An inexact proximal generalized alternating direction method of multipliers

V.A. Adona * M.L.N. Gonçalves * J.G. Melo *

September 18, 2019

Abstract

This paper proposes and analyzes an inexact variant of the proximal generalized alternating direction method of multipliers (ADMM) for solving separable linearly constrained convex optimization problems. In this variant, the first subproblem is approximately solved using a relative error condition whereas the second one is assumed to be easy to solve. It is important to mention that in many ADMM applications one of the subproblems has a closed-form solution; for instance, ℓ_1 regularized convex composite optimization problems. The proposed method possesses iteration-complexity bounds similar to its exact version. More specifically, it is shown that, for a given tolerance $\rho > 0$, an approximate solution of the Lagrangian system associated to the problem under consideration is obtained in at most $\mathcal{O}(1/\rho^2)$ (resp. $\mathcal{O}(1/\rho)$ in the ergodic case) iterations. Numerical experiments are presented to illustrate the performance of the proposed scheme.

2000 Mathematics Subject Classification: 47H05, 49M27, 90C25, 90C60, 65K10.

Key words: generalized alternating direction method of multipliers, convex program, relative error criterion, pointwise iteration-complexity, ergodic iteration-complexity.

1 Introduction

Recently, there has been a growing interest in the study of the alternating direction method of multipliers (ADMM) and its variants for solving the separable linearly constrained optimization problem

$$\min\{f(x) + g(y) : Ax + By = b\}, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^p \to \mathbb{R}$ are convex functions, $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times p}$, and $b \in \mathbb{R}^m$. The ADMM is an augmented Lagrangian type method that explores the separable structure of problem (1) in such a way that the augmented Lagrangian subproblem is solved alternately. The first ones to consider this scheme were Glowinski and Marroco in [17] and Gabay and Mercier in [16]. An important class of problems that can be fit into the above setting is the following composite convex optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) + g(Qx), \tag{2}$$

*IME, Universidade Federal de Goias, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (E-mails: vandoadona@gmail.com, maxlng@ufg.br and jefferson@ufg.br). The work of these authors was supported in part by CAPES, FAPEG/GO, and CNPq Grants 406975/2016-7 and 302666/2017-6.

where $Q \in \mathbb{R}^{n\times n}$. Indeed, this can be done by considering an artificial variable $y = Qx$ and setting $A = -Q$, $B = I$, and $b = 0$. A special instance of (2) consists of $Q = I$ and g as the l_1 regularization $g(\cdot) = \mu\|\cdot\|_1$, where μ is a regularization parameter. The latter instance with $f(x) = \|Dx - d\|^2$, where $D \in \mathbb{R}^{m\times n}$, corresponds to the popular LASSO problem. Problems such as (2) and, more generally, (1) have found many applications in different areas; see, for example, [7] and references therein for a throughout discussion on these problems as well as the use of the ADMM and some variants to solve them.

In [11], Eckstein and Bertsekas proposed the following generalized ADMM (for short $\mathcal{G}$-ADMM) for solving (2): let $(y_0, \gamma_0) \in \mathbb{R}^p \times \mathbb{R}^m$, $\beta > 0$ and $\alpha \in (0, 2)$ be given and consider two summable sequences $\{\mu_k\} \subset \mathbb{R}_+$ and $\{\nu_k\} \subset \mathbb{R}_+$; for $k = 1, 2, \ldots$ do

$$x_k \approx \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) - \langle \gamma_{k-1}, y_{k-1} - Qx\rangle + \frac{\beta}{2}\|y_{k-1} - Qx\|^2 \right\}, \tag{3}$$

$$y_k \approx \arg\min_{y\in\mathbb{R}^p} \left\{ g(y) - \langle \gamma_{k-1}, y - Qx_k\rangle + \frac{\beta}{2}\|\alpha(y - Qx_k) + (1-\alpha)(y - y_{k-1})\|^2 \right\}, \tag{4}$$

$$\gamma_k = \gamma_{k-1} - \beta\left[\alpha(y_k - Qx_k) + (1-\alpha)(y_k - y_{k-1})\right], \tag{5}$$

where the approximate solutions x_k and y_k are such that $\|x_k - x_k^e\| \le \mu_k$ and $\|y_k - y_k^e\| \le \nu_k$, with x_k^e and y_k^e being the exact solutions of (3) and (4), respectively. More recently (see [1, 8, 15]), several authors have studied the following proximal version of the exact $\mathcal{G}$-ADMM ($\mu_k = \nu_k = 0$) to solve (1):

$$x_k \in \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) - \langle \gamma_{k-1}, Ax\rangle + \frac{\beta}{2}\|Ax + By_{k-1} - b\|^2 + \frac{1}{2}\|x - x_{k-1}\|_G^2 \right\}, \tag{6}$$

$$y_k \in \arg\min_{y\in\mathbb{R}^p} \left\{ g(y) - \langle \gamma_{k-1}, By\rangle + \frac{\beta}{2}\|\alpha(Ax_k + By - b) + (1-\alpha)B(y - y_{k-1})\|^2 + \frac{1}{2}\|y - y_{k-1}\|_H^2 \right\}, \tag{7}$$

$$\gamma_k = \gamma_{k-1} - \beta\left[\alpha(Ax_k + By_k - b) + (1-\alpha)B(y_k - y_{k-1})\right], \tag{8}$$

where G and H are symmetric positive semidefinite matrices, and $\|\cdot\|_G^2 := \langle G\cdot, \cdot\rangle$, etc. In particular, iteration-complexity bounds have been established in [1, 15] under different assumptions. Note that the standard ADMM corresponds to the above method with $(G, H) = (0, 0)$ and $\alpha = 1$. As has been observed by many authors (see, e.g., [1, 6, 15, 23]), the use of the relaxation parameter $\alpha > 1$ in (7)–(8) may considerably improve the numerical performance of the method.

This paper proposes and analyzes an inexact variant of the proximal $\mathcal{G}$-ADMM (6)–(8) for solving (1). The method is interesting in applications in which subproblem (7) is easy to solve whereas (6) is not, being necessary therefore to use iterative methods to approximately solve it. The proposed scheme allows inexact solutions of the following inclusion (derived from the first-order optimality condition for (6) with $G = I$)

$$0 \in \partial f^1(x) - A^*(\gamma_{k-1} - \beta(Ax + By_{k-1} - b)) + (x - x_{k-1})/\beta, \tag{9}$$

such that a relative error condition is satisfied. The error condition used here is similar to the one studied in [29, 30] in the context of a hybrid proximal extragradient method. It is shown that the new

[1] The ε-subdifferential of a convex function $h : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\partial_\varepsilon h(x) := \{u \in \mathbb{R}^n : h(\tilde{x}) \ge h(x) + \langle u, \tilde{x} - x\rangle - \varepsilon, \ \forall\, \tilde{x} \in \mathbb{R}^n\} \qquad \forall\, x \in \mathbb{R}^n.$$

When $\varepsilon = 0$, then $\partial_0 h(x)$ is denoted by $\partial h(x)$ and is called the subdifferential of f at x.

inexact method possesses iteration-complexity bounds similar to its exact version. More specifically, consider the Lagrangian system associated to (1)

$$0 \in \partial f(x) - A^*\gamma, \qquad 0 \in \partial g(y) - B^*\gamma, \qquad 0 = Ax + By - b. \tag{10}$$

We show that, for a given tolerance $\rho > 0$, an approximate solution $(\bar{x}, \bar{y}, \bar{\gamma})$ of (10) with residue $(v_{\bar{x}}, v_{\bar{y}}) \in \mathbb{R}^m \times \mathbb{R}^p$ satisfying

$$v_{\bar{x}} \in \partial f(\bar{x}) - A^*\bar{\gamma}, \quad v_{\bar{y}} \in \partial g(\bar{y}) - B^*\bar{\gamma}, \quad \max\{\|A\bar{x} + B\bar{y} - b\|, \|v_{\bar{x}}\|, \|v_{\bar{y}}\|\} \leq \rho$$

is obtained in at most $\mathcal{O}(1/\rho^2)$ iterations. We also show that, in the ergodic case, an approximate solution $(\tilde{x}, \tilde{y}, \tilde{\gamma})$ of (10) with residues $(v_{\tilde{x}}, v_{\tilde{y}}) \in \mathbb{R}^m \times \mathbb{R}^p$ and $(\varepsilon_{\tilde{x}}, \varepsilon_{\tilde{y}}) \in \mathbb{R}_+ \times \mathbb{R}_+$ satisfying

$$v_{\tilde{x}} \in \partial_{\varepsilon_{\tilde{x}}} f(\tilde{x}) - A^*\tilde{\gamma}, \quad v_{\tilde{y}} \in \partial_{\varepsilon_{\tilde{y}}} g(\tilde{y}) - B^*\tilde{\gamma}, \quad \max\{\|A\tilde{x} + B\tilde{y} - b\|, \|v_{\tilde{x}}\|, \|v_{\tilde{y}}\|, \varepsilon_{\tilde{x}}, \varepsilon_{\tilde{y}}\} \leq \rho$$

can be obtained in at most $\mathcal{O}(1/\rho)$ iterations. Some numerical experiments are presented in order to illustrate the performance of the new method. In particular, it is verified that the use of the relaxation parameter $\alpha > 1$, specially $\alpha \approx 1.9$, improves considerably its numerical behavior.

