A new class of potential affine algorithms for linear convex programming

Pinto, A. W. M.,[*] Oliveira, P. R. [†]and Cruz Neto, J. X.[‡]

04/21/2002

Abstract

We obtain a new class of primal affine algorithms for the linearly constrained convex programming. It is constructed from a family of metrics generated the $-r$ power, $r \geq 1$, of the diagonal iterate vector matrix. We obtain the so called weak convergence. That class contains, as particular cases, the multiplicative Eggermont algorithm for the minimization of a convex function on the positive orthant, when $r = 1$, and the affine scaling Gonzaga and Carlos direction for the general problem, corresponding to $r = 2$. The last author obtained some weaker properties, and the weak convergence for Eggermont method was obtained by Iussem.

1 Introduction

We consider the resolution of

$$\begin{array}{rl} \min & f(x) \\ \text{s. to} & Ax = b \\ & x \geq 0, \end{array} \tag{1.1}$$

where $f : R^n \to R$ is a convex and continuously differentiable function. Also, we assume that the feasible set, $F = \{x \in R^n : \ Ax = b, \ x \geq 0\}$, where A is a

[*]apihngora@bol.com.br
[†]poliveir@cos.ufrj.br
[‡]jxavier@ufpi.br

$m \times n$ matrix, $m \leq n$, is nonempty and nondegenerate. We are interested in *affine scaling potential primal algorithms.* If we restrain ourselves to primal interior point methods, which is the class we are interested, we mention the affine scaling algorithm for linear and quadratic cost functions by [5], the polynomial logarithmic barrier approach for convex selfconcordant functions, as can be seen in the book of [4], and the affine-scaling algorithm for convex functions by [12]. As our algorithm can be easily particularized to the case without the linear constraints, that is, the minimization over the positive orthant, we mention the multiplicative method by [7], after analyzed by [13].

In our approach, we let the positive orthant R^n_{++} as a Riemannian manifold. We observe that the knowledge of Riemannian geometry concepts are not necessary to understand the paper, but the interested people can see, e.g.,[6], for Riemannian geometry, and the books by [18] and [23], for that specific connection with optimization. Associated to R^n_{++}, we define a metrics through the diagonal matrix $G(x) = X^{-r}$, for $r \geq 1$, X being the diagonal matrix whose nonnull elements are the $x \in R^n_{++}$ entries. Now, we take $M = \left\{x \in R^n_{++} : \ Ax = b\right\}$ as a submanifold of R^n_{++}, with the induced metrics. As we will show, considering f defined on M, the opposite of its gradient is the affine-scaling direction corresponding to each $r \geq 1$. That fact allows the analysis of the overall class as the study of gradient algorithms. The connection between geometry and mathematical programming comes from [15]. Since then, some research has been object of consideration, by [11], [14], [1], [2], [18], [19], [20], [8], [3], [9], [10], [23], [16], and etc. 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 works in the usual way: the line search is performed along the direction. In the same lines of this paper, we proposed classes of algorithms based on the metrics G given above, for the nonlinear complementarity problem, in [22], and, for the minimization over the positive orthant, within a proximal method setting, in [17].

The directions of our class, corresponding to $r = 1$ and $r = 2$, are, respectively, the [7] and [12] directions. The convergence result we have got is the so called *weak-convergence*, as obtained by [13] for the multiplicative method. They are stronger than those obtained by [12]. Finally, let's mention the power affine-scaling method by [21], for linear programming, whose convergence results are also valid for the degenerate case.

The paper is organized as follows.

2 Background

In this section, we present the algorithm direction as a gradient of f, within a certain framework. We limit ourselves to present the theoretical results we need to construct that direction. Those interested on the detailed description of the concepts presented here, can see, e.g., [23]. We let $M = \left\{x \in R^n_{++} : \ Ax = b\right\} \neq \emptyset$, a submanifold of $\left(R^n_{++}, X^{-r}\right)$, for $r \geq 1$, with the induced metrics. We assume the following :

Hypothesis 1 rank $A = m$

Hypothesis 2 1.1 is a nondegenerate. problem.

