Noname manuscript No.
(will be inserted by the editor)

Regularized Interior Proximal Alternating Direction Method for Separable Convex Optimization Problems

Felipe Alvarez · **Julio López** · **Natalia Ruiz**

Received: date / Accepted: date

Abstract In this article we present a version of the proximal alternating direction method for a convex problem with linear constraints and a separable objective function, in which the standard quadratic regularizing term is replaced with an interior proximal metric for those variables that are required to satisfy some additional convex constraints. Moreover, the proposed method has the advantage that the iterates are computed only approximately. Under standard assumptions, global convergence of the primal-dual sequences computed by the algorithm is established. Finally, we report some numerical experiments applied to statistical learning problems to illustrate the behavior of our algorithm.

Keywords Alternating direction methods · Decomposition · Interior-point methods · Statistical learning problem.

Mathematics Subject Classification (2000) MSC 49M27 · MSC 49M37 · MSC 90C51

1 Introduction

In this paper we develop a special alternating direction method (ADM) for solving approximately a convex optimization problem with separable structure as follows:

$$(P) \qquad \min_{x,z} \left\{ f(x) + g(z) \mid Ax + Bz = b,\ x \in \overline{C} \right\},$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are closed convex proper functions, A and B are $p \times n$ and $p \times m$ real matrices, respectively, and $b \in \mathbb{R}^p$ is a given vector. Here $\overline{C}$

F. Alvarez
Mathematical Engineering Department and Center for Mathematical Modeling, University of Chile, Beauchef 851, Santiago, Chile
E-mail: falvarez@dim.uchile.cl

J. López
Faculty of Engineering, University Diego Portales, Ejército 441, Santiago, Chile
E-mail: julio.lopez@udp.cl

N.Ruiz (✉)
Mathematical Engineering Department, University of Chile, Beauchef 851, Santiago, Chile
E-mail: nruiz@dim.uchile.cl

denotes the closure of $C \subset \mathbb{R}^n$, an open nonempty convex set; a typical example being the strictly positive orthant $C = \mathbb{R}^n_{++}$ so that $x \in \overline{C} = \mathbb{R}^n_+$ corresponds to $x \geq 0$.

Under the following natural condition: $\mathrm{dom} f \cap C \neq \emptyset$, we are interested in designing a method which exploits the separable structure of (P). With such a motivation, we propose the following iterative algorithm for (P): given $(x^k, z^k, y^k) \in C \times \mathbb{R}^m \times \mathbb{R}^p$, compute consecutively

$$x^{k+1} \approx \mathrm{argmin}_{x\in\overline{C}}\ f(x) + \langle y^k, Ax\rangle + \frac{\lambda}{2}\|Ax + Bz^k - b\|^2 + \frac{1}{2\lambda}d(x, x^k), \tag{1}$$

$$z^{k+1} \approx \mathrm{argmin}_{z\in\mathbb{R}^m}\ g(z) + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|Ax^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2, \tag{2}$$

$$y^{k+1} = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{3}$$

Here $\lambda > 0$ is a positive scalar and $d(\cdot, x^k)$ is a given *proximal distance* with respect to C (see section 2, definition 1). The latter enforces that $x^{k+1} \in C$, that is, the x-iterate lies in the interior of the additional constraints set $\overline{C}$. Two popular choices for d include either a Bregman distance (see, e.g., Burachick and Iusem (1998), Solodov and Svaiter (2000)) or a proximal distance based on second order homogeneous kernels (see, e.g., Auslender et al. (1999), Teboulle et al. (1999)). Another possible choice is the double regularization technique introduced by Silva and Eckstein (2006), which extends the notion of second order homogeneous kernels. Any of these choices will be covered by our analysis.

In order to explain the ideas behind the iterative scheme given by (1)-(3), let us introduce the restricted Lagrangian function $\ell : \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ given by

$$\ell(x,z,y) = f(x) + g(z) + \langle y, Ax + Bz - b\rangle, \tag{4}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product in $\mathbb{R}^p$. Throughout this paper, we assume the following: There exists a saddle point $(x^*, z^*, y^*) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ of ℓ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$, that is,

$$\ell(x^*, z^*, y) \leq \ell(x^*, z^*, y^*) \leq \ell(x, z, y^*),\ \ \forall (x,z,y) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p. \tag{5}$$

In this setting, if we introduce the augmented Lagrangian $\ell_\lambda : \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ by

$$\ell_\lambda(x,z,y) = \ell(x,z,y) + \frac{\lambda}{2}\|Ax + Bz - b\|^2, \tag{6}$$

then, from a given $(x^k, z^k, y^k) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$, the classical ADM produces iterates via the following consecutive steps:

$$x^{k+1} \in \mathrm{Argmin}_{x\in\overline{C}}\ell_\lambda(x, z^k, y^k) = \mathrm{Argmin}_{x\in\overline{C}}\ f(x) + \langle y^k, Ax\rangle + \frac{\lambda}{2}\|Ax + Bz^k - b\|^2, \tag{7}$$

$$z^{k+1} \in \mathrm{Argmin}_{z\in\mathbb{R}^m}\ell_\lambda(x^{k+1}, z, y^k) = \mathrm{Argmin}_{z\in\mathbb{R}^m}\ g(z) + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|Ax^{k+1} + Bz - b\|^2, \tag{8}$$

$$y^{k+1} = \mathrm{argmax}_{y\in\mathbb{R}^p}\ \ell_\lambda(x^{k+1}, z^{k+1}, y) - \frac{1}{2\lambda}\|y - y^k\|^2 = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{9}$$

ADM was originally proposed by Gabay and Mercier (1976), and since then, it has received intensive attention for others researchers; see, e.g., Tseng (1991). This method is closely related to the Douglas-Rachford operator splitting algorithm which solves monotone inclusion problems; see, e.g., Douglas and Rachford (1956), Combettes and Pesquet (2007), Svaiter (2011), Bot and Hendrich (2013) (see Combettes (2004); Tseng (2000); Briceño-Arias and Combettes (2011) for others important operator splitting algorithms).

A substantial improvement on the ADM is to combine the classical proximal point algorithm (PPA) with the ADM, which gives rise to proximal alternating method of multipliers (PADM) introduced by Eckstein (1994), and current literature along this direction of research is dominated by the utilization of the quadratic proximal regularization. For example, Xu (2007) proposed a PADM with quadratic proximal regularization (QPADM) for variational inequalities and Attouch et al. (2009) for convex optimization in an infinite dimensional setting (see also Attouch et al. (2011)). The idea of PADM is to regularize the augmented Lagrangian by adding to it some primal quadratic proximal terms, having a stronger primal convergence theory than the standard method.

On the other hand, the conceptual primal iterations given by (1) and (2) can be equivalently written as follows:

$$x^{k+1} \approx \operatorname{argmin}_{x\in C}\ \ell_\lambda(x,z^k,y^k) + \frac{1}{2\lambda}d(x,x^k),$$
$$z^{k+1} \approx \operatorname{argmin}_{z\in\mathbb{R}^m}\ \ell_\lambda(x^{k+1},z,y^k) + \frac{1}{2\lambda}\|z-z^k\|^2.$$

In this alternate proximal procedure, minimization steps are performed consecutively on the primal variables, first with respect to x by using a *regularized interior proximal* term so that all the iteration points lie in the interior of the feasible set (see, e.g., Auslender and Teboulle (2006)), and then with respect to z using a standard quadratic proximal metric. Therefore, we call the iterative scheme given by (1), (2) and (3) the *regularized interior proximal alternating direction method* (RIPADM).

The main theoretical result in this paper is a general convergence result for a special approximate version of RIPADM (*cf.* Theorem 1), bridging two different areas of proximal algorithm theory: ADM on the one hand, and generalized nonquadratic proximal distance on the other. To the best of our knowledge, the only partial positive result in this direction is the recent work by Yu and Li (2011) (see Li et al (2013) for a version with inexact computations), where these authors investigate a particular PADM for variational inequalities under non-negativity constraints in which the regularizing proximal term in the ADM subproblems is induced by the logarithmic-quadratic proximal (LQP) kernel, previously developed in Auslender et al. (1999). In the case of optimization problems, the LQP approach can be viewed as a particular instance of our general setting.

Along a complementary line of research, an entropic proximal decomposition method (EPDM) has been introduced by Auslender and Teboulle (2001) for solving general variational inequality problems with particular separable structure. The EPDM combines the LQP theory for non-negativity constraints with a predictor-corrector proximal multiplier method developed previously by Chen and Teboulle (1994). EPDM differs from RIPADM as the predicted multiplier estimate is used to compute proximal steps *in parallel* for both primal variables, thanks to the fact that the decoupling is obtained by a sort of linearization of the squared Euclidean norm in the augmented Lagrangian at the current iterate, following the strategy proposed by Stephanopoulos and Westerberg (1975). Thus, EPDM is a parallel method in the primal variables, which is different from the ADM approach (the latter is closer to a Gauss-Seidel scheme).

Outline of the paper. The paper is organized as follows. In section 2, we recall some preliminaries on generalized proximal distances, we give an approximate version of the RIPADM (cf. (1)-(2)), in which the stopping criteria for the inner inexact computations are fairly practical to check, and we identify some examples of interior metrics satisfying appropriate conditions for our analysis. In section 3, we prove our main full convergence result for the inexact version of RIPADM. In section 4, we report the numerical results obtained when we apply the RIPADM scheme to solve some statistical learning applications. Concluding remarks are given in section 5.