Previous related works. Inexact versions of the ADMM and its variants considering absolute and/or relative error conditions have been proposed in the literature; see, for instance, [2, 12, 13, 14, 33]. In [12], the authors proposed and analyzed an augmented Lagrangian method whose subproblem is solved using a relative error condition similar to that proposed in [29] for a family of proximal point type methods. The previous study was further developed in [13] to the ADMM setting in order to solve (2). The latter reference also analyzed an ADMM whose subproblems are solved using absolute error conditions. An inexact ADMM with relative error conditions similar to the one analyzed in [13] was also studied in [33] for solving (1). Paper [14] proposed a relaxed Douglas–Rachford splitting method for solving (2) and derived, as a consequence, a variant of the ADMM which uses, in a special way, a relative error condition. Finally, [2] studied a partially inexact ADMM whose first subproblem is approximately solved using a relative error condition similar to the one considered here. The main difference between the study in [2] and the one presented in this paper is that the ADMM variants analyzed are different. Specifically, [2] considered an ADMM which contains the usually called Glowinski's stepsize parameter in the update rule of the Lagrangian multiplier whereas here we are considering the ADMM variant proposed by Eckstein and Bertsekas in [11]. Finally, in the aforementioned papers only [2] presented an iteration-complexity analysis of the method. Their iteration-complexity bounds are similar to the ones derived in this paper.

Organization of the paper. Section 2 introduces and analyzes the inexact proximal $\mathcal{G}$-ADMM. Section 3 is devoted to the numerical study of the proposed method. This section is divided into two subsections. The first one illustrates the performance of the method for solving the LASSO problem whereas the second subsection is devoted to the l_1-regularized logistic regression problem. A conclusion is presented in Section 4. The appendix contains the proof of an essential result related to the proposed method.

2 Inexact proximal $\mathcal{G}$-ADMM

In this section, we formally state the inexact proximal $\mathcal{G}$-ADMM for computing approximate solutions of (1) and present some properties as well as its pointwise and ergodic iteration-complexity bounds.

We start by describing the method.

Algorithm 1 (Inexact proximal $\mathcal{G}$-ADMM)

Input: $(x_0, y_0, \gamma_0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, $\beta > 0$, $\tau_1, \tau_2 \in [0, 1)$, $\alpha \in (0, 2 - \tau_1)$, and a symmetric positive semidefinite matrix $H \in \mathbb{R}^{p \times p}$;

for $k = 1, 2, \ldots$ **do**

1: compute $(\tilde{x}_k, v_k) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$v_k \in \partial f(\tilde{x}_k) - A^* \tilde{\gamma}_k, \quad \|\tilde{x}_k - x_{k-1} + \beta v_k\|^2 \leq \tau_1 \|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \tau_2 \|\tilde{x}_k - x_{k-1}\|^2, \tag{11}$$

where

$$\tilde{\gamma}_k = \gamma_{k-1} - \beta(A\tilde{x}_k + By_{k-1} - b); \tag{12}$$

2: compute an optimal solution $y_k \in \mathbb{R}^p$ of the subproblem

$$\min_{y \in \mathbb{R}^p} \left\{ g(y) - \langle \gamma_{k-1}, By \rangle + \frac{\beta}{2} \|\alpha(A\tilde{x}_k + By - b) + (1 - \alpha)B(y - y_{k-1})\|^2 + \frac{1}{2}\|y - y_{k-1}\|_H^2 \right\}; \tag{13}$$

3: set

$$x_k = x_{k-1} - \beta v_k, \qquad \gamma_k = \gamma_{k-1} - \beta \left[\alpha(A\tilde{x}_k + By_k - b) + (1 - \alpha)B(y_k - y_{k-1})\right]. \tag{14}$$

Some comments about Algorithm 1 are in order. First, if $\tau_1 = \tau_2 = 0$, then the inequality in (11), combined with the first relation in (14), implies that $\tilde{x}_k = x_k$ and $v_k = (x_{k-1} - x_k)/\beta$. Hence, in view of the definition of $\tilde{\gamma}_k$ in (12) and the inclusion in (11), we conclude that x_k is a solution of (9). Therefore, Algorithm 1 can be seen as a variant of the proximal $\mathcal{G}$-ADMM (6)–(8) in which its first subproblem is approximately solved using a relative error condition. Now, if x_k is a solution of the inclusion in (9), then the pair $(\tilde{x}_k, v_k) := (x_k, (x_{k-1} - x_k)/\beta)$ trivially satisfies (11). Second, it is assumed that (13) can be easily solved. On the one hand, if the matrix B in (1) is not the identity, then subproblem (13) with the usual choice $H := \xi I - \beta B^* B$ with $\xi \geq \beta \|B^* B\|$ (the symbol * stands for the transpose of a matrix) becomes a prox-subproblem

$$y_k = \arg\min_{y \in \mathbb{R}^p} \left\{ g(y) + \frac{\xi}{2}\|y - \hat{y}\|^2 \right\} \tag{15}$$

for some $\hat{y} \in \mathbb{R}^p$. In many ADMM applications, g is well-structured (e.g., the l_1-norm) and hence the latter problem is easy to solve or even has a closed-form solution. On the other hand, if $B = I$ in (1), then $H = 0$ seems to be a natural choice.

In order to establish iteration-complexity bounds for Algorithm 1, we consider the following basic assumption:

Assumption 1. There exists a solution $(x^*, y^*, \gamma^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ of the Lagrangian system associated (10).

The set of points satisfying (10) is denoted by Ω^*. It is well-known that $(x^*, y^*, \gamma^*) \in \Omega^*$ if and only if (x^*, y^*) is a solution to problem (1) and γ^* is an associated Lagrange multiplier. Let us also consider a matrix M and an operator T defined as follows

$$M = \begin{bmatrix} \frac{1}{\beta} I_n & 0 & 0 \\ 0 & (H + \frac{\beta}{\alpha} B^* B) & \frac{1-\alpha}{\alpha} B^* \\ 0 & \frac{1-\alpha}{\alpha} B & \frac{1}{\alpha\beta} I_m \end{bmatrix}, \quad T(x, y, \gamma) = \begin{bmatrix} \partial f(x) - A^* \gamma \\ \partial g(y) - B^* \gamma \\ Ax + By - b \end{bmatrix}. \tag{16}$$

Remark 2.1. *i) It can be easily verified that, for every $\beta > 0$ and $\alpha \in (0, 2)$, M is symmetric positive semidefinite. ii) Since f and g are convex functions, the operators ∂f and ∂g are maximal monotone (see [26]) which, in turn, implies that the operator T is also maximal monotone. iii) Note that a triple (x^*, y^*, γ^*) is a solution of the Lagrangian system* (10) *if and only if $0 \in T(x^*, y^*, \gamma^*)$.*

The following notion of approximate solutions of (10) will be considered:

Definition 2.2. *Given a tolerance $\rho > 0$, a triple $(x, y, \gamma) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is said to be a $\rho-$approximate solution of* (10) *with residue r if*

$$r \in T(x, y, \gamma), \quad \|r\| \leq \rho. \tag{17}$$

In view of Remark 2.1(iii), for all $\rho > 0$, any element in Ω^* is a ρ-approximate solution with residue 0. For convenience, consider the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$ defined by

$$z_{k-1} = (x_{k-1}, y_{k-1}, \gamma_{k-1}), \quad \tilde{z}_k = (\tilde{x}_k, y_k, \tilde{\gamma}_k), \qquad \forall k \geq 1. \tag{18}$$

It will be shown that, for any given $\rho > 0$, there exists an index k such that $\tilde{z}_k$ is a ρ-approximate solution of (10) with residue $r_k := M(z_{k-1} - z_k)$. To this end, let us now introduce the following quantities

$$d_0 := \inf \left\{ \|(x - x_0, y - y_0, \gamma - \gamma_0)\|_M : (x, y, \gamma) \in \Omega^* \right\}, \tag{19}$$

$$\sigma := \max \left\{ \frac{1 + \alpha\tau_1}{1 + \alpha(2 - \alpha)}, \tau_2 \right\} \quad \text{and} \quad \xi := \frac{1}{\alpha^3}[\sigma(1 + \alpha - \alpha^2) + (1 - \tau_1)\alpha - 1]. \tag{20}$$

Note that, if M is positive definite, then the quantity d_0 measures the distance in the norm $\|\cdot\|_M$ of the initial point (x_0, y_0, γ_0) to the solution set Ω^*. Furthermore, in view of the assumptions on α, τ_1 and τ_2 in Algorithm 1, we trivially have $\sigma \in (0, 1)$ and $\xi > 0$.

The next result, whose proof is presented in Appendix A, shows some important relations satisfied by the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$.

