Interior Proximal Algorithm with Variable Metric for Second-Order Cone Programming: Applications to Structural Optimization and Support Vector Machines

Felipe Alvarez, Julio López and Héctor Ramírez C. *

Abstract

In this work, we propose an inexact interior proximal type algorithm for solving convex second-order cone programs. This kind of problems consists of minimizing a convex function (possibly nonsmooth) over the intersection of an affine linear space with the Cartesian product of second-order cones. The proposed algorithm uses a distance variable metric, which is induced by a class of positive definite matrices, and an appropriate choice of regularization parameter. This choice ensures the well-definedness of the proximal algorithm and forces the iterates to belong to the interior of the feasible set. Also, under suitable assumptions, it is proven that each limit point of the sequence generated by the algorithm solves the problem. Finally, computational results applied to structural optimization and support vector machines are presented.

Keywords: Proximal method, second-order cone programming, variable metric, structural optimization, multiple loads problem, support vector machines.

1 Introduction

In this paper, we consider the following convex *second-order cone programming* (SOCP) problem

$$\text{(SOCP)} \qquad f_* = \min_{x\in\mathbb{R}^n} f(x); \quad \mathbf{B}x = \mathbf{d}, \quad w^j(x) = A^j x + b^j \in \mathcal{L}_+^{m_j},\ j = 1,\dots,J,$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function (possibly nonsmooth), $\mathbf{B}$ is a full rank $r \times n$ real matrix with $r \leq n$, $\mathbf{d} \in \mathbb{R}^r$, A^j are full rank $m_j \times n$ real matrices, $b^j \in \mathbb{R}^{m_j}$, $j = 1,\dots,J$. For an integer $m \geq 2$, the set $\mathcal{L}_+^m$ denotes the second-order cone (SOC) (also called the Lorentz cone or ice-cream cone) of dimension m defined as $\mathcal{L}_+^m = \{y = (y_1, \bar{y}) \in \mathbb{R} \times \mathbb{R}^{m-1} : \ \|\bar{y}\| \leq y_1\}$, where $\|\cdot\|$ denotes the Euclidean norm. Since the norm is not differentiable at 0, (SOCP) is not in the class of smooth convex programs. On the other hand, a Lorentz cone can be rewritten as the smooth nonconvex constraint $\mathcal{L}_+^m = \{y \in \mathbb{R}^m : y_2^2 + \dots + y_m^2 \leq y_1^2,\ y_1 \geq 0\}$. However, this constraint is not qualified at 0 (see, for instance, [9, Definition 3.20]).

In recent years, SOCP have received considerable attention because of its wide range of applications in engineering, control and robust optimization (see for instance [26, 2, 33] and the references therein). It is known that $\mathcal{L}_+^m$, like $\mathbb{R}_+^m$ and the cone $\mathcal{S}_+^m$ of $m \times m$ real symmetric positive semidefinite matrices, belongs to the class of symmetric cones to which a Jordan algebra may be associated [16]. Using this connection, interior-point methods have been developed for solving linear programs with SOC constraints [26, 38].

**Departamento de Ingeniería Matemática, Centro de Modelamiento Matemático(CNRS UMI 2807), FCFM, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile (falvarez@dim.uchile.cl, jclopez@dim.uchile.cl, hramirez@dim.uchile.cl).

In this work, we propose an inexact interior *proximal point algorithm* (PPA) with variable metric for solving a convex SOCP whose objective function is not required to be smooth. The standard (PPA) was introduced by Martinet [27] based on previous work by Moreau [28], and it was then further developed and studied by Rockafellar [37] for the problem of finding zeros of a maximal monotone operator. Later, several authors [12, 13, 15] have generalized (PPA) for convex programming with nonnegative constrains, replacing the quadratic regularization term by a Bregman distance or ϕ-divergence distance. Recently, Auslender and Teboulle [6] have dealt with general types of constraints, including SOC and Semidefinite ones, via a unified proximal distance framework. In all these works, the pseudo-distances are used to force the iterates to stay in the interior of the feasible set.

The idea of PPA with variable metric was originally studied by Quian [31] for monotone operators, and by Bonnans et al. [8]. for convex programming (see also [24]). Since then, this idea has been exploited in different articles [10, 11]. In [34], Souza et al. considered the matrix $H(x)$ =diag$(x_1^{-r}, \ldots, x_n^{-r})$, $r \geq 2$, in order to define a variable metric on $\mathbb{R}^n_+$: $\langle \cdot, \cdot \rangle^H_x$, for all $x \in \mathbb{R}^n_{++}$. They defined a new class of variable metric interior proximal point algorithm for the minimization of a continuous proper convex function on $\mathbb{R}^n_+$. This algorithm uses a regularization parameter appropriately chosen so that the iterates be interior points. Moreover, the convergence to a Karush-Khun-Tucker (KKT) point is obtained.

The purpose of this paper is to define an inexact interior proximal type algorithm for solving convex SOCP problems, where the used metric is induced by a general class of positive definite matrices.The outline of this paper is as follows. In Section 2, we recall some basic notions and properties associated with SOC. In Section 3, we present our algorithm with variable metric and prove its convergence properties. In Section 4, we present the notion of quasi-nonincreasing metrics and we prove the convergence of our method under some suitably chosen assumptions. In section 5, we describe the case of the metric induced by the Hessian of the spectral logarithm, which is not covered by the analysis in section 4. Finally, in Section 6 we consider two different applications of Linear SOCP, we discuss MATLAB implementations of the proposed algorithms and we present some computational experiments; this is an intermediate step toward more general and possibly nonsmooth convex problems, which are not addressed in this paper from the numerical point of view.

Notation

The following notation is used throughout the paper. For a closed proper convex function f, its effective domain is defined by dom$f = \{x : f(x) < +\infty\}$ and ∂f denote its subdifferential [36]. The superscript $\top$ denotes transpose operator and I_d denotes the identity matrix in $\mathbb{R}^{d\times d}$. For a symmetric matrix M, we denote its smallest and largest eigenvalues by $\boldsymbol{\lambda}_{\min}(M)$ and $\boldsymbol{\lambda}_{\max}(M)$, respectively. Given a matrix $A \in \mathbb{R}^{p\times q}$, the smallest and largest singular value of A will be denoted by $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$, respectively. If we have a finite number of matrices $A^1, \ldots, A^J$ such that each $A^j \in \mathbb{R}^{m_j \times n}$, we define $\boldsymbol{\sigma}_{min}(\mathbf{A}) = \min\{\sigma_{\min}(A^j) : j = 1, \ldots, J\}$ and $\mathbf{A} := (A^1; \ldots; A^J) \in \mathbb{R}^{q\times n}$ whose rows are those of A^1 to A^J, where $q = \sum_{j=1}^J m_j$. We also denote by $\mathcal{K} := \mathcal{L}^{m_1}_+ \times \ldots \times \mathcal{L}^{m_J}_+$. The set $\mathcal{L}^m_{++} = \{y = (y_1, \bar{y}) \in \mathbb{R} \times \mathbb{R}^{m-1} : \|\bar{y}\| < y_1\}$ is the interior of the second-order cone $\mathcal{K} = \mathcal{L}^m_+$ and the set $\partial\mathcal{L}^m_+ = \{y \in \mathcal{L}^m_+ : y_1 = \|\bar{y}\|\}$ denotes its boundary. We denote by $\mathcal{X}^*$ the optimal solution set of (SOCP). Finally, we define by $\mathbf{w}(x) := (w^1(x), \ldots, w^J(x)) \in \mathbb{R}^q$, where $w^j(x) = A^j x + b^j$ for $j = 1, \ldots, J$.

2 Algebra preliminaries

In this section, we recall some basic concepts and properties about the Jordan algebra associated with the second-order cone $\mathcal{L}^m_+$ with $m \geq 2$ that are needed for this work (see [16] for more details). For any $v = (v_1, \bar{v})$, $w = (w_1, \bar{w}) \in \mathbb{R} \times \mathbb{R}^{m-1}$, the *Jordan product* for $\mathcal{L}^m_+$ is defined by

$v \circ w = (v^\top w, v_1\bar{w} + w_1\bar{v})$. This product can be equivalently written as

$$v \circ w = \mathrm{Arw}(v)w, \tag{2.1}$$

where

$$\mathrm{Arw}(v) := \begin{pmatrix} v_1 & \bar{v}^\top \\ \bar{v} & v_1 I_{m-1} \end{pmatrix}$$

is the *arrow* matrix of vector v. It is clear that the product $\circ$ is commutative, and it follows easily from (2.1) that $(v, w) \mapsto v \circ w$ is a bilinear function where the element $e = (1, 0, \ldots, 0) \in \mathbb{R}^m$ plays the role of unit element for this algebra. On the other hand, the cone $\mathcal{L}^m_+$ is not closed under the product $\circ$. This product is not associative in general but it is power associative, that is, for all $w \in \mathbb{R}^m$, w^k can be unambiguously defined as $w^k = w^p \circ w^q \quad \forall p, q \in \mathbb{N};\ p + q = k$. If $w \in \mathcal{L}^m_+$, then there exists a unique vector in $\mathcal{L}^m_+$, which we denote by $w^{1/2}$, such that $(w^{1/2})^2 = w^{1/2} \circ w^{1/2} = w$.

We next introduce the *spectral factorization* of vectors in $\mathbb{R}^m$ associated with $\mathcal{L}^m_+$. For any $w = (w_1, \bar{w}) \in \mathbb{R} \times \mathbb{R}^{m-1}$, we can decompose w as

$$w = \lambda_1(w)u_1(w) + \lambda_2(w)u_2(w), \tag{2.2}$$

where $\lambda_1(w), \lambda_2(w)$ and $u_1(w), u_2(w)$ are the *spectral values* and *spectral vectors* of w given by

$$\lambda_i(w) = w_1 + (-1)^i \|\bar{w}\|, \tag{2.3}$$

$$u_i(w) = \begin{cases} \frac{1}{2}(1, (-1)^i \frac{\bar{w}}{\|\bar{w}\|}) & , \ \text{if } \bar{w} \neq 0, \\ \frac{1}{2}(1, (-1)^i \bar{v}) & , \ \text{if } \bar{w} = 0, \end{cases} \tag{2.4}$$

for $i = 1, 2$ with $\bar{v}$ being any unit vector in $\mathbb{R}^{m-1}$ (satisfying $\|\bar{v}\| = 1$). If $\bar{w} \neq 0$ the decomposition (2.2) is unique. Notice that $\lambda_1(w) \le \lambda_2(w)$ and the vectors $u_i(w)$, $i = 1, 2$, belong to $\partial\mathcal{L}^m_+$. In the sequel, we denote by $\lambda_{\min}(w) = \lambda_1(w)$, $\lambda_{\max}(w) = \lambda_2(w)$. For each $w = (w_1, \bar{w}) \in \mathbb{R} \times \mathbb{R}^{m-1}$, the trace and determinant of w with respect to $\mathcal{L}^m_+$ are defined as

$$\mathrm{tr}(w) = \lambda_{\min}(w) + \lambda_{\max}(w) = 2w_1\,; \det(w) = \lambda_{\min}(w)\lambda_{\max}(w) = w_1^2 - \|\bar{w}\|^2. \tag{2.5}$$

These definitions can be viewed as the analogues of the trace and determinant of matrices. Note that, in order to avoid any misleading, the smallest and largest eigenvalue of a symmetric matrix M has been denoted by bold symbols $\boldsymbol{\lambda}_{\min}(M)$ and $\boldsymbol{\lambda}_{\max}(M)$, respectively.

A vector $w = (w_1, \bar{w}) \in \mathbb{R} \times \mathbb{R}^{m-1}$ is said to be nonsingular if $\det(w) \neq 0$. If w is nonsingular, then there exists a unique $v = (v_1, \bar{v}) \in \mathbb{R} \times \mathbb{R}^{m-1}$ such that $w \circ v = v \circ w = e$. We call this v the inverse of w and denote it by w^{-1}. Direct calculations yields

$$w^{-1} = \frac{1}{w_1^2 - \|\bar{w}\|^2}(w_1, -\bar{w}) = \frac{1}{\det(w)}(\mathrm{tr}(w)e - w). \tag{2.6}$$

The spectral decomposition entails some basic properties, which are summarized below (see, for instance, [16, 17]).

Proposition 2.1. *For any $w = (w_1, \bar{w}) \in \mathbb{R} \times \mathbb{R}^{m-1}$ the spectral values $\lambda_{\min}(w), \lambda_{\max}(w)$ and spectral vectors $u_1(w), u_2(w)$ given by (2.3) and (2.4), have the following properties:*
a) $u_1(w)$ and $u_2(w)$ are orthogonal for the Jordan product: $u_1(w) \circ u_2(w) = 0$, and $\|u_1(w)\| = \|u_2(w)\| = \frac{1}{\sqrt{2}}$.
b) $u_1(w)$ and $u_2(w)$ are idempotent for the Jordan product: $u_i(w) \circ u_i(w) = u_i(w)$, for $i = 1, 2$.
c) $\lambda_{\min}(w), \lambda_{\max}(w)$ are nonnegative (resp. positive) if and only if $w \in \mathcal{L}^m_+$ (resp. $w \in \mathcal{L}^m_{++}$).
d) The Euclidean norm of w can be represented as $\|w\|^2 = \frac{1}{2}(\lambda_{\min}(w)^2 + \lambda_{\max}(w)^2)$.

The next result provides some interesting inequalities; for a proof see, for instance, [4, Proposition 3.1].

Proposition 2.2. *Let $v, w \in \mathbb{R}^m$, then $\lambda_{\min}(v) + \lambda_{\min}(w) \leq \lambda_{\min}(v+w) \leq \lambda_{\min}(v) + \lambda_{\max}(w)$, and $\lambda_{\max}(v) + \lambda_{\min}(w) \leq \lambda_{\max}(v+w) \leq \lambda_{\max}(v) + \lambda_{\max}(w)$.*

Following [25], for any function $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ we consider the *spectrally defined* function $\Phi_g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ given by

$$\Phi_g(w) = g(\lambda_{\min}(w)) + g(\lambda_{\max}(w)), \tag{2.7}$$

if $\lambda_{\min}(w), \lambda_{\max}(w) \in \operatorname{dom}(g)$ and $\Phi_g(w) = +\infty$ otherwise. Clearly $\Phi_g(w) = \operatorname{tr}(g^{soc}(w))$ if $\lambda_{\min}(w), \lambda_{\max}(w) \in \operatorname{dom}(g)$, where $g^{soc}(w) = g(\lambda_{\min}(w))u_1(w) + g(\lambda_{\max}(w))u_2(w)$, for all $w \in \mathbb{R} \times \mathbb{R}^{m-1}$ is the corresponding SOC function (see [30, 17] for more details). If for example, we consider the *logarithm barrier function* $g(t) = -\ln(t), \quad \operatorname{dom}(g) = \mathbb{R}_{++}, \quad g(t) = +\infty$ otherwise, then its function spectrally defined is given by

$$\Phi_g(w) = -\ln(w_1^2 - \|\bar{w}\|^2) = -\ln(\det(w)) \ \text{ if } w \in \mathcal{L}^m_{++};\ +\infty \text{ otherwise.}$$

Proposition 2.3. *([30, lemma 2.10] and [17, proposition 5.2]) Let g be continuously differentiable on $int(\operatorname{dom}(g))$. Then*

(a) Φ_g is continuously differentiable on $int(\operatorname{dom}(\Phi_g))$ and $\nabla \Phi_g(w) = 2(g')^{soc}(w)$, $\forall w \in \mathcal{L}^m_{++}$.