2 Preliminaries

2.1 Generalized proximal distances

Definition 1 (Auslender and Teboulle, 2006, Definition 2.1) A function $d:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}_+\cup\{+\infty\}$ is called a proximal distance with respect to an open nonempty convex set $C\subset\mathbb{R}^n$ if for each $v\in C$ we have the following properties:

(d_1) $d(\cdot,v)$ is proper, closed, convex and continuously differentiable on C;
(d_2) $\mathrm{dom}d(\cdot,v)\subset\overline{C}$ and $\mathrm{dom}\partial_1 d(\cdot,v)=C$, where $\partial_1 d(\cdot,v)$ denotes the subgradient map of the function $d(\cdot,v)$ with respect to the first variable;
(d_3) $d(\cdot,v)$ is level bounded on $\mathbb{R}^n$, i.e., $\lim_{\|u\|\to\infty} d(u,v)=+\infty$;
(d_4) $d(v,v)=0$.

We denote by $\mathscr{D}(\mathscr{C})$ the family of functions d satisfying Definition 1. The following proposition guarantees the existence of solutions to the internal problems of the RIPADM method, that is, due to next proposition the RIPADM method will be well defined.

Proposition 1 *(Auslender and Teboulle, 2006, Proposition 2.1) Let $F:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be a proper, closed and convex function. Suppose that $F_*=\inf_{u\in\overline{C}}F(u)>-\infty$ and $\mathrm{dom}F\cap C\neq\emptyset$. Let $d\in\mathscr{D}(\mathscr{C})$, and for all $v\in C$ consider the optimization problem*

$$(P(v))\qquad F_*(v)=\inf_{u\in\overline{C}}F(u)+d(u,v).$$

Then the optimal set $S(v)$ of $P(v)$ is nonempty and compact, and for each $\varepsilon\geq 0$ there exist $u(v)\in C$, $g\in\partial_\varepsilon F(u(v))$ such that

$$g+\nabla_1 d(u(v),v)=0,$$

where $\partial_\varepsilon F(u(v))$ denotes the ε-subdifferential of F at $u(v)$. For such a $u(v)\in C$ we have

$$F(u(v))+d(u(v),v)\leq F_*(v)+\varepsilon.$$

2.2 The inexact RIPADM algorithm

Consider problem (P) and assume that $\mathrm{dom}f\cap C\neq\emptyset$. Let $d\in\mathscr{D}(\mathscr{C})$ and λ be a positive scalar. Starting from a point $(x^0,z^0,y^0)\in C\times\mathbb{R}^m\times\mathbb{R}^p$, we generate the sequence $\{(x^k,z^k,y^k)\}\subset C\times\mathbb{R}^m\times\mathbb{R}^p$, and sequences of errors $\{a^k\}$ and $\{b^k\}$ via the following steps:

Step 1. Find $(x^{k+1},a^{k+1})\in C\times\mathbb{R}^n$ with $r^{k+1}\in\partial f(x^{k+1})$ solving:

$$a^{k+1}=r^{k+1}+A^t[y^k+\lambda(Ax^{k+1}+Bz^k-b)]+\frac{1}{2\lambda}\nabla_1 d(x^{k+1},x^k),\tag{10}$$

where the error a^{k+1} satisfies the conditions

$$\|a^{k+1}\|\leq\varepsilon_{k+1},\ \ \|a^{k+1}\|\|x^{k+1}\|\leq\eta_{k+1}.\tag{11}$$

Step 2. Find $(z^{k+1}, b^{k+1}) \in \mathbb{R}^m \times \mathbb{R}^m$ with $\bar{r}^{k+1} \in \partial g(z^{k+1})$ solving:

$$b^{k+1} = \bar{r}^{k+1} + B^t[y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b)] + \frac{1}{\lambda}(z^{k+1} - z^k), \tag{12}$$

where the error b^{k+1} satisfies the conditions

$$\|b^{k+1}\| \leq \bar{\varepsilon}_{k+1}, \ \|b^{k+1}\|\|z^{k+1}\| \leq \bar{\eta}_{k+1}. \tag{13}$$

Step 3. Compute

$$y^{k+1} = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b). \tag{14}$$

The positive scalars $\varepsilon_k, \eta_k, \bar{\varepsilon}_k, \bar{\eta}_k > 0$ satisfy the following conditions:

$$\sum_{k=1}^{\infty} \varepsilon_k < \infty, \ \sum_{k=1}^{\infty} \eta_k < \infty, \ \sum_{k=1}^{\infty} \bar{\varepsilon}_k < \infty, \ \text{and} \ \sum_{k=1}^{\infty} \bar{\eta}_k < \infty. \tag{15}$$

Remark 1 In the exact case where $a_k = 0 \in \mathbb{R}^n$ and $b_k = 0 \in \mathbb{R}^m$ for all k, steps 1 and 2 in the previous scheme amount to find exact solutions to the minimization problems in (1) and (2), respectively. Due to Proposition 1, such an exact RIPADM is well defined.

2.3 Compatible proximal pairs (d,H) and examples

Next, we associate with $d \in \mathscr{D}(\mathscr{C})$ a corresponding proximal distance H satisfying some desirable properties needed to analyze the convergence of RIPADM.

Definition 2 Given an open nonempty convex set $C \subset \mathbb{R}^n$ and $d \in \mathscr{D}(\mathscr{C})$, we say that (d,H) is a compatible proximal pair associated with $\overline{C}$ if $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$ is a finite valued function on $\overline{C} \times C$ and for each $a, b \in C$ satisfies:

(i) $H(a,a) = 0$;
(ii) $\langle c - b, \nabla_1 d(b,a) \rangle \leq H(c,a) - H(c,b) - \kappa H(b,a)$, $\forall c \in \overline{C}$, and some fixed $\kappa > 0$;
(iii) For all $c \in \overline{C}$, $H(c,\cdot)$ is level bounded on C.

We write $(d,H) \in \mathscr{F}(\overline{\mathscr{C}})$ when a triple $[\overline{C}, d, H]$ satisfies the premises of Definition 2. The requested properties for the function H emerge naturally from the analysis of the classical proximal algorithm (PA) and later extended for various specific classes of IPA (see Auslender and Teboulle (2006)).

In the literature one can find various proximal distances which are obtained compatible proximal pair (d,H). For example, a Bregman distance D_h, where h is the Bregman kernel, satisfies the Definition 2 with $d = D_h = H$. Another class of proximal distances is formed by second order homogeneous proximal distances. Here, we find the famous logarithmic-quadratic (Log-quad) proximal distance (see Auslender et al. (1999)) defined by:

$$d_\varphi(u,v) = \sum_{i=1}^{n} \mu \left(v_i^2 \log\left(\frac{v_i}{u_i}\right) + u_i v_i - v_i^2 \right) + \frac{\nu}{2}(u_i - v_i)^2, \ \forall u, v \in \mathbb{R}_{++}, \tag{16}$$

where the kernel $\varphi(t) = \mu(\log(t) + t - 1) + \frac{\nu}{2}(t-1)^2$, with $\nu \geq \mu > 0$.

Another example of proximal distances are double regularizations introduced by Silva and Eckstein (2006). Let B be a box set, that is, $B = \{x \in \mathbb{R}^n | \ a_i \leq x_i \leq b_i, \ i = 1, \ldots, n\}$, where $a_i \in [-\infty, +\infty)$ and $b_i \in (-\infty, +\infty]$, $a_i \leq b_i$. A *double regularitazion* for the box B is a separable function $\widetilde{D} : B \times \text{int}B \to \mathbb{R}$ of the form:

$$\widetilde{D}(u,v) = \sum_{i=1}^{n} \widetilde{d}_i(u_i, v_i) = \sum_{i=1}^{n} d_i(u_i, v_i) + \frac{\nu}{2}(u_i - v_i)^2, \tag{17}$$

where $\nu \geq 1$ is a scalar and functions $\widetilde{d}_i$ satisfy some assumptions (see Assumptions 2.1.1-2.1.3 of Silva and Eckstein (2006)).

Lemma 1 *(Eckstein and Silva, 2010, Lemma 1) Suposse $\widetilde{D}$ is a double regularization for the box B with regularization factor $\nu \geq 1$. Under the last assumption we have*

$$\langle c - b, \nabla_1 \widetilde{D}(b,a) \rangle \leq \frac{\nu + 1}{2} \left(\|c - a\|^2 - \|c - b\|^2 \right) - \frac{\nu - 1}{2} \|a - b\|^2, \tag{18}$$

for any $c \in B$ and $a, b \in intB$.

We see that if $\nu > 1$ and $H(u,v) = \frac{\nu+1}{2}\|u - v\|^2$ it follows that $(\widetilde{D}, H) \in \mathscr{F}(\mathscr{B})$ is a compatible proximal pair on B.

3 Full convergence of the inexact RIPADM

Let us return to problem (P). Consider the Lagrangian function ℓ defined by (4). We suppose:

Assumption A

(A_1) dom $f \cap C \neq \emptyset$.
(A_2) There exists a saddle point $(x^*, z^*, y^*) \in \overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ of ℓ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$ (see (5)).

Assumption B

There exists a compatible proximal pair (d, H) associated with $\overline{C}$ which verifies:

(B_1) If $\{u^k\} \subset \overline{C}$ and $\{v^k\} \subset C$ are sequences such that

$$\lim_k H(u^k, v^k) = 0,$$

and one of the sequences ($\{u^k\}$ or $\{v^k\}$) converges, then the other one also converges to the same limit.
(B_2) The function $H : \overline{C} \times C \to \mathbb{R}$ extends by continuity to $\overline{C} \times \overline{C}$.
(B_3) $\forall u \in \overline{C}$ and $\forall \{u^k\} \subset C$ converging u, we have $\lim_{k\to\infty} H(u, u^k) = 0$.