Proposition 2.3. *Let $\{z_k\}$ and $\{\tilde{z}_k\}$ be as in* (18)*. Consider $\{\eta_k\}_{k \geq 0}$ defined by*

$$\eta_0 = 4\xi d_0^2, \quad \eta_k = \xi \|y_k - y_{k-1}\|_H^2, \qquad \forall k \geq 1, \tag{21}$$

where d_0 and ξ are as in (19) *and* (20)*, respectively. Let also M, T and σ be as in* (16) *and* (20)*. Then, the following hold:*

a) for every $k \geq 1$, we have

$$0 \in T(\tilde{z}_k) + M(z_k - z_{k-1}), \quad \|z_k - \tilde{z}_k\|_M^2 + \eta_k \leq \sigma \|z_{k-1} - \tilde{z}_k\|_M^2 + \eta_{k-1}; \tag{22}$$

b) for every $k \geq 1$ and $z^* := (x^*, y^*, \gamma^*) \in \Omega^*$, we have

$$\|z_k - z^*\|_M^2 \leq \|z_{k-1} - z^*\|_M^2 + \eta_{k-1} - \eta_k - (1 - \sigma)\|z_{k-1} - \tilde{z}_k\|_M^2.$$

Remark 2.4. *i) Note that the inclusion in* (22) *can be interpreted as a generalized proximal inclusion where the pair $(z_k, \tilde{z}_k)$ is controlled according to the relative error condition in* (22)*. Indeed, the relations in* (22) *imply that the sequence $\{(z_k, \tilde{z}_k, \eta_k)\}_{k \geq 1}$ is an implementation of the HPE framework studied in [20]. As a consequence of the latter conclusion, the pointwise iteration-complexity bound* *in Theorem 2.5 can also be derived from [20, Theorem 3.3]. However, since its proof follows easily from Proposition 2.3, we present it here for completeness and convenience of the reader. On the other hand, although the ergodic iteration-complexity bound (see Theorem 2.6) is related to [20, Theorem 3.4], its proof does not follow immediately from the latter theorem. ii) The inequality in Proposition 2.3(b) is closely related to the well-known quasi-Fejér inequality which can be used to show that $\{z_k\}$ converges to a point in Ω^* when M is positive definite.*

We next present the main results of this paper. The first one contains a pointwise iteration-complexity bound for Algorithm 1 to obtain an approximate solution in the sense of Definition 2.2. The second one derives an iteration-complexity bound to obtain a relaxed approximate solution of (10).

Theorem 2.5. *For a given tolerance $\rho > 0$, Algorithm 1 generates a $\rho-$approximate solution $(\tilde{x}_i, y_i, \tilde{\gamma}_i)$ of* (10) *with an associated residue $r_i = M(z_{i-1} - z_i)$ in at most $\mathcal{O}(d_0^2/\rho^2)$ iterations, where $\{z_i\}$ and d_0 are as in* (18) *and* (19), *respectively.*

Proof. First note that, in view of the inclusion in (22), we have $r_k := M(z_{k-1} - z_k)$ is a residue to the inclusion in (17) associated to $\tilde{z}_k$, for every $k \geq 1$. Let λ_M be the largest eigenvalue of the matrix M in (16). Hence, combining the definition of r_k, the inequality in (22) and simple algebra, we obtain

$$\begin{aligned} \|r_k\|^2 &\leq \lambda_M \|z_{k-1} - z_k\|_M^2 \leq 2\lambda_M \left[\|z_{k-1} - \tilde{z}_k\|_M^2 + \|\tilde{z}_k - z_k\|_M^2\right] \\ &\leq 2\lambda_M \left[(\sigma + 1)\|z_{k-1} - \tilde{z}_k\|_M^2 + \eta_{k-1} - \eta_k\right], \end{aligned} \tag{23}$$

for every $k \geq 1$. It follows from Proposition 2.3(b) and (23) that, for every $z^* := (x^*, y^*, \gamma^*) \in \Omega^*$,

$$\begin{aligned} \sum_{k=1}^{i} \|r_k\|^2 &\leq \frac{2\lambda_M}{1-\sigma} \sum_{k=1}^{i} \left[(\sigma + 1)\left(\|z_{k-1} - z^*\|_M^2 - \|z_k - z^*\|_M^2\right) + 2(\eta_{k-1} - \eta_k)\right] \\ &\leq \frac{2\lambda_M}{1-\sigma} \left((\sigma + 1)\|z_0 - z^*\|_M^2 + 2\eta_0\right), \end{aligned}$$

which in turn, in view of the definitions of d_0 and η_0 given in (19) and (21), implies that there exists a scalar $c > 0$ such that

$$\sum_{k=1}^{i} \|r_k\|^2 \leq c d_0^2. \tag{24}$$

In particular, the latter inequality implies that $\{r_k\}$ converges to zero. Hence, let i be the first index in which $\|r_i\| \leq \rho$ (which is equivalent to say that $\tilde{z}_i$ is a ρ-approximate solution with residue r_i). Note that if $i = 1$, then the statement of the theorem trivially follows. Now assume that $i > 1$. It follows from (24) that

$$(i-1)\rho^2 < \sum_{k=1}^{i-1} \|r_k\|^2 \leq c d_0^2$$

and hence $i = \mathcal{O}(d_0^2/\rho^2)$, concluding the proof of the theorem. □

Theorem 2.6. *Let $\{(x_k, y_k, \gamma_k, \tilde{x}_k, \tilde{\gamma}_k)\}$ be generated by Algorithm 1 and consider the sequences $\{(x_k^a, y_k^a, \gamma_k^a, \tilde{x}_k^a, \tilde{\gamma}_k^a)\}$ and $\{q_k^a\}$ defined by*

$$(x_k^a, y_k^a, \gamma_k^a, \tilde{x}_k^a, \tilde{\gamma}_k^a) = \frac{1}{k} \sum_{i=1}^{k} (x_i, y_i, \gamma_i, \tilde{x}_i, \tilde{\gamma}_i), \qquad q_k^a = \frac{1}{k}\left[(x_0, y_0, \gamma_0) - (x_k, y_k, \gamma_k)\right]. \tag{25}$$

Then, for every $k \geq 1$, there exist $\varepsilon^a_{k,x}, \varepsilon^a_{k,y} \geq 0$ such that the following relations hold

$$r^a_k := Mq^a_k \in \Big(\partial_{\varepsilon^a_{k,x}} f(\tilde{x}^a_k) - A^*\tilde{\gamma}^a_k,\ \partial_{\varepsilon^a_{k,y}} g(y^a_k) - B^*\tilde{\gamma}^a_k,\ A\tilde{x}^a_k + By^a_k - b\Big) \tag{26}$$

$$\|r^a_k\| \leq \frac{\theta d_0}{k}, \qquad \max\{\varepsilon^a_{k,x}, \varepsilon^a_{k,y}\} \leq \frac{\theta d_0^2}{k}, \tag{27}$$

where M and d_0 are as in (16) *and* (19)*, respectively, and θ is a positive scalar depending on (α, τ_1, τ_2) and the largest eigenvalue of M.*

Proof. First of all, define $(v_i, u_i, w_i) = M(z_{i-1} - z_i)$ for every $i \geq 1$. Hence, it follows from Proposition 2.3(a) and (16) that

$$v_i + A^*\tilde{\gamma}_i \in \partial f(\tilde{x}_i), \qquad u_i + B^*\tilde{\gamma}_i \in \partial g(y_i), \qquad w_i = A\tilde{x}_i + By_i - b. \tag{28}$$

On the one hand, from the above equality and (25), we have

$$w^a_k := \frac{1}{k}\sum_{i=1}^{k} w_i = A\tilde{x}^a_k + By^a_k - b. \tag{29}$$

Now, in view of the inclusions in (28), it follows from (25) and [19, Theorem 2.1] that the sequences $\{\varepsilon^a_{k,x}\}$ and $\{\varepsilon^a_{k,y}\}$ defined by

$$\varepsilon^a_{k,x} := \frac{1}{k}\sum_{i=1}^{k} \langle v_i + A^*\tilde{\gamma}_i, \tilde{x}_i - \tilde{x}^a_k\rangle, \qquad \varepsilon^a_{k,y} := \frac{1}{k}\sum_{i=1}^{k} \langle u_i + B^*\tilde{\gamma}_i, y_i - y^a_k\rangle, \tag{30}$$

are nonnegative and

$$\frac{1}{k}\sum_{i=1}^{k} v_i \in \partial_{\varepsilon^a_{k,x}} f(\tilde{x}^a_k) - A^*\tilde{\gamma}^a_k, \quad \frac{1}{k}\sum_{i=1}^{k} u_i \in \partial_{\varepsilon^a_{k,y}} g(y^a_k) - B^*\tilde{\gamma}^a_k. \tag{31}$$

The inclusion in (26) follows from (29) and (31) and the fact that $\sum_{i=1}^{k}(v_i, u_i, w_i) = M(z_0 - z_k)$. Therefore, the proof of the existence of the elements $\varepsilon^a_{k,x}, \varepsilon^a_{k,y} \geq 0$ such that (26) holds is completed.

Let us now prove that (27) holds for r^a_k, $\varepsilon^a_{k,x}$ and $\varepsilon^a_{k,y}$ as defined above. Using (18) and the definition of q^a_k in (25), we have

$$kq^a_k = z_0 - z_k = (z_0 - z^*) + (z^* - z_k),$$

where $z^* = (x^*, y^*, \gamma^*) \in \Omega^*$. Thus, from Proposition 2.3(b), we obtain

$$k^2\|q^a_k\|^2_M \leq 2(\|z^* - z_0\|^2_M + \|z^* - z_k\|^2_M) \leq 4(\|z^* - z_0\|^2_M + \eta_0). \tag{32}$$

Since $r^a_k = Mq^a_k$, we obtain $\|r^a_k\|^2 \leq \lambda_M \|q^a_k\|^2_M$, where λ_M is the largest eigenvalue of M. Hence, using (32) and the definitions of d_0 and η_0 in (19) and (21), respectively, we conclude that the bound on $\|r^a_k\|$ in (27) holds with $\theta = \theta_1 := 2\sqrt{\lambda_M(1 + 4\xi)}$.