(b) The Hessian of Φ_g at $w \in \mathcal{L}^m_{++}$ is given by the formula $\nabla^2 \Phi_g(w) = 2g''(\bar{w})I$ if $\bar{w} = 0$, and otherwise is given by

$$\nabla^2 \Phi(w) = 2 \begin{pmatrix} b & c\bar{w}^\top/\|\bar{w}\| \\ c\bar{w}/\|\bar{w}\| & aI_{m-1} + (b-a)\bar{w}\bar{w}^\top/\|\bar{w}\|^2 \end{pmatrix}, \quad \bar{x}_2 \neq 0$$

where $a = \frac{g'(\lambda_2) - g'(\lambda_1)}{\lambda_2 - \lambda_1}$, $b = \frac{g''(\lambda_1) + g''(\lambda_2)}{2}$, $c = \frac{g''(\lambda_2) - g''(\lambda_1)}{2}$. If $g''(t) > 0$ for all $t \in \operatorname{dom}(g)$, then $\nabla^2 \Phi_g(z)$ is positive definite for all $z \in \operatorname{dom}(\Phi(g))$.

If in particular, we consider the logarithm barrier, we get $\nabla \Phi_g(w) = -2w^{-1}$, $w \in \mathcal{L}^m_{++}$. Also, we have an explicit expression for the Hessian of Φ_g in $w \in \mathcal{L}^m_{++}$ given by $\nabla^2 \Phi_g(w) = 2(\mathcal{Q}_w)^{-1}$, where

$$\mathcal{Q}_w = \begin{pmatrix} \|w\|^2 & 2w_1\bar{w}^\top \\ 2w_1\bar{w} & \det(w)I_{m-1} + 2\bar{w}\bar{w}^\top \end{pmatrix}. \tag{2.8}$$

As $g''(t) > 0$, it follows that $\mathcal{Q}_w$ is positive definite $\forall w \in \mathcal{L}^m_{++}$. The matrix $\mathcal{Q}_w$ is called the *quadratic representation* of w, which exists for any $w \in \mathbb{R}^m$. The following theorem gives some interesting properties of $\mathcal{Q}_w$ (see e.g. [2, Theorem 3, Theorem 9]), which will be used later on.

Theorem 2.4. *Let $w \in \mathbb{R}^m$ be arbitrary.*

1. *If w is decomposed as in (2.2) then $\lambda^2_{\min}(w)$ and $\lambda^2_{\max}(w)$ are eigenvalues of $\mathcal{Q}_w$. Furthermore, if $\lambda_{\min}(w) \neq \lambda_{\max}(w)$ then each one has multiplicity 1. In addition, $\det(w)$ is an eigenvalue of $\mathcal{Q}_w$, and has multiplicity $m-2$ when w is nonsingular and $\lambda_{\min}(w) \neq \lambda_{\max}(w)$.*
2. *If w is nonsingular, then $\mathcal{Q}_w(\mathcal{L}^m_+) = \mathcal{L}^m_+$; likewise, $\mathcal{Q}_w(\mathcal{L}^m_{++}) = \mathcal{L}^m_{++}$.*

From this theorem, one has in particular that $\mathcal{Q}_w$ is nonsingular if and only if w is nonsingular.

Lemma 2.5. *Let $w \in \mathcal{L}^m_+$. Then, there exists a matrix $\mathcal{Q}_{w^{1/2}}$ which maps e to w (that is $\mathcal{Q}_{w^{1/2}}e = w$), given explicitly by*

$$\mathcal{Q}_{w^{1/2}} = \begin{pmatrix} w_1 & \bar{w}^\top \\ \bar{w} & \det(w)^{1/2}I_{m-1} + \frac{\bar{w}\bar{w}^\top}{\det(w)^{1/2} + w_1} \end{pmatrix}. \tag{2.9}$$

This matrix is positive semidefinite and satisfies that $\mathcal{Q}_{w^{1/2}} = \mathcal{Q}_w^{1/2}$. Moreover, when $w \in \mathcal{L}^m_{++}$, $\mathcal{Q}_{w^{1/2}}$ turns out to be a positive definite matrix.

If in addition $\bar{w} \neq 0$, then the matrix $\mathcal{Q}_{w^{1/2}}$ can be written as follows:

$$\mathcal{Q}_{w^{1/2}} = Arw(w) - \begin{pmatrix} 0 & 0 \\ 0 & (w_1 - \det(w)^{1/2})\left(I - \frac{\bar{w}\bar{w}^\top}{\|\bar{w}\|^2}\right) \end{pmatrix}. \tag{2.10}$$

Proof. This result is obtained from [38, Proposition 2.1]. □

We end this section by establishing a technical lemma that will be useful in the sequel of the paper.

Lemma 2.6. *For any $x \in \mathbb{R}^m$, $x \in \mathcal{L}^m_{++}$ if and only if $\langle x, y\rangle > 0$, $\forall y \in \mathcal{L}^m_+$, $y \neq 0$.*

Proof. For any $x = (x_1, \bar{x}) \in \mathcal{L}^m_{++}$ and $y = (y_1, \bar{y}) \in \mathcal{L}^m_+$ with $y \neq 0$, we know that $\|\bar{x}\| < x_1$ and $\|\bar{y}\| \leq y_1$. Then $\langle x, y\rangle = x_1 y_1 + \bar{x}^\top \bar{y} \geq x_1 y_1 - \|\bar{x}\|\,\|\bar{y}\| \geq x_1 y_1 - \|\bar{x}\|\, y_1 = y_1(x_1 - \|\bar{x}\|) > 0$,where the first inequality follows from the Cauchy-Schwartz inequality. Now, we suppose that $\langle x, y\rangle > 0$, $\forall y \in \mathcal{L}^m_+$ with $y \neq 0$. Taking $y = e$ we deduce that $x_1 > 0$. If $\bar{x} = 0$, the result follows. On the other hand, if $\bar{x} \neq 0$, we set $y = (1, -\frac{\bar{x}}{\|\bar{x}\|})$. It is clear that $y \in \mathcal{L}^m_+$ and $y \neq 0$. Hence, $0 < \langle x, y\rangle = x_1 - \|\bar{x}\| = \lambda_{\min}(x)$. This means that $x \in \mathcal{L}^m_{++}$. □

3 Proximal Algorithm using Variable Metric

Let $\mathcal{F} = \{x \in \mathbb{R}^n : w^j(x) = A^j x + b^j \in \mathcal{L}^{m_j}_{++},\ j = 1, \ldots, J\}$, $\mathcal{B} = \{x \in \mathbb{R}^n : \mathbf{B}x = \mathbf{d}\}$ and $C = \mathcal{B} \cap \mathcal{F}$. The feasible set of (SOCP) is $\bar{C}$, the closure of C in $\mathbb{R}^n$. From now on we suppose that the following assumptions hold true:

(A1) $f_* > -\infty$.

(A2) $\mathrm{dom} f \cap C \neq \emptyset$ (Slater's condition).

3.1 Algorithm PAVM

We denote by $\mathbf{M} = \mathrm{diag}(M^1, \ldots, M^J)$ a block diagonal matrix with $M^j \in \mathbb{R}^{m_j \times m_j}$ being symmetric and positive definite for each $j = 1, \ldots, J$. We suppose that $\mathbf{A}$ has rank n. Set $\langle \cdot, \cdot\rangle_{\mathbf{M}} := \langle \mathbf{A}^\top \mathbf{M}\mathbf{A}\cdot, \cdot\rangle$, and let us define the following induced norms $\|u\|^2_{\mathbf{M}} := \langle u, u\rangle_{\mathbf{M}} = \langle \mathbf{M}\mathbf{A}u, \mathbf{A}u\rangle$ and $\|u\|^{*2}_{\mathbf{M}} := \langle (\mathbf{A}^\top \mathbf{M}\mathbf{A})^{-1} u, u\rangle$, $\forall u \in \mathbb{R}^n$.

The Proximal Algorithm with Variable Metric (PAVM) for solving the problem (SOCP) is defined as follows:

For each $k = 1, 2, \ldots$, take $\delta_k > 0$ and $\eta_k > 0$ with $\sum_{k=1}^\infty \delta_k < \infty$ and $\sum_{k=1}^\infty \eta_k < \infty$.

Step 0: Start with some initial point $x^0 \in C$, $g^0 \in \partial f(x^0)$ and block diagonal matrix $\mathbf{M}_0$. Set $k = 0$

Step 1: Given $x^k \in C$, $g^k \in \partial f(x^k)$ and an appropriate matrix $\mathbf{M}_k$ and suitable parameter $\gamma_k > 0$, find x^{k+1}, $g^{k+1} \in \mathbb{R}^n$ and $\omega^{k+1} \in \mathbb{R}^r$ such that

$$g^{k+1} \in \partial f(x^{k+1}), \tag{3.1}$$

$$g^{k+1} + \gamma_k \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k) + \mathbf{B}^\top \omega^{k+1} = \epsilon^{k+1}, \tag{3.2}$$

$$\mathbf{B}x^{k+1} = \mathbf{d}, \tag{3.3}$$

where the associated error ϵ^{k+1} satisfies the following conditions:

$$\|\epsilon^{k+1}\| \leq \delta_k, \quad \|\epsilon^{k+1}\| \max(\|x^{k+1}\|, \|x^k\|) \leq \eta_k. \tag{3.4}$$

Step 2: If x^{k+1} satisfies a prescribed stopping rule, then stop.

Step 3: Update $\mathbf{M}_{k+1}$. Replace k by $k+1$ and go to step 1.

Remark 3.1. We remark that this algorithm is a generic numerical method since we do not specify a procedure for computing the next iterate x^{k+1}. In particular one may compute x^{k+1} by using the *bundle method* or by applying some iterations of a standard descent method for the unconstrained minimization of the strongly convex function $F_k(x) := f(x) + \frac{1}{2}\gamma_k\|x - x^k\|^2_{\mathbf{M_k}}$. Moreover, since f is a closed proper convex function, it directly follows that F_k has bounded sublevel sets. Therefore, the optimal set of $\inf\{F_k(x) : \mathbf{B}x = \mathbf{d}\}$ is nonempty and compact. Hence (3.1)-(3.4) hold with $\epsilon^{k+1} = 0$, and thus the sequence generated by PAVM is well defined.

Remark 3.2. The second condition on $\{\epsilon^k\}$ in (3.4), is similar to IPA1 in [6]. This is motivated by an implementable inexact minimization of the auxiliary function $F_k(x) := f(x) + \frac{1}{2}\gamma_k\|x - x^k\|^2_{\mathbf{M_k}}$.

3.2 Interior point iterates

In this section we prove that an appropriate choice of γ_k in the PAVM guarantees that the iterates x^k remain in C for all $k \geq 1$ (see Proposition 3.1 below), obtaining thus an *interior* proximal point algorithm. To this end, we need to introduce a piece of notation. The largest eigenvalue of $\mathbf{M} = \mathrm{diag}(M^1, \ldots, M^J)$ is given by $\boldsymbol{\lambda}_{\max}(\mathbf{M}) = \max\{\boldsymbol{\lambda}_{\max}(M^j) : j = 1, \ldots, J\}$. Moreover, for any element $\mathbf{z} \in \mathbb{R}^m$, we denote by $\mathbf{Q_z} = \mathrm{diag}(\mathcal{Q}_{z_1}, \ldots, \mathcal{Q}_{z_J})$, where $\mathcal{Q}_{z_j} \in \mathbb{R}^{m_j \times m_j}$ is defined in (2.8). Thus

$$\boldsymbol{\lambda}_{\max}(\mathbf{Q_z}) = \max\{\boldsymbol{\lambda}_{\max}(\mathcal{Q}_{z_j}) : j = 1, \ldots, J\}. \tag{3.5}$$

So, when $z \in \mathcal{L}^m_{++}$, Part 1 of Theorem 2.4 implies that $\boldsymbol{\lambda}_{\max}(\mathcal{Q}_z) = \lambda^2_{\max}(z)$, obtaining then

$$\boldsymbol{\lambda}_{\max}(\mathbf{Q_z}) = \max\{\lambda^2_{\max}(z_j) : j = 1, \ldots, J\}. \tag{3.6}$$

Finally, we denote by $\boldsymbol{\lambda}_{\max}(\mathbf{Q}_z^{-1}\mathbf{M}^{-1}) = \max\{\boldsymbol{\lambda}_{\max}(\mathcal{Q}_{z_j}^{-1/2}{M^j}^{-1}\mathcal{Q}_{z_j}^{-1/2}) : j = 1, \ldots, J\}$. Similar definitions can be stated for the smallest eigenvalue $\boldsymbol{\lambda}_{\min}(\cdot)$. These last relations help us to prove our first result related to our PAVM algorithm.

Proposition 3.1. *Suppose that for every $k = 0, 1, \ldots,$ the parameter γ_k satisfies*

$$\gamma_k > \sqrt{2}(\boldsymbol{\sigma}_{min}(\mathbf{A}))^{-1}\boldsymbol{\lambda}_{\max}(\mathbf{Q}_{\mathbf{w}(x^k)})^{1/2}\boldsymbol{\lambda}_{\max}(\mathbf{Q}^{-1}_{\mathbf{w}(x^k)}\mathbf{M}_k^{-1})[\|g^k\| + \delta_k] \tag{3.7}$$

Then the sequence $\{x^k\}$ generated by PAVM is contained in C.

Proof. By induction. This is true for $k = 0$ by step 0. Now, assume that $x^k \in C$. By construction, x^{k+1} satisfies $\mathbf{B}x^{k+1} = \mathbf{d}$. On the other hand, from the monotonicity of ∂f, it follows that $\langle g^{k+1} - g^k, x^{k+1} - x^k\rangle \geq 0$, which together with (3.2) yields to

$$\langle \gamma_k \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k) + \mathbf{B}^\top\omega^{k+1}, x^{k+1} - x^k\rangle \leq \langle g^k, x^k - x^{k+1}\rangle + \langle \epsilon^{k+1}, x^{k+1} - x^k\rangle.$$

From (3.3) and the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} \gamma_k\langle \mathbf{M}_k\mathbf{A}(x^{k+1} - x^k), \mathbf{A}(x^{k+1} - x^k)\rangle &\leq \|g^k\|\|x^k - x^{k+1}\| + \|\epsilon^{k+1}\|\|x^{k+1} - x^k\| \\ &\leq [\|g^k\| + \delta_k]\|x^{k+1} - x^k\|, \end{aligned} \tag{3.8}$$

where we used (3.4). As each M^j_k is a positive definite matrix, we have that

$$\begin{aligned} &\langle \mathbf{M}_k\mathbf{A}(x^{k+1} - x^k), \mathbf{A}(x^{k+1} - x^k)\rangle \\ &\qquad = \sum_{j=1}^J \langle \mathcal{Q}^{1/2}_{w^j(x^k)} M^j_k A^j(x^{k+1} - x^k), \mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\rangle \\ &\qquad \geq \sum_{j=1}^J \boldsymbol{\lambda}_{\min}(\mathcal{Q}^{1/2}_{w^j(x^k)} M^j_k \mathcal{Q}^{1/2}_{w^j(x^k)})\|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\|^2. \end{aligned} \tag{3.9}$$

By Lemma 2.5, $\mathcal{Q}^{1/2}_{w^j(x^k)}$ is positive definite. Thus $\|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\| \geq \boldsymbol{\lambda}_{\min}(\mathcal{Q}^{-1/2}_{w^j(x^k)})\|A^j(x^{k+1} - x^k)\| = \boldsymbol{\lambda}_{\min}(\mathcal{Q}^{-1}_{w^j(x^k)})^{1/2}\|A^j(x^{k+1} - x^k)\|$. Using this lower bound once in (3.9) it follows that $\langle \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k), \mathbf{A}(x^{k+1} - x^k)\rangle \geq \sum_{j=1}^J \boldsymbol{\lambda}_{\min}(\mathcal{Q}^{-1}_{w^j(x^k)})^{1/2} \boldsymbol{\lambda}_{\min}(\mathcal{Q}^{1/2}_{w^j(x^k)} M^j_k \mathcal{Q}^{1/2}_{w^j(x^k)})\|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\|\|A^j(x^{k+1} - x^k)\|$. From (3.8) together with (3.5) and the well-known property $\boldsymbol{\lambda}_{\max}(M^{-1}) = \boldsymbol{\lambda}_{\min}(M)^{-1}$ for any symmetric nonsingular matrix M, it follows that

$$\begin{aligned} \gamma_k \sum_{j=1}^J \|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\|\|A^j(x^{k+1} - x^k)\| \\ \leq \boldsymbol{\lambda}_{\max}(\mathbf{Q}_{\mathbf{w}(x^k)})^{1/2} \boldsymbol{\lambda}_{\max}(\mathbf{Q}^{-1}_{\mathbf{w}(x^k)} \mathbf{M}_k^{-1})[\|g^k\| + \delta_k]\|x^{k+1} - x^k\|. \end{aligned} \tag{3.10}$$

Moreover, since each A^j is full rank , we get

$$\|A^j(x^{k+1} - x^k)\| \geq \frac{1}{\|A^{j\dagger}\|_{spec}}\|x^{k+1} - x^k\|,$$

where $A^{j\dagger}$ denotes the pseudoinverse of Moore-Penrose of A^j and $\|A\|_{spec} = \sigma_{\max}(A)$ denotes the spectral norm of a given matrix A. By the identity $\sigma_{\max}(A^{j\dagger}) = (\sigma_{\min}(A^j))^{-1}$ (see for instance [21], page 421, exercise 7), we get from (3.10) that $\gamma_k \sum_{j=1}^J \sigma_{\min}(A^j)\|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\| \leq \boldsymbol{\lambda}_{\max}(\mathbf{Q}_{\mathbf{w}(x^k)})^{1/2} \boldsymbol{\lambda}_{\max}(\mathbf{Q}^{-1}_{\mathbf{w}(x^k)} \mathbf{M}_k^{-1})[\|g^k\| + \delta_k]$, which implies that

$$\begin{aligned} \sum_{j=1}^J \|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\| \\ \leq \frac{1}{\gamma_k}(\boldsymbol{\sigma}_{min}(\mathbf{A}))^{-1} \boldsymbol{\lambda}_{\max}(\mathbf{Q}_{\mathbf{w}(x^k)})^{1/2} \boldsymbol{\lambda}_{\max}(\mathbf{Q}^{-1}_{\mathbf{w}(x^k)} \mathbf{M}_k^{-1})[\|g^k\| + \delta_k] \\ < \frac{1}{\sqrt{2}}. \end{aligned}$$

For the last inequality we have used (3.7).