If $d = \widetilde{D}$ is a double regularization, according to the lemma 1 of Silva and Eckstein is possible to take $H = \frac{\nu+1}{2}\|u - v\|^2$ with $\nu > 1$, then the pair (d, H) satisfies the assumption B. On the other hand, Solodov and Svaiter (2000) proved that the Bregman distances satisfy the condition (B_1) (see Theorem 2.4 of Solodov and Svaiter (2000)). Moreover, the Bregman distances satisfy the condition (B_3) but not necessarily the condition (B_2).

We will prove some results related to a primal-dual sequence generated by the RIPADM method. Under Assumption A, the primal-dual sequence (x^k, z^k, y^k) is bounded. Additionally, if the proximal pair (d, H) verify (B_1), all limit points of $\{(x^k, z^k)\}$ are solutions of the problem (P), for example this is the case of Bregman distances. Moreover, if the proximal pair (d, H) satisfy the Assumption B (for example double regularization), the sequence globally converges to a saddle point of the Lagrangian function ℓ of the problem (P). Below we state our principal result:

Theorem 1 *Let $\{(x^k,z^k,y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$ be a primal-dual sequence generated by the RIPADM method (10)-(14).*

(i) Under Assumptions A and (B_1), all limit points of the primal sequence $\{(x^k,z^k)\}$ are solutions of the problem (P).
(ii) If besides the Assumptions (B_2) and (B_3) are satisfied, the primal-dual sequence $\{(x^k,z^k,y^k)\}$ globally converges to a saddle point $(x^\infty,z^\infty,y^\infty)$ of the Lagrangian ℓ of the problem (P).

We have divided the proof of Theorem 1 into a sequence of lemmas and propositions. We begin stating two lemmas, the first one is a classical result on scalar sequences, and the second one gives us two saddle point type inequalities. These inequalities are used in the proof of the Proposition 2 below which asserts that the sequence $\{(x^k,z^k,y^k)\}$ generated by RIPADM is bounded if at least one saddle point (x^*,z^*,y^*) of ℓ exists. Finally, under Assumptions A and B, we prove that there exists a unique limit point $(x^\infty,z^\infty,y^\infty)$ of the primal-dual sequence $\{(x^k,z^k,y^k)\}$ and so we get the global convergence of our algorithm.

Lemma 2 *(Polyak, 1987, Section 2.2) Suppose $\{w_k\}$, $\{\beta_k\} \subset \mathbb{R}$ are sequences such that $\{w_k\}$ is bounded below, $\sum_{k=0}^\infty \beta_k$ exists and is finite, and the recursion $w_{k+1} \le w_k + \beta_k$ holds for all k. Then $\{w_k\}$ is convergent.*

Lemma 3 *Let $(d,H) \in \mathscr{F}(\overline{\mathscr{C}})$. Consider the sequence $\{(x^k,z^k,y^k)\} \subset C \times \mathbb{R}^m \times \mathbb{R}^p$, and errors sequences $\{a^k\} \subset \mathbb{R}^n$ and $\{b^k\} \subset \mathbb{R}^m$ being generated by the RIPADM algorithm given by (10)-(14). Under Assumption (A_1), for all $x \in dom(f)$, $z \in dom(g)$, $y \in \mathbb{R}^p$, we have:*

$$\begin{aligned}\ell(x^{k+1},z^{k+1},y^k) - \ell(x,z,y^k) \le\ & \frac{\lambda}{2}(\|Ax+Bz^k-b\|^2 - \|Ax^{k+1}+Bz^k-b\|^2 \\ & -\|Ax^{k+1}-Ax\|^2) + \frac{\lambda}{2}(\|Ax^{k+1}+Bz-b\|^2 \\ & -\|Ax^{k+1}+Bz^{k+1}-b\|^2 - \|Bz^{k+1}-Bz\|^2) \\ & +\frac{1}{2\lambda}\left(H(x,x^k) - H(x,x^{k+1}) - \kappa H(x^{k+1},x^k)\right) \\ & +\frac{1}{2\lambda}\left(\|z-z^k\|^2 - \|z-z^{k+1}\|^2 - \|z^{k+1}-z^k\|^2\right) \\ & +\langle a^{k+1},x^{k+1}-x\rangle + \langle b^{k+1},z^{k+1}-z\rangle.\end{aligned} \tag{19}$$

$$\ell(x^{k+1},z^{k+1},y) - \ell(x^{k+1},z^{k+1},y^k) = \frac{1}{2\lambda}\left(\|y-y^k\|^2 - \|y-y^{k+1}\|^2 + \|y^{k+1}-y^k\|^2\right). \tag{20}$$

Proof Lemma 3. From steps (10) and (12) of our algorithm, since $r^{k+1} \in \partial f(x^{k+1})$ and $\bar{r}^{k+1} \in \partial g(z^{k+1})$ we have:

$$f(x^{k+1}) + \left\langle \frac{-1}{2\lambda}\nabla_1 d(x^{k+1},x^k) - A^t[y^k + \lambda(Ax^{k+1}+Bz^k-b)] + a^{k+1}, x - x^{k+1}\right\rangle \le f(x), \tag{21}$$

$$g(z^{k+1}) + \left\langle \frac{-1}{\lambda}(z^{k+1}-z^k) - B^t[y^k + \lambda(Ax^{k+1}+Bz^{k+1}-b)] + b^{k+1}, z - z^{k+1}\right\rangle \le g(z), \tag{22}$$

for all $x \in \text{dom}(f)$ and $z \in \text{dom}(g)$. Adding and rearranging these inequalities, we obtain

$$\begin{aligned} & f(x^{k+1}) + g(z^{k+1}) + \langle y^k, Ax^{k+1} + Bz^{k+1} \rangle \\ & \quad - \Big(f(x) + g(z) + \langle y^k, Ax + Bz \rangle \Big) \le \lambda \langle Ax^{k+1} + Bz^k - b, Ax - Ax^{k+1} \rangle \\ & \qquad + \lambda \langle Ax^{k+1} + Bz^{k+1} - b, Bz - Bz^{k+1} \rangle \\ & \qquad + \frac{1}{2\lambda} \langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1} \rangle \\ & \qquad + \frac{1}{\lambda} \langle z^{k+1} - z^k, z - z^{k+1} \rangle \\ & \qquad + \langle a^{k+1}, x^{k+1} - x \rangle + \langle b^{k+1}, z^{k+1} - z \rangle. \end{aligned}$$

Considering the definition of the Lagrangian function ℓ and using the relation $\|m-n\|^2 = \|m\|^2 + \|n\|^2 - 2\langle m, n \rangle$ in the last inequality, we obtain

$$\begin{aligned} \ell(x^{k+1}, z^{k+1}, y^k) - \ell(x, z, y^k) \le & \ \frac{\lambda}{2} (\|Ax + Bz^k - b\|^2 - \|Ax^{k+1} + Bz^k - b\|^2 \\ & - \|Ax^{k+1} - Ax\|^2) + \frac{\lambda}{2} (\|Ax^{k+1} + Bz - b\|^2 \\ & - \|Ax^{k+1} + Bz^{k+1} - b\|^2 - \|Bz^{k+1} - Bz\|^2) \\ & + \frac{1}{2\lambda} \langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1} \rangle \\ & + \frac{1}{2\lambda} \Big(\|z - z^k\|^2 - \|z^{k+1} - z^k\|^2 - \|z - z^{k+1}\|^2 \Big) \\ & + \langle a^{k+1}, x^{k+1} - x \rangle + \langle b^{k+1}, z^{k+1} - z \rangle. \end{aligned} \tag{23}$$

Using (ii) of Definition 2 with $a = x^k$, $b = x^{k+1}$ and $c = x$, we have

$$\langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1} \rangle \le H(x, x^k) - H(x, x^{k+1}) - \kappa H(x^{k+1}, x^k). \tag{24}$$

From (23) and (24), we obtain the inequality (19).

Finally we prove the equality (20). We have

$$\ell(x^{k+1}, z^{k+1}, y) - \ell(x^{k+1}, z^{k+1}, y^k) = \langle y - y^k, Ax^{k+1} + Bz^{k+1} - b \rangle, \tag{25}$$

but from (14) we have $Ax^{k+1} + Bz^{k+1} - b = \frac{1}{\lambda}(y^{k+1} - y^k)$, then

$$\begin{aligned} \ell(x^{k+1}, z^{k+1}, y) - \ell(x^{k+1}, z^{k+1}, y^k) &= \frac{1}{\lambda} \langle y - y^k, y^{k+1} - y^k \rangle \\ &= \frac{1}{2\lambda} \Big(\|y - y^k\|^2 - \|y - y^{k+1}\|^2 + \|y^{k+1} - y^k\|^2 \Big). \end{aligned}$$

□

Proposition 2 *Suppose that the hypothesis of Lemma 3 holds.*

(i) For any saddle point $(\bar{x}, \bar{z}, \bar{y})$ of ℓ, the sequence $\{E_k(\bar{x}, \bar{z}, \bar{y})\}$ is convergent where

$$E_k(\bar{x}, \bar{z}, \bar{y}) = \frac{1}{2\lambda} \Big(H(\bar{x}, x^k) + \|z^k - \bar{z}\|^2 + \|y^k - \bar{y}\|^2 \Big) + \frac{\lambda}{2} \|B(z^k - \bar{z})\|^2. \tag{26}$$

(ii) If at least one saddle point $(\bar{x},\bar{z},\bar{y})$ of ℓ exists then the sequence $\{(x^k,z^k,y^k)\}$ is bounded in $C\times\mathbb{R}^m\times\mathbb{R}^p$, the quantities $H(x^{k+1},x^k)$, $\|z^{k+1}-z^k\|^2$ and $\|Ax^{k+1}+Bz^k-b\|^2$ are summable, hence they vanish as k goes to $+\infty$.