Now, from (30), we have

$$\varepsilon^a_{k,x} + \varepsilon^a_{k,y} = \frac{1}{k}\sum_{i=1}^{k} \Big(\langle v_i, \tilde{x}_i - \tilde{x}^a_k\rangle + \langle u_i, y_i - y^a_k\rangle + \langle \tilde{\gamma}_i, A\tilde{x}_i - A\tilde{x}^a_k + By_i - By^a_k\rangle\Big)$$
$$= \frac{1}{k}\sum_{i=1}^{k} \Big(\langle v_i, \tilde{x}_i - \tilde{x}^a_k\rangle + \langle u_i, y_i - y^a_k\rangle + \langle \tilde{\gamma}_i, w_i - w^a_k\rangle\Big),$$

where the last equality is due to the definitions of w_i and w^a_k in (28) and (29), respectively. Additionally, the definitions of w_i, w^a_k and $\tilde{\gamma}^a_k$ imply that

$$\frac{1}{k}\sum_{i=1}^{k}\langle \tilde{\gamma}_i, w_i - w^a_k\rangle = \frac{1}{k}\sum_{i=1}^{k}\langle \tilde{\gamma}_i - \tilde{\gamma}^a_k, w_i - w^a_k\rangle = \frac{1}{k}\sum_{i=1}^{k}\langle w_i, \tilde{\gamma}_i - \tilde{\gamma}^a_k\rangle.$$

Therefore, since $z_i = (x_i, y_i, \gamma_i)$ and $M(z_{i-1} - z_i) = (v_i, u_i, w_i)$, we obtain

$$\varepsilon^a_{k,x} + \varepsilon^a_{k,y} = \frac{1}{k}\sum_{i=1}^{k}\langle M(z_{i-1} - z_i), \tilde{z}_i - \tilde{z}^a_k\rangle, \tag{33}$$

where $\tilde{z}^a_k := (\tilde{x}^a_k, y^a_k, \tilde{\gamma}^a_k)$. On the other hand, observe that for every $z \in \mathbb{R}^{n+m+p}$, we have

$$\|z - z_i\|^2_M - \|z - z_{i-1}\|^2_M = \|\tilde{z}_i - z_i\|^2_M - \|\tilde{z}_i - z_{i-1}\|^2_M + 2\langle M(z_{i-1} - z_i), z - \tilde{z}_i\rangle$$
$$\leq (\sigma - 1)\|\tilde{z}_i - z_{i-1}\|^2_M + \eta_{i-1} - \eta_i + 2\langle M(z_{i-1} - z_i), z - \tilde{z}_i\rangle$$

where the last inequality is due to (22). Thus, since $\sigma < 1$ (see (20)), we find

$$\sum_{i=1}^{k} 2\langle M(z_{i-1} - z_i), \tilde{z}_i - z\rangle \leq \|z - z_0\|^2_M - \|z - z_k\|^2_M + \eta_0 - \eta_k \leq \|z - z_0\|^2_M + \eta_0.$$

Letting $z = \tilde{z}^a_k$ in the above inequality and using (33), we obtain

$$2k(\varepsilon^a_{k,x} + \varepsilon^a_{k,y}) \leq \|\tilde{z}^a_k - z_0\|^2_M + \eta_0 \leq \frac{1}{k}\sum_{i=1}^{k}\|\tilde{z}_i - z_0\|^2_M + \eta_0, \leq \max_{i=1,\ldots,k}\|\tilde{z}_i - z_0\|^2_M + \eta_0, \tag{34}$$

where the second inequality is due to the definition of $\tilde{z}^a_k$ and the convexity of the function $\|\cdot\|^2_M$. Now, since $\|z + z' + z''\|^2_M \leq 3\left(\|z\|^2_M + \|z'\|^2_M + \|z''\|^2_M\right)$ for all z, z', z'', we have, for every $i \geq 1$ and $z^* \in \Omega^*$,

$$\|\tilde{z}_i - z_0\|^2_M \leq 3\left[\|\tilde{z}_i - z_i\|^2_M + \|z^* - z_i\|^2_M + \|z^* - z_0\|^2_M\right]$$
$$\leq 3\left[\sigma\|\tilde{z}_i - z_{i-1}\|^2_M + \eta_{i-1} + \|z^* - z_i\|^2_M + \|z^* - z_0\|^2_M\right],$$

where the last inequality is due to (22). Hence, using Proposition 2.3(b), we obtain

$$\|\tilde{z}_i - z_0\|^2_M \leq 3\left[\sigma\|\tilde{z}_i - z_{i-1}\|^2_M + \eta_{i-1} + \|z^* - z_{i-1}\|^2_M + \eta_{i-1} + \|z^* - z_0\|^2_M\right]$$
$$\leq 3\left[\sigma\|\tilde{z}_i - z_{i-1}\|^2_M + 2(\|z^* - z_{i-1}\|^2_M + \eta_{i-1}) + \|z^* - z_0\|^2_M\right]$$
$$\leq 3\left[\sigma\|\tilde{z}_i - z_{i-1}\|^2_M + 3\|z^* - z_0\|^2_M + 2\eta_0\right],$$

which, combined with (34), yields

$$2k\left(\varepsilon_{k,x}^{a}+\varepsilon_{k,y}^{a}\right)\leq 3\left[3\left(\|z^{*}-z_{0}\|_{M}^{2}+\eta_{0}\right)+\sigma\max_{i=1,\ldots,k}\|\tilde{z}_{i}-z_{i-1}\|_{M}^{2}\right].$$

From Proposition 2.3(b), we also have

$$(1-\sigma)\|\tilde{z}_{i}-z_{i-1}\|_{M}^{2}\leq\|z^{*}-z_{i-1}\|_{M}^{2}+\eta_{i-1}\leq\|z^{*}-z_{0}\|_{M}^{2}+\eta_{0}.$$

Combining the previous inequalities, we obtain

$$\varepsilon_{k,x}^{a}+\varepsilon_{k,y}^{a}\leq\frac{3(3-2\sigma)}{2(1-\sigma)k}\left(\|z^{*}-z_{0}\|_{M}^{2}+\eta_{0}\right).$$

Hence, using the definitions of d_0 and η_0 in (19) and (21), respectively, we conclude that the second inequality in (27) holds with $\theta=\theta_2:=3(3-2\sigma)(1+4\xi)/2(1-\sigma)$. Therefore, the estimations in (27) trivially follow by defining $\theta=\max\{\theta_1,\theta_2\}$. □

Remark 2.7. *It follows from Theorem 2.6 that, for a given tolerance $\rho>0$, in at most $k=\mathcal{O}(\max\{d_0,d_0^2\}/\rho)$ iterations, the triple $(\tilde{x}_k^a,y_k^a,\tilde{\gamma}_k^a)$, together with r_k^a, satisfies the inclusion in* (26) *with $\varepsilon_{k,x}^a,\varepsilon_{k,y}^a\geq 0$ and $\max\{\|r_k^a\|,\varepsilon_{k,x}^a,\varepsilon_{k,y}^a\}\leq\rho$. Hence, the triple $(\tilde{x}_k^a,y_k^a,\tilde{\gamma}_k^a)$ can be seen as an approximate solution of* (10) *with residue r_k^a in the sense that the inclusions in* (10) *are relaxed by using the ε-subdifferential operator instead of the subdifferential. It should be mentioned that the quantities $\varepsilon_{k,x}^a$ and $\varepsilon_{k,y}^a$ can be explicitly computed* (*see* (30))*. Their expressions are not explicitly stated in order to simplify the statement of the theorem.*

In the next section, we present some numerical experiments to illustrate the performance of Algorithm 1 under different choices of the relaxation parameter α.

3 Numerical experiments

This section reports the numerical performance of Algorithm 1 to solve two classes of problems, namely, LASSO and l_1–regularized logistic regression. Different values of the relaxation parameter α were considered in order to illustrate its effect and show that, similarly to the exact $\mathcal{G}$-ADMM, the performance of the algorithm improves considerably when $\alpha>>1$, specially $\alpha\approx 1.9$. The proposed algorithm was compared with two other schemes: the partially inexact ADMM [13, Algorithm 2] and a generalized version of the "exact" ADMM considered in [11], denoted here as Algorithm 2. The latter algorithm corresponds to the scheme (6)–(8) with $(G,H)=(0,0)$ and x_k being such that there exists a residue v_k satisfying

$$v_k\in\partial f(x_k)-A^{*}\left[\gamma_{k-1}+\beta\left(Ax_k+By_{k-1}-b\right)\right],\qquad\|v_k\|\leq 10^{-8}.$$

Note that the above inclusion with $v_k=0$ is the one derived from the first-order optimality condition for (6) with $G=0$. It should be mentioned that the applications considered here are such that the solution of the second subproblem of the three analyzed algorithms can be explicitly computed.

The algorithms were tested using six non-simulated data sets from the Elvira biomedical repository [9]. Each one of them is represented by a matrix $D\in\mathbb{R}^{m\times n}$ and a vector $d\in\mathbb{R}^{m}$ as follows:

1. Colon tumor gene expression [3] with $m=62$ and $n=2000$;

2. Central nervous system (CNS) data [25] with $m = 60$ and $n = 7129$;

3. Leukemia cancer-ALLMLL data [18] with $m = 38$ and $n = 7129$;

4. Lung cancer-Michigan data [5] with $m = 96$ and $n = 7129$;

5. Lymphoma-Harvard data [27] with $m = 77$ and $n = 7129$;

6. Prostate cancer data [28] with $m = 102$ and $n = 12600$.

In addition, for the second class of problems, we also selected the Madelon data set (see [21]) from the ICU Machine Learning Repository [10], which has dimensions $m = 2000$ and $n = 500$. All experiments were performed on MATLAB R2015a using an Intel(R) Core i7 2.4GHz computer with 8GB of RAM.

