International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V71I1P223
Volume 71 Issue 1, 257-270, January 2023

Original Article

A Symmetric Cone Proximal Multiplier Algorithm

Erik Alex Papa Quiroz[1], Miguel Angel Cano Lengua[2], Julio Cesar Lopez Luis[3],
Rolando Ichpas Tapia[4]

[1] Universidad Privada del Norte, Universidade Federal de Goiás, Goiania, Brazil
[1,2] Universidad Nacional Mayor de San Marcos, Universidad Tecnológica del Perú, Lima, Perú
[3]Universidad Diego Portales, Chile
[4] Intendencia Nacional de Estrategias y Riesgos, Lima, Perú

[2]Corresponding Author : mcanol@unmsm.edu.pe

Received: 15 November 2022 Revised: 14 January 2023 Accepted: 17 January 2023 Published: 23 January 2023

Abstract - *This paper introduces a proximal multipliers algorithm to solve separable convex symmetric cone minimization problems subject to linear constraints. The algorithm is motivated by the method proposed by Sarmiento et al. (2016, optimization v.65, 2, 501-537), but we consider in the finite-dimensional vectorial spaces, further to an inner product, a Euclidean Jordan Algebra. Under some natural assumptions on convex analysis, it is demonstrated that all accumulation points of the primal-dual sequences generated by the algorithm are solutions to the problem and assuming strong assumptions on the generalized distances; we obtain the global convergence to a minimize point. To show the algorithm's functionality, we provide an application to find the optimal hyperplane in Support Vector Machine (SVM) for binary classification.*

Keywords - *Symmetric convex cone optimization, Separable techniques, Proximal distances, Proximal method of multipliers, Support vector machine.*

1. Introduction

In this work, we present the development of the symmetric cone proximal multiplier algorithm (SC - PMA), the same one that solves an optimization problem with separable structures; we will give a particular application to a support vector machine model related to data classification.

Consider $\mathbb{V}_1$ and $\mathbb{V}_2$ be two linear vectorial spaces of finite dimensions on $\mathbb{R}$, with $\mathbb{R}$ denoting the Euclidean space. In this space, we define two inner products: $\langle\cdot,\cdot\rangle_{\mathbb{V}_1}$ for $\mathbb{V}_1$ and $\langle\cdot,\cdot\rangle_{\mathbb{V}_2}$ for $\mathbb{V}_2$ with the Jordan product $\circ_1$ and $\circ_2$, respectively. Based on these tools, we can define the following Euclidean Jordan algebras $\mathbb{V}_1 = (\mathbb{V}_1, \circ_1, \langle\cdot,\cdot\rangle_{\mathbb{V}_1})$ and $\mathbb{V}_2 = (\mathbb{V}_2, \circ_2, \langle\cdot,\cdot\rangle_{\mathbb{V}_2})$, see subsection 2.1 for a strict definition of this concept.

In this paper, we are interested in studying an optimization algorithm for solving the following convex *symmetric cone optimization (CSCO) problem:*

$$\min\{f(x) + g(z) : \mathbb{A}x + \mathbb{B}z = b, x \in k_1, z \in k_2\}, \quad (P)$$

where $f: \mathbb{V}_1 \to \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{V}_2 \to \mathbb{R} \cup \{+\infty\}$ are proper, closed and convex functions (possibly nonsmooth), $\mathbb{A}: \mathbb{V}_1 \to \mathbb{R}^m$ and $\mathbb{B}: \mathbb{V}_2 \to \mathbb{R}^m$ are linear applications, $b \in \mathbb{R}^m$, and $k_1 := \{x \circ_1 x : x \in \mathbb{V}_1\}$ and $k_2 := \{z \circ_2 z : z \in \mathbb{V}_2\}$ are the sets of square elements (symmetric cones) in $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively.

The model (P) recovers a wide variety of applications in the fields of science and engineering (for example, in the economy, game theory, and management science, see, for instance, [1, 2, 16, 17, 25] and the references of those papers. In particular, it includes a lot of applications of current interest (for more details, see 1 and 2 of Section 3).

Many research, taking advantage of the separable structure of the objective function, have introduced several decomposition methods to solve the problem (P). Between the most recognized methods, we can find the alternating direction of the multipliers method [9], the partial inverse method [22], and the predictor-corrector proximal multiplier (PCPM) method [7]. In this present research, we will focus on the (PCPM) using proximal distances.

The steps of the PCPM method to solve (P) without conic constraints are the following.

$$\begin{cases} p^{k+1} = y^k + \lambda_k(Ax^k + Bz^k - b), \\ x^{k+1} = \arg\min_{x\in\mathbb{R}^n} \left\{f(x) + \langle p^{k+1}, Ax\rangle + \frac{1}{2\lambda_k}\|x - x^k\|^2\right\}, \\ z^{k+1} = \arg\min_{x\in\mathbb{R}^n} \left\{g(z) + \langle p^{k+1}, Bx\rangle + \frac{1}{2\lambda_k}\|z - z^k\|^2\right\}, \\ y^{k+1} = y^k + \lambda_k(Ax^k + Bz^k - b) \end{cases} \quad (1)$$

where $\{\lambda_k\}_{k\in\mathbb{N}}$ is a sequence of positive parameters. The notation $y = \arg\min_{x\in\mathbb{R}^n}\{h(x)\}$, where h is a proper function, which means that y is the global minimum of f on $\mathbb{R}^n$. Auslender and Teboulle [3] developed a proximal decomposition algorithm using the logarithmic-quadratic distance, and Kyono and Fukushima [11] introduced a proximal decomposition algorithm using Bregman distance [15]. Sarmiento et al. [20] developed an extension of the PCPM method to solve (P) using regularized proximal distances [4].

With the intention of recovering more applications, such as separable second-order cones and semidefinite optimization problems, this research aims to extend the PCPM on Euclidean Jordan algebras using proximal distances. We prove the convergence of the sequences generated by the proposed algorithm to a minimum point of the problem (P). We also present an application to find the optimal hyperplane in Support Vector Machine (SVM) for binary classification and give computational results after appropriately implementing the algorithm.

The following sections organize the paper:

Section 2 evokes certain basic on Jordan Euclidean algebras. Then, we present the definition of proximal distances defined in symmetrical cones. Section 3 details the proposed algorithm to solve (P) and establish its global convergence. Section 4 presents an application to find the optimal hyperplane in SVM and shows the algorithm's implementation. For that, a linear generator program for determining random data for the algorithm is generated using MATLAB software.

2. Preliminaries

Below, we present the notation and terminology related to convex analysis and linear algebra needed in this paper. Given a closed proper convex function f, we denote the effective domain as $dom\,(f) = \{x \in \mathbb{V}: f(x) < +\infty\}$, with $\mathbb{V}$ denoting as a finite-dimensional vectorial space with an inner product. For $\varepsilon \geq 0$, the following set

$$\partial_\varepsilon f(x) = \{p \in \mathbb{V}: f(x) + \langle p, z - x\rangle - \varepsilon \leq f(z), \forall z \in \mathbb{V}\}$$

is called the ε-Fenchel subdifferential at x, and if $\varepsilon = 0$, then we denote $\partial f = \partial_0 f$.

For a given set $S \subset \mathbb{R}^n$, the function $\delta_S(.)$ is the indicator function of S, that is $\delta_S(x) = 0$, if $x \in S$; and $\delta_S(x) = \infty$, if $x \notin S$. The set $\mathcal{N}_S(x)$ is the normal cone to S at $x \in S$. Given a set $\mathcal{K}$, $int(\mathcal{K})$ and $bd(\mathcal{K})$, denote the interior and the boundary of $\mathcal{K}$, respectively. Let $\mathbb{A}: \mathbb{V} \to \mathbb{R}^n$ be a linear application; we use the notation $\mathbb{A}^*$ by its adjoint defined by $\langle \mathbb{A}x, y\rangle = \langle x, \mathbb{A}^* y\rangle$, for all $x \in \mathbb{V}, y \in \mathbb{R}^m$.

2.1. Euclidean Jordan Algebra

This subsection concerns providing elemental tools about Euclidean Jordan algebras, being of utmost importance for the study of canonical optimization; we recommend [32] and [12] for exhaustive revision.

A Euclidean Jordan algebra consists of a real vectorial space doted by an inner product $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and a bilinear mapping $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \circ \boldsymbol{y} : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$ satisfying the following three conditions:

(i) $x \circ y = y \circ x$ for all $x, y \in \mathbb{V}$.
(ii) $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$, for all $x, y \in \mathbb{V}$ where $x^2 = x \circ x$.
(iii) $\langle x \circ y, z\rangle_{\mathbb{V}} = \langle y, x \circ z\rangle_{\mathbb{V}}$, for all $x, y, z \in \mathbb{V}$;
and exists an (unique) unitary element $e \in \mathbb{V}$, such that $x \circ e = x$, for all $x \in \mathbb{V}$.

With the above properties, we said that $\mathbb{V}$ is a Euclidean Jordan algebra and $x \circ y$ is the *Jordan product* of x and y.

2.2. Proximal Distances

Definition 2.1. *An extended-valued function* $H: \mathbb{V} \times \mathbb{V} \to \mathbb{R} \cup \{+\infty\}$ *is a proximal distance related to* $int(\mathcal{K})$ *if it satisfies the following properties:*

(P1) $dom(H(\cdot,\cdot)) = \mathcal{C}_1 \times \mathcal{C}_2$ *with* $int(\mathcal{K}) \times int(\mathcal{K}) \subseteq \mathcal{C}_1 \times \mathcal{C}_2 \subseteq \mathcal{K} \times \mathcal{K}$.

(P2) $H(u, v) \geq 0\ \forall u, v \in \mathbb{V}$, *and* $H(v, v) = 0,\ \forall v \in int(\mathcal{K})$.