On the other hand, it holds from Lemma 2.5 that $\mathcal{Q}^{-1/2}_{w^j(x^k)} w^j(x^k) = e_j$, which yields $\|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\| = \|\mathcal{Q}^{-1/2}_{w^j(x^k)}(w^j(x^{k+1}) - w^j(x^k))\| = \|\mathcal{Q}^{-1/2}_{w^j(x^k)} w^j(x^{k+1}) - e_j\|$, and by virtue of Proposition 2.1(d) it follows that $\|\mathcal{Q}^{-1/2}_{w^j(x^k)} A^j(x^{k+1} - x^k)\| \geq \frac{1}{\sqrt{2}}|\lambda_{\min}(\mathcal{Q}^{-1/2}_{w^j(x^k)} w^j(x^{k+1}) - e_j)|$ for all $j = 1, \ldots, J$. Therefore, for each $j = 1, \ldots, J$, we get

$$|\lambda_{\min}(\mathcal{Q}^{-1/2}_{w^j(x^k)} w^j(x^{k+1}) - e_j)| < 1,$$

which implies that $-1 < \lambda_{\min}(\mathcal{Q}^{-1/2}_{w^j(x^k)} w^j(x^{k+1}) - e_j) < 1; \quad \forall j = 1, \ldots, J$. By using Weyl's Theorem (cf. Proposition 2.2) in both inequalities, we get $0 < \lambda_{\min}(\mathcal{Q}^{-1/2}_{w^j(x^k)} w^j(x^{k+1})) < 2$, $\forall j = 1, \ldots, J$. This implies that $\mathcal{Q}^{-1/2}_{w^j(x^k)} w^j(x^{k+1}) \in \mathcal{L}^{m_j}_{++}$, that is, $w^j(x^{k+1}) \in \mathcal{Q}_{w^j(x^k)^{1/2}}(\mathcal{L}^{m_j}_{++})$ for all $j = 1, \ldots, J$. Therefore, by Theorem 2.4 and (3.3), it follows that $x^{k+1} \in C$, and the proof is complete. □

Remark 3.3. Note that the matrix $\mathbf{M}_k$ defines the shape of the level curves of the variable metric considered in our PAVM Algorithm while the regularization parameter γ_k decides indirectly the step length of the next iterate taking into account this choice of $\mathbf{M}_k$. Indeed, if for instance we rescale our metric by using $\alpha\mathbf{M}_k$, for some $\alpha > 0$, instead of $\mathbf{M}_k$, we can multiply the right hand side of (3.7) by α and the equation (3.2) of Step 1 of our algorithm remains unchanged.

3.3 Boundedness and some related results

Let us recall a technical lemma which will be useful in the sequel (see [29]).

Lemma 3.2. (i) *Let* $\{v_k\}$ *and* $\{\alpha_k\}$ *be nonnegative real sequences satisfying* $v_{k+1} \leq v_k + \alpha_k$ *for* $\sum \alpha_k < \infty$. *Then the sequence* $\{v_k\}$ *converges.*
(ii) *Let* $\{\lambda_k\}$ *be a sequence of positive numbers,* $\{a_k\}$ *a real sequence and* $b_n = \sigma_n^{-1} \sum_{k=0}^{n} \lambda_k a_k$, *where* $\sigma_n = \sum_{k=0}^{n} \lambda_k$. *If* $\sigma_n \to \infty$, *one has* $\liminf a_n \leq \liminf b_n \leq \limsup b_n \leq \limsup a_n$.

Proposition 3.3. *Let* $\{x^k\} \subset C$ *be a sequence generated by PAVM under* (3.7). *Then the following hold:*

- *(i)* $\{f(x^k)\}$ *converges.*
- *(ii) If* $\mathcal{X}^*$ *is nonempty and bounded, then the sequence* $\{x^k\}$ *is bounded.*
- *(iii)* $\sum\limits_{k=0}^{\infty} \left(\gamma_k \sum_{j=1}^{J} \|x^{k+1} - x^k\|^2_{M_k^j}\right) < \infty$.

Proof. *(i)* From (3.2)-(3.3), and since $g^{k+1} \in \partial f(x^{k+1})$ we have $f(x^k) + \langle \epsilon^{k+1}, x^{k+1} - x^k \rangle \geq f(x^{k+1}) + \gamma_k \sum_{j=1}^{J} \|x^{k+1} - x^k\|^2_{M_k^j} \geq f(x^{k+1})$. By (3.4), and using $\langle \epsilon^{k+1}, x^{k+1} - x^k \rangle \leq \|\epsilon^{k+1}\|(\|x^k\| + \|x^{k+1}\|) \leq 2\|\epsilon^{k+1}\| \max(\|x^{k+1}\|, \|x^k\|)$, we obtain

$$f(x^{k+1}) + \gamma_k \sum_{j=1}^{J} \|x^{k+1} - x^k\|^2_{M_k^j} \leq f(x^k) + 2\eta_k. \tag{3.11}$$

Thus $0 \leq f(x^{k+1}) - f_* \leq f(x^k) - f_* + 2\eta_k$. Hence, using Lemma 3.2(i) we deduce that the sequence $\{f(x^k)\}$ converges.
(ii) Summing (3.11) over $k = 0, \dots, l$, one has $f(x^{l+1}) - f(x^0) \leq 2\sum_{k=0}^{l} \eta_k$. Since $\sum_{k=1}^{\infty} \eta_k$ exists, it follows that for some $\bar{\eta} \geq 0$ we have $f(x^{l+1}) \leq f(x^0) + 2\bar{\eta} < \infty$, for all $l \geq 0$. As $\mathcal{X}^*$ is bounded, f is level bounded over $\overline{C}$. Thus, one has that $\{x^k\}$ is a bounded sequence.
(iii) From (3.11) we get $\sum_{k=0}^{N} \left(\gamma_k \sum_{j=1}^{J} \|x^{k+1} - x^k\|^2_{M_k^j}\right) \leq f(x^0) - f(x^{N+1}) + 2\sum_{k=0}^{N} \eta_k \leq f(x^0) - f_* + 2\sum_{k=1}^{N+1} \eta_k$. Letting $N \to +\infty$, we obtain the desired result. □

Remark 3.4. As consequence of above proposition, it follows that $\{g^k\}$ is bounded when the function f is defined everywhere.

The next result gives is similar to [13, Lemma 3.2].

Lemma 3.4. *Let* $\{x^k\}$ *be a sequence generated by (PAVM). Then for all* $x \in \bar{C} \cap \operatorname{dom} f$ *the following inequality holds*

$$\frac{2}{\gamma_k}(f(x^{k+1}) - f(x)) \leq \|x - x^k\|^2_{\mathbf{M}_k} - \|x - x^{k+1}\|^2_{\mathbf{M}_k} - \|x^{k+1} - x^k\|^2_{\mathbf{M}_k} + \frac{2}{\gamma_k}\langle \epsilon^{k+1}, x^{k+1} - x \rangle.$$

Proof. For any $x \in \bar{C}$, because $g^{k+1} \in \partial f(x^{k+1})$, we have $f(x^{k+1}) + \langle g^{k+1}, x - x^{k+1} \rangle \leq f(x)$. Using (3.2)-(3.3) and the inequality above, we get

$$f(x^{k+1}) - f(x) \leq \langle \epsilon^{k+1}, x^{k+1} - x \rangle - \gamma_k \langle \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k), x^{k+1} - x \rangle. \tag{3.12}$$

Since $\mathbf{M}_k$ is symmetric, we have $\|x - x^k\|^2_{\mathbf{M}_k} = \|x - x^{k+1}\|^2_{\mathbf{M}_k} + \|x^{k+1} - x^k\|^2_{\mathbf{M}_k} + 2\langle \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k), x - x^{k+1} \rangle$. Then the result follows directly from (3.12). □

4 Quasi-nonincreasing metrics

We consider the following hypotheses on matrices M_k^j:

(H-i) The sequences $\{M_k^{j^{-1}}\}$ are bounded, for each $j = 1, \ldots, J$.

(H-ii) For each $j = 1, \ldots, J$, there exists a nonnegative sequence $\{\nu_k^j\}$ such that $(M_k^j - M_{k+1}^j + \nu_k^j I) \in \mathcal{S}_+^{m_j}$ and $\sum_{k=1}^{\infty} \nu_k^j < \infty$.

Remark 4.1. *Since each M_k^j is positive definite,* **(H-i)** *is equivalent to saying that there exists a $\underline{\eta}_j > 0$ such that $\boldsymbol{\lambda}_{\min}(M_k^j) > \underline{\eta}_j$, for all $k \in \mathbb{N}$ and $j = 1, \ldots, J$.*

Remark 4.2. *Note that* **(H-ii)** *implies that sequences $\{M_k^j\}$ are bounded.*

Lemma 4.1. *Let $\{x^k\}$ be a sequence generated by the PAVM under* (3.7) *for some $\beta_k \geq \underline{\beta} > 0$. Assume that* **(H-$i$)** *holds. Then $\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty$.*

Proof. By using the facts that each M_k^j is positive definite and each A^j is full rank, one has $\|x^{k+1} - x^k\|^2_{M_k^j} \geq \boldsymbol{\lambda}_{\min}(M_k^j)\|A^j(x^{k+1} - x^k)\|^2 \geq \boldsymbol{\lambda}_{\min}(M_k^j)\sigma_{\min}(A^j)^2\|x^{k+1} - x^k\|^2$, whence $\|x^{k+1} - x^k\|^2_{\mathbf{M}_k} = \sum_{j=1}^{J} \|x^{k+1} - x^k\|^2_{M_k^j} \geq \sum_{j=1}^{J} \boldsymbol{\lambda}_{\min}(M_k^j)\sigma_{\min}(A^j)^2\|x^{k+1} - x^k\|^2$. Now, by the boundedness of the sequence $\{M_k^j\}$ for each $j = 1, \ldots, J$, there exists $\underline{\eta}_j > 0$ such that $\boldsymbol{\lambda}_{\min}(M_k^j) > \underline{\eta}_j$, for all $j = 1, \ldots, J$. Taking $\underline{\eta} = \underline{\beta} \min_{j=1,\ldots,J} \underline{\eta}_j \sigma_{\min}(A^j)^2$, we obtain $\sum\limits_{k=0}^{\infty} \left(\gamma_k \sum_{j=1}^{J} \|x^{k+1} - x^k\|^2_{M_k^j} \right) \geq J\underline{\eta} \sum\limits_{k=0}^{\infty} \|x^{k+1} - x^k\|^2$, and the result follows from Proposition 3.3(iii). □

Remark 4.3. As consequence of this result, it follows that $\lim\limits_{k \to +\infty} \|x^{k+1} - x^k\| = 0$.

From now on we define: $\sigma_n = \sum\limits_{j=0}^{n-1} \gamma_j^{-1}$, for all $n \in \mathbb{N}$.

Lemma 4.2. *Let $\{x^k\}$ be the sequence generated by algorithm (PAVM). Assume that* **(H-ii)** *holds. For any $x \in \bar{C} \cap \operatorname{dom} f$ the following hold:*

$$\begin{aligned} -2\sigma_n f(x) + \sum_{k=0}^{n-1} \frac{2}{\gamma_k} f(x^{k+1}) \quad &\leq \quad \|x - x^0\|^2_{\mathbf{M}_0} - \|x^n - x\|^2_{\mathbf{M}_n} - \sum_{k=0}^{n-1} \big(\|x^{k+1} - x^k\|^2_{\mathbf{M}_k} \\ &\quad - \frac{2}{\gamma_k}\langle \epsilon^{k+1}, x^{k+1} - x \rangle - \sum_{j=1}^{J} \nu_k^j \|A^j(x^{k+1} - x)\|^2 \big). \end{aligned} \tag{4.1}$$

Proof. Since $(M_k^j - M_{k+1}^j + \nu_k^j I) \in \mathcal{S}_+^{m_j}$, one has $\|x^{k+1} - x\|^2_{M_k^j} + \nu_k^j\|A(x^{k+1} - x)\|^2 \geq \|x^{k+1} - x\|^2_{M_{k+1}^j}$. By using this inequality in the estimate of Lemma 3.4, we have

$$\begin{aligned} \frac{2}{\gamma_k}(f(x^{k+1}) - f(x)) \quad &\leq \quad \|x - x^k\|^2_{\mathbf{M}_k} - \|x^{k+1} - x\|^2_{\mathbf{M}_{k+1}} - \|x^{k+1} - x^k\|^2_{\mathbf{M}_k} + \\ &\quad \frac{2}{\gamma_k}\langle \epsilon^{k+1}, x^{k+1} - x \rangle + \sum_{j=1}^{J} \nu_k^j \|A^j(x^{k+1} - x)\|^2. \end{aligned}$$

Summing for $k = 0, \ldots, n-1$ the result follows immediately. □

Theorem 4.3. *Let $\{x^k\}$ be the sequence generated by algorithm (PAVM) under*

$$\gamma_k \geq \sqrt{2}(\boldsymbol{\sigma}_{min}(\mathbf{A}))^{-1}\boldsymbol{\lambda}_{\max}(\mathbf{Q}_{\mathbf{w}(x^k)})^{1/2}\boldsymbol{\lambda}_{\max}(\mathbf{Q}^{-1}_{\mathbf{w}(x^k)}\mathbf{M}_k^{-1})[\|g^k\| + \delta_k] + \beta_k \tag{4.2}$$

for some $\beta_k \geq \underline{\beta} > 0$. Assume that **(H-ii)** *holds and that $\mathcal{X}^*$ is nonempty and bounded. If*

$$\lim_{n\to\infty} \sigma_n = +\infty,$$

then the following hold:

(*i*) *The sequence $\{f(x^k)\}$ converge to f_*.*

(*ii*) *The limit points of $\{x^k\}$ belong to $\mathcal{X}^*$.*

(*iii*) *The sequence $\{\|x^k - u\|^2_{\mathbf{M}_k}\}$ converge for all $u \in \mathcal{X}^*$.*