For all tests, we set $(x_0, y_0, \gamma_0) = (0, 0, 0)$ and $\beta = 1$, and used the same overall termination condition

$$\|M(z_{k-1} - z_k)\|_\infty \le 10^{-4}, \tag{35}$$

where M and z_k are as in (16) and (18), respectively. In Algorithm 1, the remaining initialization data were $\tau_1 = 0.99 \times (2 - \alpha)$, $\tau_2 = 1 - 10^{-8}$ and $H = 0$, and a hybrid inner stopping criterion was used; specifically, the inner-loop terminates when v_k satisfies either the inequality in (11) or $\|v_k\| \le 10^{-8}$. The latter strategy was also used in [13, 14, 33] and it is motivated by the fact that, close to a solution, the former condition seems to be more restrictive than the latter. We mention that the performance of the partially inexact ADMM [13, Algorithm 2] was basically the same as the one of Algorithm 1 with $\alpha = 1$, and hence only the results of the latter scheme are displayed in the tables.

3.1 LASSO problem

Our first test problem is the LASSO [31, 32]

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Dx - d\|^2 + \mu\|x\|_1,$$

where $D \in \mathbb{R}^{m \times n}$, $d \in \mathbb{R}^m$, and $\mu > 0$ is a regularization parameter. In our experiment, the matrix D and the vector d were set as listed in the beginning of this section. Moreover, we scaled d and the columns of D in order to have unit l_2-norm, and set $\mu = 0.1\|D^*d\|_\infty$. It is easy to see that the above problem can be rewritten as an instance of (1) in which $f(x) = \frac{1}{2}\|Dx - d\|^2$, $g(y) = \mu\|y\|_1$, $A = -I$, $B = I$ and $b = 0$. In this case, the pair $(\tilde{x}_k, v_k)$ in (11) can be obtained by computing an approximate solution $\tilde{x}_k$ with a residual v_k of the following linear system

$$(D^*D + \beta I)x = (D^*d + \beta y_{k-1} - \gamma_{k-1}).$$

For approximately solving the above linear system, we used the conjugate gradient method [24] with starting point $D^*d + \beta y_{k-1} - \gamma_{k-1}$. Note also that subproblem (13) has a closed-form solution

$$y_k = \mathcal{S}_{\frac{\mu}{\beta}}\left(\alpha\tilde{x}_k + (1-\alpha)y_{k-1} + \frac{1}{\beta}\gamma_{k-1}\right),$$

where $\mathcal{S}_\kappa : \mathbb{R}^n \to \mathbb{R}^n$ is the shrinkage operator [4] defined as

$$\mathcal{S}_\kappa^i(w) = \operatorname{sign}(w^i)\max(0, |w^i| - \kappa) \quad i = 1, 2, \ldots, n, \tag{36}$$

Table 1: **LASSO problem**

Data set	$\alpha = 1.0$ Alg. 1	$\alpha = 1.0$ Alg. 2	$\alpha = 1.3$ Alg. 1	$\alpha = 1.3$ Alg. 2	$\alpha = 1.5$ Alg. 1	$\alpha = 1.5$ Alg. 2	$\alpha = 1.7$ Alg. 1	$\alpha = 1.7$ Alg. 2	$\alpha = 1.9$ Alg. 1	$\alpha = 1.9$ Alg. 2
	Number of outer iterations									
Colon	116	114	88	89	78	77	69	69	63	63
CNS	319	321	248	249	217	217	194	194	182	182
Leukemia	600	600	431	431	370	370	329	330	320	320
Lung	535	535	412	412	357	357	315	315	282	282
Lymphoma	331	331	255	255	222	222	196	196	176	176
Prostate	431	430	331	331	287	287	254	254	227	227
	Total number of inner iterations									
Colon	2136	4656	1607	3639	1450	3149	1308	2822	1216	2576
CNS	10060	16064	7818	12466	6871	10862	6203	9712	6024	9108
Leukemia	11351	17365	8033	12478	6909	10715	6196	9556	6195	9263
Lung	12516	22836	9622	17588	8373	15240	7475	13451	6881	12048
Lymphoma	8619	15182	6522	11703	5850	10180	5208	8998	4796	8072
Prostate	19562	35002	15083	26944	13088	23374	11906	20700	11003	18478
	CPU time in seconds									
Colon	16.4	23.3	12.3	18.2	10.9	17.0	9.7	14.4	9.2	13.1
CNS	754.4	944.4	584.6	743.4	515.6	643.1	472.9	576.7	449.0	538.7
Leukemia	1119.2	1290.4	789.0	927.8	679.4	797.0	606.1	710.5	600.4	689.4
Lung	1114.7	1470.9	872.3	1159.5	762.5	998.5	670.9	880.5	607.6	788.8
Lymphoma	769.7	931.0	601.8	728.1	489.0	634.6	433.3	564.1	393.1	504.1
Prostate	4325.1	5926.5	3509.2	4494.2	3083.7	3900.2	2664.4	3438.1	2343.7	3103.0

and $\mathrm{sign}(\cdot)$ denotes the sign function.

Table 1 displays the numerical results obtained. In order to compare the algorithms, we consider the number of outer iterations, the total number of accumulated inner iterations and the CPU time in seconds. In Figure 1, we plot the arithmetic mean of the latter three comparisons criteria for each algorithm for solving the six LASSO problem instances. From these results, one can see that the number of outer iterations of Algorithm 1 and Algorithm 2 are basically the same for every considered relaxation parameter α. In particular, the numerical advantage of using $\alpha > 1$, specially $\alpha \approx 1.9$, is also verified for Algorithm 1. Algorithm 1 performed at least 33% less inner iterations than Algorithm 2, reaching, in some instances, 50% less inner iterations. Note that this performance improvement also reflected favorably in terms of CPU time.

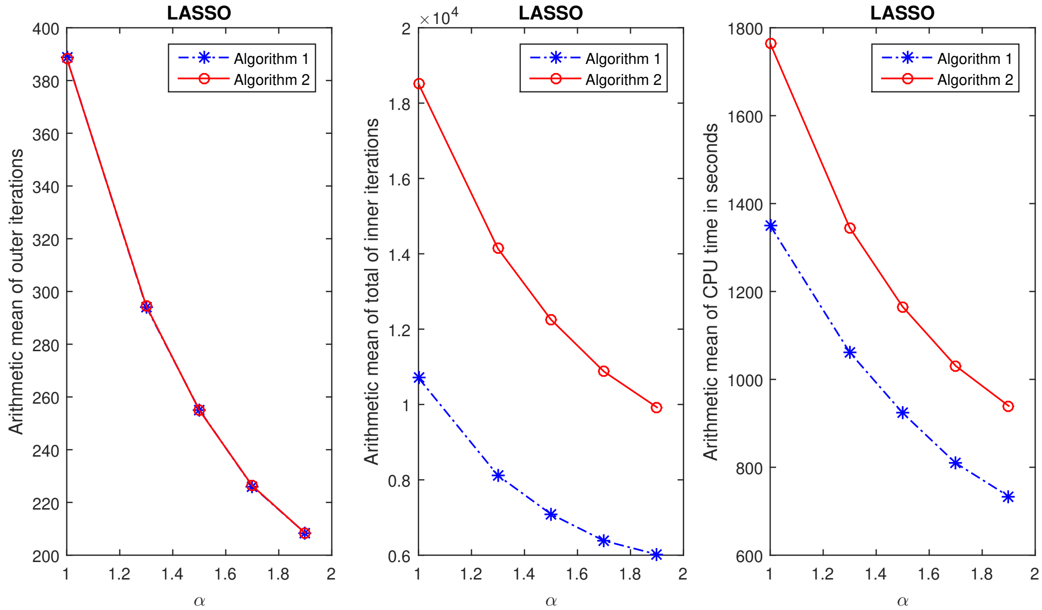

Figure 1: Arithmetic mean of the LASSO problem results

3.2 l_1-Regularized logistic regression problem

In this subsection, we consider the l_1-regularized logistic regression problem [22]

$$\min_{t\in\mathbb{R},u\in\mathbb{R}^n} \frac{1}{m}\sum_{i=1}^{m} \log\Big(1+\exp\big(-d^i(\langle D_i,u\rangle+t)\big)\Big)+\mu\,\|u\|_1\,, \tag{37}$$

where $D_i \in \mathbb{R}^n$ are the rows of a matrix $D \in \mathbb{R}^{m\times n}$, $d^i \in \{-1,+1\}$ are the coordinates of a vector $d \in \mathbb{R}^m$ and μ is a regularization parameter. In our experiment, the matrix D and the vector d were chosen as described in the beginning of this section. We scaled the columns of D in order to have unit l_2-norm and set $\mu = 0.5\lambda_{\max}$, where $\lambda_{\max}$ is as defined in [22, Subsection 2.1]. By defining $z^{i:j} := (z^i,\ldots,z^j) \in \mathbb{R}^{j-i+1}$, problem (37) can be rewritten as an instance of (1) in which

$$f(x)=\frac{1}{m}\sum_{i=1}^{m}\log\Big(1+\exp\big(-d^i(\langle D_i,x^{2:n+1}\rangle+x^1)\big)\Big),\qquad g(y)=\mu\|y^{2:n+1}\|_1,$$

$$A=-I,\quad B=I,\quad \text{and}\quad b=0.$$

In order to compute a pair $(\tilde{x}_k, v_k)$ as in (11), we implemented the limited-memory BFGS method [24, Algorithm 7.5] with starting point equal to $(0,\ldots,0)$. Similarly to the previous subsection, (13) has a closed-form solution $y_k := (y_k^1, y_k^{2:n+1})$ given by

$$y_k^1=\alpha\tilde{x}_k^1+(1-\alpha)y_{k-1}^1+\frac{1}{\beta}\gamma_{k-1}^1,\quad y_k^{2:n+1}=\mathcal{S}_{\frac{\mu}{\beta}}\left(\alpha\tilde{x}_k^{2:n+1}+(1-\alpha)y_{k-1}^{2:n+1}+\frac{1}{\beta}\gamma_{k-1}^{2:n+1}\right),$$

where $\mathcal{S}$ is the shrinkage operator as defined in (36).