(P3) *Given an arbitrary* $v \in int(\mathcal{K})$, $H(\cdot, v)$ *is a continuous function and strictly convex on* $\mathcal{C}_1$, *and it is continuously differentiable on* $int(\mathcal{K})$ *with* $dom(\nabla_1 H(\cdot, v)) = int(\mathcal{K})$, *where* $\nabla_1 H(\cdot, v)$ *denotes the gradient of* $H(\cdot, v)$ *with respect to the first variable.*

(P4) The set $\{u \in \mathcal{C}_1: H(u, v) \leq \gamma\}$ is bounded for all $\gamma \in \mathbb{R}$, *and* $\forall v \in \mathcal{C}_2$.

The above definition has been considered by [19] to define proximal distances on the interior of the second-order cone. Observe that the above definition has a small difference from Definition 2.1 from [4] due that in [4], the function $H(\cdot, y)$ should be strictly convex in $\mathcal{C}_1$ for all $y \in \text{int}(K)$. Let us denote by $\mathcal{D}(\text{int}(K))$ the family of functions H satisfying the properties given in Definition 2.1.

We give some extra conditions on $H \in \mathcal{D}(\text{int}(K))$ Which will be useful for the convergence of the algorithm.

(B1) For all $u, v \in \text{int}(K)$ and all $w \in \mathcal{C}_1, \langle \nabla_u H(u, v), w -$

$u\rangle \leq H(w,v) - H(w,u) - \gamma H(u,v)$,
for some $\gamma \in (0,1]$.

(B1´) For all $u, v \in \text{int}(K)$ and all $w \in \mathcal{C}_2$, $\langle \nabla_u H(u,v), w - u\rangle \leq H(v,w) - H(u,w) - \gamma' H(u,v)$,
for some $\gamma' \in (0,1]$.

(B2) For each $u \in \mathcal{C}_1$, the function $H(u,\cdot)$ is level bounded on $\mathcal{C}_2$.

(B3) For any $\{v^k\}_{k\in\mathbb{N}} \subseteq \text{int}(K)$: $v^k \to v^*$, and $c \in \mathcal{C}_1$, we have that $H(c, v^k) \to H(c, v^*)$.

(B3') For all $\{v^k\}_{k\in\mathbb{N}} \subseteq \text{int}(K)$: $v^k \to v^*$, and $c \in \mathcal{C}_2$, it holds $H(v^k, c) \to H(v^*, c)$.

(B4) Given $\{v^k\}_{k\in\mathbb{N}} \subseteq \text{int}(K)$ such that $\{v^k\}_{k\in\mathbb{N}}$ converges to $v^* \in$ K, it holds $H(v^*, v^k) \to 0$

(B4') Given $\{v^k\}_{k\in\mathbb{N}} \subseteq \text{int}(K)$ such that $\{v^k\}_{k\in\mathbb{N}}$ converges to $v^* \in K$, we have that $H(v^k, v^*) \to 0$.

(B5) For $v \in \mathcal{C}_1, \{v^k\}_{k\in\mathbb{N}} \subseteq \mathcal{C}_2$ bounded where $H(v, v^k) \to 0$ it holds $v^k \to v$.

(B5') For $v \in \mathcal{C}_2, \{v^k\}_{k\in\mathbb{N}} \subseteq \mathcal{C}_1$ bounded where $H(v^k, v) \to 0$, it holds $v^k \to v$.

Examples of proximal distances on symmetric cones can be found in López and Papa Quiroz [18].

3. Methodology

In this section, we present the conditions on the problem (P), show that support vector machine (SVM) for binary classification and sparse inverse covariance selection (SICS) can be expressed as (P), and introduce the proposed algorithm. We prove the global convergence of the sequences generated by the algorithm.

3.1. The Problem

We are interested in solving problem (P)
$\min\{f(x) + g(z) : \mathbb{A}x + \mathbb{B}z = b, x \in k_1, z \in k_2\}$,

Where $f: \mathbb{V}_1 \to \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{V}_2 \to \mathbb{R} \cup \{+\infty\}$ are two closed proper convex functions defined on the Euclidean Jordan algebra $\mathbb{V}_1 = (\mathbb{V}_1, \circ_1, \langle\cdot,\cdot\rangle_1)$ and $\mathbb{V}_2 = (\mathbb{V}_2, \circ_2, \langle\cdot,\cdot\rangle_2)$, respectively, $\mathbb{A}: \mathbb{V}_1 \to \mathbb{R}^m$ and $\mathbb{B}: \mathbb{V}_2 \to \mathbb{R}^m$ are two linear mappings, $b \in \mathbb{R}^m$ and $\mathcal{K}_1 := \{x \circ_1 x : x \in \mathbb{V}_1\}$ and $\mathcal{K}_2 := \{z \circ_2 z : z \in \mathbb{V}_2\}$ denoting the sets of square elements in $\mathbb{V}_1$ and $\mathbb{V}_2$, respectively.

Next, we give some examples of applications that fall into the optimization problem (P).

3.1.1. Support Vector Machines (SVM) for Binary Classification

Given a set of instances with their respective labels $(x^1, y^1), (x^2, y^2), \dots, (x^m, y^m)$, where each $x^i \in \mathbb{R}^n$, and $y^i \in \{-1, +1\}$, $i = 1, \dots,$ the SVM is based on the determination of an optimal hyperplane of the form

$$H(\boldsymbol{w}.\alpha) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{w}^T \boldsymbol{x} + \alpha = 0\},$$

where $w \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, which separates the given points. It can be proved, see [8], that the above optimal hyperplane is obtained by solving the following Quadratic Programming problem:

$$\min_{w,\alpha,\xi} g(w,\alpha,\xi) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \xi_i \qquad (3.1)$$
$$s.t.\ y^i(\boldsymbol{w}^T x^i + \alpha) \geq 1 - \xi_i, \xi_i \geq 0, \ \ i = 1, \dots, m$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m) \in \mathbb{R}^m$ and $C > 0$ is a penalty parameter.

We will show that (3.1) can be expressed as the problem (P). In fact, let us denote by

$$\boldsymbol{z} = (\boldsymbol{w}, \alpha) \in \mathbb{R}^{n+1},$$

and by

$$Y = Diag(\boldsymbol{y}) = Diag(y^1, y^2, \dots, y^m),$$

The diagonal matrix where the main diagonal is given by the elements of the vector $\boldsymbol{y}$. Consider also **e** being a vector of ones in $\mathbb{R}^m$, and by

$$\hat{X} = \begin{pmatrix} (x^1)^T & 1 \\ \vdots & \vdots \\ (x^m)^T & 1 \end{pmatrix} \in \mathbb{R}^{m\times(n+1)},$$

Then, the first inequality in (3.1) can be expressed as

$$Y\hat{X}\boldsymbol{z} + \boldsymbol{\xi} - \boldsymbol{e} \geq 0.$$

Now, let $\boldsymbol{u} = Y\hat{X}\boldsymbol{z} + \boldsymbol{\xi} - \boldsymbol{e} \geq 0$ and define $\boldsymbol{v} = (\boldsymbol{\xi}, \boldsymbol{u}) \in \mathbb{R}^{2m}$. We use the two variables z and $\boldsymbol{v}$ to obtain the problem (3.1). Define $f(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{w}\|^2$, $g(\boldsymbol{v}) = C\boldsymbol{e}^T\boldsymbol{\xi}$ and the matrices

$$A = Y\hat{X} \in \mathbb{R}^{m\times(n+1)}, \qquad B = (I \quad -I) \in \mathbb{B}^{m\times 2m}$$

Then, it is easy to verify that the problem (3.1) can be rewritten as

$$\min_{\boldsymbol{z}.\boldsymbol{v}}\{f(\boldsymbol{z}) + g(\boldsymbol{v}): A\boldsymbol{z} + B\boldsymbol{v} = b, \boldsymbol{v} \geq 0\} \qquad (3.2)$$

considering x=z and y=v. Observe that the dimension of z is n+1 and the dimension of v is 2m.

3.1.2. Sparse Inverse Covariance Selection

Gaussian Graphical model is a line of research of great interest in statistical learning [10, 24] that conditional independence between several different nodes is assigned zero in the inverse of the covariance matrix related to the Gauss distribution. This problem is associated with solving the semidefinite convex optimization problem:

$$\min_{X\in\mathcal{S}^n}\{\langle\mathcal{S},X\rangle-\ln\det(X)+\rho\|X\|_1 : X\in\mathcal{S}_+^n\} \quad (3.3)$$

where $\mathcal{S}_+^n$ is the set of the symmetric positive semidefinite matrix, $\rho>0$, $\mathcal{S}\in\mathcal{S}_+^n$, and $\|X\|_1$ the l_1-norm of the matrix X defined by

$$\|X\|_1 \coloneqq \sum_{i=1}^n\sum_{j=1}^n |x_{ij}|.$$

Defining the functions on $\mathcal{S}_+^n$:

$f(X)=\langle\mathcal{S},X\rangle$-In det (X) and $g(X)=\rho\|X\|_1$,

We obtain that (3.3) is equivalent to

$\min_{X\in\mathcal{S}^n}\{f(X)+g(Y): , X\in\mathcal{S}_+^n\}$,

and it can be rewritten as:

$$\min_{X,Y\in\mathcal{S}^n}\{f(X)+g(Y): X-Y=0, X,Y\in\mathcal{S}_+^n\} \quad (3.4)$$

Thus (3.3) is a particular case of (3.1) if we fixe $\mathbb{V}_1=\mathbb{V}_2=\mathcal{S}^n, \mathcal{K}_1=\mathcal{K}_2=\mathcal{S}_+^n$, and $\mathbb{A}$, $\mathbb{B}:\mathcal{S}^n\to\mathbb{R}^{\widetilde{m}}$, where $\widetilde{m}=\frac{1}{2}n(n+1), \mathbb{A}(X)=svec(X)$, and $\mathbb{B}(Y)=-svec(Y)$ with $svec(X)=$ $\left(x_{11},\sqrt{2}x_{12},\dots,\sqrt{2}x_{1n},x_{22},\sqrt{2}x_{23},\dots,\sqrt{2}x_{2n},\dots,x_{nn}\right)$

(see [14]).