Proof Proposition 2. Let $(\bar{x},\bar{z},\bar{y})\in\overline{C}\times\mathbb{R}^m\times\mathbb{R}^p$ be a saddle point of ℓ on $\overline{C}\times\mathbb{R}^m\times\mathbb{R}^p$, that is,

$$\ell(\bar{x},\bar{z},y)\leq\ell(\bar{x},\bar{z},\bar{y})\leq\ell(x,z,\bar{y}),\forall x\in\overline{C},\ \forall z\in\mathbb{R}^m,\ \forall y\in\mathbb{R}^p.$$

Then we have:

$$\ell(\bar{x},\bar{z},y^k)-\ell(x^{k+1},z^{k+1},\bar{y})\leq 0. \tag{27}$$

Adding the inequality (19) with $x=\bar{x}$ and $z=\bar{z}$, and (27) we get

$$\begin{aligned}\ell(x^{k+1},z^{k+1},y^k)-\ell(x^{k+1},z^{k+1},\bar{y}) \ &\leq \frac{\lambda}{2}(\|A\bar{x}+Bz^k-b\|^2-\|Ax^{k+1}+Bz^k-b\|^2)\\ &+\frac{\lambda}{2}(-\|Ax^{k+1}-A\bar{x}\|^2+\|Ax^{k+1}+B\bar{z}-b\|^2)\\ &+\frac{\lambda}{2}(-\|Ax^{k+1}+Bz^{k+1}-b\|^2-\|Bz^{k+1}-B\bar{z}\|^2)\\ &+\frac{1}{2\lambda}\Big(H(\bar{x},x^k)-H(\bar{x},x^{k+1})-\kappa H(x^{k+1},x^k)\Big)\\ &+\frac{1}{2\lambda}(\|\bar{z}-z^k\|^2-\|z^{k+1}-z^k\|^2-\|\bar{z}-z^{k+1}\|^2)\\ &+\langle a^{k+1},x^{k+1}-\bar{x}\rangle+\langle b^{k+1},z^{k+1}-\bar{z}\rangle.\end{aligned} \tag{28}$$

From the equality (20) of the Lemma 3 with $y=\bar{y}$ we have

$$\ell(x^{k+1},z^{k+1},y^k)-\ell(x^{k+1},z^{k+1},\bar{y}) = \frac{1}{2\lambda}\Big(\|\bar{y}-y^{k+1}\|^2-\|\bar{y}-y^k\|^2-\|y^{k+1}-y^k\|^2\Big). \tag{29}$$

Replacing the left side of (28) with (29), and since $A\bar{x}+B\bar{z}=b$ we obtain:

$$\begin{aligned}&\frac{1}{2\lambda}(\|\bar{y}-y^{k+1}\|^2-\|\bar{y}-y^k\|^2\\ &\qquad-\|y^{k+1}-y^k\|^2)\leq\frac{\lambda}{2}\Big(\|Bz^k-B\bar{z}\|^2-\|Ax^{k+1}+Bz^k-b\|^2\Big)\\ &\qquad+\frac{\lambda}{2}\Big(-\|Ax^{k+1}+Bz^{k+1}-b\|^2-\|Bz^{k+1}-B\bar{z}\|^2\Big)\\ &\qquad+\frac{1}{2\lambda}\Big(H(\bar{x},x^k)-H(\bar{x},x^{k+1})-\kappa H(x^{k+1},x^k)\Big)\\ &\qquad+\frac{1}{2\lambda}\Big(\|\bar{z}-z^k\|^2-\|z^{k+1}-z^k\|^2-\|\bar{z}-z^{k+1}\|^2\Big)\\ &\qquad+\langle a^{k+1},x^{k+1}-\bar{x}\rangle+\langle b^{k+1},z^{k+1}-\bar{z}\rangle,\end{aligned}$$

and using the equation (14), i.e., $y^{k+1}-y^k=\lambda(Ax^{k+1}+Bz^{k+1}-b)$, results:

$$\begin{aligned}\frac{1}{2\lambda}(\|\bar{y}-y^{k+1}\|^2-\|\bar{y}-y^k\|^2) \leq\ & \frac{\lambda}{2}\left(\|Bz^k-B\bar{z}\|^2-\|Ax^{k+1}+Bz^k-b\|^2\right) \\ & -\frac{\lambda}{2}\|Bz^{k+1}-B\bar{z}\|^2 \\ & +\frac{1}{2\lambda}\left(H(\bar{x},x^k)-H(\bar{x},x^{k+1})-\kappa H(x^{k+1},x^k)\right) \\ & +\frac{1}{2\lambda}\left(\|\bar{z}-z^k\|^2-\|z^{k+1}-z^k\|^2-\|\bar{z}-z^{k+1}\|^2\right) \\ & +\langle a^{k+1},x^{k+1}-\bar{x}\rangle+\langle b^{k+1},z^{k+1}-\bar{z}\rangle,\end{aligned} \tag{30}$$

Set

$$E_k(\bar{x},\bar{z},\bar{y})=\frac{1}{2\lambda}\left(H(\bar{x},x^k)+\|z^k-\bar{z}\|^2+\|y^k-\bar{y}\|^2\right)+\frac{\lambda}{2}\|Bz^k-B\bar{z}\|^2.$$

Rearranging the last inequality we obtain:

$$\begin{aligned}E_{k+1}(\bar{x},\bar{z},\bar{y})+\frac{\kappa}{2\lambda}H(x^{k+1},x^k)+\frac{1}{2\lambda}\|z^{k+1}-z^k\|^2 & \\ +\frac{\lambda}{2}\|Ax^{k+1}+Bz^k-b\|^2 \leq E_k(\bar{x},\bar{z},\bar{y})&+\langle a^{k+1},x^{k+1}-\bar{x}\rangle \\ &+\langle b^{k+1},z^{k+1}-\bar{z}\rangle.\end{aligned} \tag{31}$$

The sum $\sum_{k=0}^{\infty}\langle a^{k+1},x^{k+1}-\bar{x}\rangle$ exists and is finite. In effect, we have that $\langle a^{k+1},x^{k+1}-\bar{x}\rangle = \langle a^{k+1},x^{k+1}\rangle-\langle a^{k+1},\bar{x}\rangle$ and

$$0\leq\sum_{k=0}^{\infty}|\langle a^{k+1},x^{k+1}\rangle|\leq\sum_{k=0}^{\infty}\|a^{k+1}\|\|x^{k+1}\|=\sum_{k=1}^{\infty}\|a^k\|\|x^k\|\leq\sum_{k=1}^{\infty}\eta_k,$$

$$0\leq\sum_{k=0}^{\infty}|\langle a^{k+1},\bar{x}\rangle|\leq\sum_{k=0}^{\infty}\|a^{k+1}\|\|\bar{x}\|=\|\bar{x}\|\sum_{k=1}^{\infty}\|a^k\|\leq\|\bar{x}\|\sum_{k=1}^{\infty}\varepsilon_k. \tag{32}$$

Because of $\sum_{k=1}^{\infty}\varepsilon_k<\infty$ and $\sum_{k=1}^{\infty}\eta_k<\infty$, we get $\sum_{k=0}^{\infty}|\langle a^{k+1},x^{k+1}\rangle|$ and $\sum_{k=0}^{\infty}|\langle a^{k+1},\bar{x}\rangle|$ are finite. Then $\sum_{k=0}^{\infty}\langle a^{k+1},x^{k+1}\rangle$ and $\sum_{k=0}^{\infty}\langle a^{k+1},\bar{x}\rangle$ exist and are finite. In similar way, we can prove that $\sum_{k=0}^{\infty}\langle b^{k+1},z^{k+1}-\bar{z}\rangle$ is convergent.
Now, we apply Lemma 2 to (31) with $w_k=E_k(\bar{x},\bar{y},\bar{z})$ and $\beta_k=\langle a^{k+1},x^{k+1}-\bar{x}\rangle+\langle b^{k+1},z^{k+1}-\bar{z}\rangle$ establishing that $\{E_k(\bar{x},\bar{y},\bar{z})\}$ is convergent (with that we have proved (i)) and then $\{E_k(\bar{x},\bar{y},\bar{z})\}$ is bounded. Moreover, from Definition 2 (iii) we have that $H(\bar{x},\cdot)$ is level bounded on C and therefore $\{(x^k,z^k,y^k)\}$ is bounded. From (31) we have that the squares of the quantities $\|z^{k+1}-z^k\|$ and $\|Ax^{k+1}+Bz^k-b\|$, and $H(x^{k+1},x^k)$ are summable, hence vanish as k goes to $+\infty$. The proof is complete. □

Now, we are ready to give the proof of our main result.

Proof Theorem 1.

(i) By Assumption (A_2) we know that there exists a saddle point (x^*,z^*,y^*) of ℓ on $\overline{C}\times\mathbb{R}^m\times\mathbb{R}^n$. From Proposition 2 (ii) it follows that $\{(x^k,z^k,y^k)\}$ is bounded and then this sequence has a subsequence $\{(x^{k_j},z^{k_j},y^{k_j})\}$ which converges to some $(x^\infty,z^\infty,y^\infty)\in\overline{C}\times\mathbb{R}^m\times\mathbb{R}^p$ (as $x^{k_j}\in C$ in the limit we get $x^\infty\in\overline{C}$).