Table 2: l_1**-regularized logistic regression problem**

	$\alpha = 1.0$		$\alpha = 1.3$		$\alpha = 1.5$		$\alpha = 1.7$		$\alpha = 1.9$	
Data set	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2	Alg. 1	Alg. 2
	Number of outer iterations									
Colon	370	337	253	259	216	224	196	197	175	176
CNS	278	278	216	213	186	185	163	163	144	145
Leukemia	625	624	481	480	416	416	367	367	328	328
Lung	551	513	435	400	375	347	378	380	548	528
Lymphoma	375	375	289	287	251	248	223	219	197	195
Prostate	882	879	678	676	585	585	512	516	457	462
Madelon	1935	1953	1480	1502	1269	1302	1105	1148	975	1027
	Total number of inner iterations									
Colon	9912	18645	7033	14460	5784	12334	5515	10949	4883	9688
CNS	8758	15515	6781	11881	5969	10259	5086	9068	4528	8077
Leukemia	15402	27859	11763	21486	10354	18560	8951	16271	7925	14538
Lung	15744	28487	13005	22329	10642	18813	10559	20320	16208	28931
Lymphoma	11191	21638	8666	16485	7443	14248	6546	12590	5826	11228
Prostate	37327	68770	28419	52865	24842	45705	22902	40480	21267	36160
Madelon	19857	38698	14859	29584	11898	25871	9806	22601	8159	20371
	CPU time in seconds									
Colon	21.8	48.3	13.5	37.4	10.3	31.8	9.8	28.3	8.7	24.7
CNS	107.9	302.0	88.9	232.2	79.0	199.0	68.9	177.4	61.5	159.2
Leukemia	168.6	417.1	131.8	337.4	110.1	279.7	93.4	243.0	91.4	215.8
Lung	352.2	844.1	292.5	638.4	239.8	539.1	242.9	572.5	363.7	822.8
Lymphoma	190.0	527.5	156.3	402.5	134.3	351.9	121.8	308.7	108.1	276.0
Prostate	1246.5	3844.6	918.9	2950.4	807.8	2562.0	761.8	2271.1	782.1	2036.8
Madelon	817.6	1589.2	605.6	1205.2	461.3	1065.1	390.6	887.6	332.2	809.4

Table 2 displays the numerical results obtained. As in Subsection 3.1, the methods were compared in terms of the number of outer iterations, the total number of inner iterations and the CPU time in seconds. In Figure 2, we plot the arithmetic mean of the latter three comparison criteria for each method for solving the seven l_1-regularized logistic regression problem instances. By analyzing Table 2 and Figure 2, one can see that Algorithm 1 performed, basically, the same number of outer iterations than Algorithm 2. Regarding the total number of inner iterations, Algorithm 1 performed at least 41% less than Algorithm 2, reaching, in some instances, 60% less inner iterations. Note that the saving with respect to CPU times was very expressive. Specifically, Algorithm 1 was at least 48% faster than Algorithm 2. The reason lies in the difficulty to solve (6) for the l_1-regularized logistic regression problem.

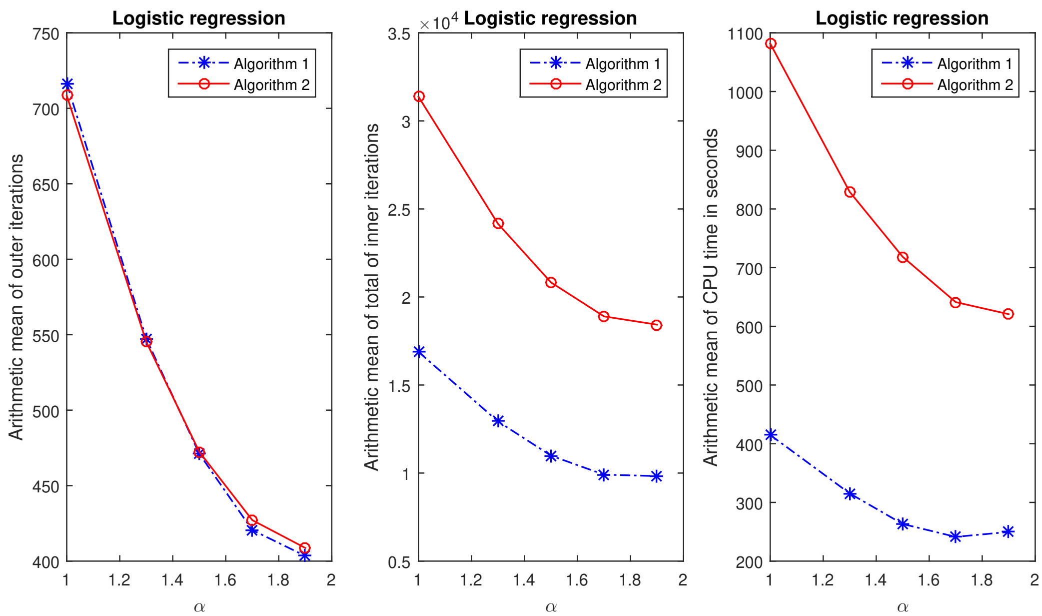

Figure 2: Arithmetic mean of the l_1-regularized logistic regression problem results

4 Conclusions

This paper proposed and analyzed an inexact proximal generalized ADMM for computing approximate solutions of linearly constrained convex optimization problems. The proposed method is a variant of the generalized ADMM proposed by Bertsekas and Eckstein in [11]. It basically consists of approximately solving a prox-inclusion associated to the first generalized ADMM subproblem using a relative error condition. The second generalized ADMM subproblem is regularized by a proximal term and assumed to be easy to solve. It was shown that the proposed inexact method has pointwise and ergodic iteration-complexity bounds similar to its exact version. Some numerical experiments were carried out in order to illustrate the numerical behavior of the method. They indicate that the proposed scheme represents an useful tool for solving some real-life applications that can be formulated as linearly constrained convex optimization problems.

A Proof of Proposition 2.3

We start by showing that the inclusion in Proposition 2.3(a) holds. First, it follows from the definitions of $\tilde{\gamma}_k$ and γ_k as in (12) and (14), respectively, that

$$\tilde{\gamma}_k - \gamma_{k-1} = \frac{\beta}{\alpha} B(y_k - y_{k-1}) + \frac{1}{\alpha}\left(\gamma_k - \gamma_{k-1}\right), \quad \forall k \geq 1. \tag{38}$$

Proof of the inclusion in Proposition 2.3(a): From the inclusion in (11) and the first relation in (14), we have

$$\frac{1}{\beta}(x_{k-1} - x_k) = v_k \in \partial f(\tilde{x}_k) - A^*\tilde{\gamma}_k. \tag{39}$$

Now, the first-order optimality condition for (13) and the definition of γ_k in (14) imply that

$$0 \in \partial g(y_k) - B^*\gamma_k + H(y_k - y_{k-1}). \tag{40}$$

On the other hand, it follows from (38) that

$$\gamma_k = \tilde{\gamma}_k - \frac{1-\alpha}{\alpha}(\gamma_k - \gamma_{k-1}) - \frac{\beta}{\alpha}B(y_k - y_{k-1}),$$

which, combined with (40), yields

$$(H + \frac{\beta}{\alpha}B^*B)\,(y_{k-1} - y_k) + \frac{1-\alpha}{\alpha}B^*\,(\gamma_{k-1} - \gamma_k) \in \partial g(y_k) - B^*\tilde{\gamma}_k. \tag{41}$$

From the second equality in (14), we obtain

$$\frac{1-\alpha}{\alpha}B\,(y_{k-1} - y_k) + \frac{1}{\alpha\beta}\,(\gamma_{k-1} - \gamma_k) = A\tilde{x}_k + By_k - b. \tag{42}$$

Therefore, the desired inclusion now follows by combining (39), (41), (42) and the definitions of M and T in (16). $\square$

In order to prove the remaining statements of Proposition 2.3, we need to establish two technical results. Note first that the relation in (38) implies that

$$\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 = \frac{\beta}{\alpha^2}\,\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_S^2\,, \quad \text{where} \quad S = \begin{bmatrix} \beta B^*B & B^* \\ B & \frac{1}{\beta}I \end{bmatrix}. \tag{43}$$

For simplicity, we also consider the following symmetric matrices

$$N = \begin{bmatrix} [1+\alpha(2-\alpha)]\,\beta B^*B & (1+\alpha-\alpha^2)B^* \\ (1+\alpha-\alpha^2)B & \frac{1}{\beta}I \end{bmatrix}, \quad P = \begin{bmatrix} \beta B^*B & (1-\alpha)B^* \\ (1-\alpha)B & \frac{(1-\alpha)^2}{\beta}I \end{bmatrix}. \tag{44}$$

It is easy to verify that S, N and P are positive semidefinite for every $\beta > 0$ and $\alpha \in (0, 2)$.