3.2. Proximal Multiplier Algorithm

Let $H_1:\mathbb{V}_1\times\mathbb{V}_1\to\mathbb{R}\cup\{+\infty\}$ be a function satisfying $H_1\in\mathcal{D}\big(int(\mathcal{K}_1)\big)$. Consider another function $H_2:\mathbb{V}_2\times\mathbb{V}_2\to\mathbb{R}\cup\{+\infty\}$ with $H_2\in\mathcal{D}\big(int(\mathcal{K}_2)\big)$ and $\theta_1,\theta_2>0$ positive parameters. We define

$$H_{\theta_1}(x_1,x_2)\coloneqq H_1(x_1,x_2)+\frac{\theta_1}{2}\|x_1-x_2\|^2 \quad (3.5)$$

$$H_{\theta_2}(z_1,z_2)\coloneqq H_2(z_1,z_2)+\frac{\theta_2}{2}\|z_1-z_2\|^2 \quad (3.6)$$

It is easy to show that for each $i=1,2, H_{\theta_i}$ is also a proximal distance with respect to $int(\mathcal{K}_i)$, that is $H_{\theta_i}\in\mathcal{D}\big(int(\mathcal{K})\big)$.

The proposed algorithm, called **SC-PMA,** which means Symmetric cone Proximal Multiplier Algorithm, for solving the problem (P) is defined by:

3.3. Algorithm SC-PMA

Let $H_i\in\mathcal{D}\big(int(\mathcal{K}_i)\big), \theta_i>0, i=1{,}2, tol>0$ and $\{\varepsilon_k\},\{\zeta_k\},\{\lambda_k\}$ be sequences of positive scalars.

Step 0: Start with some initial point $\omega^0=(x^0,z^0,y^0)\in int(\mathcal{K}_1)\times int(\mathcal{K}_2)\times\mathbb{R}^m$. Set $k=0$.

Step 1: Compute

$$p^{k+1}=y^k+\lambda_k(\mathbb{A}x^k+\mathbb{B}z^k-b), \quad (3.7)$$

Step 2: Find $(x^{k+1},z^{k+1})\in int(\mathcal{K}_1)\times int(\mathcal{K}_2)$ and $(g^{k+1},g^{k+2})\in\mathbb{V}_1\times\mathbb{V}_2$, such that

$$g_1^{k+1}\in\partial_{\varepsilon_k}f(x^{k+1})$$
$$g_1^{k+1}+\mathbb{A}^*p^{k+1}+\frac{1}{\lambda_k}\nabla_x H_{\theta_1}(x^{k+1},x^k)=0 \quad (3.8)$$
$$g_2^{k+1}\in\partial_{\zeta_k}g(z^{k+1})$$
$$g_2^{k+1}+\mathbb{B}^*p^{k+1}+\frac{1}{\lambda_k}\nabla_z H_{\theta_2}(z^{k+1},z^k)=0 \quad (3.9)$$

Step 3: Compute

$$y^{k+1}=y^k+\lambda_k(\mathbb{A}x^{k+1}+\mathbb{B}z^{k+1}-b) \quad (3.10)$$

Step 4: Set $w^{k+1}=(x^{k+1},z^{k+1},y^{k+1})$. If $\|w^{k+1}-w^k\|\le tol$, stop; otherwise replace k by $k+1$ and go to Step 1.

The next lemma is a well know property related to proximal point algorithms; see Theorem 2.1 of [4] or Lemma 2.2 of [13] for proof of that result.

Lemma 3.1. Let $F:\mathbb{V}\to\mathbb{R}\cup\{+\infty\}$ be a closed proper convex function, $H\in\mathcal{D}\big(int(\mathcal{K})\big)$ satisfying (B1) or (B1') and $\lambda_k>0$. If $\{(u^k,g_1^k)\}_{k\in\mathbb{N}}$ and $\{(v^k,g_2^k)\}_{k\in\mathbb{N}}$ are two sequences satisfying

$$g_1^{k+1}\in\partial_{\varepsilon_k}F(u^{k+1})$$
$$g_1^{k+1}+\lambda_k^{-1}\nabla_x H(u^{k+1},u^k)=0$$
$$\tilde{g}_2^{k+1}\in\partial_{\zeta_k}F(v^{k+1})$$
$$\tilde{g}_2^{k+1}+\lambda_k^{-1}H(v^{k+1}-v^k)=0$$

Then, for any $k\ge 0$, we obtain

$\lambda_k\big(F(u^{k+1})-F(u)\big)\le H(u,u^k)-H(u,u^{k+1})+\lambda_k\varepsilon_k, \forall u\in C_1$, if (B1) holds;

$\lambda_k\big(F(u^{k+1}) - F(u)\big) \le H(u^k, u) - H(u^{k+1}, u) + \lambda_k \varepsilon_k, \forall u \in C_2$, if (B1') holds;

$$2\lambda_k\big(F(v^{k+1}) - F(v)\big) \le \|v^k - v\|^2 - \|v^{k+1} - v\|^2 - \|v^{k+1} - v^k\|^2 + 2\lambda_k \zeta_k, \forall v \in \mathbb{R}^n$$

3.4. Convergence Analysis of Algorithm SC-PMA

In this section, we prove the convergence of the iterates $\{(x^{k+1}, z^{k+1}\}$, to an optimal solution of the primal problem (P) as also the convergence of $\{y^{k+1}\}$ to the optimal Lagrangian multiplier of (P).

We impose the following assumptions:

(A) Problem (P) admits at least an optimal solution (x^*, z^*).

(B) $\exists x \in int(\mathcal{K}_1) \cap ri\big(dom(f)\big), z \in int(\mathcal{K}_2) \cap ri\big(dom(g)\big)$ Such that $\mathbb{A}x + \mathbb{B}z = b$.

(C) The sequences $\{\zeta_k\}_{k\in\mathbb{N}}$ and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ are nonnegative, and $\sum_{k=0}^{\infty}(\zeta_k + \varepsilon_k) < \infty$.

Remark 3.2. Observe that (B) *guarantee the existence of an optimal dual Lagrange multiplier* y^*. *Hence, under conditions (A), (B)* (x^*, z^*, y^*) *is a saddle point of* L, *that is,*

$$L(x^*, z^*, y) \le L(x^*, z^*, y^*) \le L(x, z, y^*), \forall x \in \mathcal{K}_1 \cap dom(f), z \in \mathcal{K}_2 \cap dom(g), y \in \mathbb{R}^m$$

where

$$L(x, z, y) = f(x) + g(z) + \langle y, \mathbb{A}x + \mathbb{B}z \pm b\rangle,$$
$$x \in \mathcal{K}_1, z \in \mathcal{K}_2, y \in \mathbb{R}^m$$

denotes the Lagrangian function for the problem (P).

The following result shows that the algorithm SC-PMA is well-defined

Proposition 3.1. *Under assumptions (A), (b),* $H_i \in \mathcal{D}\big(int(\mathcal{K}_i)\big), i = 1,2,$ *and* $(x^k, z^k, y^k) \in int(\mathcal{K}_1) \times int(\mathcal{K}_2) \times \mathbb{R}^m$, *there exists a unique point* $(x^{k+1}, z^{k+1}) \in int(\mathcal{K}_1) \times int(\mathcal{K}_2)$ *satisfying (3.8)-(3.9).*

Proof. Similar to Theorem 4.1 of [20].

The next result establishes estimates for the sequences generated by the SC-PMA.

Proposition 3.2. *Let* $\{x^k, z^k, y^k, p^k\}_{k\in\mathbb{N}}$ *be the sequence generated by SC-PMA,* (x^*, z^*) *be an optimal solution of (P) and* y^* *be a corresponding Lagrange multiplier. Suppose that* $H_i \in \mathcal{D}\big(int(\mathcal{K}_i)\big)$, *for* $i = 1,2$, *and (B1) or (B1') holds. Then, for all* $k \ge 0$, we have

$$H_{\theta_1}(x^*, x^{k+1}) + H_{\theta_2}(z^*, z^{k+1}) \le H_{\theta_1}(x^*, x^k) + H_{\theta_2}(z^*, z^k) - \gamma_1 H_{\theta_1}(x^{k+1}, x^k) - \gamma_2 H_{\theta_2}(z^{k+1}, z^k) - \lambda_k\langle p^{k+1} - y^*, \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b\rangle + \lambda_k(\zeta_k + \varepsilon_k), \quad (3.12)$$

if (B1) holds;

and

$$H_{\theta_1}(x^{k+1}, x^*) + H_{\theta_2}(z^{k+1}, z^*) \le H_{\theta_1}(x^k, x^*) + H_{\theta_2}(z^k, z^*) - \gamma_1' H_{\theta_1}(x^{k+1}, x^k) - \gamma_2' H_{\theta_2}(z^{k+1}, z^k) - \lambda_k\langle p^{k+1} - y^*, \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b\rangle + \lambda_k(\zeta_k + \varepsilon_k), if\ (B1')\ holds \quad (3.13)$$