We will demostrate that (x^∞,z^∞) is a feasible point for (P) and then we'll prove that $f(x^\infty)+g(z^\infty)\le f(x^*)+g(z^*)$. From Proposition 2 (ii) we have that the quantities $H(x^{k+1},x^k)$ and $\|Ax^{k+1}+Bz^k-b\|$ vanish as k goes to $+\infty$. Then $H(x^{k_j+1},x^{k_j})\to 0$, and due to (B_1) we have $x^{k_j+1}\to x^\infty$. The quantity $\|Ax^{k_j}+Bz^{k_j}-b\|$ converge to zero because $\|x^{k_j+1}-x^{k_j}\|\to 0$ and $\|Ax^{k_j+1}+Bz^{k_j}-b\|\to 0$. Then, $0=\lim_{j\to+\infty}\|Ax^{k_j}+Bz^{k_j}-b\|=\|Ax^\infty+Bz^\infty-b\|$ and so $Ax^\infty+Bz^\infty=b$. Therefore, (x^∞,z^∞) is a feasible point for (P). Considering the inequality (19) of the Lemma 3 on the sequence $\{(x^{k_j},z^{k_j},y^{k_j})\}$, and with $x=x^*$ and $z=z^*$ we have:

$$\begin{aligned}\ell(x^{k_j+1},z^{k_j+1},y^{k_j})-\ell(x^*,z^*,y^{k_j})\le\ &\frac{\lambda}{2}(\|Ax^*+Bz^{k_j}-b\|^2-\|Ax^{k_j+1}+Bz^{k_j}-b\|^2\\&-\|Ax^{k_j+1}-Ax^*\|^2)+\frac{\lambda}{2}(\|Ax^{k_j+1}+Bz^*-b\|^2\\&-\|Ax^{k_j+1}+Bz^{k_j+1}-b\|^2-\|Bz^{k_j+1}-Bz^*\|^2)\\&+\frac{1}{2\lambda}\left(H(x^*,x^{k_j})-H(x^*,x^{k_j+1})-\kappa H(x^{k_j+1},x^{k_j})\right)\\&+\frac{1}{2\lambda}\left(\|z^*-z^{k_j}\|^2-\|z^*-z^{k_j+1}\|^2-\|z^{k_j+1}-z^{k_j}\|^2\right)\\&+\langle a^{k_j+1},x^{k_j+1}-x^*\rangle+\langle b^{k_j+1},z^{k_j+1}-z^*\rangle.\end{aligned}$$

From the definition of $E_k(x^*,z^*,y^*)$ (cf. (26)) and since $Ax^*+Bz^*=b$, the last inequality is equivalent to:

$$\begin{aligned}\ell(x^{k_j+1},z^{k_j+1},y^{k_j})-\ell(x^*,z^*,y^{k_j})\le\ &E_{k_j}(x^*,z^*,y^*)-E_{k_j+1}(x^*,z^*,y^*)\\&-\frac{1}{2\lambda}\left(\|y^{k_j}-y^*\|^2-\|y^{k_j+1}-y^*\|^2\right)\\&-\frac{\lambda}{2}\left(\|Ax^{k_j+1}+Bz^{k_j}-b\|^2+\|Ax^{k_j+1}+Bz^{k_j+1}-b\|^2\right)\\&-\frac{\kappa}{2\lambda}H(x^{k_j+1},x^{k_j})-\frac{1}{2\lambda}\|z^{k_j+1}-z^{k_j}\|^2\\&+\langle a^{k_j+1},x^{k_j+1}-x^*\rangle+\langle b^{k_j+1},z^{k_j+1}-z^*\rangle.\end{aligned}\tag{33}$$

The right side of inequality (33) goes to zero as j tends to $+\infty$. Indeed, from the Proposition 2 (ii) we have that $\|z^{k+1}-z^k\|\to 0$, and $z^{k_j}\to z^\infty$, then $z^{k_j+1}\to z^\infty$ as j goes to $+\infty$. Moreover, since $y^{k_j+1}=y^{k_j}+\lambda(Ax^{k_j+1}+Bz^{k_j+1}-b)$, $y^{k_j}\to y^\infty$ and $Ax^\infty+Bz^\infty=b$, we have $y^{k_j+1}\to y^\infty$. Therefore, the subsequence $\{(x^{k_j+1},z^{k_j+1},y^{k_j+1})\}$ converges to $(x^\infty,z^\infty,y^\infty)$ as j goes to $+\infty$. On the other hand, from the Proposition 2(ii) we have $E_{k_j}(x^*,z^*,y^*)-E_{k_j+1}(x^*,z^*,y^*)\to 0$. Also, the quantities $\|Ax^{k_j+1}+Bz^{k_j}-b\|^2$ and $\|Ax^{k_j+1}+Bz^{k_j+1}-b\|^2$ converge to $\|Ax^\infty+Bz^\infty-b\|^2$ and since $Ax^\infty+Bz^\infty-b=0$, they vanish as j goes to $+\infty$. Moreover, by the Cauchy-Schawrz we have:

$$|\langle a^{k_j+1},x^{k_j+1}-x^*\rangle|\le\|a^{k_j+1}\|\|x^{k_j+1}-x^*\|.$$

From (11), (13) and (15) we have that $\|a^{k+1}\| \to 0$. Then

$$|\langle a^{k_j+1}, x^{k_j+1} - x^* \rangle| \to 0,$$

as j goes to $+\infty$. In the similar way, we can prove that: $|\langle b^{k_j+1}, z^{k_j+1} - z^* \rangle| \to 0$.
Taking limit as j goes to $+\infty$ in the inequality (33) we have that the right side goes to zero, and due to f and g are closed functions we obtain:

$$f(x^\infty) + g(z^\infty) \leq f(x^*) + g(z^*).$$

(ii) We will demonstrate that $(x^\infty, z^\infty, y^\infty)$ is a saddle point of ℓ on $\overline{C} \times \mathbb{R}^m \times \mathbb{R}^p$, that is,

$$\ell(x^\infty, z^\infty, y) \leq \ell(x^\infty, z^\infty, y^\infty) \leq \ell(x, z, y^\infty), \forall x \in \overline{C},\ \forall z \in \mathbb{R}^m,\ \forall y \in \mathbb{R}^p. \quad (34)$$

From part (i), we know (x^∞, z^∞) is a feasible point of (P), then

$$\ell(x^\infty, z^\infty, y) = \ell(x^\infty, z^\infty, y^\infty),\ \forall y \in \mathbb{R}^p. \quad (35)$$

Now we'll prove the rigth inequality of (34). Consider the inequality (19) in Lemma 3 where (x^k, z^k, y^k) and (x^{k+1}, z^{k+1}) are replaced by $(x^{k_j}, z^{k_j}, y^{k_j})$ and (x^{k_j+1}, z^{k_j+1}) respectively. Moreover, applying the Cauchy-Schwarz inequality we get

$$\begin{aligned}
\ell(x^{k_j+1}, z^{k_j+1}, y^{k_j}) - \ell(x, z, y^{k_j}) \leq\ & \frac{\lambda}{2}(\|Ax + Bz^{k_j} - b\|^2 - \|Ax^{k_j+1} + Bz^{k_j} - b\|^2 \\
& - \|Ax^{k_j+1} - Ax\|^2) + \frac{\lambda}{2}(\|Ax^{k_j+1} + Bz - b\|^2 \\
& - \|Ax^{k_j+1} + Bz^{k_j+1} - b\|^2 - \|Bz^{k_j+1} - Bz\|^2) \\
& + \frac{1}{2\lambda}\left(H(x, x^{k_j}) - H(x, x^{k_j+1}) - \kappa H(x^{k_j+1}, x^{k_j})\right) \\
& + \frac{1}{2\lambda}\left(\|z - z^{k_j}\|^2 - \|z^{k_j+1} - z^{k_j}\|^2 - \|z - z^{k_j+1}\|^2\right) \\
& + \|a^{k_j+1}\|\|x^{k_j+1} - x\| + \|b^{k_j+1}\|\|z^{k_j+1} - z\|. \quad (36)
\end{aligned}$$

Taking $\liminf$ as $j \to \infty$ on both sides of (36), and since $Ax^\infty + Bz^\infty - b = 0$ we obtain

$$\liminf \left\{ \ell(x^{k_j+1}, z^{k_j+1}, y^{k_j}) - \ell(x, z, y^{k_j}) \right\} \leq \liminf \left\{ \frac{1}{2\lambda}\left(H(x, x^{k_j}) - H(x, x^{k_j+1})\right) \right\}.$$

From the assumption (B_2) we have that $\lim H(x, x^{k_j}) = H(x, x^\infty) = \lim H(x, x^{k_j+1})$, and since f and g are closed functions we obtain:

$$\ell(x^\infty, z^\infty, y^\infty) \leq \ell(x, z, y^\infty),\ \forall x \in \mathrm{dom}(f), \forall z \in \mathrm{dom}(g). \quad (37)$$

From (35) and (37) we have $(x^\infty, z^\infty, y^\infty)$ is a saddle point of ℓ.