Hypothesis 3 f is a continuously differentiable (C^1) and convex function.

Now, we are going to obtain the gradient of f, restricted to M.

The first element we need is the tangent space to M, the usual null space of A, given by

$$T_x M = TM = \{d \in R^n : \ Ad = 0\},$$

for any x (it would be x dependent for nonlinear constraints). In order to define the gradient, we need the expression of the projection operator on TM. Thus, denoting I as the identity $n \times n$ matrix, the projection on TM, under any (positive definite matrix) metrics G is

$$P_G(x) = I - G^{-1}(x) A^T \left(AG^{-1}(x) A^T\right)^{-1} A.$$

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

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

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

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

for any $x \in R^n_{++}$, where ∇ represents the usual (Euclidean) gradient. Therefore, the gradient of f restricted to (M, X^{-r}) is the projection of $\nabla_{R^n_{++}} f(x)$ on TM, that is,

$$\nabla f_M(x) = P(x) X^r \nabla f(x)$$

Clearly, we have the Cauchy-Schwartz inequality: particularly, for all $d \in TM$ such that $\|d\|_G := \langle d, Gd \rangle^{1/2} = 1$, it is true that

$$\langle d, \nabla f_M(x) \rangle_G \leq \|d\|_G \|\nabla_M f(x)\|_G = \|\nabla_M f(x)\|_G .$$

Consequently, $-\nabla_M f(x) / \|\nabla_M f(x)\|_G$ is the steepest descent direction of f at x. Also, an obvious fact is that $\nabla_M f(x) \in TM$.

Now, suppose that f is convex. Then x^* is a solution for 1.1, if, and only if, there exist Lagrangian multipliers $y^* \in R^m$, $s^* \in R^n_+$ such that

$$\begin{gathered} A^T y^* + s^* = \nabla f(x^*) \\ Ax^* = b \\ x_i^* s_i^* = 0, i = 1, ..., n \\ x^* \geq 0. \end{gathered}$$

From those conditions, some straightforward calculations lead to

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

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

$$y(x) = \left(AX^r A^T\right)^{-1} AX^r \nabla f(x) \tag{2.1}$$

$$s(x) = \nabla f(x) - A^T y(x) = P^T(x) \nabla f(x) \tag{2.2}$$

$$d(x) = Xs(x) = P(x) X^r \nabla f(x) = \nabla_M f(x). \tag{2.3}$$

Observe that from the nondegeneracy hypothesis, all of above functions can be extended to the closure of M.

3 The algorithm and its convergence

We apply the gradient direction just computed to define the following algorithm.

3.1 The r-projection algorithm

Denote by $\chi(v)$ the largest entry of v . Given $x^0 \in M,\ \beta > 0$ and $0 < \delta < 1$, set $k = 0$,

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

As an immediate consequence of above construction is the fact that the algorithm is an interior point method. Indeed, for $x^k \in M$, for $k \geq 0$, it holds that

$$\left(X^{-1^k} x^{k+1}\right)_i = 1 - t^k \left(x_i^k\right)^{-1} d_i^k \geq 1, \text{ if } d_i^k \leq 0,$$

or, we have

$$t^k \left(x_i^k\right)^{-1} d_i^k \leq \frac{\delta\mu^k}{\beta + \mu^k} < 1,$$

otherwise. Therefore one gets $1 - t^k \left(x_i^k\right)^{-1} d_i^k > 0$, as aimed. Clearly, in order to assure a large search interval $\left[0, \alpha^k\right]$, the parameter δ must be near 1, and β near 0.

3.2 Convergence

The following assumption will be considered from now on:

Hypothesis 4 The starting level set, given by $L_{x^0} = \{x \in M : f(x) \leq f(x^0)\}$ is bounded.