(*iv*) *Furthermore, if* **(H-i)** *holds then $\{x^k\}$ converges to some $x^* \in \mathcal{X}^*$.*

Proof. (*i*) Let $\theta_{k+1}(x) = \langle \epsilon^{k+1}, x^{k+1} - x\rangle$ and $\vartheta^j_{k+1}(x) = \nu^j_k\|A^j(x^{k+1} - x)\|^2$ for $j = 1,\ldots,J$. Using (3.4), one has $\theta_{k+1}(x) \leq \varrho_{k+1}$ where $\varrho_{k+1} = \eta_k + \|x\|\delta_k$, which satisfies $\sum_{k=0}^{\infty}\theta_{k+1}(x) < \infty$ and as $\gamma_k \geq \underline{\beta}$, $\sum_{k=0}^{\infty}\theta_{k+1}(x)\gamma_k^{-1} < \infty$. On the other hand, by boundedness of $\{x^{k+1}\}$ (see Proposition 3.3), there exists $\tau > 0$ such that $\|A^j(x^{k+1} - x)\| \leq \sigma_{\max}(A^j)(\tau + \|x\|)$, for all $j = 1,\ldots,J$ and therefore $\sum_{k=0}^{\infty}\sum_{j=1}^{J}\vartheta^j_{k+1}(x) < \infty$. Then, dividing (4.1) by σ_n and invoking to Lemma 3.2(ii), we get from (4.1) that $\liminf f(x^n) \leq f(x)$ for each $x \in \bar{C}$ so that $\liminf f(x^n) \leq f_*$, which together with the fact that $f(x^n) \geq f_*$ implies that $\liminf f(x^n) = f_*$. Hence, using the Proposition 3.3 it follows that the sequence $\{f(x^k)\}$ converges to f_*.
(*ii*) From Proposition 3.3 we have that $\{x^k\}$ is bounded. Since f is lsc, passing to the limit and reminding that $\{x^k\} \subset C$, it follows that each limit point is an optimal solution.
(*iii*) For all $u \in \mathcal{X}^*$, from above inequality we obtain

$$\|x^{k+1} - u\|^2_{\mathbf{M}_{k+1}} \leq \|u - x^k\|^2_{\mathbf{M_k}} + \frac{2}{\gamma_k}\langle \epsilon^{k+1}, x^{k+1} - u\rangle + \sum_{j=1}^{J}\nu^j_k\|A^j(x^{k+1} - x)\|^2.$$

By part (*i*), we get

$$\|x^{k+1} - u\|^2_{\mathbf{M}_{k+1}} \leq \|u - x^k\|^2_{\mathbf{M_k}} + \frac{2}{\underline{\beta}}\varrho_{k+1} + \sum_{j=1}^{J}\sigma_{\max}(A^j)^2(\tau + \|u\|)^2\nu^j_k.$$

Then, from the nonnegativity of $\|x^k - u\|^2_{\mathbf{M}_k}$, we can apply Lemma 3.2 for establishing the convergence of $\|x^k - u\|^2_{\mathbf{M}_k}$ for all $u \in \mathcal{X}^*$.
(*iv*) From part (*iii*) we obtain that the sequences $\{\|x^k - u\|_{M^j_k}\}$ converge to some $c(u) \in \mathbb{R}^+$, $\forall u \in \mathcal{X}^*$ and for each $j = 1,\ldots,J$. Let x^∞ be a limit point of $\{x^k\}$. Take a subsequence $\{x^{k_i}\}$ of $\{x^k\}$ such that $x^{k_i} \to x^\infty \in \mathcal{X}^*$ (by (*ii*)). From hypothesis (**H-ii**), $\{M^j_k\}$ is bounded, for each $j = 1,\ldots,J$. Passing onto a subsequence, if necessary, we can suppose that $M^j_{k_i} \to \overline{M}^j$, for each $j = 1,\ldots,J$. Then $\|x^{k_i} - x^\infty\|^2_{M^j_{k_i}} \to 0$. So that $c(x^\infty) = 0$. Moreover, since $\|x^k - x^\infty\|^2_{M^j_k} \geq \lambda_{\min}(M^j_k)\sigma_{\min}(A^j)^2\|x^k - x^\infty\|^2$ and **(H-i)** holds true, we get that $x^k \to x^\infty$. ☐

The following result yields a global rate of convergence estimate, which is similar to the one obtained for proximal-type algorithms in convex minimization problems.

Proposition 4.4. *Let $\{x^k\}$ be the sequence generated by (PAVM). Assume hypotheses* **(H)** *hold and that $\mathcal{X} \neq \emptyset$. Then, there exists $\tau > 0$ such that for all $u \in \mathcal{X}^*$, we have*

$$\begin{aligned} f(x^n) - f(u) \quad \leq \quad & \frac{\|u - x^0\|^2_{\mathbf{M}_0} - \|u - x^n\|^2_{\mathbf{M}_n}}{2\sigma_n} - \frac{1}{2\sigma_n}\sum_{k=0}^{n-1} \gamma_k(\sigma_k + \sigma_{k+1})\|x^{k+1} - x^k\|^2_{\mathbf{M}_k} + \\ & \frac{1}{2\sigma_n}\sum_{k=0}^{n-1}\left(\frac{2\eta_k + \|u\|\delta_k}{\gamma_k} + 4\sigma_k\eta_{k+1} + \sum_{j=1}^{J} \nu_k^j \sigma^2_{\max}(A^j)(\tau + \|u\|)^2\right). \end{aligned} \tag{4.3}$$

Proof. Let $u \in \mathcal{X}^*$. Setting x^k for x in (3.12), multiplying the resulting inequality by σ_k and using the fact that $\sigma_{k+1} = \frac{1}{\gamma_k} + \sigma_k$ (with $\sigma_0 = 0$), we get

$$\sigma_{k+1} f(x^{k+1}) - \sigma_k f(x^k) - \frac{1}{\gamma_k} f(x^{k+1}) \leq \sigma_k \langle \epsilon^{k+1}, x^{k+1} - x^k\rangle - \sigma_k\gamma_k\|x^{k+1} - x^k\|^2_{\mathbf{M}_k}.$$

Summing the last inequality over $k = 0, \ldots, n-1$, noting $\sigma_0 = 0$ and using (3.11), one has

$$\sigma_n f(x^n) - \sum_{k=0}^{n-1} \frac{1}{\gamma_k} f(x^{k+1}) \leq 2\sum_{k=0}^{n-1} \sigma_k\eta_k - \sum_{k=1}^{n-1} \sigma_k\gamma_k\|x^{k+1} - x^k\|^2_{\mathbf{M}_k}. \tag{4.4}$$

Adding twice (4.4) to (4.1), we have

$$\begin{aligned} 2\sigma_n(f(x^n) - f(u)) \leq \; & \|u - x^0\|^2_{\mathbf{M}_0} - \|u - x^n\|^2_{\mathbf{M}_n} - \sum_{k=0}^{n-1} \|x^{k+1} - x^k\|^2_{\mathbf{M}_k} \\ & -2\sum_{k=0}^{n-1} \sigma_k\gamma_k\|x^{k+1} - x^k\|^2_{\mathbf{M}_k} \\ & + \sum_{k=0}^{n-1}\left(\frac{2}{\gamma_k}\langle \epsilon^{k+1}, x^{k+1} - u\rangle + 4\sigma_k\eta_{k+1} + \sum_{j=1}^{J} \nu_k^j\|A^j(x^{k+1} - u)\|^2\right). \end{aligned}$$

Because $\langle \epsilon^{k+1}, x^{k+1} - u\rangle \leq \eta_k + \|u\|\delta_k$, $\|A^j(x^{k+1} - u)\| \leq \sigma_{\max}(A^j)(\tau + \|u\|)$, for some $\tau > 0$, the above inequality can be written as

$$\begin{aligned} 2\sigma_n(f(x^n) - f(u)) \quad \leq \quad & \|u - x^0\|^2_{\mathbf{M}_0} - \|u - x^n\|^2_{\mathbf{M}_n} - \sum_{k=0}^{n-1} \gamma_k(\sigma_k + \sigma_{k+1})\|x^{k+1} - x^k\|^2_{\mathbf{M}_k} \\ & + \sum_{k=0}^{n-1}\left(\frac{2(\eta_k + \|u\|\delta_k)}{\gamma_k} + 4\sigma_k\eta_{k+1} + \sum_{j=1}^{J} \nu_k^j \sigma^2_{\max}(A^j)(\tau + \|u\|)^2\right). \end{aligned}$$

Dividing by $2\sigma_n$, we get the desired inequality. □

Remark 4.4. *Ignoring the negative terms in the estimate of proposition above we obtain*

$$f(x^n) - f(u) \leq \frac{\|u - x^0\|^2_{\mathbf{M}_0}}{2\sigma_n} + \frac{1}{\sigma_n}\sum_{k=0}^{n-1}\left(\frac{\eta_k + \|u\|\delta_k}{\gamma_k} + 2\sigma_k\eta_{k+1} + \sum_{j=1}^{J} \frac{\nu_k^j\sigma^2_{\max}(A^j)}{2}(\tau + \|u\|)^2\right).$$

The following result is an extension direct of [19, Theorem 3.1] to our algorithm. For simplicity we suppose that $\eta_k = \|\epsilon^k\| = 0$, $\forall k \geq 1$.

Theorem 4.5. *Let $\{x^k\}$ be the sequence generated by PAVM algorithm under (4.2) for some $\beta_k \geq \underline{\beta} > 0$. Assume that hypotheses* **(H)** *hold, that $\mathcal{X}^*$ is nonempty and bounded and that $\eta_k = \|\epsilon^k\| = 0$, $\forall k \geq 1$. If $\lim_{n\to\infty} \sigma_n = +\infty$ then $\sigma_n(f(x_n) - f^*) \to 0$.*

Proof. By Theorem 4.3-(iv), x^k converges to some $x^* \in \mathcal{X}^*$. We denote by $\zeta_k = f(x^k) - f(x^*)$. From (3.11) we have

$$\zeta_k - \zeta_{k+1} = f(x^k) - f(x^{k+1}) \geq \gamma_k \|x^{k+1} - x^k\|^2_{\mathbf{M}_k}. \tag{4.5}$$

Setting x^* for x in (3.12), we obtain

$$\begin{aligned} \zeta_{k+1} = f(x^{k+1}) - f(x^*) &\leq -\gamma_k \langle \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k), x^{k+1} - x^* \rangle \\ &= -\gamma_k \langle \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k), x^k - x^* \rangle - \gamma_k \|x^{k+1} - x^k\|^2_{\mathbf{M}_k} \\ &\leq -\gamma_k \langle \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k), x^k - x^* \rangle. \end{aligned}$$

But $|\langle \mathbf{A}^\top \mathbf{M}_k \mathbf{A}(x^{k+1} - x^k), x^k - x^* \rangle| \leq \|x^{k+1} - x^k\|_{\mathbf{M}_k} \|x^* - x^k\|_{\mathbf{M}_k}$, so from the inequality above one has $\zeta_{k+1} \leq \gamma_k \|x^{k+1} - x^k\|_{\mathbf{M}_k} \|x^* - x^k\|_{\mathbf{M}_k}$ or equivalently $\|x^{k+1} - x^k\|_{\mathbf{M}_k} \geq \frac{\zeta_{k+1}}{\gamma_k \|x^* - x^k\|_{\mathbf{M}_k}}$. Using this inequality in (4.5), we have $\zeta_k \geq \zeta_{k+1} + \frac{\zeta^2_{k+1}}{\gamma_k} (\|x^* - x^k\|^2_{\mathbf{M}_k})^{-1} = \zeta_{k+1} \left(1 + \frac{\zeta_{k+1}}{\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k}}\right)$, whence

$$\zeta_k^{-1} \leq \zeta_{k+1}^{-1} \left(1 + \frac{\zeta_{k+1}}{\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k}}\right)^{-1}. \tag{4.6}$$

On the other hand, setting x^* for x in the estimate of Lemma 3.4 we obtain

$$f(x^{k+1}) \leq f(x^{k+1}) + \frac{\gamma_k}{2} \|x^{k+1} - x^k\|^2_{\mathbf{M}_k} \leq f(x^*) + \frac{\gamma_k}{2} \|x^* - x^k\|^2_{\mathbf{M}_k},$$

which yields to $0 \leq \frac{\zeta_{k+1}}{\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k}} \leq \frac{1}{2}$. Moreover, the function $(1+t)^{-1}$ is convex for $t > -1$, hence $(1+t)^{-1} \leq 1 - \frac{2}{3} t$, for $t \in [0, \frac{1}{2}]$. This last inequality together with (4.6) implies that $\zeta_k^{-1} \leq \zeta_{k+1}^{-1} \left(1 - \frac{2}{3} \frac{\zeta_{k+1}}{\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k}}\right) = \zeta_{k+1}^{-1} - \frac{2}{3} \frac{1}{\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k}}$. Summing this for $k = 0, \ldots, n-1$, we get $\zeta_n^{-1} \geq \zeta_n^{-1} - \zeta_0^{-1} \geq \frac{2}{3} \sum_{k=0}^{n-1} \frac{1}{\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k}}$, obtaining $\zeta_n = f(x^n) - f(x^*) \leq \frac{3}{2} \frac{1}{\sum_{k=0}^{n-1} (\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k})^{-1}}$. Multiplying this inequality by σ_n gives $\sigma_n (f(x^n) - f(x^*)) \leq \frac{3}{2} \frac{1}{\sigma_n^{-1} \sum_{k=0}^{n-1} (\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k})^{-1}}$. By Theorem 4.3-($iv$), we have that $\|x^* - x^k\|^{-1}_{\mathbf{M}_k} \to \infty$. Therefore, using the Silverman-Toeplitz Theorem (see for instance [23, pag. 76]), the series $\sigma_n^{-1} \sum_{k=0}^{n-1} (\gamma_k \|x^* - x^k\|^2_{\mathbf{M}_k})^{-1} \to \infty$ also. In consequence, the results follows. □

5 Metric induced by the Hessian of the spectral Log

In this section, we will consider

$$\mathbf{M}_k = 2\mathbf{Q}^{-1}_{\mathbf{w}(x^k)}$$

a block diagonal matrix, where each block is given by the inverses of the $m_j \times m_j$ matrix $\mathcal{Q}_{w^j(x)}$, defined in (2.8). This choice is a natural extension to the case SOC of the algorithm proposed by Souza et al. [34]. Notice that (4.2) reduces to

$$\gamma_k \geq \frac{\sqrt{2}}{2} (\boldsymbol{\sigma}_{min}(A))^{-1} \boldsymbol{\lambda}_{\max}(\mathbf{Q}_{\mathbf{w}(x^k)})^{1/2} (\|g^k\| + \delta_k) + \beta_k. \tag{5.1}$$

From now on we suppose that f is defined in all $\mathbb{R}^n$. Lemma 4.2 and Theorem 4.3 do not apply here because (**H-ii**) fails. Souza et al. show the convergence of their algorithm following the ideas of [22, Theorem 2] relying on some componentwise comparison arguments. Such an argument is not valid for spectral values, and the convergence is an open problem in general. However, we are able to establish the convergence for the specific case of a linear objective function.

5.1 Algorithm PAVM-Log

For each $k = 1, 2, \dots$, let $\beta_k > 0$, $\delta_k > 0$ and $\eta_k > 0$ with $\beta_k \in (\underline{\beta}, 1)$ where $\underline{\beta} > 0$, $\sum \delta_k < \infty$ and $\sum \eta_k < \infty$.