Now we prove that the sequence $\{(x^k, z^k, y^k)\}$ globally converges to the saddle point $(x^\infty, z^\infty, y^\infty)$. From Proposition 2.(i) the following limit exists: $E = \lim_{k\to\infty} E_k(x^\infty, z^\infty, y^\infty)$. But for the subsequence $\{k_j\}$ we have that $(x^{k_j}, z^{k_j}, y^{k_j}) \to (x^\infty, z^\infty, y^\infty)$ and by Assumption (B_3) we have $\lim_{k\to\infty} H(x^\infty, x^{k_j}) = 0$. Hence $\lim_{j\to\infty} E_{k_j}(x^\infty, z^\infty, y^\infty) = 0$. Then $E = \lim_{j\to\infty} E_{k_j}(x^\infty, z^\infty, y^\infty) = 0$, and so $H(x^\infty, x^k) \to 0$. Moreover, due to (B_1) we obtain $x^k \to x^\infty$, consequently, we get $(x^k, z^k, y^k) \to (x^\infty, z^\infty, y^\infty)$, and the proof is complete. □

4 Numerical Experiences

In this section, we study three types of problems in the form (P) with $C = \mathbb{R}^n_{++}$, that is, of the form:

$$(P) \qquad \min_{x,z} \{f(x) + g(z) \mid Ax + Bz = b,\ x \in \mathbb{R}^n_+\}.$$

The first problem is the constrained lasso problem, the second one is a modification of the constrained lasso problem, and the third one is a learning machine problem. In order to solve each of these problems, we apply three methods: our approach (RIPADM), the proximal method of multipliers (PMM), and the classical ADM (cf. (7)-(9)). Recall that the PMM (proposed by Rockafellar (1976)) is obtained by applying the proximal algorithm to the primal-dual system, that is, it generates a sequence $\{(x^k, y^k, z^k)\}$ via the following iterates

$$(x^{k+1}, z^{k+1}) = \arg\min_{x \geq 0, z} \ell_\lambda(x, z, y^k) + \frac{1}{2\lambda}(\|x - x^k\|^2 + \|z - z^k\|^2) \tag{38}$$

$$y^{k+1} = y^k + \lambda(Ax^{k+1} + Bz^{k+1} - b), \tag{39}$$

where ℓ_λ denotes the Lagrangian augmented of the problem (P) (cf. (6)). On the other hand, in the RIPADM scheme a non-quadratic proximal distance is used to treat positivity constraint $x \geq 0$. In our numerical experiments will use the Log-quad distance with $\mu = 1$ and $\nu = 2$ (cf. (16)).

The numerical experiments were done on a laptop with Intel Core i3 processor, 1.80 GHz, 4 GB RAM. The operating system of the laptop is Windows 8, and the program used is MATLAB version 8.2.0.701 (R2013b).

In our experiments, the following stopping rule was taken: $|val_k - val^*| < 10^{-5}$, where val_k denotes the value of the objective function at the iteration k of the appropriate algorithm, and val^* denotes the optimal value obtained by CVX solver, which is available in `http://cvxr.com/cvx`. This solver also was used for solving the inner problems that appear in the three methods: RIPADM, PMM and ADM.

4.1 Constrained Lasso problem

Consider the standard linear regression model:

$$d = Dz + \varepsilon,$$

where $D \in \mathbb{R}^{r \times m}$ is a matrix of predictors, $d \in \mathbb{R}^r$ is a response vector, $z \in \mathbb{R}^m$ is a vector of regression coefficients, and $\varepsilon \in \mathbb{R}^r$ is a vector of random noises. In high-dimensional setting where the number of responses is much smaller than the number of regression coefficients, $r \ll m$, the traditional least-squares method does not perform well. To overcome this difficulty, certain sparsity conditions are assumed on the vector of regression coefficients, that is, one consider the following problem

$$\min_z \left\{ \frac{1}{2}\|Dz - d\|_2^2 + \gamma\|z\|_1 \right\},$$

where $\gamma > 0$ is a running parameter, and $\|\cdot\|_1$ denotes the 1-norm in $\mathbb{R}^m$. This problem is known as *Lasso problem* and it can be interpreted as finding a sparse solution to a least squares or linear regression problem, that is, the lasso problem is L_1-regularized linear regression (Tibshirani (1996)).

Inspired by significant applications such as portfolio selection (Fan et al. (2012)) and monotone regression (James et al. (2012)), the following problem was proposed recently in James et al. (2012):

$$\min_z \left\{ \frac{1}{2}\|Dz-d\|^2 + \gamma\|z\|_1 \mid Bz \leq b \right\}, \tag{40}$$

where $D \in \mathbb{R}^{r\times m}$, $d \in \mathbb{R}^r$, $B \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$, and $\gamma > 0$ are problem data. This problem is known as *constrained Lasso problem*.

By introducing a slack variable $x \in \mathbb{R}^n$, we can rewrite the problem (40) like (P) with $f(x) = 0$, $g(z) = \frac{1}{2}\|Dz-d\|^2 + \gamma\|z\|_1$, and $A = I$. Thus, the RIPADM, PMM, and ADM are applicable.

RIPADM: This iterative scheme consists of the following steps:

$$\begin{aligned}
x^{k+1} &\approx \operatorname{argmin}_x \langle y^k, x\rangle + \frac{\lambda}{2}\|x + Bz^k - b\|^2 + \frac{1}{2\lambda} d(x, x^k),\\
z^{k+1} &\approx \operatorname{argmin}_z \frac{1}{2}\|Dz-d\|^2 + \gamma\|z\|_1 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2,\\
y^{k+1} &= y^k + \lambda(x^{k+1} + Bz^{k+1} - b).
\end{aligned} \tag{41}$$

Remark 2 If we replace the distance d by the Log-quad distance (16), the x-problem is rewrite as:

$$x^{k+1} \approx \operatorname{argmin}_x \langle y^k, x\rangle + \frac{\lambda}{2}\|x + q^k\|^2 + \frac{1}{2\lambda}\sum_{i=1}^n \left[\mu\left((x_i^k)^2 \log\left(\frac{x_i^k}{x_i}\right) + x_i x_i^k - (x_i^k)^2\right) + \frac{\nu}{2}(x_i - x_i^k)^2\right],$$

where $q^k = Bz^k - b$. The optimality condition of this problem give us the following equations:

$$y_i^k + \lambda(x_i^{k+1} + q_i^k) + \frac{1}{2\lambda}\left(-\mu\frac{(x_i^k)^2}{x_i^{k+1}} + \mu x_i^k + \nu x_i^{k+1} - \nu x_i^k\right) = 0, \quad i = 1, \ldots, n.$$

Then, from this, we can obtain the following closed-form solution

$$x_i^{k+1} = \frac{-\tilde{b}_i + \sqrt{\tilde{b}_i^2 - 4ac_i}}{2a}, \quad i = 1, \ldots, n, \tag{42}$$

where $a = \lambda + \frac{\nu}{2\lambda}$, $\tilde{b}_i = y_i^k + \lambda q_i^k + ((\mu - \nu)/2\lambda)x_i^k$, and $c_i = (-\mu/2\lambda)(x_i^k)^2$.

PMM: Here, we compute the sequences $\{(x^k, z^k, y^k)\}$ via (38)-(39) with augmented Lagrangian

$$\ell_\lambda(x, y, z) = \frac{1}{2}\|Dz-d\|^2 + \gamma\|z\|_1 + \langle y, x + Bz - b\rangle + \frac{\lambda}{2}\|x + Bz - b\|^2. \tag{43}$$

ADM: This method computes the sequences $\{(x^k, z^k, y^k)\}$ via (7)-(9) with augmented Lagrangian given by (43).

r	n		RIPADM	PMM	ADM
10	30	CPU time(s)	**192.29**	291.33	225.80
		Objective Value	1.309512	1.309514	1.309513
		Iterations	231	299	**178**
30	50	CPU time(s)	**79.21**	113.56	149.75
		Objective Value	3.343769	3.343757	3.343770
		Iterations	93	**88**	90
50	100	CPU time(s)	140.17	**90.47**	162.56
		Objective Value	4.103247	4.103251	4.103239
		Iterations	123	**51**	72
70	200	CPU time(s)	**337.70**	388.60	648.58
		Objective Value	6.354824	6.354806	6.354807
		Iterations	158	**102**	128
100	300	CPU time(s)	765.18	**725.76**	1168.11
		Objective Value	7.855478	7.855494	7.855478
		Iterations	151	**73**	100

Table 1 Numerical results for the constrained lasso problem with $\lambda = 1$.

In Table 1 we present the numerical results obtained when we apply the RIPADM, PMM, and ADM schemes for solving the problem (40). For each scheme, we take the same starting point $(x^0, z^0, y^0) = (\mathbf{1}, \mathbf{1}, \mathbf{3}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$, $\gamma = 1$, and set $m = n$. The problem data were randomly generated from the uniform distribution and with a single stream fixed, that is, `s=RandStream('mt19937ar','Seed',1)`, D`=rand(s,r,m)`, B`=rand(s,n,m)`, d`=rand(s,r,1)`, and b`=rand(s,n,1)`.

Table 1 shows the average of runtime in seconds (run five times), the objective values and the number of iterations of the RIPADM, PMM, and ADM schemes apply to Problem (40) for different values of r and n. The best results in terms of CPU time and number of iterations is highlighted in bold type. From Table 1, we observe that the PMM scheme uses fewer iterations (four out of five experiments) in comparison with the other two methods, however our approach uses less CPU time in three out of five experiments.

4.2 Constrained Lasso problem with a cost function

In this section, we consider a cost function associated to the problem (40), specifically, we consider the following problem

$$\min_{x,z} \left\{ \frac{1}{2}\|Dz-d\|^2 + \gamma\|z\|_1 + \frac{\beta}{2}\|x\|^2 \mid x + Bz = b,\ x \geq 0 \right\}, \tag{44}$$

with $\beta > 0$. Clearly, this problem has the form of (P), thus the RIPADM, PMM, and ADM are again applicable.