Now, the following lemma shows some useful relations:

Lemma 3.1 *Let suppose given $x \in M$ and $d \in TM$. Then it holds:*
(i) $\nabla f(x)^T d = s(x)^T d$
(ii) $\nabla f(x) d(x) = \left\| X^{r/2} s(x) \right\|^2 = \left\| X^{-r/2} d(x) \right\|^2$,
where $s(x)$ and $d(x)$ are given, respectively, by, 2.2 and 2.3.

Proof. (i) Following 2.2, we get, successively,

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

which is (i), the last equality being a consequence of the fact that $d \in TM$ implies $Ad = 0$.

(ii) Reasoning as above, and using 2.3, we have

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

which is the first result. Now, again from 2.3, substitute $s(x) = X^{-r} d(x)$ in last expression to get the remainder result:

$$\nabla f(x)^T d(x) = \left\| X^{-r/2} s(x) \right\|^2.$$

$\blacksquare$

Lemma 3.2 *The sequence generated by the algorithm satisfies* $\lim_{k\to\infty} X^k s^k = 0$.

Proof. First observe that $X^k s^k \to 0$ if and only if $X^{r/2^k} s^k \to 0$. This is a consequence of the following facts: from the hypothesis, L_{x^0} is bounded, also $x^k \in L_{x^0}$, and, due to the nondegeneracy hypothesis 2, $s(x)$ is continuous in the closure of M. Therefore, $\{x^k\}$ and $\{s^k\}$ are both bounded. Now, use the contradiction hypothesis that $\lim_{k\to\infty} X^{r/2^k} s^k \neq 0$. In that case, there exists some subsequence $\{x^{k_j}\}$ and $\varepsilon > 0$ such that $\left\| X^{r/2^{k_j}} s^{k_j} \right\| \geq \varepsilon$. On the other hand, the boundedness of $\{x^k\}$ allows, without loss of generality, to admit that $x^{k_j} \to \overline{x}$, for some $\overline{x} \in M$. Additionally, the degeneracy hypothesis leads to $s^{k_j} \to s(\overline{x}) = \overline{s}$. At this point, we set

$$L = \sup_{x \in L_{x0}} \left\{ \chi \left(X^{-1} d\left(x\right), 0 \right) \right\} = \sup_{x \in L_{x0}} \left\{ \chi \left(X^{r^{-1}} s\left(x\right), 0 \right) \right\}.$$

Clearly, L is well defined and finite. Then, the sequence $\left\{ \alpha^k \right\}$ 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

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

that is, through a remainder Taylor term, $f\left(x^{k_j+1}\right) \leq f\left(x^{k_j}\right) - \alpha \nabla f\left(x^{k_j}\right)^T d^{k_j} + R\left(x^{k_j}, \alpha d^{k_j}\right)$, or

$$f\left(x^{k_j+1}\right) - f\left(x^{k_j}\right) \leq -\alpha \nabla f\left(x^{k_j}\right)^T d^{k_j} + R\left(x^{k_j}, \alpha d^{k_j}\right). \tag{3.1}$$

Now, we utilize the boundedness of L_{x^0} and the monotonicity of $\left\{ f\left(x^k\right) \right\}$, to obtain $f\left(x^{k_j+1}\right) - f\left(x^{k_j}\right) \to 0$. Besides, due to the preceding lemma, number (ii), and the working hypothesis, the following is true: $\nabla f\left(x^{k_j}\right)^T d^{k_j} \to \left\| \overline{X}^{r/2} s\left(\overline{x}\right) \right\|^2 \geq \varepsilon$. We can take the limits in 3.1, getting

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

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

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

but this contradicts the differentiability of f at $\overline{x}$, thus showing the aimed result. ∎

Corollary 3.3 *The directions generated by the algorithm verify* $\lim_{k \to \infty} d^k = 0$.