Proof. Assume that (B1) holds. Using Lemma 3.1, Part (i) with $F(\cdot) := f(\cdot) + \langle p^{k+1}, \mathbb{A}\,\cdot\rangle$ and with $F(\cdot) := g(\cdot) + \langle p^{k+1}, \mathbb{B}\,\cdot\rangle$ we have that, for all $x \in \mathcal{C}_1$:

$$\lambda_k(f(x^{k+1}) + \langle p^{k+1}, \mathbb{A}x^{k+1}\rangle - f(x) - \langle p^{k+1}, \mathbb{A}x\rangle) \le H_{\theta_1}(x, x^k) - H_{\theta_1}(x, x^{k+1}) - \gamma_1 H_{\theta_1}(x^{k+1}, x^k) + \lambda_k \varepsilon_k,$$

and for all $z \in \mathcal{C}_1$:

$$\lambda_k(g(z^{k+1}) + \langle p^{k+1}, \mathbb{B}z^{k+1}\rangle - g(z) - \langle p^{k+1}, \mathbb{B}z\rangle) \le H_{\theta_2}(z, z^k) - H_{\theta_2}(z, z^{k+1}) - \gamma_2 H_{\theta_2}(z^{k+1}, z^k) + \lambda_k \zeta_k.$$

Adding the inequalities above, we obtain

$$\lambda_k\big(L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\big) \le H_{\theta_1}(x, x^k) - H_{\theta_1}(x, x^{k+1}) - \gamma_1 H_{\theta_1}(x^{k+1}, x^k) + H_{\theta_2}(z, z^k) - H_{\theta_2}(z, z^{k+1}) - \gamma_2 H_{\theta_2}(z^{k+1}, z^k) + \lambda_k(\zeta_k + \varepsilon_k) \quad (3.14)$$

The other side, as (x^*, z^*, y^*) is a saddle point of L we have

$$\lambda_k\big(L(x^*, z^*, p^{k+1}) - L(x^{k+1}, z^{k+1}, y^*)\big) \le 0.$$

Using (3.11) with $x = x^*, z = z^*$, and adding the above inequality and after rearranging terms, we obtain (3.9). The inequality (3.10) is obtained by using the same arguments but using the property (B1').

Proposition 3.3. *Let* $\{x^k, z^k, y^k, p^k\}_{k\in\mathbb{N}}$ *be the sequence generated by algorithm (SC-PMA),* (x^*, z^*) *be an optimal solution of (P) and* y^* *be a corresponding Lagrange multiplier. Then, the following inequalities hold for all* $k \ge 0$

$$\lambda_k\langle \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b, y^* - y^{k+1}\rangle \le \tfrac{1}{2}(\|y^k - y^*\|^2 - \|y^{k+1} - y^*\|^2 - \|y^{k+1} - y^k\|^2), \quad (3.15)$$

$\lambda_k\langle \mathbb{A}x^{k+1} + \mathbb{B}z^{k+1} - b, y^{k+1} - p^{k+1}\rangle \leq \frac{1}{2}(\|y^k - y^{k+1}\|^2 - \|p^{k+1} - y^{k+1}\|^2 - \|p^{k+1} - y^k\|^2)$ (3.16)

Proof. The inequalities (3.15)-(3.16) follow directly from [3, Proposition 2] or [11, Lemma 4.1].

For any vector $w_1 = (x_1, z_1, y_1) \in \mathcal{C}_1 \times \mathcal{C}_1 \times \mathbb{R}^m$ and $w_2 = (x_2, z_2, y_2) \in \mathcal{C}_2 \times \mathcal{C}_2 \times \mathbb{R}^m$ we define $\widehat{H}_\theta(w_1, w_2) = H_{\theta_1}(x_1, x_2) + H_{\theta_2}(z_1, z_2) + \frac{1}{2}\|y_1 - y_2\|^2$. (3.17)

Lemma 3.2. *Let $\{x^k, z^k, y^k, p^k\}_{k\in\mathbb{N}}$ be the sequence generated by the algorithm (SC-PMA) and let $w^* = (x^*, z^*, y^*)$ with (x^*, z^*) an optimal solution of (P) and y^* its corresponding Lagrange multiplier. Assume that $H_i \in \mathcal{D}(int(k_i))$, for $i = 1,2$. If (B1) holds, then*

$$\begin{aligned}\widehat{H}_\theta(w^*, w^{k+1}) \leq \widehat{H}_\theta(w^*, w^k) &- \frac{1}{2}(\theta_1\gamma_1 - 4\lambda_k^2\|\mathbb{A}\|^2)\|x^{k+1} - x^k\|^2 \\ &- \frac{1}{2}(\gamma_2\theta_2 - 4\lambda_k^2\|\mathbb{B}\|^2)\left\|z^{k+1} - \mathrm{z}^k\right\|^2 \\ &- \frac{1}{2}\|p^{k+1} - y^{k+1}\|^2 - \frac{1}{2}\|p^{k+1} - y^k\|^2 \\ &+ \lambda_k(\zeta_k + \varepsilon_k); \qquad (3.18)\end{aligned}$$

And if (B1') holds, then

$$\widehat{H}_\theta(w^{k+1}, w^*) \leq \widehat{H}_\theta(w^k, w^*) - \frac{1}{2}(\theta_1 - 4\lambda_k^2\|\mathbb{A}\|^2)\|x^{k+1} - x^k\|^2 - \frac{1}{2}(\theta_2 - 4\lambda_k^2\|\mathbb{B}\|^2)\|z^{k+1} - z^k\|^2 - \frac{1}{2}\|p^{k+1} - y^{k+1}\|^2 - \frac{1}{2}\|p^{k+1} - y^k\|^2 + \lambda_k(\zeta_k + \varepsilon_k); \qquad (3.19)$$

Let $w^{k+1} = (x^{k+1}, z^{k+1}, y^{k+1})$ and $w^k = (x^k, z^k, y^k)$.

Then, adding (3.12) to the above inequality, we get

$$\widehat{H}_\theta(w^*, w^{k+1}) \leq \widehat{H}_\theta(w^*, w^k) - \frac{\gamma_1\theta_1}{2}\|x^{k+1} - x^k\|^2 - \frac{\gamma_2\theta_2}{2}\|z^{k+1} - z^k\|^2 - \frac{1}{2}(\|p^{k+1} - y^{k+1}\|^2 - \|p^{k+1} - y^k\|^2) + \rho_k + \lambda_k(\zeta_k + \varepsilon_k); \qquad (3.20)$$

where $\rho_k = \lambda_k\langle y^{k+1} - p^{k+1}, \mathbb{A}(x^{k+1} - x^k) + \mathbb{B}(z^{k+1} - z^k)\rangle$. Now, by using (3.9) and (3.10), it follows that

$$\rho_k = \lambda_k^2\|\mathbb{A}(x^{k+1} - x^k) + \mathbb{B}(z^{k+1} - z^k)\|^2 \leq 2\lambda_k^2(\|\mathbb{A}\|^2\|x^{k+1} - x^k\|^2 + \|\mathbb{B}\|^2\|z^{k+1} - z^k\|^2)$$

Hence, the result follows by employing this inequality in (3.20). The inequality (3.19) follows by applying the same arguments above and the property (B1 ′).

Theorem 3.1. *Let $\{x^k, z^k, y^k, p^k\}_{k\in\mathbb{N}}$ be the sequence generated by the algorithm (SC-PMA) and let $w^* = (x^*, z^*, y^*)$ with (x^*, z^*) an optimal solution of (1) and y^* its corresponding Lagrange multiplier. Assume that $H_i \in \mathcal{D}(int(k_i))$, for $i = 1,2$, and (B1)-(B3) or (B1´),(B3´) hold. If $\{\lambda_k\}$ satisfies*

$$\lambda_k\|\mathbb{A}\| \leq \frac{1}{2}(\gamma_1\theta_1 - \vartheta)^{\frac{1}{2}},$$

$$\lambda_k\|\mathbb{B}\| \leq \frac{1}{2}(\theta_2 - \vartheta)^{\frac{1}{2}}, \forall k \geq 0, \qquad (3.21)$$

For some $\eta > 0$ and $0 < \vartheta < min\{\theta_1, \theta_2\}$, then the following hold:

The sequence $w^k = (x^k, z^k, y^k)$ is bounded, and every limit point of w^k is a saddle point of the Lagrangian.

Furthermore, if (B4)-(B5) or (B4´)-(B5´) hold, then the sequence $\{(x^k, z^k, y^k)\}$ globally converges to a solution to the problem (P).