RIPADM: This iterative scheme consists of the following steps:

$$\begin{aligned}
x^{k+1} &\approx \operatorname{argmin}_x \frac{\beta}{2}\|x\|^2 + \langle y^k, x\rangle + \frac{\lambda}{2}\|x + Bz^k - b\|^2 + \frac{1}{2\lambda} d(x, x^k), \\
z^{k+1} &\approx \operatorname{argmin}_z \frac{1}{2}\|Dz-d\|^2 + \gamma\|z\|_1 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2, \\
y^{k+1} &= y^k + \lambda(x^{k+1} + Bz^{k+1} - b).
\end{aligned}$$

Remark 3 In a way similar to Remark 2, if we replace the distance d by the Log-quad distance (16), we can obtain the following closed-form solution for the x-problem

$$x_i^{k+1} = \frac{-\tilde{b}_i + \sqrt{\tilde{b}_i^2 - 4ac_i}}{2a}, \quad i = 1, \ldots, n, \tag{45}$$

where $a = \beta + \lambda + \frac{\nu}{2\lambda}$, $\tilde{b}_i = y_i^k + \lambda q_i^k + ((\mu - \nu)/2\lambda)x_i^k$, and $c_i = (-\mu/2\lambda)(x_i^k)^2$.

PMM: Here, we compute the sequences $\{(x^k, z^k, y^k)\}$ via (38)-(39) with augmented Lagrangian

$$\ell_\lambda(x, y, z) = \frac{1}{2}\|Dz - d\|^2 + \gamma\|z\|_1 + \frac{\beta}{2}\|x\|^2 + \langle y, x + Bz - b\rangle + \frac{\lambda}{2}\|x + Bz - b\|^2. \tag{46}$$

ADM: This method computes the sequences $\{(x^k, z^k, y^k)\}$ via (7)-(9) with augmented Lagrangian given by (46).

In Table 2 we summarize the numerical results obtained when we apply the RIPADM, PMM, and ADM schemes for solving the problem (44). For each scheme, we take the same starting point $(x^0, z^0, y^0) = (\mathbf{1}, \mathbf{1}, \mathbf{3}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$, $\gamma = \beta = 1$, and set $m = n$. The problem data were randomly generated in similar way to the above case.

r	n		RIPADM	PMM	ADM
10	30	CPU time(s)	**232.71**	355.69	479.47
		Objective Value	3.715823	3.715823	3.715823
		Iterations	**324**	326	315
30	50	CPU time(s)	**41.10**	133.51	111.32
		Objective Value	6.855128	6.855116	6.855122
		Iterations	**51**	104	67
50	100	CPU time(s)	**140.34**	249.80	286.16
		Objective Value	10.501275	10.501275	10.501275
		Iterations	**131**	141	143
70	200	CPU time(s)	**220.39**	427.66	505.65
		Objective Value	14.609375	14.609376	14.609376
		Iterations	**100**	112	114
100	300	CPU time(s)	**658.07**	1152.91	1560.74
		Objective Value	23.198968	23.198968	23.198968
		Iterations	**137**	139	146

Table 2 Numerical results for the constrained lasso problem with cost function and $\lambda = 1$.

It can be seen in Table 2 that the RIPADM scheme is much faster than the other two methods (see highlighted in bold type) for the different values of r and n. For instance, for $r = 70$, $n = 200$ the PMM scheme needs almost twice the CPU time of the RIPADM one, while the ADM scheme requires more than twice of CPU time of our approach. This can be see in Figure 1, where we show the evolution of objective function values with respect to CPU time for the three schemes.

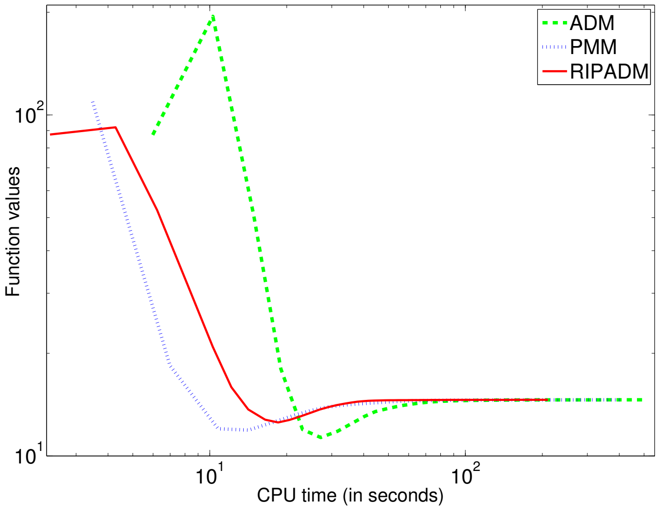

Fig. 1 Objective function values versus CPU time (in seconds).

4.3 Twin support vector machine classifier

Support vector machines (SVM) is a new machine learning method which is developed on the basic of statistical learning theory and structural risk minimization Cortes and Vapnik (1995); Vapnik (1995). One of the most popular SVM in classification is the "maximum margin" one that attempts to find the optimal separating hyperplane maximizing the margin between two disjoint half planes (associated to positive and negative samples). The resulting optimization task involves the minimization of a convex quadratic function subject to linear inequality constraints.

In order to reduce the computational cost of SVM, Jayadeva et al. (2007) proposed a nonparallel hyperplane classifier, called Twin support vector machines (TWSVM) for binary classification. TWSVMs constructs two nonparallel hyperplanes such that each one is closer to one of the two training dataset and is as far as possible from the other. The above hyperplanes are obtaining by solving two small quadratic programming problems.

On the other hand, an important task in classification is to identify a subset of features which contribute most to classification. The benefit of feature selection is crucial for achieving good classification accuracy in the presence of redundant features. To select important groups of features automatically and simultaneously, Zou and Yuan (2009) proposed to consider the F_∞-norm SVM.

Motivated from the works on TWSVM, and F_∞-norm SVMs, for linearly separable case, we propose to consider the norm-mixed TWSVM given by the following problems

$$\begin{aligned} \min_{w_1,t_1} \ & \|D_1 w_1 + \mathbf{e}_1 t_1\|_\infty + \frac{c_1}{2}(\|w_1\|^2 + t_1^2) \\ \text{s.t.} \quad & -(D_2 w_1 + \mathbf{e}_2 t_1) \geq \mathbf{e}_2, \end{aligned} \tag{47}$$

and

$$\begin{aligned} \min_{w_2,t_2} \; & \|D_2 w_2 + \mathbf{e}_2 t_2\|_\infty + \frac{c_2}{2}(\|w_2\|^2 + t_2^2) \\ \text{s.t.} \quad & (D_1 w_2 + \mathbf{e}_1 t_2) \geq \mathbf{e}_1, \end{aligned} \tag{48}$$

where $\|\cdot\|_\infty$ denotes the infinity-norm in Euclidean space, $D_1 \in \mathbb{R}^{m_1\times n}$ and $D_2 \in \mathbb{R}^{m_2\times n}$ are matrices containing the training dataset of positive and negative class, respectively, c_1, and c_2 are positive parameters, and $\mathbf{e}_1 \in \mathbb{R}^{m_1}$ and $\mathbf{e}_2 \in \mathbb{R}^{m_2}$ are vectors of ones.

We will focus on solving the problem (47). Set $z = [w_1^\top, t_1]^\top \in \mathbb{R}^{n+1}$. Then the problem (47) can be write in the form:

$$\min_z \left\{ \|[D_1 \; \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 | -[D_2 \; \mathbf{e}_2]z \geq \mathbf{e}_2 \right\}. \tag{49}$$

By introducing a slack variable $x \in \mathbb{R}^{m_2}$ to the inequality constraints of the above problem, we can reformulate it in the form of (P) with $f(x) = 0$, $g(z) = \|[D_1 \; \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2$, $A = I$, $B = [D_2 \; \mathbf{e}_2] \in \mathbb{R}^{m_2\times n+1}$ and $b = -\mathbf{e}_2$. Then, the RIPADM, PMM, and ADM are applicable.

RIPADM: This iterative scheme consists of the following steps:

$$\begin{aligned} x^{k+1} &\approx \operatorname{argmin}_x \langle y^k, x\rangle + \frac{\lambda}{2}\|x + Bz^k - b\|^2 + \frac{1}{2\lambda} d(x, x^k), \\ z^{k+1} &\approx \operatorname{argmin}_z \|[D_1 \; \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 + \langle y^k, Bz\rangle + \frac{\lambda}{2}\|x^{k+1} + Bz - b\|^2 + \frac{1}{2\lambda}\|z - z^k\|^2, \\ y^{k+1} &= y^k + \lambda(x^{k+1} + Bz^{k+1} - b). \end{aligned}$$

PMM: Here, the sequences $\{(x^k, z^k, y^k)\}$ are compute via (38)-(39) with

$$\ell_\lambda(x,y,z) = \|[D_1 \; \mathbf{e}_1]z\|_\infty + \frac{c_1}{2}\|z\|^2 + \langle y, x + Bz - b\rangle + \frac{\lambda}{2}\|x + Bz - b\|^2. \tag{50}$$

ADM: This method computes the sequences $\{(x^k, z^k, y^k)\}$ via (7)-(9) with augmented Lagrangian given by (50).

In order to solve numerically the norm-mixed TWSVM problem (49) with the RIPADM, PMM, and ADM schemes, we use four real data sets taken from the UCI Repository (Asuncion and Newman (2007)): Liver Disorders (BUPA), Australian Credit (AUS), Wisconsin Breast Cancer (WBC), and Pima Indians Diabetes (DIA). A brief information regarding each of the data sets is given below:

- **BUPA**: It contains $m = 345$ samples of patients, divided into $m_1 = 145$ and $m_2 = 200$, with $n = 6$ attributes.
- **AUS**: It concerns credit card applications and contains $m = 690$ samples, divided into $m_1 = 145$ and $m_2 = 200$, with $n = 14$ features.
- **WBC**: It contains $m = 569$ observations of tissue samples ($m_1 = 212$ diagnosed as malignant and $m_2 = 357$ as benign tumors) described by $n = 30$ features.
- **DIA**: It contains $m = 768$ samples of patients, divided into $m_1 = 268$ (tested positive) and $m_2 = 500$ (tested negative), with $n = 8$ attributes.