Proof. That is a direct consequence of the foregoing lemma and the application of equality 2.3, that is $d^k = X^{r^k} s^k$. ∎

Corollary 3.4 *The iterates verify* $\lim_{k\to\infty} \left(x^{k+1} - x^k\right) = 0$.

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

Lemma 3.1 says that any accumulation point $\overline{x}$ of the iteration sequence $\left\{x^k\right\}$ verifies the complementarity condition $\overline{X}s\left(\overline{x}\right) = 0$. We are going to show that $\overline{s} = s\left(\overline{x}\right) \geq 0$, if f is convex, so proving that any limit point of $\left\{x^k\right\}$ is a solution to 1.1. In that direction, fix a certain accumulation point $\overline{x}$, and let $\overline{s} = s\left(\overline{x}\right)$. We can consider a partition B, N of the set $\{1, 2, ..., n\}$, for $N := \{i : \overline{s}_i \neq 0\}$. Then, decomposing $\overline{x}$ as $\overline{x} = \overline{x}_B + \overline{x}_N$, we get $\overline{x}_N = 0$. Now, observe that if there exists another limit point $\widetilde{x}$, it holds that $f\left(\overline{x}\right) = f\left(\widetilde{x}\right)$, because $\left\{f\left(x^k\right)\right\}$ is a convergent sequence. As an outcome, $\overline{s}$ will be an optimal multiplier to 1.1 if $\overline{s}_N > 0$. Those remarks will induce us in defining the following set:

$$\Omega = \left\{x \in \overline{M} : f\left(x\right) = f\left(y\right) \text{ and } x_N = 0 \right\},$$

where $\overline{M}$ is the closure of M. A first property for Ω is its compactness, as a consequence of the boundedness of L_{x^0}. A second and obvious fact is that the accumulation points such that $x_N = 0$ are necessarily in Ω. Finally, we can present the main steps to achieve our aim. First, we dilate Ω, through the following definition:

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

Next, we show that there exists some dilation such that, if $\overline{s}_N$ has some negative component, then the corresponding sequence has a limit point outside the dilated set. But that contradicts last corollary.

A simple but essential property about those sets is that if $x \in \overline{M}$ and $x \notin \Omega_\delta$, then the point-set distance $d\left(x, \Omega\right) > \delta$. Now, we show that

Lemma 3.5 *The set* Ω *is convex.*

Proof. Take u and v in Ω, and $t \in [0, 1]$. Then

$$(tu + (1-t)\,v)_N = tu_N + (1-t)\,v_N = 0.$$

On the other hand, the convexity of f implies

$$\begin{aligned} f\,(tu + (1-t)\,v) &\le tf\,(u) + (1-t)\,f\,(v) \\ &\le tf\,(\overline{x}) + (1-t)\,f\,(\overline{x}) = f\,(\overline{x})\,. \end{aligned} \tag{3.2}$$

Additionally, observe that $tu + (1-t)\,v - \overline{x} \in TM$, so, using lemma 3.1, it holds

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

the last equality due to $s_B\,(\overline{x}) = 0$ and $(tu + (1-t)\,v - \overline{x})_N = 0$. Make use again of the convexity of f to get:

$$\begin{aligned} f\,(tu + (1-t)\,v) \ge f\,(\overline{x}) + \nabla f\,(\overline{x})^T\,(tu + (1-t)\,v - \overline{x}) \\ = f\,(\overline{x})\,. \end{aligned}$$

It follows, comparing with 3.2, that $f\,(tu + (1-t)\,v) = f\,(\overline{x})$, so Ω is convex.
■

The next is a known convex analysis result. For the sake of completeness, we prove it.

