedron. Define the sets

$$B = \left\{j : \boldsymbol{s}_j^* = 0\right\}$$
$$N = \left\{j : \boldsymbol{s}_j^* > 0\right\}.$$

Thus, $\mathbf{A}_B^T \boldsymbol{y}^* = \boldsymbol{c}_B$. As all dual solutions are nondegenerate, $\mathbf{A}_B$ has full column rank m.
From part (c) of the Proposition 3.6, let $\boldsymbol{x}^*$ be any cluster point of the sequence $\{\boldsymbol{x}^k\}$. Since, $\boldsymbol{x}_j^k \boldsymbol{s}_j^k \to 0$, from part (d) of the Proposition 3.6, then $\boldsymbol{x}_{\boldsymbol{j}}^* = 0$ for each $j \notin B$. Thus,

$$\boldsymbol{b} = \mathbf{A}\boldsymbol{x}^* = \mathbf{A}_B \boldsymbol{x}_B^*.$$

Since, $\mathbf{A}_B$ has full column rank m, this system has, at most, one solution. But each cluster point $\boldsymbol{x}^*$ of $\{\boldsymbol{x}^k\}$ solves this system, thus, the sequence has only one cluster point $\boldsymbol{x}^*$, and so

$$\boldsymbol{x}^k \to \boldsymbol{x}^*.$$

Now, assume that $j \in B$. If $\boldsymbol{x}_j^* < 0$, there exists an integer $L \geq 1$ such that $\boldsymbol{x}_j^k < 0$, $\forall k \geq L$. But, $0 > \boldsymbol{x}_j^k = -(\boldsymbol{s}_j^k)^{-r} \boldsymbol{ds}_j^k$, which implies $\boldsymbol{ds}_j^k > 0$. Thus,

$$\boldsymbol{s}_j^{k+1} = \boldsymbol{s}_j^k + \alpha_k \boldsymbol{ds}_j^k > \boldsymbol{s}_j^k$$

So $\boldsymbol{s}_j^k \nrightarrow 0$, which contradicts $j \in B$. Hence, $\boldsymbol{x}^* \geq 0$. Since, $\boldsymbol{y}^*$ and $\boldsymbol{s}^*$ are feasible for the dual, and $\boldsymbol{x}^*$ is feasible for the primal and they satisfy the complementary slackness conditions, the proof is finished. □

4 Obtained Results

4.1 *Initial Solution and Stopping Criterion*

Algorithms GPAS and GDAS require that an initial interior feasible solution be provided. To get one such solution, we use the Big-M scheme for GPAS and the Phase I/Phase II scheme proposed in [20] for GDAS. In those schemes, the algorithm is initially applied to an artificial problem with a modified stopping criterion. In this stage, if no interior feasible solution exists, the algorithm finds a solution that satisfies the stopping criterion for problem (P).

In our computational experiments, the algorithms terminated when we had a small relative improvement of the objective function, i.e.,

$$|f(\boldsymbol{x}^k) - f(\boldsymbol{x}^{k-1})|/\max\{1, |f(\boldsymbol{x}^{k-1})|\} < \epsilon \text{ in GPAS case,}$$

and,

$$|\boldsymbol{b}^T\boldsymbol{y}^k - \boldsymbol{b}^T\boldsymbol{y}^{k-1}|/\max\{1, |\boldsymbol{b}^T\boldsymbol{y}^{k-1}|\} < \epsilon \text{ in GDAS case,}$$

where ϵ is a small given positive tolerance. We set $\epsilon = 10^{-8}$ for both algorithms.

4.2 *Tested Problems*

With the objective of observing the behaviour in r parameter and of comparing with the classic algorithm (when $r = 1$ or $r = 2$), we implement the families of GPAS and GDAS algorithms proposed, applied to some problems. Tests with small-scale linear programs of NETLIB library have been made. In GPAS case, we also test some quadratic programming described in the Maros and Mészáros repository [21]. In this case, the objective function is given by

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{c}_0,$$

where $\mathbf{Q}$ is a symmetric and positive semidefinite $n \times n$ matrix.

Table 1 supplies the data of the NETLIB library problems and Table 2 furnishes the data of the quadratic programming problems. The dimensions of the problems are given in columns 2 and 3, and include fixed or slacks variables. The number of rows does not include the objective function. In column 4, the number of nonzero elements of A is displayed. Column 5 gives the optimal value supplied by the libraries.

All the tests were made in *Pentium* 866MHz and implemented in *Matlab* version 6.0 within *Windows 98*.