Proof. (i) Assume that (B1)-(B2) hold. Since λ_k satisfies (3.18), from (3.15), we have that

$$\widehat{H}_\theta(w^*, w^{k+1}) \leq \widehat{H}_\theta(w^*, w^k) - \frac{\vartheta}{2}(\|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2) - \frac{1}{2}(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) + \lambda_k(\zeta_k + \varepsilon_k) \qquad (3.22)$$

This implies that $\{w^k\}_{k\in\mathbb{N}} \subseteq \{w \in \mathrm{int(k)} \times \mathbb{R}^p \times \mathbb{R}^m : \widehat{H}_\theta(w^*, w) \leq \bar{\alpha}\}$, with $\bar{\alpha} = \widehat{H}_\theta(w^*, w^0) + \sum_{k=0}^{\infty}\lambda_k(\zeta_k + \varepsilon_k)$. By assumption (B2) and the fact that $\sum_{k=0}^{\infty}\lambda_k(\zeta_k + \varepsilon_k) < \infty$ (cf. Assumption (A3) and (3.18)), it follows that the sequence $\{w^k\}_{k\in\mathbb{N}}$ is bounded. Moreover, (3.19), together with $\widehat{H}_\theta(w^*, w^{k+1}) \geq 0$ implies that there exists $l(w^*) \geq 0$ such that

$$\lim_{k\to\infty}\widehat{H}_\theta(w^*, w^k) = l(w^*). \qquad (3.23)$$

Therefore, by taking the limits on both sides of (3.21), we obtain

$$\|x^{k+1} - x^k\| \to 0, \|z^{k+1} - z^k\| \to 0,$$
$$\|p^{k+1} - y^{k+1}\| \to 0, \|p^{k+1} - y^k\| \to 0. \qquad (3.24)$$

On the other hand, since $(w^k)_{k\in\mathbb{N}}$ is bounded, there exists a subsequence $\{w^{kj} = (x^{kj}, z^{kj}, y^{kj})\}_{j\in\mathbb{N}}$ and a limit point $w^\infty = (x^\infty, z^\infty, y^\infty)$ such that $w^{kj} \to w^\infty$. We now proceed to show that w^∞ is a saddle point of L. First, since $\lambda_k \geq \eta$, passing to the limit in (3.14) on the subsequence, and using (B3) and (3.24), we obtain

$$L(x^\infty, z^\infty, y^\infty) \leq L(x, z, y^\infty), \quad \forall x \in \mathcal{C}_1, \ \forall z \in \mathcal{C}_2. \qquad (3.25)$$

Second, by applying Lemma 3.1, Part (iii) (in its exact form) with $F(\cdot) \coloneqq -L(x^{k+1}, z^{k+1}, \cdot)$, we have

$$\lambda_k\left(L(x^{k+1}, z^{k+1}, y) - L(x^{k+1}, z^{k+1}, y^{k+1})\right) \\ \leq \frac{1}{2}(\|y^k - y\|^2 \|y^{k+1} - y\|^2), \\ \forall y \in \mathbb{R}^m.$$

Taking the limit on the subsequence in the above inequality and using (3.23), we have

$$L(x^\infty, z^\infty, y) \leq L(x^\infty, z^\infty, y^\infty), \quad \forall y \in \mathbb{R}^m. \quad (3.26)$$

It follows from (3.23) that $Ax^\infty + Bz^\infty = b$. Finally, since $\{x^k\}_{k\in\mathbb{N}} \subset \mathrm{int}(k_1)$ and $\{z^k\}_{k\in\mathbb{N}} \subset \mathrm{int}(k_2)$, passing to the limit one has $(x^\infty, z^\infty) \in k_1 \times k_2$. Hence, the result follows from (3.22)-(3.23). In the case that (B1') (B3') hold, the proof is similar to above and by using (P4) of Definition 2.1 instead of (B2).

(ii) Suppose that (B4)-(B5) holds. Let w^∞ be the limit of a subsequence $\{w^{kj}\}_{j\in\mathbb{N}}$ of $\{w^k\}_{k\in\mathbb{N}}$, that is, $w^{kj} \to w^\infty$. Then, by (B4), we have

$$\lim_{j\to\infty} \widehat{H}_\theta(w^\infty, w^{kj}) = 0. \quad (3.27)$$

Since w^∞ is a saddle point of L (by Part (i)), it follows from (3.2.3) and (3.27) that $l(w^\infty) = 0$. Hence, by using (B5), we obtain that the sequence $\{w^k\}_{k\in\mathbb{N}}$ converges to w^∞. Now, if (B4')- (B5') keeps the result same result obtained previously.

4. Numerical Experiment

This section presents an implementation of the proposed SC-PMA applied to find linear hyperplanes in binary classification in SVM.

Given a set of points $x^1, x^2, \dots, x^m \in \mathbb{R}^n$ with their respective labels $y^1, y^2, \dots, y^m \in \{-1, +1\}$, we form the m-upla $(x^1, y^1), (x^2, y^2), \dots, (x^m, y^m)$. The objective of the SVM is to determine an optimal hyperplane

$$H(w.\alpha) = \{x \in \mathbb{R}^n : w^T x + \alpha = 0\},$$

where $w \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, which separates the given points.

As we observed in Subsection A of Section III, that class of problems could be expressed as our model (3.1), that is,

$$\min_{z.v}\{f(z) + g(v) : Az + Bv = b, v \geq 0\},$$

Where the variables are:

$z = (w, \alpha) \in \mathbb{R}^{n+1}$ and $v = (\xi, u) \in \mathbb{R}^{2m}$ with

$$\xi = (\xi_1, \dots, \xi_m) \in \mathbb{R}^m \ with\ \xi_i \geq 0$$

$$u = (u_1, \dots, u_m) \in \mathbb{R}^m \ with\ u_i \geq 0$$

The separable functions are

$f(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{w}\|^2$ and $g(\boldsymbol{v}) = C\boldsymbol{e}^T\boldsymbol{\xi}$,

The matrix A and B are the following

$$A = Y\hat{X} \in \mathbb{R}^{m\times(n+1)}, \qquad B = (I - I) \in \mathbb{B}^{m\times 2m}$$

where

$$Y = Diag(\boldsymbol{y}) = Diag(y^1, y^2, \dots, y^m)$$

and

$$\hat{X} = \begin{pmatrix} x^{1\,T} & 1 \\ \vdots & \vdots \\ x^m & 1 \end{pmatrix} \in \mathbb{R}^{m\times(n+1)},$$

Finally, the vector b is given by $b = \boldsymbol{e} = (1,1,\dots,1) \in \mathbb{R}^m$.

We use MATLAB Software (R2017a), a computer 8th Gen Intel (R) Core (TM) *i5-8250U CPU*, 1.60 GHz, 1.80 GHz, 4.00 GB, Windows 1064 bits.

We will give three examples of finding optimal hyperplanes using the SC-PMA with the amount of data of m=10, m=50 and m=100, respectively, obtained by an implementation using the function Rand of MATLAB.

4.1. Subheadings

The results and discussion may be presented separately or in one combined section and optionally divided into headed subsections.

The parameters to enter are:
Data: Linearly separable data set

MAX_ITER: Maximum number of iterations (the number of maximum interactions given to the algorithm)

GRAF_CON : (1) to graph and (0) to not graph (with (1) it shows the graph of the plane in another case, (0) it does not show the separating hyperplane)

ITER_CON : (1) to show convergence and (0) to not show (with (1) it shows the graph of the convergence of the optimal point also of "z" and "v" another case (0) does not present the separating hyperplane)

Rho: Acceleration parameter

lambda: Step size in each iteration

initial point $\omega^0 = (z^0, v^0, y^0) \in \mathbb{R}^{n+1} \times \mathbb{R}^{2m} \times \mathbb{R}^m$ an arbitrary starting point.

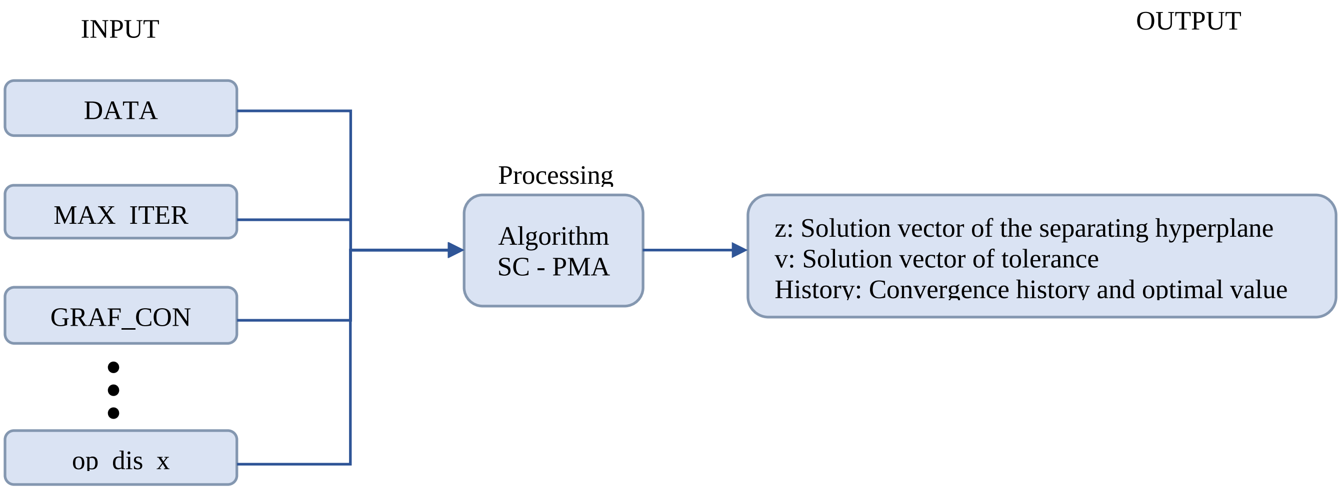

Fig. 1 SC-PMA Processing Diagram

op_dis_f : Proximal distance for f

op_dis_g : Proximal distance for g

op_dis_x : Proximal Distance Type

In the SC-PMA, it is necessary to have two proximal distances. The first distance related to the variable $\boldsymbol{z} = (\boldsymbol{w}, \boldsymbol{\alpha}) \in \mathbb{R}^{\boldsymbol{n+1}}$ (associated with op_disf):

$$H(z, y) = \sqrt{\sum_{i=1}^{n+1} (z_i - y_i)^2}$$

The other distance is related to the variable $\boldsymbol{v} = (\boldsymbol{\xi}, \boldsymbol{u}) \in \mathbb{R}^{\boldsymbol{2m}}$(associated to op_dis_g). As this variable has conditions of nonnegativity, that is, $\boldsymbol{\xi_i} \geq \boldsymbol{0}$ and $\boldsymbol{u_i} \geq \boldsymbol{0}$, for i=1,2,...,m; we can choose the implementation of any of the following distances:

Choosing the Kullback–Leibler Bregman distance:

$$H(v, y) = \sum_{i=1}^{2m} \left(v_i log\left(\frac{v_i}{y_i}\right) + y_i - v_i \right).$$

Itakura saito proximal distance:

$$H(v, y) = \sum_{i=1}^{2m} \left(\frac{v_i}{y_i} - \ log\left(\frac{v}{y_i}\right) \right) - 1.$$

Second-order homogeneous proximal distance

$$H(z, y) = \sum_{i=1}^{2m} \frac{y}{2}\ (z_i - y_i)^2 + \sigma(y_i^2 log\frac{y_i}{z_i} + z_i y_i - y_i^2).$$

We should observe that these distances are used to measure the separation of the points to the separating hyperplane. Furthermore, as the implementation of the algorithm depends on the employed proximal distance, these distances are also used to compare which of them converges better.