Lemma 3.6 *Let be given f, verifying hypothesis 3, and a convex set $\Omega \subseteq R^n$. Suppose that f is constant in Ω. Then ∇f is also constant in Ω.*

Proof. The theorem is trivial if Ω has an unique point. Therefore, we can admit that it has more than one element. Now, we consider a fact derived from the convex analysis. Denote by $\mathrm{ri}(\Omega)$ the relative interior of the set Ω, then

$$\text{If } \overline{x} \in \mathrm{ri}\,(\Omega)\,,\text{ then } \nabla f\,(\overline{x})^T\,(y - \overline{x}) = 0, \text{ for all } y \in \Omega.$$

Indeed, for $y \in \Omega$ and $\overline{x} \in \mathrm{ri}(\Omega)$, there exists some $\varepsilon > 0$ such that $(1-t)\,\overline{x} + ty \in M$, for all $t \in [-\varepsilon, 1]$. Let's set $g\,(t) := f\,((1-t)\,\overline{x} + ty)$. Clearly, g is constant in $[-\varepsilon, 1]$, and so, for any $t \in (-\varepsilon, 1]$, it is true that

$$0 = g^{'}(t) = \nabla f\left((1-t)\,\overline{x} + ty\right)^T (y - \overline{x})\,.$$

Then, particularly, for $t = 0$, we achieve the aimed result: $\nabla f(\overline{x})^T (y - \overline{x}) = 0$.

Next, we fix $\overline{x} \in$ri(Ω), and define the function $h(x) := f(x) - \nabla f(\overline{x})^T (x - \overline{x}) = 0$. h is a C^1 convex function, and $\nabla h(x) = \nabla f(x) - \nabla f(\overline{x})$, so $\nabla h(\overline{x}) = 0$. Use again the convexity and the C^1 property of h, to see the optimality condition given by last equality, that is, we have shown that $\overline{x}$ is a global minimizer to h. On the other hand, taking $y \in \Omega$, we have $h(y) = f(y) - \nabla f(\overline{x})^T (y - \overline{x})$. Due to the property shown above, $h(y) = f(y) = f(\overline{x}) = h(\overline{x})$, therefore any element of Ω is a global minimizer to h. Then $\nabla h(y) = 0$, for all $y \in \Omega$, that is $\nabla f(y) = \nabla f(\overline{x})$, for all $y \in \Omega$, which is the aimed result. ■

Lemma 3.7 *Let be given $x \in \Omega$. Then, $s(x) = \overline{s} = s(\overline{x})$.*

Proof. From 2.2, we have $s(x) = \left(I - A^T \left(AX^r A^T\right)^{-1} AX^r\right) \nabla f(x)$. As in Ω, $\nabla f(x) = \nabla f(\overline{x}) = \overline{s} + A^T \overline{y}$, it follows that

$$s(x) = \left(I - A^T \left(AX^r A^T\right)^{-1} AX^r\right) \left(\overline{s} + A^T \overline{y}\right).$$

Notice that the facts that $x \in \Omega$ implies $x_N = 0$, and, $\overline{s}_B = 0$ lead to $X^r \overline{s} = 0$. Therefore, developing the previous expression for $s(x)$, we get

$$\begin{aligned} s(x) = \overline{s} + A^T \overline{y} - A^T \left(AX^r A^T\right)^{-1} AX^r \overline{s} - A^T \left(AX^r A^T\right)^{-1} AX^r A^T \overline{y} = \\ = \overline{s} + A^T \overline{y} - A^T \overline{y} = \overline{s}, \end{aligned}$$

as wanted. ■

We are going to show the existence of a dilation Ω_δ from Ω, such that

$$\operatorname{sign}\left(d_i(x)\right) = \operatorname{sign}\left(\overline{s}_i\right), \text{ for all } i \in N.$$

We need the following parameter: $\theta := 1/2 \min_{i \in N} |\overline{s}_i|$.