Table 1. NETLIB problems data

Problem	Row	Col	NZ(A)	Optimal Value
adlittle	56	138	424	2.2549496316e+05
afiro	27	51	102	-4.6475314286e+02
bandm	305	472	2494	-1.5862801845e+02
blend	74	114	522	-3.0812149846e+01
israel	174	316	2443	-8.9664482186e+05
kb2	52	77	331	-1.7499001299e+03
sc105	105	163	340	-5.2202061212e+01
sc205	205	317	665	-5.2202061212e+01
sc50a	50	78	160	-6.4575077059e+01
sc50b	50	78	148	-7.0000000000e+01
scagr7	129	185	465	-2.3313892548e+06
share1b	117	253	1179	-7.6589318579e+04
share2b	96	162	777	-4.1573224074e+02
stocfor1	117	165	501	-4.1131976219e+04

Table 2. Quadratic problems data

Problem	Row	Col	NZ(A)	Optimal Value
genhs28	8	20	24	9.2717369e-01
hs118	47	62	39	6.6482045e+02
hs21	5	7	2	-9.9960000e+01
hs35	1	4	3	1.1111111e-01
hs76	3	7	10	-4.6818182e+00
lotschd	7	12	54	2.3984159e+03
qptest	3	5	4	4.3718750e+00
zecevic2	4	6	4	-4.1250000e+00

4.3 *Results for GPAS*

In all the tests, the variation of the r parameter was from $r = 1.00$ until $r = 2.00$ with an increment of 0.05. The *safety factor* parameters were set to $\delta = 0.99$ and $\beta = 0.001$.

Tables 3 and 4 aim to supply an analysis of the behaviour of the family of algorithms implemented, according to r parameter, and to compare it with the existing classic algorithm when $r = 1$ or $r = 2$. In them, we supply the number of iterations for some values of r and the number of iterations in Phase I, $N1$, where we use $r = 1.5$.

Tables 5 and 6 supply one more succinct analysis of the results obtained. In them, we supply the number of iterations of the classic method ($r = 1.00$ or $r = 2.00$), the number of minimum iterations (n_{min}), the values of r where one gets the lesser number of iterations (r_{min}). The value $\overline{r_{min}}$, given in column r_{min}, is the average of the averages of r_{min} values for each problem. Column 3 gives the number of iterations for r value nearest to $\overline{r_{min}}$. Let t be the number of iterates generated by the algorithm. Columns 7 and 8 present detailed results of GPAS algorithm found in iterate t for $r = 1.5$. Primal value found is given

Table 3. Iterations number of GPAS algorithm in some r values and in Phase I with $r = 1.5$ for NETLIB problems.

	r										
Problem	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	N1
adlittle	2221	62	42	35	30	29	31	521	151	890	97	7
afiro	217	21	14	13	14	15	19	26	38	58	89	9
blend	905	60	42	35	51	38	113	92	133	95	164	10
kb2	1463	138	110	81	80	83	100	122	77	88	109	15
sc105	619	79	65	102	129	157	155	192	233	271	303	8
sc50a	369	41	37	39	44	56	73	97	134	180	240	8
sc50b	203	34	33	36	41	50	65	87	122	169	225	7
share2b	2132	71	64	81	86	125	58	151	113	164	210	15
stocfor1	1098	44	41	44	60	63	77	66	78	89	71	14
totals	9227	550	448	466	535	616	691	1354	1079	2004	1508	

Table 4. Iterations number of GPAS algorithm in some r values and in Phase I with $r = 1.5$ for quadratic problems.

	r										
Problem	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	N1
genhs28	15	15	15	15	15	15	15	15	15	15	16	6
hs118	613	34	27	24	23	23	27	31	35	40	39	7
hs21	43	17	17	20	26	39	60	91	131	175	215	7
hs35	23	25	59	112	39	81	671	881	1171	2472	1696	6
hs76	62	14	24	48	68	169	204	494	901	1484	1960	6
lotschd	384	19	14	13	12	11	12	13	14	18	23	6
qpcblend	1117	137	68	55	76	108	188	307	596	1121	2038	10
qptest	9	8	9	15	10	16	22	38	70	128	222	5
zecevic2	46	6	5	5	17	38	77	100	74	200	237	6
totals	2312	275	238	307	286	500	1276	1970	3007	5653	6446	

in column 7, and column 8 presents the normalized duality gap (N.D.G.),

$$|(\boldsymbol{x}^t)^T \mathbf{Q}\boldsymbol{x}^t + \boldsymbol{c}^T\boldsymbol{x}^t - \boldsymbol{b}^T\boldsymbol{y}^t| / \max(1, |\frac{1}{2}(\boldsymbol{x}^t)^T\mathbf{Q}\boldsymbol{x}^t + \boldsymbol{c}^T\boldsymbol{x}^t + \boldsymbol{c}_0|).$$

4.4 *Results for GDAS*

In all the tested problems the variation of the r parameter was from $r = 1.3$ until $r = 3.3$ with an increment of 0.1. The *safety factor* parameter was set to $\gamma = 0.99$.