Step 0: Start with some initial point $x^0 \in C$. Set $k = 0$

Step 1: Given $x^k \in C$, $g^k \in \partial f(x^k)$ and γ_k satisfying (5.1), solve

$$g^{k+1} \in \partial f(x^{k+1}), \tag{5.2}$$
$$g^{k+1} + 2\gamma_k \mathbf{A}^\top \mathbf{Q}^{-1}_{\mathbf{w}(x^k)} \mathbf{A}(x^{k+1} - x^k) + \mathbf{B}^\top \omega^{k+1} = \epsilon^{k+1}, \tag{5.3}$$
$$\mathbf{B}x^{k+1} = \mathbf{d}, \tag{5.4}$$

for some $\omega^{k+1} \in \mathbb{R}^r$, where

$$\|\epsilon^{k+1}\| \le \delta_k, \quad \|\epsilon^{k+1}\| \max(\|x^{k+1}\|, \|x^k\|) \le \eta_k. \tag{5.5}$$

Step 2: If x^{k+1} satisfies a prescribed stopping rule, then stop.

Step 3: Replace k by $k+1$ and go to step 1.

A direct consequence of Proposition 3.3 is the following result.

Corollary 5.1. *Assume that $\mathcal{X}^*$ is nonempty and bounded. Then the sequence $\{\gamma_k\}$ can be chosen to be bounded (it suffices to take the equality in (5.1)).*

Lemma 5.2. *Let $\{x^k\}$ be sequence generated by (PAVM-Log) and assume that $\mathcal{X}^*$ is nonempty and bounded. Then*

$$\lim_{k \to +\infty} \|g^{k+1} + \mathbf{B}^\top \omega^{k+1}\|^{*2}_{\mathbf{M}_k} = 0, \tag{5.6}$$

where $\mathbf{M}_k = 2\mathbf{Q}^{-1}_{\mathbf{w}(x^k)}$.

Proof. By using (5.3) one has

$$\begin{aligned} \gamma_k \|x^{k+1} - x^k\|^2_{\mathbf{M}_k} &= \frac{1}{\gamma_k}[\|\epsilon^{k+1}\|^{*2}_{\mathbf{M}_k} + \|g^{k+1} + \mathbf{B}^\top \omega^{k+1}\|^{*2}_{\mathbf{M}_k} \\ &\quad -2\langle g^{k+1} + \mathbf{B}^\top \omega^{k+1}, (\mathbf{A}^\top \mathbf{M}_k \mathbf{A})^{-1} \epsilon^{k+1}\rangle] \end{aligned}$$

Due to the above Corollary and Proposition 3.3, the result follows. □

Remark 5.1. If we assume that $\mathcal{X}^*$ is nonempty and bounded, and that f is defined everywhere. Then, $\{(x^{k+1}, g^{k+1}, \gamma_k)\}$ is bounded. Thus, there exists a subsequence $\{(x^{k_j+1}, g^{k_j+1}, \gamma_{k_j})\}$ and a point $(u, \tilde{g}, \tilde{\gamma})$ such that $(x^{k_j+1}, g^{k_j+1}, \gamma_{k_j}) \to (u, \tilde{g}, \tilde{\gamma})$ as $j \to +\infty$. Moreover, since $\mathbf{B}$ is onto, the subsequence ω^{k_j} of $\{\omega^k\}$ defined in (5.3) can be written as $\omega^{k_j+1} = (\mathbf{B}\mathbf{B}^\top)^{-1}\mathbf{B}(\epsilon^{k_j+1} - g^{k_j+1} - 2\gamma_{k_j}\mathbf{A}^\top \mathbf{Q}^{-1}_{\mathbf{w}(x^{k_j})}\mathbf{A}(x^{k_j+1} - x^{k_j}))$. If we suppose that $u \in C$, then taking limit when $j \to +\infty$ in the above equality we get: $\lim_{j \to +\infty} \omega^{k_j+1} = -(\mathbf{B}\mathbf{B}^\top)^{-1}\mathbf{B}\tilde{g}$, where we have used Remark 4.3. Therefore, from (5.2) it follows that $0 \in \partial f(u) + \mathrm{Im}(\mathbf{B}^\top)$. This condition implies that the limit points are optimal solutions of (SOCP).

5.2 Properties of the cluster points

In this section, we establish some results about the limit points of the iterates x^k of our PAVM-Log algorithm (that is, when $\mathbf{M}_k = 2\mathbf{Q}^{-1}_{\mathbf{w}(x^k)}$). Having in mind the applications given in the next section, we assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ (that is, $q = n$).

By construction of sequence $\{x^k\}$, any limit point $u \in \mathbb{R}^n$ satisfy $\mathbf{B}u = \mathbf{d}$ and $\mathbf{w}(u) \in \mathcal{K}$. From (5.6), it follows that $\lim_{k \to +\infty} \|\mathbf{Q}^{1/2}_{\mathbf{w}(x^k)}(\mathbf{A}^\top)^{-1}(g^{k+1} + \mathbf{B}^\top \boldsymbol{\omega}^{k+1})\| = 0$. Let $(u, \tilde{g}, \tilde{\boldsymbol{\omega}}) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r$

be a limit point of $\{(x^{k+1}, g^{k+1}, \boldsymbol{\omega}^{k+1})\}$. It follows that $\mathbf{Q}^{1/2}_{\mathbf{w}(u)}(\mathbf{A}^\top)^{-1}(\tilde{g} + \mathbf{B}^\top\tilde{\boldsymbol{\omega}}) = 0$. We define $\mathbf{s} = (\mathbf{A}^\top)^{-1}(\tilde{g} + \mathbf{B}^\top\tilde{\boldsymbol{\omega}})$ limit point of the sequence $\mathbf{s}^{k+1} = (\mathbf{A}^\top)^{-1}(g^{k+1} + \mathbf{B}^\top\boldsymbol{\omega}^{k+1})$. Then we have $\mathbf{Q}^{1/2}_{\mathbf{w}(u)}\mathbf{s} = 0; \qquad \mathbf{A}^\top\mathbf{s} = \tilde{g} + \mathbf{B}^\top\tilde{\boldsymbol{\omega}}$. Now, if $\mathbf{s} = (s_1, \ldots, s_m)$ with $s_j \in \mathbb{R}^{m_j}$ for $j = 1, \ldots, J$, it follows that

$$\mathcal{Q}^{1/2}_{w_j(u)}s_j = 0, \quad j = 1, \ldots, J. \tag{5.7}$$

(Recall that $\mathbf{Q}^{1/2}_{\mathbf{w}(u)}$ =diag$(\mathcal{Q}^{1/2}_{w_1(u)}, \ldots, \mathcal{Q}^{1/2}_{w_J(u)})$). From (2.1), (2.10) and (5.7) we obtain that

$$0 = w_i(u) \circ s_j - \begin{pmatrix} 0 \\ (w_{1j}(u) - \det(w_j(u))^{1/2})\left(\bar{s}_j - \frac{\bar{w}_j(u)^\top \bar{s}_j}{\|\bar{w}_j(u)\|^2}\bar{w}_j(u)\right) \end{pmatrix}; \quad j = 1, \ldots, J,$$

where we used that $s_j = (s_{j1}, \bar{s}_j) \in \mathbb{R} \times \mathbb{R}^{m_j - 1}$. By the definition of product "$\circ$", the first component in the equation above implies that $w_j(u)^\top s_j = 0, \quad$ for $j = 1, \ldots, J$. From his discussion, we have the following proposition.

Proposition 5.3. *Assume that $q = n$ and that $\mathcal{X}^*$ is nonempty and bounded. Then any limit point $(u, \tilde{g}, \tilde{\boldsymbol{\omega}}) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r$ of $\{(x^{k+1}, g^{k+1}, \boldsymbol{\omega}^{k+1})\}$ satisfies*

$$\mathbf{A}^\top\mathbf{s} = \tilde{g} + \mathbf{B}^\top\tilde{\boldsymbol{\omega}}, \quad \mathbf{B}u = \mathbf{d}, \quad \mathbf{w}(u) \in \mathcal{K}, \quad w_j(u)^\top s_j = 0,\ j = 1, \ldots, J,$$

where $\mathbf{s} = (\mathbf{A}^\top)^{-1}(\tilde{g} + \mathbf{B}^\top\tilde{\boldsymbol{\omega}})$.

So, in order to verify that a limit point of $\{(x^{k+1}, g^{k+1}, \boldsymbol{\omega}^{k+1})\}$ satisfies the first-order optimality conditions of the problem (SOCP), we only need to prove that $s_j \in \mathcal{L}^{m_j}_+$, for $j = 1, \ldots, J$. We suppose first that $w_j(u) \in \mathcal{L}^{m_j}_{++}$, then $\mathcal{Q}^{1/2}_{w_j(u)}$ is nonsingular (by [2, Corollary 4]), and hence the equality (5.7) implies that $s_j = 0$, for $j = 1, \ldots, J$, concluding the desired result. Consider now the case when $w_j(u) \in \partial\mathcal{L}^{m_j}_+ \setminus \{0\}$. We argue by contradiction, that is, we suppose that $s_j \notin \mathcal{L}^{m_j}_+$. This mean that, for each $j = 1, \ldots, J$, one and only one of the following situations occurs: (*i*) $\lambda_{\min}(s_j), \lambda_{\max}(s_j) < 0$. (*ii*) $\lambda_{\min}(s_j) < 0$, $\lambda_{\max}(s_j) = 0$. (*iii*) $\lambda_{\min}(s_j) < 0$, $\lambda_{\max}(s_j) > 0$. The case (*i*) is discarded as follows: if $\lambda_{\min}(s_j), \lambda_{\max}(s_j) < 0$, for $j = 1, \ldots, J$, $-s_j \in \mathcal{L}^{m_j}_{++}$, for $j = 1, \ldots, J$. By Lemma 2.6, it follows that $\forall w_j(u) \in \mathcal{L}^{m_j}_+$, $w_j(u)^\top s_j < 0$, for each $j = 1, \ldots, J$, which is a contradiction.

Unfortunately, we have not been able to discard cases (*ii*) and (*iii*) for a general objective function f. However, in the next section we show that cases (*ii*) and (*iii*) can be discarded if we suppose in addition that f is linear, that is, when (SOCP) becomes a linear SOCP problem. We conjecture that for a general (SOCP) when f is not necessarily linear, the limit points of the sequence generated by PAVM-Log algorithm belong to $\mathcal{X}^*$.

5.3 Full convergence for a linear objective function

In this section we suppose that f is linear, i.e. $f(x) = c^\top x$. Also, we assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ (that is, $q = n$) and that $\mathcal{X}^*$ is nonempty and bounded. Recall that for $j = 1, \ldots, J$, $C_j = \{x \in \mathbb{R}^n : A^j x + b^j \in \mathcal{L}^{m_j}_+\}$, $\mathcal{F} = \prod_{j=1}^J C_j$, $\mathcal{B} = \{x \in \mathbb{R}^n : \mathbf{B}x = \mathbf{d}\}$ and $C = \mathcal{B} \cap \mathcal{F}$. Our convergence result for this particular case is stated below. It is based on the recession analysis for optimization problems (see [36, 5]).

Proposition 5.4. *Assume that f is linear, that $q = n$ and that $\mathcal{X}^*$ is nonempty and bounded. If the following inclusion holds for each $j = 1, \ldots, J$*

$$A^j(\operatorname{Ker}\mathbf{B}) \supseteq \mathcal{L}^{m_j}_+, \tag{5.8}$$

then any limit point of $\{x^k\}$ belong to $\mathcal{X}^*$.

Proof. Let u be a limit point of $\{x^k\}$. We proceed to show that u satisfies the first-order optimality conditions of (SOCP). Since the results established in the last section, it only left to prove that $\mathbf{s} = (s_1, \ldots, s_J) \in \mathcal{K}$. It is well known that $\mathcal{X}^*$ be nonempty and bounded iff (see [5])

$$f_\infty(d) > 0 \qquad \forall\, d \in C_\infty,\, d \neq 0. \tag{5.9}$$

Now, note that the recession function of f is given by $f_\infty(d) = c^\top d$, for all $d \in \mathbb{R}^n$, and the recession set of feasible set is given by $C_\infty = \{d \in \mathbb{R}^n : A^j d \in \mathcal{L}_+^{m_j},\ j = 1, \ldots, J,\ \mathbf{B}d = 0\}$. Then, as A^j is full rank, condition (5.9) is rewritten as $c^\top d > 0$,$\forall\, d \neq 0$; $A^j d \in \mathcal{L}_+^{m_j}$, $j = 1, \ldots, J$, $\mathbf{B}d = 0$. Since $c^\top d = [A^j ({A^j}^\top A^j)^{-1}(c + \mathbf{B}^\top \bar{\omega})]^\top A^j d$ for any $j = 1, \ldots, J$ and inclusion (5.8), this condition implies that $[A^j ({A^j}^\top A^j)^{-1}(c + \mathbf{B}^\top \bar{\omega})]^\top v > 0$, for all $v \in \mathcal{L}_+^{m_j} \setminus \{0\}$, $j = 1, \ldots, J$. Then, the Lemma 2.6 implies that $s_j = A^j ({A^j}^\top A^j)^{-1}[c + \mathbf{B}^\top \bar{\omega}] \in \mathcal{L}_{++}^{m_j}$, for $j = 1, \ldots, J$. Hence, the limit point satisfies the first-order optimality conditions of (SOCP), that is, $u \in \mathcal{X}^*$. □

6 Computational experiments on some specific applications

6.1 Preliminaries

In this section we discuss some MATLAB implementations of PAVM-Log, the algorithm described in the previous section, and we report some computational results on some specific instances of two classes of SOCP. More precisely, we consider multiple load models in truss structural optimization, and robust classification by hyperplanes under data uncertainty in support vector machines. Our main goal is to illustrate how our algorithm works in practice and to verify empirically that it produces correct results for these special problems.

On purpose we have chosen two well-known applications that can be formulated as a Linear SOCP (LSOCP). This allows us to compare our results with those obtained by standard solvers for LSOCP. In fact, our results for both applications are compared with the solutions obtained by SeDuMi 1.1R2 (Self-Dual Minimization) toolbox for MATLAB, which is considered one of the best implementations of a primal-dual interior point method for solving LSOCPs; see [35] for more details.

Since by construction our algorithm forces the iterates x^k to be in the interior of the feasible set, we use the result obtained by SeDuMi as a benchmark for the optimal value of the objective function f, in the sense that a small difference between $f(x^k)$ and that benchmark will ensure the correctness of our solution, up to some relative error tolerance of course.

All numerical experiments were performed on a Toshiba Tecra laptop with an Intel Pentium M 740 CPU 1.73GHz processor and 512MB of RAM, running Microsoft Windows XP operating system. The computer codes were all written in MATLAB 7.3, Release 2006b.

6.2 Truss structural optimization

6.2.1 The multiple load problem

By a *truss* we mean a mechanical structure composed of thin elastic bars, connecting some pairs of nodal points in $\mathbb{R}^d$ with either $d = 2$ or $d = 3$. When subjected to a given *load* (distribution of external forces applied at the nodes) the structures deformates, until the reaction forces caused by deformations of the bars compensate the external load. The deformated truss stores a certain amount of potential energy, and this energy, named the *compliance*, measures the stiffness of the truss, that is, its ability to withstand the load; the less is compliance, the more rigid is the truss with respect to the load; see, for instance, [1, 7] and the references therein.

In the usual Truss Topology Design problem we are given the nodal set and one (single-load) or several (multi-load) loads, along with total volume of the bars. The displacements of some of the nodes are completely or partially fixed. Some of the bars can get zero volume, i.e., be eliminated from the resulting construction, so that in fact the topology of the construction is optimized as well.