Lemma 3.8 *There exists $\delta > 0$ such that, for all $x \in \Omega_\delta$, the following is true:*

(1) $\|s(x) - \overline{s}\| \le \theta$

(2) *If $x_N > 0$, then $s_i(x)\,\overline{s}_i > 0$, and $d_i(x)\,\overline{s}_i > 0$, for all $i \in N$.*

Proof. (1) As the optimization problem is nondegenerate, $s(x)$ is uniformly continuous in the compact Ω_1, so implying the existence of some positive δ, such that, for x and y in Ω_1, and verifying $\|x-y\| \leq \delta$, it holds that $\|s(x)-s(y)\| \leq \theta$, where using was made of the θ definition. Thus, if $x \in \Omega_\delta$, there exists some $y \in \Omega$ that satisfies the inequality $\|x-y\| \leq \delta$, and, consequently, $\|s(x)-s(y)\| \leq \theta$. However, s is constant in Ω, that is, $s(y) = s(\overline{x}) = \overline{s}$, which gives the aimed result: $\|s(x)-\overline{s}\| \leq 1/2 \min_{i\in N} |\overline{s}_i|$.

(2) If $i \in N$, the previous result allows to write

$$|s_i(x) - \overline{s}_i| \leq \|s(x) - \overline{s}\| \, 1/2 \min_{j\in N} |\overline{s}_j| \leq 1/2\, |\overline{s}_i|,$$

this meaning that $s_i(x)\,\overline{s}_i > 0$, which is the first result to achieve. Now, from 2.3, $d_i(x) = x_i^r s_i(x)$, so $d_i(x)\,\overline{s}_i > 0$, that finishes the proof. ■

In order to obtain the positivity of $\overline{s}_N$, we, first, will show that the accumulation points of the sequence that are not in Ω , are outside Ω_δ, too. After that, as already anticipated, we will see that, if some entry of $\overline{s}_N$ is negative, the sequence will possess a limit point in Ω, and a limit point out of Ω_δ, contradicting corollary 3.4.

Lemma 3.9 *Let $\widetilde{x}$ a limit point to $\{x^k\}$. Then $\widetilde{x} \in \Omega$ or $\widetilde{x} \notin \Omega_\delta$.*

Proof. Suppose that the limit point $\widetilde{x}$ fulfils the following contradiction hypothesis: $\widetilde{x} \in \Omega_\delta$, $\widetilde{x} \notin \Omega$, and $\widetilde{s} = s(\widetilde{x})$. The convergence of $\{f(x^k)\}$ leads to $f(\widetilde{x}) = f(\overline{x})$, where $\overline{x}$ is the accumulation point we have fixed. Now, by hypothesis, $\widetilde{x} \notin \Omega$, then, there exists some $i \in N$ such that $\widetilde{x}_i > 0$. Additionally, the last lemma says that $\widetilde{s}_i\overline{s}_i > 0$, therefore, $\widetilde{s} \neq 0$. Thus, we have $\widetilde{x}_i\widetilde{s}_i \neq 0$, contradicting the complementarity condition. ■

Lemma 3.10 *Let $\overline{s} = s(\overline{x})$ possess some negative entry. Then $\{x^k\}$ has an accumulation point $\widetilde{x} \notin \Omega_\delta$.*

Proof. First observe that if all accumulation points are in Ω_δ, they were forcefully in Ω, which is contained in the relative interior of Ω_δ. That is a trivial consequence of the previous lemma. We have more: in that case, the whole sequence would be contained in Ω_δ, from some iteration, let's say, k_0. Now, suppose that there exists $i \in N$ such that $(\overline{s}_N)_i < 0$, and all accumulation points of $\{x^k\}$ are in Ω_δ, so in Ω. Then, for $k \geq k_0$, it would be true that $x^k \in \Omega_\delta$ and $s\left(x^k\right)_i \overline{s}_i > 0$, which implies $s\left(x^k\right)_i < 0$, and $\left(d^k\right)_i < 0$, due to lemma 3.8. In that case, it holds that

$$x_i^{k+1} = x_i^k - t^k d_i^k \geq x_i^k \geq x_i^{k_0} > 0,$$

from which we get the nondecreasing property of $\left\{x_i^k\right\}$. But, that contradicts the fact that $\liminf x_i^k = 0$. ■

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

Lemma 3.11 *For any accumulation point of the algorithm iterations, it holds that* $\overline{s}_N > 0$.