The outputs after the process of the SC-PMA are the following:

z: Solution vector of the separating hyperplane

v: Solution vector of tolerances

History: Convergence history and optimal value

History is very important to obtain tables I, II and III, where we are going to show

-the number of iterations

-the number of inner iterations to solve each subproblem (3.8) and (3.10) denoted by $N(z_{\{k\}}), N(v_{\{k\}})$, respectively.

- The difference between consecutive points of each variable: $\|z_{\{k\}} - z_{\{k-1\}}\|$, $\|v_{\{k\}} - v_{\{k-1\}}\|$ and $\|w_{\{k\}} - w_{\{k-1\}}\|$

- The value of the objective function in each iteration.

4.2. Linear Generation

A linear generator was implemented to generate the data for the algorithm test, which randomly finds the points in the plane. We use the rand command of the Matlab software, and the number of points is chosen with the requirements described below.

The input to generate the data are the following:

- n_var: Number of variables (Depending on the dimension, it can be two-dimensional or three-dimensional.)
  In our case, we use the value n_var =3 because the points belong to this dimension.
- n_obs: Number of observations (Amount of data to be classified with the algorithm).
  The value of n_obs change depending on the amount of data used. In our implementation, we use the values of 10, 50 and 100.
- Minimo: Minimum value of the squares (Lower bound of the interval)
- Maximo: Maximum value of the squares (Highest bound of the interval)
- GRAF : (1) for graphic(Presents the data set in three-dimensional space) and (0) for no graphic(Does not present the data set in three-dimensional space).

The output process of the SC-PMA is:

- Data: Generated data that is linearly separable.

We will generate points in $\mathbb{R}^3$, so the value of n in the model (3.2) is 3, that is,$n = 3$. The implementation of this linear generation is shown in the appendix.

4.3. Sub-program Main

This program integrates the linear generator with the SC-PMA. After generating the data through the linear generator, the main subprogram of the implementation calls the SC-PMA. It proceeds to find the separating plane, the number of iterations and the execution time.

Thus, we have as results the separation plane, the optimal value, as well as the solution points "z" and "v". As $z = (w, \alpha) \in \mathbb{R}^{n+1}$, then we will obtain the optimal hiperplane defined by

$$H(w.\alpha) = \{x \in \mathbb{R}^n : w^T x + \alpha = 0\}.$$

To execute the SC-PMA is necessary to solve the problem in each iteration of the subproblems (3.9) and (3.10). In this paper, we use the fmincon command.

We present three numerical examples of data created by an aleatory linear generator. For this presentation, we consider the second-order homogeneus proximal distance:

$$H_\rho(u, v) = \sum_{i=1}^{n} \rho \left(v_i^2 \ln \frac{v_i}{u_i} + u_i v_i - v_i^2 \right) + \frac{\rho}{2} \sum_{i=1}^{n} (u_i - v_i)^2$$

and different values for the proximal parameter: $\lambda = 0.09$ for the first example, $\lambda = 0.05$. For the second example and $\lambda = 0.01$ for the third example. We observe that as λ decreases, the iterations of the algorithm are smaller.

4.3.1. Example 1

For this example, we will consider the following data set, which is linearly separable, as shown in Fig 2. Also, we consider

$$\lambda_k = 0.09, \rho = 1, tol = 10^{-3}$$

and the maximum number of iterations 1500.

In the following figure, Fig 3, we see the convergence of the objective function $f + g$, the variable z and v. When we write convergence of the objective function, we refer to the difference in absolute value between the objective function $f + g$ obtained by the SC-PMA and the objective function $f + g$ obtained by the CVX function of Matlab.

Figure 4 shows the set of 10 observations distributed in three-dimensional space and are linearly separable.

Elapsed time is 2344.122166 seconds; the convergence occurred in iteration 967. Fig 4 shows the data set separated by a plane, whose coefficients of the equation are:

$-0.2913; -0.1007; -0.4574; 0.0251;$

and its respective hyperplane is :
$-0.2913X - 0.1007Y - 0.4574Z + 0.0251 = 0.$

Table I shows the computational results obtained after processing the data. Thus we see the convergence to the optimal value, the number of iterations, the module of the point difference of z_k with respect to the previous z_{k-1} in the same way, for v_k with respect to the previous v_{k-1} and so also from w_k with respect to the previous w_{k-1}.

Table 1. Computational Results for Example 1

Iteration	$N(z_{\{k\}})$	$N(v_{\{k\}})$	$\lVert z_{\{k\}} - z_{\{k-1\}} \rVert$	$\lVert v_{\{k\}} - v_{\{k-1\}} \rVert$	$\lVert w_{\{k\}} - w_{\{k-1\}} \rVert$	**Objective function**
1	4	14	0.78167	0.32408	1.43995	3.18318
2	9	22	0.48174	0.23141	1.25799	2.56337
3	8	19	0.87878	0.25250	0.87878	1.51274
4	8	24	0.93322	0.25200	0.93322	1.00437
5	8	20	0.72358	0.20552	0.72358	1.05220
⋮	⋮	⋮	⋮	⋮	⋮	⋮
966	5	33	0.00014	0.00087	0.00104	0.50203
967	5	32	0.00015	0.00093	0.00098	0.50188

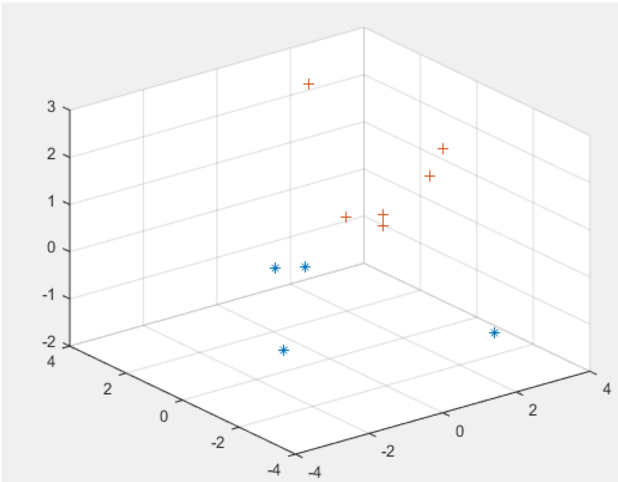

Fig. 2 Linear separable data

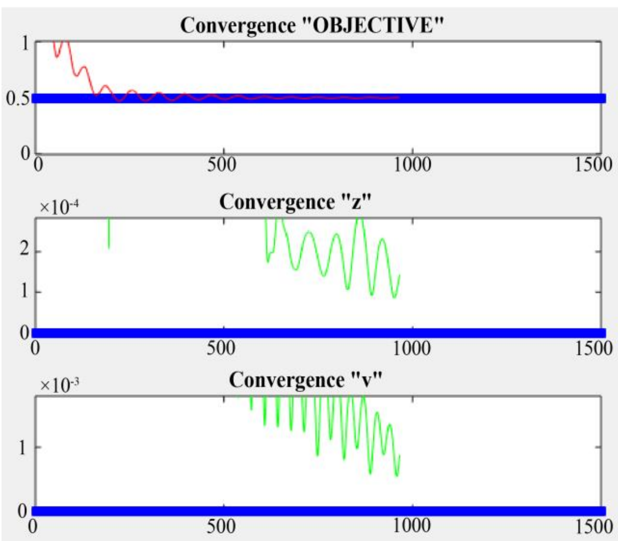

Fig. 3 Convergence of the objective function and the norms $\|z_k - z_{k-1}\|$ y $\|v_k - v_{k-1}\|$