Lemma A.1. *Let $\{z_k\}$ and $\{\tilde{z}_k\}$ be as in* (18)*. Then, for every $k \geq 1$, the following hold:*

$$\|\tilde{z}_k - z_{k-1}\|_M^2 \geq \frac{1}{\beta}\,\|\tilde{x}_k - x_{k-1}\|^2 + \frac{1}{\alpha^3}\,\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_N^2 \tag{45}$$

and

$$\|\tilde{z}_k - z_k\|_M^2 = \frac{1}{\beta}\,\|\tilde{x}_k - x_k\|^2 + \frac{1}{\alpha^3}\,\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_P^2\,, \tag{46}$$

where the matrices M, N and P are as in (16) *and* (44)*.*

Proof. Using the fact that $\tilde{z}_k - z_{k-1} = (\tilde{x}_k - x_{k-1}, y_k - y_{k-1}, \tilde{\gamma}_k - \gamma_{k-1})$ and the definition of M in (16), we obtain

$$\begin{aligned}\|\tilde{z}_k - z_{k-1}\|_M^2 =& \frac{1}{\beta}\|\tilde{x}_k - x_{k-1}\|^2 + \|y_k - y_{k-1}\|_H^2 + \frac{\beta}{\alpha}\|B(y_k - y_{k-1})\|^2 \\ &+ \frac{2(1-\alpha)}{\alpha}\langle B(y_k - y_{k-1}), \tilde{\gamma}_k - \gamma_{k-1}\rangle + \frac{1}{\alpha\beta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2.\end{aligned}$$

On the other hand, equality (38) implies that

$$\langle B(y_k - y_{k-1}), \tilde{\gamma}_k - \gamma_{k-1}\rangle = \frac{\beta}{\alpha}\|B(y_k - y_{k-1})\|^2 + \frac{1}{\alpha}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle,$$

and

$$\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 = \frac{\beta^2}{\alpha^2}\|B(y_k - y_{k-1})\|^2 + \frac{2\beta}{\alpha^2}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle + \frac{1}{\alpha^2}\|\gamma_k - \gamma_{k-1}\|^2.$$

Combining the last three equalities, we find

$$\begin{aligned}\|\tilde{z}_k - z_{k-1}\|_M^2 \geq& \frac{1}{\beta}\|\tilde{x}_k - x_{k-1}\|^2 + \left(\frac{1}{\alpha} + \frac{2(1-\alpha)}{\alpha^2} + \frac{1}{\alpha^3}\right)\beta\|B(y_k - y_{k-1})\|^2 \\ &+ \left(\frac{2(1-\alpha)}{\alpha^2} + \frac{2}{\alpha^3}\right)\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle + \frac{1}{\alpha^3\beta}\|\gamma_k - \gamma_{k-1}\|^2.\end{aligned}$$

Thus, (45) follows from the last equality and the definition of N in (44).

Let us now prove (46). Using $\tilde{z}_k - z_k = (\tilde{x}_k - x_k, 0, \tilde{\gamma}_k - \gamma_k)$ (see (18)) and the definition of M in (16), we have

$$\|\tilde{z}_k - z_k\|_M^2 = \frac{1}{\beta}\|\tilde{x}_k - x_k\|^2 + \frac{1}{\alpha\beta}\|\tilde{\gamma}_k - \gamma_k\|^2.$$

It follows from (38) and some algebraic manipulations that

$$\|\tilde{\gamma}_k - \gamma_k\|^2 = \frac{\beta^2}{\alpha^2}\|B(y_k - y_{k-1})\|^2 + \frac{2(1-\alpha)\beta}{\alpha^2}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle + \frac{(1-\alpha)^2}{\alpha^2}\|\gamma_k - \gamma_{k-1}\|^2.$$

Therefore, the desired equality now follows by combining the last two equalities and the definition of P in (44). □

Lemma A.2. *Let $\{(x_k, y_k, \gamma_k)\}$ be generated by Algorithm 1. Then, the following hold:*

(a) $2\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle \geq \|y_1 - y_0\|_H^2 - 4d_0^2$, *where d_0 is as in (19);*

(b) $2\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle \geq \|y_k - y_{k-1}\|_H^2 - \|y_{k-1} - y_{k-2}\|_H^2$, *for every $k \geq 2$.*

Proof. (a) Consider z_0, z_1 and $\tilde{z}_1$ as in (18), and let an arbitrary $z^* := (x^*, y^*, \gamma^*) \in \Omega^*$ (see Assumpiton 1). Note that, in view of the definition of d_0 in (19), in order to establish (a), it is sufficient to prove that

$$\Theta := \|y_1 - y_0\|_H^2 - 2\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle \leq 4\|z^* - z_0\|_M^2, \tag{47}$$

where M is as in (16). Let us then show (47). From the definitions of M and $\{z_k\}$, we have

$$\begin{aligned}\|z_1 - z_0\|_M^2 &= \frac{1}{\beta}\|x_1 - x_0\|^2 + \|y_1 - y_0\|^2_{H+\frac{\beta}{\alpha}B^*B} + \frac{2(1-\alpha)}{\alpha}\langle B(y_1 - y_0), \gamma_1 - \gamma_0\rangle + \frac{1}{\alpha\beta}\|\gamma_1 - \gamma_0\|^2 \\ &= \frac{1}{\beta}\|x_1 - x_0\|^2 + \Theta + \left\|\frac{\sqrt{\beta}}{\sqrt{\alpha}}B(y_1 - y_0) + \frac{1}{\sqrt{\alpha\beta}}(\gamma_1 - \gamma_0)\right\|^2.\end{aligned}$$

Hence, we obtain

$$\Theta \le \|z_1 - z_0\|_M^2 \le 2\left(\|z^* - z_1\|_M^2 + \|z^* - z_0\|_M^2\right), \tag{48}$$

where the last inequality is due to $\|z + z'\|_M^2 \le 2\left(\|z\|_M^2 + \|z'\|_M^2\right)$ for all z, z'. We will now prove that

$$\|z^* - z_1\|_M^2 \le \|z^* - z_0\|_M^2. \tag{49}$$

Since we have already proved that the inclusion in Proposition 2.3(a) holds, we have $M(z_0 - z_1) \in T(\tilde{z}_1)$ where M and T are as in (16). Thus, using that $0 \in T(z^*)$ and T is monotone, we obtain $\langle M(z_0 - z_1), z^* - \tilde{z}_1\rangle \le 0$. Hence,

$$\begin{aligned}\|z^* - z_1\|_M^2 - \|z^* - z_0\|_M^2 &= \|(z^* - \tilde{z}_1) + (\tilde{z}_1 - z_1)\|_M^2 - \|(z^* - \tilde{z}_1) + (\tilde{z}_1 - z_0)\|_M^2 \\ &= \|\tilde{z}_1 - z_1\|_M^2 + 2\langle M(z_0 - z_1), z^* - \tilde{z}_1\rangle - \|\tilde{z}_1 - z_0\|_M^2 \\ &\le \|\tilde{z}_1 - z_1\|_M^2 - \|\tilde{z}_1 - z_0\|_M^2.\end{aligned}$$

Using (46), the inequality in (11), and the first equality in (14) (all with $k = 1$), we have

$$\|\tilde{z}_1 - z_1\|_M^2 \le \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 + \frac{\tau_2}{\beta}\|\tilde{x}_1 - x_0\|^2 + \frac{1}{\alpha^3}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_P^2,$$

where P is as in (44). Now, (45) with $k = 1$ becomes

$$\|\tilde{z}_1 - z_0\|_M^2 \ge \frac{1}{\beta}\|\tilde{x}_1 - x_0\|^2 + \frac{1}{\alpha^3}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_N^2$$

where N is as in (44). Combining the last three inequalities and the fact that $\tau_2 < 1$ (see Algorithm 1), we find

$$\begin{aligned}\|z^* - z_1\|_M^2 - \|z^* - z_0\|_M^2 &\le \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 + \frac{1}{\alpha^3}\left(\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_P^2 - \|(y_1 - y_0, \gamma_1 - \gamma_0)\|_N^2\right) \\ &= \frac{\tau_1}{\beta}\|\tilde{\gamma}_1 - \gamma_0\|^2 - \frac{2-\alpha}{\alpha^2}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_S^2,\end{aligned} \tag{50}$$

where the last equality is due to the fact that $P - N = -\alpha(2-\alpha)S$, with S given in (43). The last inequality, (43) with $k = 1$ and the fact that $\alpha \in (0, 2 - \tau_1)$ yield

$$\|z^* - z_1\|_M^2 - \|z^* - z_0\|_M^2 \le \frac{\alpha + \tau_1 - 2}{\alpha^2}\|(y_1 - y_0, \gamma_1 - \gamma_0)\|_S^2 \le 0,$$

which implies that (49) holds. Therefore, (a) now follows by combining (48) and (49).

(b) From the first-order optimality condition for (13) and the second relation in (14), we obtain

$$B^*\gamma_j - H(y_j - y_{j-1}) \in \partial g(y_j) \qquad \forall\, j \ge 1.$$

Hence, for every $k \geq 2$, using the above inclusion with $j \leftarrow k$ and $j \leftarrow k-1$ and the monotonicity of ∂g , we have

$$\begin{aligned}\langle B^*(\gamma_k - \gamma_{k-1}), y_k - y_{k-1}\rangle &\geq \|y_k - y_{k-1}\|_H^2 - \langle H(y_{k-1} - y_{k-2}), y_k - y_{k-1}\rangle \\ &\geq \frac{1}{2}\|y_k - y_{k-1}\|_H^2 - \frac{1}{2}\|y_{k-1} - y_{k-2}\|_H^2 ,\end{aligned}$$

where the last inequality is due to the fact that $2\langle Hy, y'\rangle \leq \|y\|_H^2 + \|y'\|_H^2$ for all y, y'. Therefore, (b) follows trivially from the last inequality. □

We are now ready to prove the remaining statements of Proposition 2.3.