Let $n = d \cdot N - s$ be the number of degrees of freedom of a *ground structure* consisting of N nodes, where s denote the number of fixed nodal coordinate directions. Let $m \geq n$ be the number of potential bars. We denote by $x_i = a_i \ell_i \geq 0$ the volume of the i-th bar with $i \in \{1, \dots, m\}$, where a_i is its cross-sectional area and ℓ_i its length. Due to the truss model, we assume that external loads apply only at nodal points, which in global reduced coordinates are described by a vector $f \in \mathbb{R}^n$. Under the assumption that each bar is subject to only axial tension or compression, the mechanical response of the truss is described by the condition of elastic equilibrium $K(x)u = f$, where $u \in \mathbb{R}^n$ is the nodal displacements vector in global reduced coordinates and $K(x)$ is the *global stiffness matrix* of the truss, and is defined by

$$K(x) = \sum_{i=1}^{m} x_i K_i.$$

Here, $x \geq 0$ is the volume vector and $K_i \in \mathbb{R}^{n \times n}$ is the *specific stiffness matrix* of the i-th bar, and is given by

$$K_i = \frac{E_i}{\ell_i^2} \zeta_i \zeta_i^\top, \tag{6.1}$$

where E_i is Young' modulus for the material of the i-th bar and $\zeta_i \in \mathbb{R}^n$ is a vector that contains the cosines and sines that describing the orientation of i-th bar. Since each K_i has form dyadic, the matrix $K(x)$ is positive semidefinite for $x \geq 0$. Moreover, if we suppose that

$$\operatorname{span}\{\zeta_1, \dots, \zeta_m\} = \mathbb{R}^n \tag{6.2}$$

and $x > 0$, then $K(x)$ is positive definite. We assume that the ground structure satisfies (6.2). The problem of finding a truss with the minimum possible compliance [1, 7] for a given volume $V > 0$ of material is formulated by

$$\min_{(x,u) \in \mathbb{R}^m \times \mathbb{R}^n} \frac{1}{2} f^\top u\,; \quad K(x)u = f, \quad \sum_{i=1}^{m} x_i = V,\ x_i \geq 0,\ i = 1, \dots, m. \tag{6.3}$$

This formulation is so-called *single load problem*.

Optimal solutions using a single load model may be unstable, even under small perturbations in the principal load. In fact, there are several examples that show some optimal structures giving infinite compliance under small perturbations. An alternative to deal with this inconvenient is to consider a *multiload model* instead of the single load one, by minimizing a weighted average of the compliances associated with k different loading scenarios $f_1, \dots, f_k \in \mathbb{R}^n$ (see [1, 3]):

$$\begin{aligned} \min_{(x,\hat{u}) \in \mathbb{R}^m \times \mathbb{R}^{kn}} \frac{1}{2} \sum_{j=1}^{k} \lambda_j f_j^\top u_j\,; \quad & K(x)u_j = f_j,\ j = 1, \dots, k, \\ & \sum_{i=1}^{m} x_i = V,\ x_i \geq 0,\ i = 1, \dots, m, \end{aligned} \tag{6.4}$$

where $\hat{u} = (u_1, \dots, u_k) \in \mathbb{R}^{kn}$ and $\lambda_j > 0$, $j = 1, \dots, k$, denote suitable weights on the individual compliance values. Defining $\hat{K}(x) = \sum_{i=1}^{m} \lambda_i \hat{K}_i$ with $\hat{K}_i = \operatorname{diag}(\lambda_1 K_i, \dots, \lambda_k K_i) \in \mathbb{R}^{nk \times nk}$, for $i = 1, \dots, m$, $\hat{f} = (\lambda_1 f_1, \dots, \lambda_k f_k) \in \mathbb{R}^{nk}$, (6.4) can be formulated as

$$\min_{(x,\hat{u}) \in \mathbb{R}^m \times \mathbb{R}^{kn}} \frac{1}{2} \hat{f}^\top \hat{u}\,; \quad \hat{K}(x)\hat{u} = \hat{f}, \quad \sum_{i=1}^{m} x_i = V,\ x_i \geq 0,\ i = 1, \dots, m. \tag{6.5}$$

Although this problem is analogous to (6.3), the increase of the dimension is evident and, more importantly, the matrix $\hat{K}_i$ looses the *dyadic* structure (6.1).

6.2.2 LSOCP formulation

Following [26] (see also [7, §3.4.3]), we now present a Linear SOCP optimization problem that is equivalent to (6.4). Using (6.1), the first constraint of (6.4) is written as $\sum_{i=1}^m x_i \frac{E_i}{\ell_i^2}\zeta_i^\top u_j \zeta_i = f_j$. Let

$$y_{ij} = x_i \frac{\sqrt{E_i}}{\ell_i}\zeta_i^\top u_j, \quad i = 1, \dots, m,\ j = 1, \dots, k.$$

Then $\sum_{i=1}^m y_{ij}\frac{\sqrt{E_i}}{\ell_i}\zeta_i = f_j$, for $j = 1, \dots, k$. Thus, $\sum_{i=1}^m y_{ij}^2/x_i = f_j^\top u_j$, for i such that $x_i > 0$. Hence, the problem (6.4) is formulated as

$$\begin{aligned} \min_{x\in\mathbb{R}^m, y_{ij}\in\mathbb{R}} \frac{1}{2}\sum_{j=1}^k \sum_{i:x_i>0} \lambda_j y_{ij}^2/x_i; \sum_{i=1}^m x_i = V,\ f_j = \sum_{i=1}^m y_{ij}\frac{\sqrt{E_i}}{\ell_i}\zeta_i,\ j = 1, \dots, k,\\ x_i \geq 0,\ i = 1, \dots, m. \end{aligned} \tag{6.6}$$

By introducing additional auxiliary variables t_{ij}, (6.4) can be expressed as

$$\begin{aligned} \min_{\substack{x\in\mathbb{R}^m, y_{ij}\in\mathbb{R}\\ t_{ij}\in\mathbb{R}}} \frac{1}{2}\sum_{j=1}^k \sum_{i=1}^m \lambda_j t_{ij}; \sum_{i=1}^m x_i = V,\ f_j = \sum_{i=1}^m y_{ij}\frac{\sqrt{E_i}}{\ell_i}\zeta_i,\ j = 1, \dots, k,\\ y_{ij}^2 \leq t_{ij}x_i, \quad t_{ij},\ x_i \geq 0,\ i = 1, \dots, m,\ j = 1, \dots, k. \end{aligned} \tag{6.7}$$

The last inequalities are equivalent to

$$\left\| \begin{pmatrix} 2y_{ij} \\ x_i - t_{ij} \end{pmatrix} \right\| \leq x_i + t_{ij} \ \Leftrightarrow\ (x_i + t_{ij}, 2y_{ij}, x_i - t_{ij}) \in \mathcal{L}_+^3,\ i = 1, \dots, m,\ j = 1, \dots, k.$$

Therefore, (6.4) can be equivalently written as:

$$\begin{aligned} \min_{\substack{x\in\mathbb{R}^m, y_{ij}\in\mathbb{R}\\ t_{ij}\in\mathbb{R}}} \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^k \lambda_j t_{ij}; \sum_{i=1}^m x_i = V,\ f_j = \sum_{i=1}^m y_{ij}\frac{\sqrt{E_i}}{\ell_i}\zeta_i,\ j = 1, \dots, k,\\ (x_i + t_{ij}, 2y_{ij}, x_i - t_{ij}) \in \mathcal{L}_+^3,\ i = 1, \dots, m,\ j = 1, \dots, k. \end{aligned} \tag{6.8}$$

6.2.3 Specialization of PAVM-Log

Let us denote the decision variable by $z = (\mathbf{x}, \mathbf{t}, \mathbf{y}) \in \mathbb{R}^{(2k+1)m}$, where $\mathbf{x} = (x_1, \dots, x_m)$, $\mathbf{t} = (t_{11}, t_{21}, \dots, t_{m1}, \dots, t_{1k}, t_{2k} \dots, t_{mk})$ and $\mathbf{y} = (y_{11}, y_{21} \dots, y_{m1}, \dots, y_{1k}, y_{2k} \dots, y_{mk})$, so the objective function is $p(z) = \frac{1}{2}\sum_{i=m+1}^{(k+1)m} \bar{\lambda}_i z_i$ with $\bar{\lambda} = (\lambda_1, \dots, \lambda_1, \dots, \lambda_k, \dots, \lambda_k) \in \mathbb{R}^{km}$, whose gradient is $\nabla p(z) = \frac{1}{2}(\mathbf{0}_m, \bar{\lambda}, \mathbf{0}_{km})$, where $\mathbf{0}_m \in \mathbb{R}^m$ is a column vector of zeros in $\mathbb{R}^m$. Set

$$w^{ij}(z) = (z_i + z_{jm+i}, 2z_{(k+j)m+i}, z_i - z_{jm+i}) = (x_i + t_{ij}, 2y_{ij}, x_i - t_{ij}),$$

for $i = 1, \dots, m$ and $j = 1, \dots, k$. We define $D_\Gamma = \mathsf{blkdiag}^1(\Gamma_\ell^\top, \dots, \Gamma_\ell^\top) \in \mathbb{R}^{kn\times km}$, with $\Gamma_\ell^\top \in \mathbb{R}^{n\times m}$ the matrix whose $i-$column is $\frac{\sqrt{E_i}}{\ell_i}\zeta_i$, and set

$$\mathbf{B} = \begin{pmatrix} \mathbf{1}^T & \mathbf{0}_{km} & \mathbf{0}_{km} \\ \mathbf{0}_{kn\times m} & \mathbf{0}_{kn\times km} & D_\Gamma \end{pmatrix} \in \mathbb{R}^{kn+1\times(2k+1)m}, \quad \bar{f} = \begin{pmatrix} V \\ f_{truss} \end{pmatrix} \in \mathbb{R}^{kn+1}$$

where $\mathbf{1} \in \mathbb{R}^m$ is a column vector of ones in $\mathbb{R}^m$ and $f_{truss} = (f_1, \dots, f_k) \in \mathbb{R}^{kn}$.

[1] Note that we used the notation of Matlab, that is, blkdiag is a block diagonal matrix.

Since the objective function in $(LSOCP_1)$ is linear, the proximal step in PAVM-Log corresponds to the unconstrained stationary condition for a quadratic function, which amounts to solving exactly (choosing $\delta_k = \eta_k = 0$) a system of linear equations:

$$\begin{aligned} 2\gamma_k \mathbf{A}^\top \mathbf{Q}^{-1}_{\mathbf{w}(z^k)} \mathbf{A} z^{k+1} + \mathbf{B}^\top \boldsymbol{\omega}^{k+1} &= 2\gamma_k \mathbf{A}^\top \mathbf{Q}^{-1}_{\mathbf{w}(z^k)} \mathbf{A} z^k - \frac{1}{2}(\mathbf{0}_m, \mathbf{1}, \mathbf{0}_m) \\ \mathbf{B} z^{k+1} &= \bar{f} \end{aligned} \tag{6.9}$$

for some $\boldsymbol{\omega}^{k+1} \in \mathbb{R}^{kn+1}$. In this context, condition (5.1) to generate an interior proximal point $z^{k+1} \in C$ reduces to

$$\gamma_k \geq \frac{\sqrt{m}}{2} \max_{i=1,\ldots,m, j=1,\ldots,k} \{\lambda_{\max}(w^{ij}(z^k))\} + \beta_k.$$

As a larger stepsize γ_k^{-1} means a smaller parameter γ_k, then to speed up convergence it is natural to take γ_k to be equal to the right-hand side of this condition. But in practical computations such a choice may be too conservative. Instead, we implement the following relaxed version: given $z^k \in C$ define

$$\gamma_k(\ell) = \frac{1}{2^\ell} \left[\frac{\sqrt{m}}{2} \max_{i=1,\ldots,m, j=1,\ldots,k} \{\lambda_{\max}(w^{ij}(z^k))\} + \beta_k \right], \tag{6.10}$$

and denote by $z(\ell)$ the proximal point corresponding to the regularization parameter $\gamma_k(\ell)$, that is, $z(\ell)$ is the solution of (6.9) with $\gamma_k(\ell)$ instead of γ_k. Notice that the linearity of system (6.9) implies that $z(\ell) = z^k + \gamma_k(\ell)^{-1} \Delta z^k$, where $\Delta z^k \in \mathbb{R}^{(2k+1)m}$ solves the following linear system which does not depend on ℓ:

$$\begin{aligned} 2\mathbf{A}^\top \mathbf{Q}^{-1}_{\mathbf{w}(z^k)} \mathbf{A} \Delta z^k + \mathbf{B}^\top \tilde{\boldsymbol{\omega}}^{k+1} &= -\frac{1}{2}(\mathbf{0}_m, \mathbf{1}, \mathbf{0}_m) \\ \mathbf{B} \Delta z^k &= 0 \end{aligned}$$

for some $\tilde{\boldsymbol{\omega}}^{k+1} \in \mathbb{R}^{kn+1}$. Then we set $z^{k+1} = z(\ell_k^*) = z^k + \gamma_k(\ell_k^*)^{-1} \Delta z^k$ where $\ell_k^* = \max\{0, ..., \ell_{max} : z(\ell) \in C\}$. That is, we take the regularization parameter as the smaller of the form (6.10) for $\ell \in \{0, \ldots, \ell_{max}\}$ in such a way that the updated proximal point belongs to the interior of the feasible set.

Finally, as the stopping rule we take

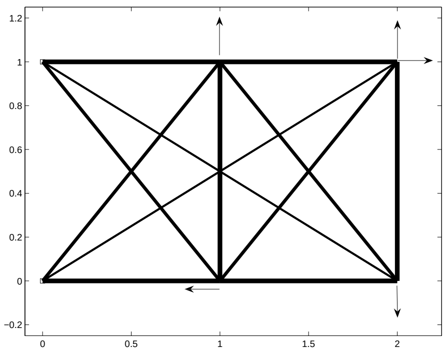
$$\frac{\|z^{k+1} - z^k\|}{\|z^{k+1}\|} \leq \text{Tol},$$

where Tol is a prescribed relative tolerance.

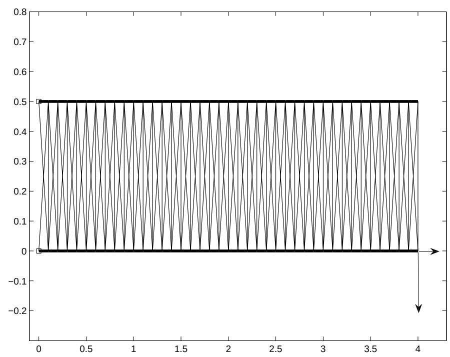
6.2.4 Computational results

We implemented in MATLAB the previously described version of PAVM-Log algorithm and we run it for solving three instances of classic examples of multiload truss optimization [1], which are illustrated in Figure 1.

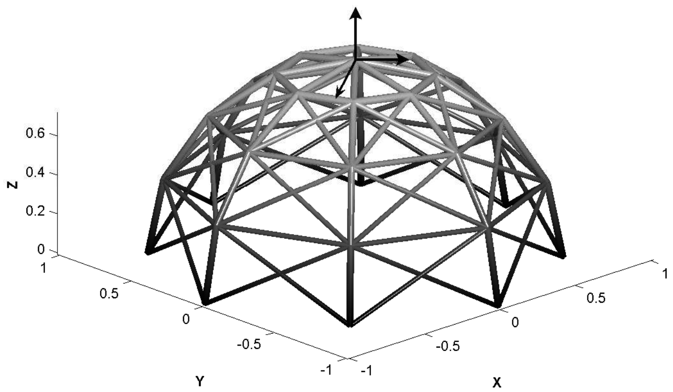
The loads applied on the structures Michell 2x1 and 2D Cantilever are shown in Figure 1 (a) and (b), respectively. These are modeled as two scenarios, one with only horizontal loads and the other one with only vertical loads, with component values between 1 and 10 and weights given by $\lambda = (\frac{1}{2}, \frac{1}{2})$. In the case of the Dome, we consider one vertical load and two orthogonal loads which are applied just on the top, with component values between 10 and 20 and weights $\lambda = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. In all cases, the total volume V is normalized, that is, $V = 1$. See Table 1 for information on the sizes of these problems.



(a) Michel 2×1



(b) 2D Cantilever



(c) Dome

Figure 1: Ground structures for the 3 examples

Table 1: Dimensions of the truss design test problems.