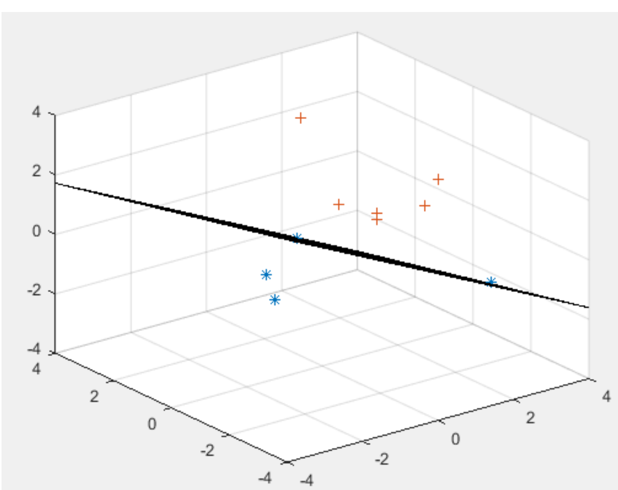

Fig. 4 The data set is separated by a plane

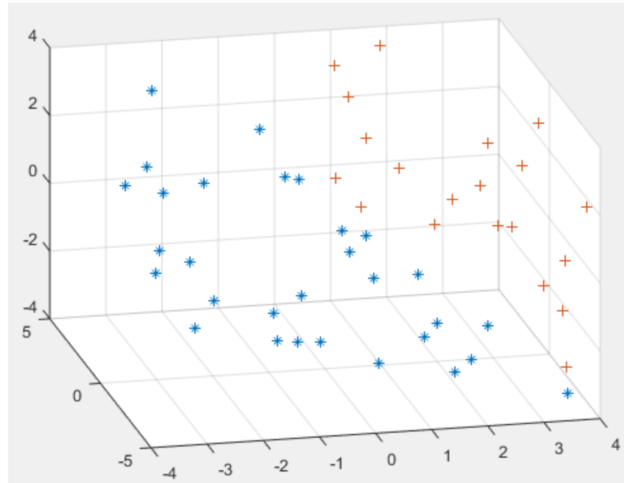

Fig. 5 Linear separable data

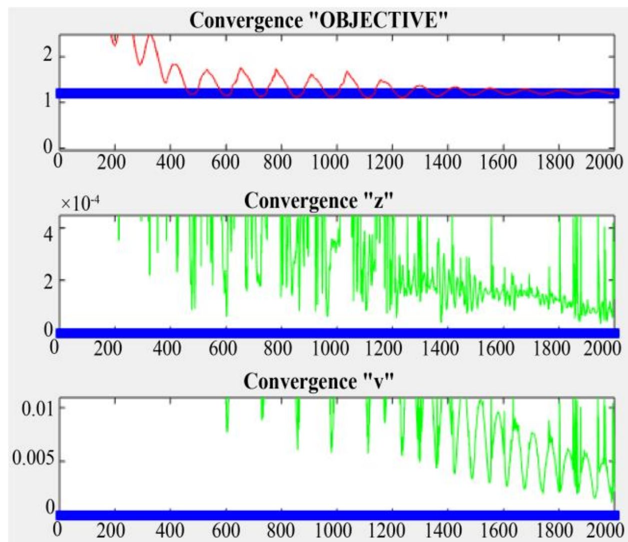

Fig. 6 Convergence of the objective function and the norms $\|z_k - z_{k-1}\|$ y $\|v_k - v_{k-1}\|$

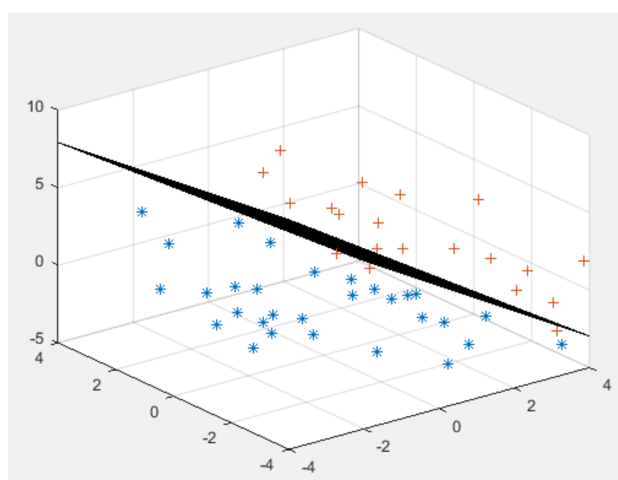

Fig. 7 The Data set is separated by a plane

Table 2. Computational Results for Example 2

Iteration	$N(z_{\{k\}})$	$N(v_{\{k\}})$	$\lVert z_{\{k\}} - z_{\{k-1\}} \rVert$	$\lVert v_{\{k\}} - v_{\{k-1\}} \rVert$	$\lVert w_{\{k\}} - w_{\{k-1\}} \rVert$	Objective function
1	12	29	2.73490	0.44568	3.48678	13.55714
2	10	22	1.78088	0.33687	2.36787	8.60897
3	9	29	2.31439	0.33125	2.31439	4.92598
4	10	29	1.37369	0.25792	1.37369	4.92598
5	8	28	0.34034	0.20416	1.01441	4.96926
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1999	6	29	0.00014	0.00265	0.00430	1.19902
2000	4	29	0.00013	0.00220	0.00423	1.19951

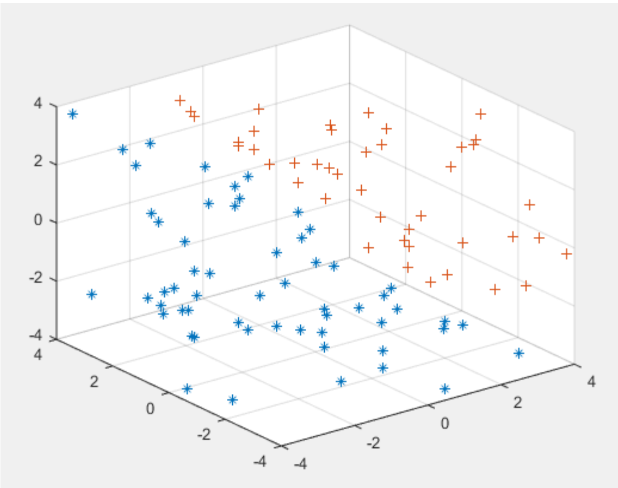

Fig. 8 Linear separable data

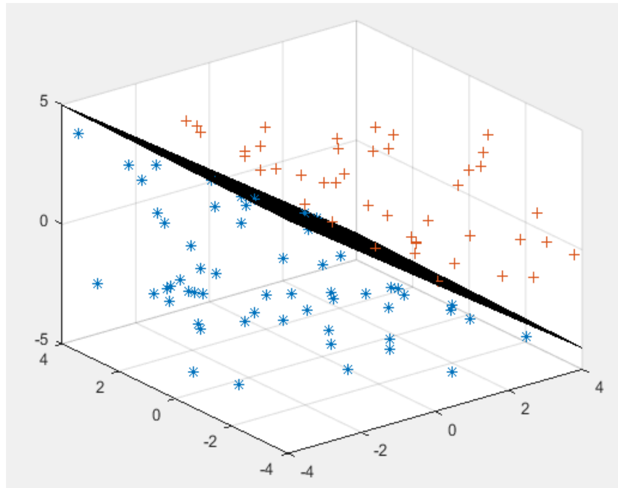

Fig. 10 Data set is separated by a plane

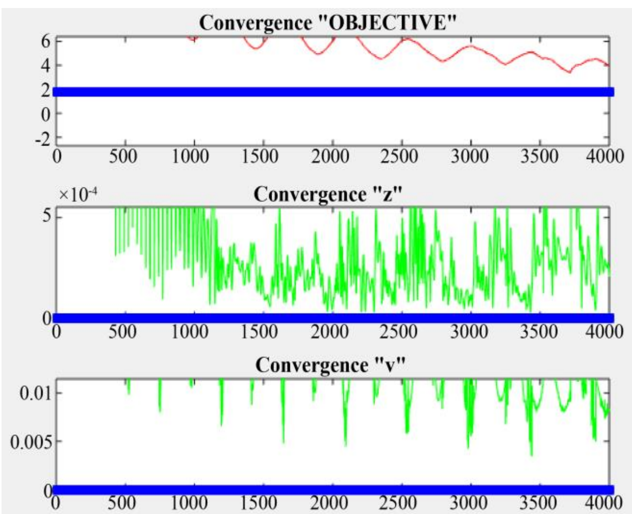

Fig. 9 Convergence of the objective function and the norms $\lVert z_k - z_{k-1} \rVert$ y $\lVert v_k - v_{k-1} \rVert$

4.3.2. Example 2

For this example, we will consider $\lambda = 0.05, \rho = 1, tol = 10^{-3}$ as also Max.Iterations $= 2000$. Fig 5 shows the set of 50 observations distributed in three-dimensional space and are linearly separable. In Fig 6, we see converging to the objective function and the variables "z" and "v" to an optimal point.

It can be observed that the convergence of the optimal value is fast in the same way as z. Still, it is slower to the parameter v. As in the previous example, the convergence of the objective function refers to the difference in absolute value between the objective function $f + g$ obtained by the SC-PMA and the objective function $f + g$ obtained by the CVX function of Matlab.

Elapsed time is 4930.763008 seconds, and Fig 7 shows the data set separated by a plane, whose coefficients of the equation are: -0.7259; -0.1075; -0.4545; 1.1248; its respective plane is:

$$-0.7259X - 0.1075Y - 0.4545Z + 1.1248 = 0$$

Table II shows the computational results for data from 50 observations,

4.3.3. Example 3

For this example, we will consider $\lambda = 0.01, \rho = 1, tol = 10^{-3}$ as also Max.Iterations $= 4000$. Fig 8 shows the set of 100 observations distributed in three-dimensional space and are linearly separable.

In Fig 9, we see converging to the objective function, the variable "z" and "v", respectively. As in the previous example, the convergence of the objective function refers to -

Table 3. Computational Results for Example 3

Iteration	$N(z_{\{k\}})$	$N(v_{\{k\}})$	$\|z_{\{k\}} - z_{\{k-1\}}\|$	$\|v_{\{k\}} - v_{\{k-1\}}\|$	$\|w_{\{k\}} - w_{\{k-1\}}\|$	Objective function
1	14	14	1.77667	0.14247	1.77667	15.20005
2	15	14	1.48693	0.12906	1.48693	20.64252
3	14	14	1.13701	0.11379	1.40472	26.19932
4	11	13	0.75213	0.09911	1.57305	30.44459
5	13	14	0.36797	0.08776	1.64786	32.42724
⋮	⋮	⋮	⋮	⋮	⋮	⋮
3999	8	14	0.00022	0.00814	0.00835	3.99335
4000	14	29	0.00020	0.00817	0.00830	3.98625

the difference in absolute value between the objective function $f + g$ obtained by the SC-PMA and the objective function $f + g$ obtained by the CVX function of Matlab.

Elapsed time is 30687.695116 seconds. The coefficients of the equation are -0.6274; 0.0156; -0.5672; 0.2475. Fig 10 shows the data set separated by a plane, whose equation of its respective plane is $-0.6274X + 0.0156 - 0.5672Z + 0.2475 = 0$.