Proof of the inequality in Proposition 2.3(a): Using (46) and the first relation in (14), we have

$$\begin{aligned}\|\tilde{z}_k - z_k\|_M^2 &= \frac{1}{\beta}\|\tilde{x}_k - x_{k-1} + \beta v_k\|^2 + \frac{1}{\alpha^3}\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_P^2 \\ &\leq \frac{\tau_1}{\beta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \frac{\tau_2}{\beta}\|\tilde{x}_k - x_{k-1}\|^2 + \frac{1}{\alpha^3}\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_P^2 ,\end{aligned}$$

where the inequality is due to the second condition in (11). It follows from the last inequality, (45) and the fact that $\sigma \geq \tau_2$ (see (20)) that

$$\sigma\|\tilde{z}_k - z_{k-1}\|_M^2 - \|\tilde{z}_k - z_k\|_M^2 \geq a_k \tag{51}$$

where

$$a_k := -\frac{\tau_1}{\beta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 + \frac{1}{\alpha^3}\left(\sigma\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_N^2 - \|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_P^2\right).$$

We will show that $a_k \geq \eta_k - \eta_{k-1}$, where the sequence $\{\eta_k\}$ is defined in (21). From (43), we find

$$\frac{\tau_1}{\beta}\|\tilde{\gamma}_k - \gamma_{k-1}\|^2 = \frac{1}{\alpha^3}\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_{\alpha\tau_1 S}^2 ,$$

which, combined with definition of a_k, yields

$$a_k = \frac{1}{\alpha^3}\|(y_k - y_{k-1}, \gamma_k - \gamma_{k-1})\|_{\sigma N - \alpha\tau_1 S - P}^2 .$$

Hence, using the definitions of N, S and P in (43) and (44), we obtain

$$a_k = \frac{1}{\alpha^3}\left(\hat{\xi}\beta\|B(y_k - y_{k-1})\|^2 + 2\xi\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle + \frac{\tilde{\xi}}{\beta}\|\gamma_k - \gamma_{k-1}\|^2\right), \tag{52}$$

where

$$\hat{\xi} = \sigma(1 + \alpha(2-\alpha)) - \alpha\tau_1 - 1, \quad \xi = \sigma(1 + \alpha - \alpha^2) + (1-\tau_1)\alpha - 1, \quad \tilde{\xi} = \sigma - \alpha\tau_1 - (1-\alpha)^2. \tag{53}$$

Now, from the definition of σ given in (20), we obtain $\sigma \geq (1 + \alpha\tau_1)/(1 + \alpha(2-\alpha))$. Hence, $\hat{\xi} \geq 0$ and

$$\tilde{\xi} \geq \frac{1 + \alpha\tau_1}{1 + \alpha(2-\alpha)} - \alpha\tau_1 - (1-\alpha)^2 = \frac{\alpha^2(2 - \tau_1 - \alpha)(2-\alpha)}{1 + \alpha(2-\alpha)} > 0,$$

where the last inequality is due to the fact that $\alpha \in (0, 2-\tau_1)$. Moreover, since $\sigma \in (0,1)$ (see (20)), we find

$$\xi = \sigma(1+\alpha-\alpha^2) + \alpha - \tau_1\alpha - 1 > \sigma(1+\alpha(2-\alpha)) - \alpha\tau_1 - 1 = \hat{\xi}.$$

Thus, $\xi > \hat{\xi} \geq 0$, and $\tilde{\xi} \geq 0$. Hence, from (52) and Lemma A.2, it follows that

$$a_k \geq \frac{2\xi}{\alpha^3}\langle B(y_k - y_{k-1}), \gamma_k - \gamma_{k-1}\rangle \geq \begin{cases} \frac{1}{\alpha^3}\left(\xi\, \|y_1 - y_0\|_H^2 - 4\xi d_0^2\right), & k=1, \\ \frac{1}{\alpha^3}\left(\xi\, \|y_k - y_{k-1}\|_H^2 - \xi\, \|y_{k-1} - y_{k-2}\|_H^2\right), & k \geq 2, \end{cases}$$

which, combined with the definitions of $\{\eta_k\}$ in (21), yields $a_k \geq \eta_k - \eta_{k-1}$ for every $k \geq 1$. Hence, the desired inequality now follows from (51). □

Proof of Proposition 2.3(b): First, for every $z^* = (x^*, y^*, \gamma^*) \in \Omega^*$, we have

$$\begin{aligned} \|z^* - z_k\|_M^2 - \|z^* - z_{k-1}\|_M^2 &= \|(z^* - \tilde{z}_k) + (\tilde{z}_k - z_k)\|_M^2 - \|(z^* - \tilde{z}_k) + (\tilde{z}_k - z_{k-1})\|_M^2 \\ &= \|\tilde{z}_k - z_k\|_M^2 - \|\tilde{z}_k - z_{k-1}\|_M^2 + 2\langle M(z_{k-1} - z_k), z^* - \tilde{z}_k\rangle. \end{aligned}$$

Now, since $M(z_{k-1} - z_k) \in T(\tilde{z}_k)$ (see (22)), $0 \in T(z^*)$, and T is monotone, we trivially obtain $\langle M(z_{k-1} - z_k), \tilde{z}_k - z^*\rangle \geq 0$. Therefore, combining the last two inequalities and (22), we obtain

$$\|z^* - z_k\|_M^2 - \|z^* - z_{k-1}\|_M^2 \leq \eta_{k-1} - \eta_k - (1-\sigma)\|\tilde{z}_k - z_{k-1}\|_M^2,$$

which is equivalent to the desired inequality. □

References

[1] V. A. Adona, M. L. N. Gonçalves, and J. G. Melo. Iteration-complexity analysis of a generalized alternating direction method of multipliers. *J. Glob. Optim. https://doi.org/10.1007/s10898-018-0697-z*, 2018.

[2] V. A. Adona, M. L. N. Goncalves, and J. G. Melo. A Partially Inexact Alternating Direction Method of Multipliers and its Iteration-Complexity Analysis. *arXiv preprint arXiv:1805.07336*, 2018.

[3] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. U. S. A.*, 96(12):6745–6750, 1999.

[4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[5] D. G. Beer, S. L. R. Kardia, C. Huang, T. J. Giordano, A. M. Levin, D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.*, 8(8):816, 2002.

[6] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods.* Academic Press, New York, 1982.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.

[8] K. Bredies and H. Sun. A proximal point analysis of the preconditioned alternating direction method of multipliers. *J. Optim. Theory Appl.*, 173(3):878–907, 2017.

[9] A. Cano, A. Masegosa, and S. Moral. ELVIRA biomedical data set repository. http://leo.ugr.es/elvira/DBCRepository/, 2005.

[10] D. Dheeru and E. K. Taniskidou. UCI machine learning repository. http://archive.ics.uci.edu/ml/datasets/madelon, 2018.

[11] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55(3, Ser. A):293–318, 1992.

[12] J. Eckstein and P. J. S. Silva. A practical relative error criterion for augmented Lagrangians. *Math. Programming*, 141(1):319–348, 2013.

[13] J. Eckstein and W. Yao. Approximate ADMM algorithms derived from Lagrangian splitting. *Comput. Optim. Appl.*, 68(2):363–405, 2017.

[14] J. Eckstein and W. Yao. Relative-error approximate versions of Douglas–Rachford splitting and special cases of the ADMM. *Math. Programming*, 170(2):417–444, 2018.

[15] E. X. Fang, H. Bingsheng, H. Liu, and Y. Xiaoming. Generalized alternating direction method of multipliers: new theoretical insights and applications. *Math. Prog. Comp.*, 7(2):149–187, 2015.

[16] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.*, 2:17–40, 1976.

[17] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *R.A.I.R.O.*, 9(R2):41–76, 1975.

[18] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, 1999.

[19] M. L. N. Gonçalves, M. M. Alves, and J. G. Melo. Pointwise and ergodic convergence rates of a variable metric proximal alternating direction method of multipliers. *J. Optim. Theory Appl.*, 177(2):448–478, 2018.

[20] M. L. N. Gonçalves, J. G. Melo, and R. D. C. Monteiro. Extending the ergodic convergence rate of the proximal ADMM. *arXiv preprint arXiv:1611.02903*, 2016.

[21] I. Guyon, S. Gunn, A. B. Hur, and G. Dror. Result analysis of the NIPS 2003 feature selection challenge. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Adv. Neural Inf. Process. Syst.*, pages 545–552. MIT Press, 2005.

[22] K. Koh, S. J. Kim, and S. Boyd. An interior-point method for large-scale l_1-regularized logistic regression. *J. Mach. Learn. Res.*, 8:1519–1555, 2007.

[23] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. I. Jordan. A general analysis of the convergence of ADMM. *arXiv preprint arXiv:1502.02009*, 2015.

[24] J. Nocedal and S. J. Wright. *Numerical Optimization 2nd.* Springer, New York, 2006.

[25] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. H. Kim, L. C. Goumnerova, P. M. Black, C. Lau, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 415(6870):436–442, 2002.

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

[27] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, Geraldine S. Pinkus, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.*, 8(1):68–74, 2002.

[28] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, et al. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1(2):203–209, 2002.

[29] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[30] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[31] R. Tibshirani. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B*, 58(1):267–288, 1996.

[32] R. J. Tibshirani. The Lasso problem and uniqueness. *Electron. J. Statist.*, 7:1456–1490, 2013.

[33] J. Xie, A. Liao, and X. Yang. An inexact alternating direction method of multipliers with relative error criteria. *Optim. Lett.*, 11(3):583–596, 2017.