Type of problem	# bars (m)	# nodes (N)	# degrees of freedom (n)
Michell 2×1	12	6	8
Dome	104	33	75
2D Cantilever	200	82	160

In our experiments, we use the following definitions and values:

- The starting point z^0 is given by $y^0_{\cdot j} = \Gamma_\ell(\Gamma_\ell^\top \Gamma_\ell)^{-1} f_j$, for $j = 1, \ldots, k$, $x^0_i = 1/m$ and $t^0_{ij} = \frac{{y^0_{ij}}^2}{x^0_i} + 2.5$, for $i = 1, \ldots, m$ and $j = 1, \ldots, k$.
- $\beta_k = 0.1$ and $\delta_k = \eta_k = 0$.
- In the relaxed version (6.10) for the regularization parameter, we take $\ell_{max} = 10$.
- In the stopping rule, we use Tol$= 10^{-2}$ and Tol$= 10^{-3}$.

Table 2 reports the results of our experiments and provide some comparisons with SeDuMi 1.1R2 toolbox for MATLAB. In this Table, the second and fifth columns show the number of proximal iterations to fulfill the stopping rule $\frac{\|z^{k+1}-z^k\|}{\|z^{k+1}\|} \leq$ Tol, the third and sixth columns report the CPU time by using our implementation in MATLAB of this specialized version of PAVM-Log, the fourth and seventh columns provides the relative difference between the value of the objective function (compliance) at the output solution obtained by PAVM-Log algorithm, and the optimal compliance given by SeDuMi, denoted by c_{pavm} and c_{sdm} respectively. Finally, the last column shows the CPU time required by SeDuMi toolbox using its default configuration.

Table 2: Total number of iterations, CPU time and objective value comparisons for tolerances Tol= $10^{-2}, 10^{-3}$.

	Tol= 10^{-2}			Tol= 10^{-3}			
Type of problem	# main iter.	CPU Time	$\frac{c_{pavm}-c_{sdm}}{c_{sdm}}$	# main iter.	CPU Time	$\frac{c_{pavm}-c_{sdm}}{c_{sdm}}$	CPU Time SeDuMi
Michell 2×1	09	00'00".11	0.005230	22	00'00".16	0.002425	00'00".62
Dome	10	00'41".36	0.051599	64	04'29".48	0.019733	00'01".27
2D Cantilever	18	03'07".28	0.014832	56	09'57".51	0.005664	00'01".40

For the 2D Cantilever, the largest problem, we can observe that depending on the accuracy of the stopping rule, PAVM-Log provides output solutions with an optimality gap of 0.6% or 1.5% when compared with the benchmark given by SeDuMi. In the case of the Dome, the same difference varies between 2.0% and 5.2%. With the exception of the toy example Michell 2x1, SeDuMi is much faster than PAVM-Log.

We can reduce the tolerance in the stopping rule to improve the accuracy of the solutions. For instance, with Tol=10^{-4} algorithm PAVM-Log provides output solutions with an optimality gap between 0.03% and 0.35% when compared with the benchmarks given by SeDuMi, as shown in the Table 3. Notice that PAVM-Log's CPU time increases considerably for medium-size problems.

Table 3: Computational results for tolerance Tol= 10^{-4}.

	Tol= 10^{-4}			
Type of problem	# main iter.	CPU Time	$\frac{c_{pavm}-c_{sdm}}{c_{sdm}}$	CPU Time SeDuMi
Michell 2×1	129	00'00".86	0.000308	00'00".62
Dome	485	33'40".14	0.003537	00'01".27
2D Cantilever	254	45'11".69	0.000854	00'01".40

6.3 Support vector machines under uncertainty

6.3.1 Linear classification by hyperplanes

Let us consider the following general binary classification problem in machine learning: from some training data points in $\mathbb{R}^n$, each of which belongs to one of two classes, the goal is to determine some way of deciding which class new data points will be in. Suppose that the training data consists of two sets of points whose elements are labeled by either 1 or -1 to indicate the class they belong to. If there exists a strictly separating $(n-1)$-dimensional hyperplane between the two data sets, namely

$$H(\mathbf{w}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^\top \mathbf{x} - b = 0\},$$

then the standard Support Vector Machine (SVM) approach is based on constructing a *linear classifier* according to the function

$$f(x) = \mathrm{sgn}(\mathbf{w}^\top \mathbf{x} - b).$$

As there might be many hyperplanes that classify the data, in order to minimize misclassification, it is natural to pick the hyperplane which maximizes the separation (margin) between the two classes, so that the distance from the hyperplane to the nearest data point is maximized. In fact, if we have a set $\mathcal{T} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$ of m training data points in $\mathbb{R}^n \times \{-1, 1\}$, the *maximum-margin* hyperplane problem can be formulated as follows (see, for instance, [14] for all details):

$$\min_{\mathbf{w},b} \|\mathbf{w}\|; \; y_i(\mathbf{w}^\top \mathbf{x}_i - b) \geq 1, \quad i = 1, \dots, m. \tag{6.11}$$

If this problem is feasible then we say that the training data set $\mathcal{T}$ is *linearly separable*. The linear equations $\mathbf{w}^\top \mathbf{x} - b = 1$ and $\mathbf{w}^\top \mathbf{x} - b = -1$ describe the so-called *supporting* hyperplanes. Of course, without changing optimal solutions, (6.11) can be reformulated as a Quadratic Programming (QP) optimization problem:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2; \; y_i(\mathbf{w}^\top \mathbf{x}_i - b) \geq 1, \quad i = 1, \dots, m. \tag{6.12}$$

Alternatively, (6.11) can be written directly as a LSOCP by introducing the auxiliary scalar variable $t \in \mathbb{R}$:

$$\min_{t,\mathbf{w},b} \; t; \; t \geq \|\mathbf{w}\|, \, y_i(\mathbf{w}^\top \mathbf{x}_i - b) \geq 1, \quad i = 1, \dots, m. \tag{6.13}$$

6.3.2 Robust linear classification under uncertainty

Next, following [32, 33], suppose that $\mathbf{X}_1$ and $\mathbf{X}_2$ are random vector variables that generate samples of the positive and negative classes respectively. In order to construct a maximum margin linear classifier such that the false-negative and false-positive error rates do not exceed $\eta_1 \in (0, 1]$ and $\eta_2 \in (0, 1]$ respectively, let us consider the following Quadratic Chance-Constrained Programming (QCCP) problem:

$$\begin{aligned} \min_{\mathbf{w},b} \; \frac{1}{2}\|\mathbf{w}\|^2; \; &\mathrm{Prob}\{\mathbf{w}^\top \mathbf{X}_1 - b < 0\} \leq \eta_1, \\ &\mathrm{Prob}\{\mathbf{w}^\top \mathbf{X}_2 - b > 0\} \leq \eta_2. \end{aligned} \tag{6.14}$$

In other words, we require that the random variable $\mathbf{X}_i$ lies on the correct side of the hyperplane with probability greater than $1 - \eta_i$ for $i = 1, 2$.

Assume that for $i = 1, 2$ we *only know* the mean $\mu_i \in \mathbb{R}^n$ and covariance matrix $\Sigma_i \in \mathbb{R}^{n \times n}$ of the random vector $\mathbf{X}_i$. In this case, for each $i = 1, 2$ we want to be able to classify correctly, up to the rate η_i, even for the *worst distribution* in the class of distributions which have common mean and covariance $\mathbf{X}_i \sim (\mu_i, \Sigma_i)$, replacing the probability constraints in (6.14) with their *robust* counterparts

$$\sup_{\mathbf{X}_1 \sim (\mu_1, \Sigma_1)} \mathrm{Prob}\{\mathbf{w}^\top \mathbf{X}_1 - b < 0\} \leq \eta_1, \qquad \sup_{\mathbf{X}_2 \sim (\mu_2, \Sigma_2)} \mathrm{Prob}\{\mathbf{w}^\top \mathbf{X}_2 - b > 0\} \leq \eta_2.$$

By virtue of an appropriate application of the multivariate Chebyshev inequality, this worst distribution approach leads to the following QSOCP, which is a deterministic formulation of (6.14) (see [32] for all details):

$$\begin{aligned} \min_{\mathbf{w},b} \; \frac{1}{2}\|\mathbf{w}\|^2; \; &\mathbf{w}^\top \mu_1 - b \geq 1 + \kappa_1 \|S_1^\top \mathbf{w}\|, \\ &b - \mathbf{w}^\top \mu_2 \geq 1 + \kappa_2 \|S_2^\top \mathbf{w}\|, \end{aligned} \tag{6.15}$$

where $\Sigma_i = S_i S_i^\top$ (for instance, Cholesky factorization) for $i = 1, 2$, and η_i and κ_i are related via the formula

$$\kappa_i = \sqrt{\frac{1-\eta_i}{\eta_i}}.$$

Notice that similarly to the standard hard-margin SVM formulation (6.12)-(6.13), problem (6.15) can be written as a LSOCP:

$$\begin{aligned} \min_{\mathbf{w},b}\ t;\ t \geq \|\mathbf{w}\|,\ \mathbf{w}^\top \mu_1 - b \geq 1 + \kappa_1 \|S_1^\top \mathbf{w}\|, \\ b - \mathbf{w}^\top \mu_2 \geq 1 + \kappa_2 \|S_2^\top \mathbf{w}\|. \end{aligned} \tag{6.16}$$

Note that any feasible hyperplane must separate the means, hence the natural condition $\mu_1 \neq \mu_2$ is necessary for (6.15) to be feasible. On the other hand, since $\kappa_i \to 0$ when $\eta_i \to 1$, the problem (6.15) can be made feasible whenever $\mu_1 \neq \mu_2$ by choosing appropriate values for η_1 and η_2. By choosing $\eta_1 \neq \eta_2$ this formulation can be used for classification with *preferential bias* towards a particular class; for instance, in the case of medical diagnosis one can allow a low η_1 and a relatively high η_2 (see [32, Section 4]). Finally, we can mention that these problems can be unfeasible for some values of η_1 or η_2, for instance when we take $\eta_i \to 0$, we get $\kappa_i \to \infty$.

So far we have assumed that the mean-covariance pairs (μ_i, Σ_i) are known. However, in many practical situations we only have the training data set $\mathcal{T} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$. Assuming that $\mathcal{T}$ consists of two samples of independent observations of the random vectors $\mathbf{X}_1$ for $y = 1$ and $\mathbf{X}_2$ for $y = -1$, he idea is to replace (μ_i, Σ_i) with a statistical estimator $(\widehat{\mu}_i, \widehat{\Sigma}_i)$; this can be done by computing the sample mean and covariance for each class from the available observations.

6.3.3 Specialization of PAVM-Log

Finding an initial condition of the problem (6.15) may be difficult. Therefore, we consider the following soft-margin SVM formulation:

$$\begin{aligned} \min_{\mathbf{w},b,\xi}\ \frac{1}{2}\|\mathbf{w}\|^2 + \nu(\xi_1 + \xi_2);\ \mathbf{w}^\top \mu_1 - b \geq 1 - \xi_1 + \kappa_1 \|S_1^\top \mathbf{w}\|, \\ b - \mathbf{w}^\top \mu_2 \geq 1 - \xi_2 + \kappa_2 \|S_2^\top \mathbf{w}\|,\ \xi_1 \geq 0,\ \xi_2 \geq 0, \end{aligned} \tag{6.17}$$

where $\nu > 0$ is a sufficiently large penalty parameter. This formulation is based on the idea suggested by Cortes and Vapnik [14] for training data that are no linearly separable. Of course, if the original problem (6.15) feasible then at the optimum for (6.17) we will obtain $\xi_1 = \xi_2 = 0$; otherwise, this approach provides a numerical mean to detect unfeasibility.

Let us denote the decision variable by $z = (\mathbf{w}, b, \xi_1, \xi_2) \in \mathbb{R}^{n+3}$, so the objective function is $p(z) = \frac{1}{2}\|\mathbf{w}\|^2 + \nu(\xi_1 + \xi_2)$ with gradient $\nabla p(z) = (\mathbf{w}, 0, \nu, \nu)$. Set

$$w^i(z) = (\xi_i + (-1)^{(i+1)}(\mu_i^\top \mathbf{w} - b) - 1, \kappa_i S_i^\top \mathbf{w}) \in \mathcal{L}_+^{n+1},$$

for $i = 1, 2$. As we have also the positivity constraints $\xi_i \geq 0$, we adapt the idea of Souza et. al [34] to this situation, that is, for those constraints we consider the Hessian of the logarithm barrier function $\psi(\xi_1, \xi_2) = -\log(\xi_1) - \log(\xi_2)$.

Notice that the objective function $p(z)$ is quadratic. Then, the proximal step in PAVM-Log corresponds to the unconstrained stationary condition for a quadratic function, which amounts to solving exactly (choosing $\delta_k = \eta_k = 0$) the following system of linear equations:

$$\begin{aligned} (\tilde{I}_n + 2\gamma_k \mathbf{A}^\top \mathbf{Q}_{\mathbf{w}(z^k)}^{-1} \mathbf{A} + \gamma_k \mathrm{blkdiag}(\mathbf{0}_{n+1\times n+1}, \frac{1}{{\xi_1^k}^2}, \frac{1}{{\xi_2^k}^2})) z^{k+1} &= 2\gamma_k \mathbf{A}^\top \mathbf{Q}_{\mathbf{w}(z^k)}^{-1} \mathbf{A} z^k \\ &\quad -(\mathbf{0}_{n+1}, \nu, \nu), \end{aligned}$$

where $\tilde{I}_n = \text{blkdiag}(I_n, \mathbf{0}_{3\times3}) \in \mathbb{R}^{n+3\times n+3}$. In this context, condition (5.1) on the regularization parameter γ_k for PAVM-Log to generate an interior proximal point $z^{k+1} \in C$ reduces to

$$\gamma_k \geq \frac{(\boldsymbol{\sigma}_{min}(A))^{-1}}{\sqrt{2}} \max_{i=1,2}\{\lambda_{\max}(w^i(z^k)), \xi_i^k\}(2\nu^2 + \|\mathbf{w}^k\|)^{1/2} + \beta_k.$$

As in the previous application, we implement the following relaxed version: given $z^k \in C$ define

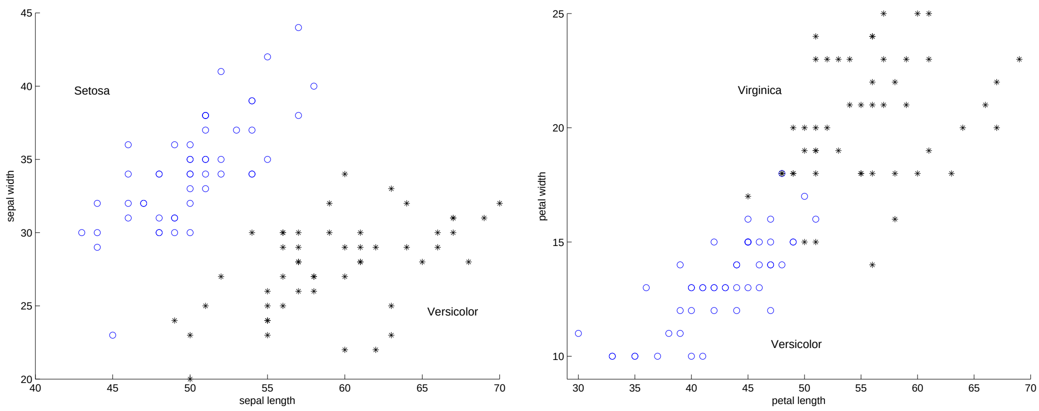
$$\gamma_k(\ell) = \frac{1}{2^\ell}\left[\frac{\sqrt{2}}{2\boldsymbol{\sigma}_{min}(A)} \max_{i=1,2}\{\lambda_{\max}(w^i(z^k)), \xi_i^k\}(2\nu^2 + \|\mathbf{w}^k\|)^{1/2} + \beta_k\right], \tag{6.18}$$

and denote by $z(\ell)$ the solution of the previous linear system with $\gamma_k(\ell)$ instead of γ_k. Then we set $z^{k+1} = z(\ell_k^*)$, where $\ell_k^* = \max\{0, \ldots, \ell_{\max} : z(\ell) \in C\}$.