The Table 7 aims to supply an analysis of the behaviour of the family of algorithms implemented, according to r parameter, and to compare it with the existing classic algorithm when $r = 2$. In it, we supply the number of iterations for some values of r and the number of iterations in Phase I, $N1$,

Table 5. Comparison between the lesser iteration number of GPAS algorithm with classical methods and primal objective value and duality gap in $r = 1.5$ for NETLIB problems.

	r						
Problem	1.00	1.35	2.00	n_{min}	r_{min}	$f(\boldsymbol{x}^t)$	N.D.G.
adlittle	2221	31	97	29	1.50	2.2549493e+05	1.61e-07
afiro	217	13	89	13	1.25 - 1.35	-4.6475315e+02	5.82e-09
blend	905	34	164	34	1.35	-3.0812126e+01	7.51e-07
kb2	1463	96	109	72	1.95	-1.7238748e+03	8.30e-07
sc105	619	86	303	65	1.20	-5.2202053e+01	1.61e-07
sc50a	369	41	240	37	1.20 - 1.25	-6.4575069e+01	1.29e-07
sc50b	203	38	225	33	1.15 - 1.20	-6.9999993e+01	9.86e-08
share2b	2132	59	210	54	1.45	-4.1573190e+02	8.24e-07
stocfor1	1098	61	71	35	1.15	-4.0975037e+04	3.30e-03
totals	9227	459	1508	372	$\overline{r_{min}}$=1.37		

Table 6. Comparison between the lesser iteration number of GPAS algorithm with classical methods and primal objective value and duality gap in $r = 1.5$ for quadratic problems.

	r						
Problem	1.00	1.25	2.00	n_{min}	r_{min}	$f(\boldsymbol{x}^t)$	N.D.G.
genhs28	15	15	16	15	1.00 - 1.95	9.2717370e-01	3.57e-05
hs118	613	26	39	22	1.45	6.6482049e+02	6.48e-08
hs21	43	18	215	17	1.10 - 1.20	-9.9960000e+01	4.28e-09
hs35	23	121	1696	23	1.00	1.1111883e-01	3.16e-04
hs76	62	38	1960	14	1.10	-4.6818159e+00	2.05e-05
lotschd	384	13	23	11	1.50	2.3984159e+03	3.42e-09
qpcblend	1117	62	2038	55	1.30	-7.8420206e-03	2.34e-06
qptest	9	9	222	6	1.05	4.3718751e+00	4.44e-05
zecevic2	46	6	237	5	1.15 - 1.20, 1.30	-4.1249998e+00	4.68e-05
totals	2312	308	6446	168	$\overline{r_{min}}$=1.25		

where we use $r = 2.0$.

Table 8 supplies one more succinct analysis of the results obtained. In it, we supply the number of iterations of the classic method ($r = 2.00$), the number of minimum iterations (n_{min}), the values of r where one gets the lesser number of iterations (r_{min}). The value $\overline{r_{min}}$, given in column r_{min}, is the average of the averages of r_{min} values of each problem. Column 2 gives the number of iterations for r value nearest $\overline{r_{min}}$. Let t be the number of iterates generated by the algorithm. Columns 6 and 7 present detailed results of GDAS algorithm found in iterate t for $r = 2.0$. Primal value found is given in column 6, and column 7 presents the normalized duality gap (N.D.G.),

$$|\boldsymbol{b}^T\boldsymbol{y}^t - \boldsymbol{c}^T\boldsymbol{x}^t|/\max(1, |\boldsymbol{c}^T\boldsymbol{x}^t|).$$

Table 7. Iterations number of GDAS algorithm in some r values and in Phase I with $r = 2.0$ for NETLIB problems.