For each scheme and data set, we take the same starting point $(x^0, z^0, y^0) = \frac{1}{10}(\mathbf{1}, \mathbf{0}, \mathbf{0}) \in \mathbb{R}^{m_2} \times \mathbb{R}^{n+1} \times \mathbb{R}^{m_1}$ and $c_1 = 1$. In Table 3, we present the results that have been obtained in the numerical tests. The best method in terms of CPU time and number of iterations is highlighted in bold type.

Data set		RIPADM	PMM	ADM
BUPA	CPU time(s)	922.73	1342.85	**684.79**
	Objective Value	1.350705	1.350714	1.350709
	Iterations	306	**192**	**192**
AUS	CPU time(s)	**486.00**	1815.14	724.12
	Objective Value	1.260135	1.260127	1.260134
	Iterations	101	148	**99**
WBC	CPU time(s)	**5434.05**	18213.59	7826.76
	Objective Value	1.496976	1.496976	1.496976
	Iterations	1407	1469	**1406**
DIA	CPU time(s)	83.83	318.25	**82.04**
	Objective Value	1.500000	1.500002	1.499990
	Iterations	13	24	**9**

Table 3 Numerical results for the learning machine problem with $\lambda = 1$.

From this Table we observe that the best results (in number of iterations) were achieved using the ADM scheme in all data sets, however, the RIPADM one has better CPU time in two data sets (AUS and WBC). Both schemes have relatively similar performance in one data set (DIA).

5 Concluding remarks

In this work, we have proposed an inexact proximal alternating direction method using a regularized interior proximal term, called RIPADM, for solving a convex problem with linear constraints and a separable objective function (see Problem (P)). Under standard assumptions, we have established that the sequences generated by the RIPADM converge globally. Then, we have applied the RIPADM, PMM and ADM scheme for solving three type of problems that appear in statistical learning applications: Lasso constrained, Lasso constrained with a cost function and twin support vector machines. The numerical experiments show that the RIPADM method achieves better results, in terms of CPU time and number of iterations (see Table 2), for the second application (cf. (44)). However, in the other two applications the RIPADM scheme has better performance (in terms of CPU time) in some tests.

These numerical results could be improved if we consider a relaxation factor for the dual sequence $\{y^k\}$, such as the works of Glowinski (1984); He et al. (2003); Xu (2007). Specifically, the idea would be to replace (14) by

$$y^{k+1} = y^k + \theta\lambda(Ax^{k+1} + Bz^{k+1} - b),$$

with $\theta \in (0, \bar{\theta})$, where $\bar{\theta}$ is a value to be determined. This analysis will be a future work.

Acknowledgements The work reported in this paper has been partially funded by Millennium Scientific Institute on Complex Engineering Systems Institute (ICM: P-05-004-F, CONICYT: FBO16).

References

Asuncion, A., D.J., Newman, UCI machine learning repository, (2007).

Attouch, H., A. Cabot, P. Frankel, J. Peypouquet, Alternating proximal algorithms for linearly constrained variational inequalities: application to domain decomposition for PDE's, *Nonlinear Anal.,* **74,** 7455–7473 (2011).

Attouch, H., M. Soueycatt, Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE'S and control, *Pac. J. Optim.,* **5,** 17–37 (2009).

Auslender, A., M. Teboulle, Interior gradient and proximal methods for convex and conic optimization *SIAM J. Optim.,* **16,** 697–725 (2006).

Auslender, A., M. Teboulle, Entropic proximal decomposition methods for convex programs and variational inequalities, *Math. Program.,* **91,** 33–47 (2001).

Auslender, A., M. Teboulle, S. Ben-Tiba, A logarithmic-quadratic proximal method for variational inequalities, *Comput. Optim. Appl.,* **12,** 31–40 (1999).

Auslender, A., M. Teboulle, S. Ben-Tiba, Interior proximal and multiplier methods based on second order homogeneous kernels, *Math. Oper. Res.,* **24,** 645–668 (1999).

Bauschke, H., J. Borwein , P. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.,* **42,** 596–636 (2003).

Boley, D., Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs, *SIAM J. Optim.,* **23,** 2183–2207 (2013).

Boyd, S., N. Parikh , E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statical learning via the alternating direction method of multipliers, *Found. Trends. Mach. Learning,* **3,** 1–122 (2010).

Bot, R., C. Hendrich, A Douglas–Rachford Type Primal-Dual Method for Solving Inclusions with Mixtures of Composite and Parallel-Sum Type Monotone Operators, *SIAM J. Optim.,* **23,** 2541–2565 (2013).

Briceño-Arias, L., P. Combettes, A monotone+ skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.,* **21,** 1230–1250 (2011).

Burachick, R., A. Iusem, A generalized proximal point algorithm for the variational inequality problem in a Hilbert Space, *SIAM J. Optim.,* **8,** 197–216 (1998).

Burachick, R., J. Dutta, Inexact proximal point methods for variational inequality problems, *SIAM J. Optim.,* **5,** 2653–2678 (2010).

Chen, G., M. Teboulle, A proximal-based method for convex minimization problems, *Math. Program.,* **64,** 81–101 (1994).

Chen, G., M. Teboulle, Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM J. Optim.,* **3,** 538–543 (1993).

Combettes, P., Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization,* **53,** 475–504 (2004).

Combettes, P., J.-C. Pesquet A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, *IEEE Journal of Selected Topics in Signal Processing,* **1,** 564–574 (2007).

Cortes, C., V. Vapnik, Support-vector networks, *Mach. Learn.,* **20,** 273–297 (1995).

Douglas, J., H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, *Transactions of the American mathematical Society,* 421–439 (1956).

Eckstein, J., Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. Oper. Res.,* **18,** 202–226 (1993).

Eckstein, J, Some saddle-function splitting methods for convex programming, *Optim. Methods Soft.,* **4,** 75–83 (1994).

Eckstein, J., P. Silva, Proximal methods for nonlinear programming: double regularization and inexact subproblems, *Comput. Optim. Appl.,* **46,** 279–304 (2010).

Fan, J., J. Zhang, K. Yu, Vast portfolio selection with gross-exposure constraints, *J. Am. Stat. Assoc.,* **107,** 592–606 (2012).

Fang, E.X., B. He, H. Liu, X. Yuan, Generalized alternating direction method of multipliers: new theoretical insights and applications, *Math. Prog. Comp.,* **7,** 149–187 (2015).

Gabay, D., B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Computers and Mathematics with Applications,* **2,** 17–40 (1976).

Glowinski, R., *Numerical Methods for Nonlinear Variational Problems.* Springer-Verlag, New York (1984).

Glowinski, R., P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics.* SIAM Studies in Applied Mathematics, Philadelphia (1989).

He, B., S. Wang, H. Yang, A modified variable-penalty alternating directions method for monotone variational inequalities, *J. Comput. Math.,* **21,** 495–504 (2003).

Iusem, A., Some properties of generalized proximal points methods for quadratic and linear programming, *J. Optim. Theory Appl.,* **85,** 593–612 (1995).

James, G.M., C. Paulson, P. Rusmevichientong, The Constrained Lasso, *Manuscript,* (2012).

Jayadeva, R. Khemchandni, S. Chandra Twin support vector machines for pattern classification, *IEEE Trans. Pattern Anal. Mach, Intell. ,* **29,** 905–910 (2007).

Kiwiel, K., Proximal minimization methods with generalized Bregman functions, *SIAM J. Control Optim.,* **35,** 1142–1168 (1997).

Li, M., L. Liao, X. Yuan, Inexact alternating direction methods of multipliers with logarithmicquadratic proximal regularization, *J. Optim. Theory Appl.,* **159,** 412–436 (2013).

Polyak, B., *Introduction to optimization.* Optimization Software Inc., New York (1987).

Rockafellar, R., *Convex Analysis.* Princeton University Press, Princeton (1970).

Rockafellar, R., Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.,* **1,** 97–116 (1976).

Silva, P., J. Eckstein, Double-regularization proximal methods with complementarity applications, *Comput. Optim. Appl.,* **33,** 115–156 (2006).

Solodov, M., B. Svaiter, An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, *Math. Oper. Res.,* **25,** 214–230 (2000).

Stephanopoulos, G., A. Westerberg, The use of Hestenes' method of multipliers to resolve dual gaps in engineering system optimization, *J. Optim. Theory Appl.,* **15,** 285–309 (1975).

Svaiter, B., On weak convergence of the Douglas-Rachford method, *SIAM J. Control Optim.*, **49**(1), 280–287 (2011).

Tibshirani, R., Regression shrinkage and selection via the lasso, *J. R. Stat. Soc., Series B*, **58**(1), 267–288 (1996).

Tseng, P., Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.,* **29,** 119–138 (1991).

Tseng, P., A modified forward-backwatd splitting method for maximal monotone mappings, *SIAM J. Control Optim.,* **38,** 431–456 (2000).

Vapnik, V., The Nature of Statistical Learning Theory, *Springer-Verlang, New York.,* (1995).

Xu, M., Proximal alternating directions method for structured variational inequalities, *J. Optim. Theory Appl.,* 107–117 (2007).

Yuan, X., M. Li., An LQP-based decomposition method for solving a class of variational inequalities, *SIAM J. Optim.,* **21,** 1309–1318 (2011).

Zou, H., M. Yuan, The F_∞-norm support vector machine, *Stat. Sinica,* **18,** 379–398 (2008).