Similarly, as stopping rule we take $\frac{\|z^{k+1} - z^k\|}{\|z^{k+1}\|} \leq \text{Tol}$, with Tol being a given relative error tolerance.

6.3.4 Computational results

In this section, we consider an example well-known called the iris data set, that can to be found in the pattern recognition literature, for instance in [39] or http://archive.ics.uci.edu/ml/datasets/Iris. These data contain 2 measures taken from a sample of 100 ornamental flowers. The data set contains four attributes of an iris, and the goal is to classify the class of iris based on these four attributes. To visualize the problem we restrict ourselves to the two features that contain the most information about the class, namely the petal length and the petal width, sepal length and sepal width. There are three species, *setosa*, *virginica* and *versicolor*, of which two are considered in each set. We look for to classify the flowers of each set in two species of the existent ones. In all the examples, $n = 2$. The distribution of the data is illustrated in Figure 2



(a) Setosa vs Versicolor (b) Versicolor vs Virginica

Figure 2: Iris data set.

In our implementation of the PAVM-Log algorithm, we use the following notations and values:

- We use the command MATLAB *mean* and *cov* on the training data $\mathcal{T}$, to compute the mean μ and covariance matrix Σ, respectively. For our data, we have the following results (rounded off with 2 decimals):

Data(Setosa vs Versicolor): Sepal (length,width)
Mean values: $\hat{\mu}_1 = (50.06, 34.28)$, $\hat{\mu}_2 = (59.36, 27.70)$.
Covariance matrices:

$$\hat{\Sigma}_1 = \begin{pmatrix} 12.42 & 9.92 \\ 9.92 & 14.37 \end{pmatrix}, \quad \hat{\Sigma}_2 = \begin{pmatrix} 26.64 & 8.52 \\ 8.52 & 9.85 \end{pmatrix}.$$

Data(Versicolor vs Virginica): Petal (length,width)
Mean values $\hat{\mu}_1 = (42.60, 13.26)$, $\hat{\mu}_2 = (55.52, 20.26)$.
Covariance matrices:

$$\hat{\Sigma}_1 = \begin{pmatrix} 22.08 & 7.31 \\ 7.31 & 3.91 \end{pmatrix}, \quad \hat{\Sigma}_2 = \begin{pmatrix} 30.46 & 4.88 \\ 4.88 & 7.54 \end{pmatrix}.$$

- The matrices S_i are computed by Cholesky factorization.
- The starting point $(\mathbf{w}^0, b^0)$ by means of: $\mathbf{w}^0 = 1.1\frac{\tilde{\mathbf{w}}}{\alpha}$, $b^0 = 1.1\frac{\tilde{b}}{\alpha}$, where $\alpha = \mu_1^\top \tilde{\mathbf{w}}^0 + \tilde{b}$, $\tilde{\mathbf{w}}^0 = (\mu_{11} - \mu_{21}, \mu_{12} - \mu_{22})$, $\tilde{b}^0 = \frac{1}{2}(\mu_{21}^2 - \mu_{11}^2 + \mu_{22}^2 - \mu_{12}^2)$. The vector $\tilde{\mathbf{w}}$ is taken as the orthogonal vector to the normal of the segment joining μ_2 and μ_1, $\tilde{b}$ as the value in the Hyperplane evaluated in the medium point of μ_2 and μ_1. And, $\xi_i^0 = \kappa(i)\|S_i^\top \mathbf{w}^0\| + 0.9$.
- The values of η_i are varied.
- $\nu = 10000$, $\beta_k = 0.12$, $\delta_k = \eta_k = 0$.
- In the relaxed version (6.18) for the regularization parameter, we take $\ell_{\max} = 10$.

Tables 4 and 5 report the results of our experiments and provide some comparisons with SeDuMi 1.1R2 toolbox for MATLAB. In this tables, the first and second columns show the error rates given, the third sixth columns show the number of proximal iterations to fulfill the stopping rule $\frac{\|z^{k+1}-z^k\|}{\|z^{k+1}\|} \leq$ Tol, the fourth and seventh columns report the CPU time by using our implementation in MATLAB of this specialized version of PAVM-Log, the fifth and eighth columns provides the relative difference between the value of the objective function at the output solution obtained by PAVM-Log algorithm, and the optimal given by SeDuMi, denoted by val_{pavm} and val_{sdm} respectively. Finally, the last column shows the CPU time required by SeDuMi toolbox using its default configuration. The value $\times$ in the table represents infeasibility of the problem. If such a case occurs, CPU times correspond to the time required by PAVM-Log algorithm to reach the prescribed tolerance obtaining an infeasible solution (i.e. when $\xi_1 \neq 0$ or $\xi_2 \neq 0$). The feasibility status of the optimization problem in SeDuMi is detected by output argument info.pinf and info.dinf.

Table 4: Numerical comparisons with SeDuMi applied to data set: Setosa vs Versicolor.

		Tol= 10^{-4}			Tol= 10^{-5}			
η_1	η_2	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	CPU time SeDuMi
0.7	0.1	27	0".027	0.006623	82	0".050	0.006441	0".210
0.5	0.1	11	0".031	0.013009	31	0".031	0.002217	0".180
0.3	0.1	13	0".010	$\times$	35	0".032	$\times$	0".050
0.1	0.1	7	0".014	$\times$	20	0".022	$\times$	0".070
0.1	0.3	12	0".022	0.019774	23	0".029	0.014521	0".130
0.3	0.3	11	0".022	0.043908	34	0".032	0.030113	0".160
0.3	0.5	18	0".027	0.021844	51	0".045	0.017778	0".200

Table 5: Numerical comparisons with SeDuMi applied to data set: Versicolor vs Virginica.

		Tol= 10^{-4}			Tol= 10^{-5}			
η_1	η_2	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	CPU time SeDuMi
0.9	0.3	7	0".022	0.015485	10	0".030	0.010218	0".170
0.7	0.3	8	0".022	0.005198	11	0".023	0.005053	0".180
0.5	0.3	8	0".017	0.003034	28	0".033	0.001307	0".220
0.3	0.3	7	0".007	0.004301	19	0".024	0.004157	0".210
0.3	0.7	14	0".018	0.012250	42	0".029	0.009152	0".140
0.1	0.3	12	0".015	$\times$	24	0".031	$\times$	0".020
0.7	0.5	7	0".014	0.019001	12	0".022	0.015555	0".150

In these experiments, we can observe that output solutions are optimal up to a gap whose range varies from 0.1% to 3.0% when compared with the benchmark given by SeDuMi, for different values of η_i. In all cases, PAVM-Log CPU time is much less than SeDuMi's. Due to the small size of the problems to be solved, we can decrease the error tolerance without much computational cost. The following Tables 6 and 7 provide the computational results with smaller tolerances for some values of η_i applied to the first data set, obtaining with PAVM-Log an optimality gap whose range varies from 0.3% to 0.7% with reasonable CPU time with respect to SeDuMi.

Table 6: Numerical comparisons with SeDuMi applied to data set: Setosa vs Versicolor.

		Tol= 10^{-6}			Tol= 10^{-7}			
η_1	η_2	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	CPU Time SeDuMi
0.7	0.1	259	0".173	0.005144	944	0".608	0.004195	0".210
0.1	0.3	60	0".044	0.009389	259	0".179	0.004317	0".130
0.3	0.5	143	0".092	0.008224	492	0".319	0.004422	0".200

Table 7: Numerical comparisons with SeDuMi applied to data set: Versicolor vs Virginica.

		Tol= 10^{-6}			Tol= 10^{-7}			
η_1	η_2	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	# Main iter.	CPU Time	$\frac{val_{pavm}-val_{sdm}}{val_{sdm}}$	CPU Time SeDuMi
0.7	0.3	14	0".039	0.004017	19	0".045	0.003424	0".180
0.3	0.7	136	0".092	0.006096	412	0".301	0.002556	0".140
0.7	0.5	19	0".029	0.011935	31	0".051	0.006506	0".150

6.4 Concluding remarks on the numerical tests

The previous numerical results show that PAVM-Log algorithm can be applied to solve a-pproximately LSCOP problems. As one should have expected, we see that SeDuMi is much faster in terms of CPU time than PAVM-Log for medium-size LSOCP test problems. For very small size problems, both algorithms are comparable with optimality gap less than 1%. The comparison with SeDuMi in terms of CPU time is not completely fair because of our rather straightforward implementation of PAVM-Log. Even so, it is natural that in the linear case an interior point method which is based on self-dual embedding and uses a primal-dual predictor-corrector scheme, performs better than our purely primal proximal-point strategy.

It is worth pointing out that PAVM-Log algorithm is not intended to compete with numerical methods for Linear SOCP such as SeDuMi, which is an efficient method in particular for large scale problems. But PAVM-Log might be considered as an alternative for small-size problems and, more importantly, for nonsmooth Convex SOCP for which is not clear how to extend SeDuMi-like approach. Indeed, convex problems can be addressed by conventional local algorithms since all critical points are global minimizers. The regularized proximal subproblem being strongly convex, we expect local algorithms to perform efficiently enough to find good approximate solutions at reasonable execution time. When the objective function is nonsmooth, we can work with ithe so called *bundle methods* [20, 24].

In this direction, the computational results presented here should be considered just as an intermediate step toward more general and possibly nonsmooth convex problems, which are not addressed in this paper from the numerical point of view. In fact, PAVM-Log algorithm as presented here is only schematic. There are a lot of theory aspects and implementation issues which should be addressed before performing and evaluating carefully designed computational experiments in the in the nonsmooth convex case. This goes beyond the scope of this paper and will be a topic for future research.

Acknowledgments

The authors are grateful to the anonymous referees for their comments on the original manuscript. Their suggestions were very useful to improve this work. The authors also wish to thank Prof. Paulo J. S. Silva for stimulating conversations and helpful comments. This work was partially supported by FONDECYT under grants 1050706 and 1070297, FONDAP in Applied Mathematics, BASAL project and the Millennium Scientific Institute on Complex Engineering Systems funded by MIDEPLAN-Chile. The second author was also supported by CONICYT.

References

[1] W. Achtziger, *Topology optimization of discrete structures: an introduction in view of computational and nosmooth aspects*, In Topology optimization in structural mechanics, vol. 374 of CISM Courses and Lectures, Springer, Vienna, pp. 57-100, 1997.

[2] F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Mathematical Programming, vol. 95 (2003), no. 1, Ser. B, pp. 3-51.

[3] F. Alvarez and M. Carrasco, *Minimization of the expected compliance as an alternative approach to multiload truss optimization*, Structural and Multidisciplinary Optim., vol 29 (2005), no. 6, pp. 470-476.

[4] A. Auslender, *Variational inequalities over the cone of semidefinite positive matrices and over the Lorentz cone*, Optimization Methods and Software, vol. 18 (2003), no. 4, pp. 359-376.

[5] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer Monographs in Mathematics, New York, 2003.

[6] ———, *Interior gradient and proximal methods for convex and conic optimization*, SIAM J. Optim., vol. 16 (2006), No 3, pp. 697-725.

[7] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization, Philadelphia, 2001.

[8] J.F. Bonnans, J.Ch. Gilbert, C. Lemacheral and C.A Sagastizábal, *A family of variable metric proximal methods*, Math. Programming, vol. 68 (1995), no. 1, Ser. A, pp. 15-47.

[9] J. Ch. Gilbert, *Optimisation Différentiable - Théorie et Algorithmes.* Syllabus de cours à I'ENSTA, 2002. Avaliable at http://www-rocq.inria.fr/~gilbert/ensta/optim.html.

[10] J.V. Burke and M. Qian, *A variable metric proximal point algorithm for monotone operators*, SIAM J. Control Optim. vol., 37 (1999), no. 2, pp. 353-375.

[11] ———, *On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating*, Math. Program. vol. 88 (2000), no. 1, Ser. A, pp. 157-181.

[12] Y. Censor and S. A. Zenios, *The proximal minimization algorithm with D-function*, J. Optim. Theory and Appl., vol. 73 (1992), no. 3, pp. 451-464.

[13] G. Chen and M. Teboulle, *Convergence analysis of proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., vol. 3 (1993), no. 3, pp. 538-543.

[14] C. Cortes and V. Vapnik. *Support-vector networks*, Machine Learning, vol. 20 (1995), pp. 273-297.

[15] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res., vol. 18 (1993), no. 1, pp. 202-216.

[16] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones*, Oxford Mathematical Monographs, Oxford University Press, New York, 1994.

[17] M. Fukushima, Z-Q. Luo and P. Teng, *Smoothing functions for second-order cone complementary problems*, SIAM J. Optimization, vol. 12 (2001/02), no. 2, pp. 436-460.

[18] G. Fung and O. L. Mangasarian, *Proximal support vector machine classifiers*, Proc. KDD-2001: Knowledge Discovery and Data Mining, F. Provost and R. Srikant, eds., 2001, pp. 77-86.

[19] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., vol. 29 (1991), no 2, pp. 403-419.

[20] J.-B. Hiriart-Urruty and C. Lemaréchal, "Convex analysis and minimization algorithms. II. Advanced theory and bundle methods". Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 306. Springer-Verlag, Berlin, 1993.

[21] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[22] A.N. Iusem, *An interior multiplicative method for optimization under positivity constraints*, Acta Applicadae Mathematicae, vol. 38 (1995), pp. 163-184.

[23] K. Knopp, *Theory and Application of Infinite Series*, Blackie & Son, Glasgow, 1951.

[24] C. Lemarèchal and C. Sagastizábal, *Variable metric bundle methods: from conceptual to implementable forms*, Math. Program., vol.76 (1997), pp. 393-410.

[25] A.S. Lewis, *Convex Analysis on the Hermitian Matrices*, SIAM J. Optimization, vol. 6 (1996), no 1, pp. 164-177.

[26] M.S. Lobo, L. Vandenberghe, S. Boyd and H. Lebret, *Applications of second-order cone programming*, Linear Algebra Appl. vol. 284 (1998), no 1-3 pp. 193-228.

[27] B. Martinet, *Pertubation des methodes d'optimisation*, Aplication, RAIRO, Analyse Numerique, vol. 12 (1978), pp. 153-171.

[28] J. J. Moreau, *Proximité et dualité dans um espace Hilbertien*, Bull. Soc. Math. France, vol. 93 (1965), pp. 273-299.

[29] R. A. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.

[30] J. Peng, C. Roos and T. Terlaky, *Primal-dual interior-point methods for second-order conic optimization based on self-regular proximities*, SIAM J. Optim., vol. 13 (2002), no 1, pp. 179-203.

[31] M. Qian, *The Variable Metric Proximal Point Algorithm: Theory and Application*, Ph.D. thesis, University of Washington, Seattle, WA, 1992.

[32] J. Saketha Nath and C. Bhattacharyya, *Maximum margin classifiers with specified false positive and false negative error rates*, Proceedings of the Seventh SIAM International Conference on Data Mining, April 26-28, 2007, Minneapolis, Minnesota. http://www.siam.org/meetings/proceedings/2007/datamining/papers/004Jagarlapudi.pdf,

[33] P.K. Shivaswamy, C. Bhattacharyya and A.J. Smola, *Second order cone programming approaches for handling missing and uncertain data*, J. Mach. Learn. Res., vol. 7 (2006), pp. 1283-1314.

[34] S. Souza and P.R. Oliveira, *A new class of interior proximal methods for optimization over the positive orthant*, Optimization online eprint, 2006.

[35] J.F. Sturm, *Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones*, Optim. Methods Softw., vol. 11 (1999), no 1-4, pp. 625-653. Current official website: http://sedumi.ie.lehigh.edu/

[36] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[37] ———, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., vol. 14 (1976), no 5, pp. 877-898.

[38] T. Tsuchiya, *A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming*, Optim. Methods Softw., vol. 11 (1999), no. 1-4, pp. 141-182.

[39] F. Vojtech and H. Vaclav, *Statistical pattern recognition toolbox for matlab user's guide*, Center for Machine Perception, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University Techniká 2, 16627 Prague 6, Czech Republic, 2004. Avaliable at http://cmp.felk.cvut.cz.