	r												
Problem	1.3	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.3	N1
adlittle	37	31	24	23	23	24	29	31	29	33	34	35	1
afiro	19	18	18	19	20	20	22	24	25	26	26	28	1
bandm	70	44	29	25	26	28	27	31	29	41	33	34	8
blend	27	25	21	21	22	23	24	26	27	27	31	28	5
israel	164	105	54	38	51	34	41	51	60	68	85	82	9
kb2	35	28	23	22	21	22	24	26	29	32	34	32	4
sc105	28	24	21	22	22	24	25	26	28	29	31	31	3
sc205	29	24	21	22	24	25	30	31	34	34	38	37	5
sc50a	21	19	18	20	20	22	25	26	26	27	29	29	1
sc50b	19	18	18	19	20	21	22	23	25	27	29	30	2
scagr7	52	40	26	23	34	25	24	25	27	30	30	31	4
share1b	60	57	58	35	32	36	35	35	46	52	54	66	7
share2b	27	23	21	21	21	23	23	25	27	28	30	28	5
stocfor1	33	30	21	21	21	22	21	23	25	29	36	30	4
totals	621	486	373	331	357	349	372	403	437	483	520	521	

Table 8. Comparison between the lesser iteration number of GDAS algorithm with classical method and primal objective value and duality gap in $r = 2.0$ for NETLIB problems.

	r					
Problem	1.80	2.00	n_{min}	r_{min}	$\boldsymbol{c}^T\boldsymbol{x}^t$	N.D.G.
adlittle	23	23	23	1.7 - 1.8, 2.0	2.2549496e+05	8.64e-10
afiro	19	20	18	1.4 - 1.7	-4.6475314e+02	1.74e-09
bandm	25	26	25	1.8	-1.5862802e+02	3.98e-09
blend	21	22	21	1.6 - 1.9	-3.0812150e+01	1.07e-09
israel	38	51	31	1.9	-8.9664482e+05	8.31e-10
kb2	22	21	21	1.7, 2.0	-1.7499001e+03	4.69e-10
sc105	22	22	21	1.6 - 1.7	-5.2202061e+01	3.34e-09
sc205	22	24	21	1.5 - 1.7	-5.2202061e+01	1.73e-09
sc50a	20	20	18	1.6	-6.4575077e+01	2.23e-09
sc50b	19	20	18	1.4 - 1.7	-7.0000000e+01	1.67e-09
scagr7	23	34	23	1.8	-2.3313898e+06	6.98e-10
share1b	35	32	30	2.3	-7.6589312e+04	8.11e-08
share2b	21	21	21	1.6 - 2.1	-4.1573224e+02	2.37e-09
stocfor1	21	21	21	1.6, 1.8 - 2.0	-4.1131976e+04	6.26e-09
totals	331	357	287	$\overline{r_{min}}$=1.77		

5 Conclusions and Possible Future Researches Suggested

We have proposed new classes of primal and dual affine scaling algorithms, and the link between those classes, with methods already known, were established. We have shown the convergence for dual family GDAS and the so-called weak convergence for primal family GPAS.

From the executed computational experiments we verify that the interior point algorithms family GPAS proposed presents, for intermediate values of r between 1 and 2, a superior performance than the one presented by the classic algorithms, when $r = 1$ or $r = 2$; while, the interior point algorithms family GDAS proposed presents, for values of r different from 2, a similar performance than the one presented by the classic algorithm, when $r = 2$. We observe, that with the adopted tolerance, the solutions obtained for some of the computational experiments, are identical to the solutions furnished by the libraries.

Moreover, we observe for both families, a convex relation between the number of iterations and the parameter r.

The facts observed, motivate us to carry out research with the view to attempt the following:

- to study the dependence of the algorithms families GPAS and GDAS in relation to the r parameter;
- to extend the results for the general case of problems without the nondegeneration assumption;
- to analyse the algorithms family GDAS for the case where the objective function is convex but not linear, as in the case of quadratic programming;
- to experiment a primal–dual affine scaling methodology;
- to improve the computational performance.

Moreover, we intend to perform tests with others problems from the above mentioned libraries, or from other reference libraries.

Acknowledgments

The work of F. G. M. Cunha and P. R. Oliveira was partially supported by CNPq/Brazil.
The work of A. W. M. Pinto and J. X. da C. Neto was partially supported by CAPES/Brazil.

References

[1] Dikin, I. I., 1967, Iterative solution of problems of linear and quadratic programming. *Soviet Mathematics Doklady,* **8**, 647–675.
[2] den Hertog, D., 1992, Interior point approach to linear, quadratic and convex programming — algorithms and complexity. PhD thesis, Delft University of Technology, Delft, Holland.
[3] Eggermont, P. P. B., 1990, Multiplicative iterative algorithms for convex program. *Linear Algebra and its Applications,* **130**, 25–42.
[4] Iussem, A. N., 1995, An interior point multiplicative methods for optimization under positivity constraints. *Acta Mathematicae,* **38**, 163–184.
[5] do Carmo, M. P., 1992, *Riemannian Geometry.* Birkhaüser, Boston.
[6] Rapcsák, T., 1997, *Smooth Nonlinear Optimization in* $\mathbb{R}^n$. Kluwer Academic Publishers, Dordrecht.