Proof. We know that $\overline{s}_B = 0$. By contradiction, suppose that there exists some $i \in N$, for which $(\overline{s}_N)_i < 0$. We are going to prove, as a consequence of that hypothesis, that exists a subsequence $\left\{x^{k_j}\right\}$ from $\left\{x^k\right\}$, which is convergent to some $u \in \Omega$, but the displaced sequence $\left\{x^{k_j+1}\right\}$ converges to some $v \in \Omega_\delta$, thus contradicting Corollary 3.4.

According to Lemma 3.10, we know that if exists some $i \in N$, for which $(\overline{s}_N)_i < 0$, the sequence $\left\{x^k\right\}$ has an accumulation point outside Ω_δ, and another, let's call it $\overline{x}$, in Ω. One concludes immediately, also using the closure property of Ω_δ, that there exists a subsequence $\left\{x^{k_j}\right\}$, such that $x^{k_j} \in \Omega_\delta$ and $x^{k_j+1} \notin \Omega_\delta$. To simplify notations, we can suppose that those sequences are convergent. Now, from Lemma 3.9, it follows that, for some $u \in \Omega$, and $v \notin \mathrm{ri}(\Omega_\delta)$, $x^{k_j} \to u$, and $x^{k_j+1} \to v$. But that is equivalent to say that $\lim_{k\to\infty} \left\|x^{k_j+1} - x^{k_j}\right\| \geq \delta$, and that is in opposition to Corollary 3.4. ■

We present the main theorem, which is a mere outcome of the foregone results.

Theorem 3.12 *Let be* $\left\{x^k\right\}$ *the sequence generated by the r-projection algorithm. Suppose that hypothesis 1, 2 and 3 are verified. Then*

(1) *Any accumulation point of* $\left\{x^k\right\}$ *solves 1.1;*

(2) $\lim_{k\to\infty} \left(x^{k+1} - x^k\right) = 0$.

Proof. (1) follows from Lemmas 3.2 and 3.11, and (2) is Corollary 3.4. ■

3.3 The r-algorithm for positivity constraints

The *r-algorithm* is easily adaptable to the problem 1.1, without the equality constraints, that is,

$$\begin{aligned} &\min f(x) \\ &\text{s.to } x \geq 0. \end{aligned}$$

Analyzing the formulas 2.1, 2.2 and 2.3, for, respectively, $y(x)$, $s(x)$ and $d(x)$, $y(x)$ is eliminated, and $s(x)$ and $d(x)$ simplify to

$$\begin{aligned} s(x) &= \nabla f(x), \\ d(x) &= X^r s(x) \end{aligned}$$

Observe that, in the framework established in the background section, we are minimizing in the whole space R^n_{++}, associated to the metrics generated by X^{-r}, so

$$d(x) = \nabla_M f(x) = X^r \nabla f(x) \in T_x M = R^n.$$

Now, the algorithm rewrites:

Given $x^o \in R^n_{++}$, $\beta > 0$ and $0 < \delta < 1$; set $k = 0$

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

About the convergence:

As both algorithms have the same structure, all the proofs can be easily adjusted for that simpler case. All the results we obtained are valid.

4 Conclusions

We have proposed a new class of projection algorithms and have shown the so called weak convergence. The link between that class and some known methods were established. The research directions of our current interest are extension of the geometric ideas to primal-dual framework, and the analysis of the influence of the parameter r.

References

[1] D. A. Bayer and J. C. Lagarias, The Nonlinear Geometry of Linear Programming I, Affine and Projective Scaling Trajectories, *Transactions of the American Mathematical Society*, vol 314, No 2, 499-526, August 1989.

[2] D. A. Bayer and J. C. Lagarias, The Nonlinear Geometry of Linear Programming II, Legendre Transform Coordinates and Central Trajectories, *Transactions of the American Mathematical Society*, vol. 314, No 2, 527-581, August 1989.