Table III presents computational results.

5. Conclusion

The present article introduces a symmetric cone proximal multiplier algorithm (SC-PMA) to solve separable optimization problems. The point of convergence of the primal-dual variables is proved to be a saddle point of the Lagrangian associated with the problem; therefore, we solve the optimization problem. Then, we apply SC-PMA to find linear hyperplanes in binary classification in support vector machines. This is the first time the SC-PMA has been implemented to solve this class of problems applied to artificial intelligence, and the obtained results motivate more investigations.

This paper is the continuation of previously published papers developed by the authors. In the papers [26, 27], we developed a systematic review of support vector machines applied to classification and regression. In the paper [18], we developed a methodology to Construct proximal distances over symmetric cones and in the proceeding paper [28], we obtained preliminary mathematical results of the SC-PMA. In this paper, we consolidated the convergence results of the SC-PMA and applied them to find linear hyperplanes to binary classification in SVM, thus complementing the result of the paper.

To improve the computational results, it is necessary to solve problems (3.9) and (3.10) efficiently. For those cases, we think that the Bundle family of methods [6, chap. 9] is perhaps the most practical computational tool for nonsmooth optimization. It may be considered future research to improve the algorithm's efficiency.

Another future research may be to compare the SC-PMA with the different classification algorithms, such as neural networks, Bayesian classifiers, and support vector machines, among others.

We will mention that there are several applications of the support vector machine technique; see [29 -31]

That is why it is important to develop this algorithm applied to a support vector machine.

References

[1] Alizadeh, F., and Goldfarb, D., “Second-Order Cone Programming,” *Mathematical Programming*, vol. 95, pp. 3-51, 2012. *Crossref*, *Crossref*, https://doi.org/10.1007/s10107-002-0339-5

[2] Alvarez, F, L´Opez, J., and Ram´Irez C., H., “Interior Proximal Algorithm with Variable Metric for Second-Order Cone Programming: Applications to Structural Optimization and Support Vector Machines,” *Optimization Methods Software*, vol. 25, no. 6, pp. 859-881, 2012. *Crossref*, https://doi.org/10.1080/10556780903483356

[3] Alfred Auslender, and Marc Teboulle, “Entropic Proximal Decomposition Methods for Convex Programs and Variational Inequalities,” *Mathematical Programming*, vol. 91, no. 1, pp. 33-47, 2001. https://doi.org/10.1007/s101070100241

[4] Alfred Auslender, and Marc Teboulle, “Interior Gradient and Proximal Methods for Convex and Conic Optimization,” *SIAM Journal on Optimization*, vol. 16, no. 3, pp. 697-725, 2006. *Crossref*, https://epubs.siam.org/doi/10.1137/S1052623403427823

[5] Alfred Auslender, Marc Teboulle, and Sami Ben-Tiba, “A Logarithmic-Quadratic Proximal Method for Variational Inequalities,” *Computational Optimization and Applications*, vol. 12, pp. 31-40, 1999. *Crossref*, https://doi.org/10.1023/A:1008607511915

[6] J. Frédéric Bonnans et al., “*Numerical Optimization: Theoretical and Practical Aspects*,” Springer, 2006.

[7] Gong Chen, and Marc Teboulle, "A Proximal-Based Decomposition Method for Convex Minimization Problems," vol. 64, pp. 81-101, 1994. *Crossref*, https://doi.org/10.1007/BF01582566

[8] Corinna Cortes, and Vladimir Vapnik "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273-297, 1995. *Crossref*, https://doi.org/10.1007/BF00994018

[9] DanielGabay, and BertrandMercier "A Dual Algorithm for the Solution of Nonlinear Variational Problems Via Finite Element Approximation," *Computers & Mathematics with Applications*, vol. 1, no. 2, pp.17 - 40, 1976. *Crossref*, https://doi.org/10.1016/0898-1221(76)90003-1

[10] Donald Goldfarb, Shiqian Ma, and Katya Scheinberg, "Fast Alternating Linearization Methods for Minimizing the Sum of Two Convex Functions," *Mathematical Programming*, vol. 141, pp. 349-382, 2013. *Crossref*, https://doi.org/10.1007/s10107-012-0530-2

[11] M. Kyono, and M. Fukushima, "Nonlinear Proximal Decomposition Method for Convex Programming," *Journal of Optimization Theory and Applications*, vol. 106, pp. 357-372, 2000. *Crossref*, https://doi.org/10.1023/A:1004655531273

[12] Jacques Faraut, and Adam Korányi, "*Analysis on Symmetric Cones, Oxford Mathematical Monographs*," *The Clarendon Press,* Oxford University Press, New York, 1994.

[13] Osman Güler, "On the Convergence of the Proximal Point Algorithm for Convex Minimization," *SIAM Journal on Control and Optimization,* vol. 29, no. 2, pp. 403-419, 1991. *Crossref*, https://doi.org/10.1137/0329022

[14] Klerk, E. De, "*Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*," Kluwer Academic Publishers, Series Applied Optimization, vol. 65, 2002.

[15] Kiwiel, K.C., "Proximal Minimization Methods with Generalized Bregman Functions," *SIAM Journal on Control and Optimization*, vol. 35, no. 4, pp.1142–1168, 1997. *Crossref*, https://doi.org/10.1137/S0363012995281742

[16] Li, A., and Tuncel, L., "Some Applications of Symmetric Cone Programming in Financial Mathematics," *Transactions in Operational Research*, vol. 17, pp. 1-19, 2006.

[17] Miguel Sousa Lobo et al., "Applications of Second-Order Cone Programming," *Linear Algebra and Its Applications*, vol. 284, pp. 193-228, 1998. *Crossref*, https://doi.org/10.1016/S0024-3795(98)10032-0

[18] López J. and Papa Quiroz E.A.,"Construction of Proximal Distances Over Symmetric Cones," *Optimization*, vol. 66, no. 8, pp. 1301-21, 2017. *Crossref*, https://doi.org/10.1080/02331934.2016.1277998

[19] ShaohuaPan, and Jein-Shan Chen, "Interior Proximal Methods and Central Paths for Convex Second-Order Cone Programming," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 73, no. 9, pp. 3083 - 3100, 2010. *Crossref*, https://doi.org/10.1016/j.na.2010.06.079

[20] Sarrmiento O, Papa Quiroz E.A, and Oliveira P.R., "A Proximal Multiplier Method for Separable Convex Minimization," *Optimization*, vol. 65, no. 2, pp. 501-537, 2015. *Crossref*, https://doi.org/10.1080/02331934.2015.1062009

[21] Abdel Radi Abdel Rahman Abdel Gadir, and Samia Abdallah Yagoub Ibrahim, "Comparison Between Some Analysis Solutions for Solving Fredholm Integral Equation of Second Kind," *International Journal of Mathematics Trends and Technology*, vol. 66, no. 6 pp. 245-260, 2020. *Crossref*, https://doi.org/10.14445/22315373/IJMTT-V66I6P525

[22] Spingarn, J.E., "Applications of the Method of Partial Inverses to Convex Programming: Decomposition*," Mathematical Programming*, vol. 32, pp.199-223, 1985.

[23] Rockafellar, R.T., "*Convex Analysis*," Princeton University Press, Princeton Mathematical Series, p. 472, 1970.

[24] Katya Scheinberg, Shiqian Ma, and Donald Goldfarb, "*Sparse Inverse Covariance Selection Via Alternating Linearization Methods,"* 2010.

[25] Lieven Vandenberghe, and Stephen Boyd, "Semidefinite Programming," *SIAM Review*, vol. 38, pp. 49-95, 1994. *Crossref*. https://doi.org/10.1137/1038003

[26] Miguel Angel Cano Lengua, and Erik Alex Papa Quiroz, "A Systematic Literature Review on Support Vector Machines Applied to Classification," *Proceedings of the 2020 IEEE Engineering International Research Conference*, *EIRCON*, 2020.

[27] Daniel Mavilo Calderon Nieto, Erik Papa, and Miguel Cano, "A Systematic Literature Review on Support Vector Machines Applied to Regression" *Proceedings of the 2021 IEEE Sciences and Humanities International Research Conference*, 2021. *Crossref*, https://doi.org/10.1109/SHIRCON53068.2021.9652268

[28] Erik Alex Papa Quiroz et al., "A Proximal Multiplier Method for Convex Separable Symmetric Cone Optimization," *Proceeding of 5th International Conference on Multimedia Systems and Signal Processing*, pp. 92-97, 2020. *Crossref*, https://doi.org/10.1145/3404716.3404734

[29] Rina Mahakud, Binod Kumar Pattanayak, and Bibudhendu Pati, "A Hybrid Multi-Class Classification Model for the Detection of Leaf Disease Using Xgboost and SVM," *International Journal of Engineering Trends and Technology*, vol. 70, no. 10, pp. 298-306, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I10P229

[30] Ibrahima Sory Keita et al., "Classification of Benign and Malignant Mris Using SVM Classifier for Brain Tumor Detection," *International Journal of Engineering Trends and Technology*, vol. 70, no. 3, pp. 234-240, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I2P226

[31] Maamar Ali Saud AL tobi et al., "Machinery Faults Diagnosis Using Support Vector Machine (SVM) and Naïve Bayes Classifiers," *International Journal of Engineering Trends and Technology*, vol. 70, no. 12, pp. 26-34, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I12P204

[32] Schmieta, S.H., and Alizadeh, F., "Extension of Primal-Dual Interior Point Algorithms to Symmetric Cones," *Mathematical Programming*, vol. 96, pp. 409-438, 2003. *Crossref*, https://doi.org/10.1007/s10107-003-0380-z