[7] Udriste, C., 1994, *Convex Functions and Optimization Methods on Riemannian Manifolds.* Kluwer Academic Publishers, Boston.

[8] Luenberger, D. G., 1972, The gradient projection method along geodesics. *Management Science,* **18**(11), 620–631.

[9] Gabay, D., 1982, Minimizing a differentiable function over a differentiable manifold. *Journal of Optimization Theory and Applications,* **37**(2), 177–219.

[10] Bayer, D. A. and Lagarias, J. C., 1989, The nonlinear geometry of linear programming: I. affine and projective scaling trajectories. *Transactions of the American Mathematical Society,* **314**(2), 499–526.

[11] Bayer, D. A. and Lagarias, J. C., 1989, The nonlinear geometry of linear programming: II. legendre transform coordinates and central trajectories. *Transactions of the American Mathematical Society,* **314**(2), 527–581.

[12] Karmarkar, N., 1990, Riemannian geometry underlying interior-point methods for linear programming. *Contemporary Mathematics,* **114**, 51–75.

[13] da Cruz Neto, J. X., Lima, L. L. and Oliveira, P. R., 1998, Geodesic algorithm in Riemannian manifold. *Balkan Journal of Geometry and Applications,* **3**(2), 89–100.

[14] Ferreira, O. P. and Oliveira, P. R., 1998, Subgradient algorithm on Riemannian manifold. *Journal of Optimization Theory and Applications,* **97**(1), 93–104.

[15] Ferreira, O. P. and Oliveira, P. R., 2002, Proximal point algorithm on Riemannian manifolds. *Optimization,* **51**(2), 257–270.

[16] Nesterov, Y. E. and Todd, M., 2002, On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics,* **2**, 333–361.

[17] da Silva, G. J. P. and Oliveira, P. R., 2005, A new class of proximal algorithms for the nonlinear complementarity problem. In: L. Qi, K. Teo and X. Yang (Eds.), *Optimization and Control with Applications, Applied optimization,* **96** (Springer), pp. 549–561.

[18] Oliveira, G. L. and Oliveira, P. R., 2002, A new class of interior-point methods for optimization under positivity constraints. Technical report ES—570/01, Department of Systems Engineering and Computer Science, COPPE, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

[19] Saigal, R., 1993, Efficient variants of the affine scaling method. Technical report 93-21, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI.

[20] Adler, I., Karmarkar, N., Resende, M. G. C. and Veiga, G., 1989, An implementation of Karmarkar's algorithm for linear programming. *Mathematical Programming,* **44**, 297–335.

[21] Maros, I. and Mészáros, C., 1999, A Repository of convex quadratic programming problems. *Optimization Methods and Software,* **11–12**, 671–681.

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

[23] Adler, I., Karmarkar, N., Resende, M. G. C. and Veiga, G., 1989, Data structures and programming techniques for the implementation of Karmarkar's algorithm. *ORSA Journal on Computing,* **1**, 84–106.

[24] Karmarkar, N. and Ramakrishnan, K., 1991, Computational results of an interior point algorithm for large scale linear programming. *Mathematical Programming,* **52**, 555–586.

[25] Monma, C. and Morton, A., 1987, Computational experiments with a dual affine variant of Karmarkar's method for linear programming. *Operations Research Letters,* **6**, 261–267.

[26] Resende, M. and Veiga, G., 1993, An efficient implementation of network interior point method. In: D.S. Johnson and C.C. McGeoch (Eds.), *Network Flows and Matching: First DIMACS Implementation Challenge, volume 12 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (Providence, RI: American Mathematical Society), pp. 299–348.

[27] Resende, M. and Veiga, G., 1990, Computational study of two implementations of the dual affine scaling algorithm. Technical report, AT&T Bell Laboratories, Murray Hill, NJ.

[28] Resende, M. and Veiga, G., 1993, An implementation of the dual affine scaling algorithm for minimun cost flow on bipartite uncapaciated networks. *SIAM Journal on Optimization,* **3**, 516–537.

[29] Gonzaga, C. C., 1991, Interior point algorithms for linear programming problems with inequality constraints. *Mathematical Programming,* **52**, 209–225.

[30] Monteiro, R. D. C., Tsuchiya, T. and Wang, Y., 1993, A simplified global convergence proof of the affine scaling algorithm. *Annals of Operations Research,* **47**, 443–482.