[3] J. X. da Cruz Neto, L. L. Lima and P. R. Oliveira, Geodesic Algorithm in Riemannian Manifold, *Balkan Journal of Geometry and Applications,* vol. 3, 2, 89-100, 1998.

[4] D. den Hertog, "Interior Point Approach to Linear, Quadratic and Convex Programming - Algorithms and Complexity ", Doctoral Thesis, Delft University, 1992.

[5] I. I. Dikin, Iterative Solution of Problems of Linear and Quadratic Programming, *Soviet Mathematics Doklady,* 8: 647-675, 1967.

[6] M. P. do Carmo, "Riemannian Geometry ", Academic Press, 1991.

[7] P. P. B. Eggermont, Multiplicative Iterative Algorithms for Convex Program,*Linear Algebra and its Applications*, 130: 25-42, 1990.

[8] L. Faybusovich, Hamiltonian Structure of Dynamical Systems which Solves Linear Programming Problems, *Physica D*, vol. 53, 217-232, 1991.

[9] O. P. Ferreira and P. R. Oliveira, Subgradient Algorithm on Riemannian Manifold, *Journal of Optimization Theory and Applications,* vol. 97, 1, 93-104.

[10] O. P. Ferreira and P. R. Oliveira, Proximal Point Algorithm on Riemannian Manifolds, *Optimization,* vol. 51, accepted.

[11] D. Gabay, Minimizing a Differentiable Function over a Differential Manifold, *Journal of Optimization Theory and Applications*, vol. 37, No2, 177-219, 1982.

[12] C. C. Gonzaga and L. A. Carlos, A Primal Affine-Scaling Algorithm for Linearly Constrained Convex Program, Tech. Report *ES* -230/90, *PESC/COPPE-Federal University of Rio de Janeiro*, 1990.

[13] A. N. Iussem, An Interior Point Multiplicative Method for Optimization Under Positivity Constraints,*Acta Mathematicae*, 38, 163-184, 1995.

[14] N. Karmarkar, Riemannian Geometry Underlying Interior-Point Methods for Linear Programming, *Contemporary Mathematics,* vol. 114, 51-75, 1990.

[15] D. G. Luenberger, The Gradient Projection Method Along Geodesics, *Management Science* 18, 620-631, 1972.

[16] Y. E. Nesterov and M. Todd, On the Riemannian Geometry Defined by self-Concordant Barriers and Interior-Point Methods, pre-print.

[17] G. L. Oliveira and P. R. Oliveira, A New Class of Interior-Point Methods for Optimization Under Positivity Constraints, Tech. Rep. *ES* 570/01, *PESC/COPPE-Federal University of Rio de Janeiro,* 2002.

[18] T. Tapcsák, "Smooth Nonlinear Optimization in R^n ", Kluwer Academic Publishers, Dordrecht, Netherlands.

[19] T. Rapcsák and T. T. Thang, Nonlinear Coordinate Representations of Smooth Optimization Problems, *Journal of Optimization Theory and Applications,*, vol. 86, 459-489, 1995.

[20] T. Rapcsák and T. T. Thang, A Class of Polynomial Variable Metric Algorithm for Linear Optimization, *Mathematical Programming*, 74, 319-331, 1996.

[21] R. Saigal, The Primal Power Affine Scaling Method, Tech, Report, No 93-21,*Dep. Ind. and Oper. Eng., University of Michigan, 1993.*

[22] G. J. P. Silva and P. R. Oliveira, A New Class of Proximal Algorithms for the Nonlinear Complementarity Problem, Tech. Rep. *ES* 560/01, *PESC/COPPE-Federal University of Rio de Janeiro,* 2001.

[23] C. Udriste, "Convex Functions and Optimization Methods in Riemannian Geometry ", Kluwer Academic Publishers, 1994.