Projective method of multipliers for linearly constrained convex minimization

Majela Pentón Machado[1] ![ORCID]

Abstract
We present a method for solving linearly constrained convex optimization problems, which is based on the application of known algorithms for finding zeros of the sum of two monotone operators (presented by Eckstein and Svaiter) to the dual problem. We establish convergence rates for the new method, and we present applications to TV denoising and compressed sensing problems.

Keywords Constrained optimization · Convex programming · Complexity · Total variation denoising · Compressed sensing

Mathematics Subject Classification 49M29 · 90C25 · 65K05 · 68Q25

1 Introduction

A broad class of problems of recent interest in image science and signal processing can be posed in the framework of convex optimization. Examples include the TV denoising model [24] for image processing and basis pursuit, which is well known for playing a central role in the theory of compressed sensing. A general subclass of such programming problems is:

$$\min_{u \in \mathbb{R}^{m_1}, v \in \mathbb{R}^{m_2}} \{ f(u) + g(v) \ : \ Mu + Cv = d \}. \tag{1}$$

Here $f : \mathbb{R}^{m_1} \to (-\infty, \infty]$ and $g : \mathbb{R}^{m_2} \to (-\infty, \infty]$ are proper closed convex functions, $M : \mathbb{R}^{m_1} \to \mathbb{R}^n$ and $C : \mathbb{R}^{m_2} \to \mathbb{R}^n$ are linear operators, and $d \in \mathbb{R}^n$.

A well-known iterative method for solving optimization problems that have a separable structure as (1) does, is the *Alternating Direction Method of Multipliers* (ADMM), which goes back to the works of Glowinski and Marrocco [13], and of Gabay

✉ Majela Pentón Machado
majela.penton@ufba.br

1 UFBA, Universidade Federal da Bahia, Salvador, BA 40170-110, Brazil

and Mercier [12]. ADMM solves the coupled problem (1) performing a sequences of steps that decouple functions f and g, making it possible to exploit the individual structure of these functions. It can be interpreted in terms of alternating minimization, with respect to u and v, of the augmented Lagrangian function associated with problem (1). The classical ADMM (i.e., with dual steplength 1) can also be viewed as an instance of the method called *Douglas-Rachford splitting* applied to the dual problem of (1), as was shown by Gabay in [11].

Other splitting schemes have been effectively applied to the dual problem of (1), which is a special case of the problem of finding a zero of the sum of two maximal monotone operators. For example, the *Proximal Forward Backward splitting* method, developed by Lions and Mercier [17], and Passty [21], corresponds to the well-known Tseng's [25] *Alternating Minimization Algorithm* (AMA) for solving (1). This method has simpler steps than ADMM, in the former one of the minimizations of the augmented Lagrangian is replaced by the minimization of the Lagrangian itself; however, it requires strong convexity of one of the objective functions.

The goal of our work is to construct an optimization scheme for solving (1) applying a splitting method to its dual problem. Specifically we are interested in the family of splitting-projective methods proposed in [9] by Eckstein and Svaiter to address inclusion problems given by the sum of two maximal monotone operators. We will apply a specific instance of these algorithms to solve a reformulation of the dual problem of (1) as the problem of finding a zero of the sum of two maximal monotone operators, which allows us to obtain a new algorithm for solving this problem. To the best of our knowledge, this is the first time that the methods in [9] are applied to the optimization problem (1). The specific instance we choose to study here is motivated by preliminary numerical results (illustrated in Sect. 2), which demonstrate the performance of the method. This iterative method will be referred to as the *Projective Method of Multipliers* (PMM). The convergence properties of the PMM will be obtained using the convergence results already established in [9]. In contrast to [9], which only studies the global convergence of the family of splitting-projective methods, we also establish in this work the iteration complexity of the PMM. Using the Karush–Kuhn–Tucker (KKT) conditions for problem (1) we give convergence rate for the PMM measured by the pointwise and ergodic iteration-complexities.

The remainder of this paper is organized as follows. Section 2 reviews some definitions and facts on convex functions that will be used in our subsequent presentation. It also briefly discusses Lagrangian duality theory for convex optimization, for more details in this subject we refer the reader to [22]. Section 3 presents the Projective Method of Multipliers (PMM) for solving the class of linearly constrained optimization problems (1). This section also presents global convergence of the PMM using the convergence analysis presented in [9]. Section 4 derives iteration-complexity results for the PMM. Finally, Sect. 5 presents some applications in image restoration and compressed sensing. This section also exhibits preliminary numerical results describing the performance of the PMM in solving these problems.

1.1 Notation

Throughout this paper, we let $\mathbb{R}^n$ denote an n-dimensional space with inner product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. For a matrix A, A^T indicates its transpose and $\|A\|_F = \sqrt{trace(AA^T)}$ its Frobenius norm. Given a linear operator M, we denote by M^* its adjoint operator. If C is a convex set we indicate by ri (C) its relative interior.

2 Preliminaries

In this section we describe some basic definitions and facts on convex analysis that will be needed along this work. We also discuss the Lagrangian formulation and dual problem of (1). This approach will play an important role in the design of the PMM for problem (1).

2.1 Generalities on convex functions

Given an extended real valued convex function $f : \mathbb{R}^n \to (-\infty, \infty]$, the domain of f is the set

$$\text{dom } f := \left\{ x \in \mathbb{R}^n \: : \: f(x) < \infty \right\}.$$

Since f is a convex function, it is obvious that dom f is convex. We say that function f is *proper* if dom $f \neq \emptyset$. Furthermore, we say that f is *closed* if it is a lower semicontinuous function.

Definition 1 Given a convex function $f : \mathbb{R}^n \to (-\infty, \infty]$ a vector $v \in \mathbb{R}^n$ is called a *subgradient* of f at $x \in \mathbb{R}^n$ if

$$f(x') \geq f(x) + \langle v, x' - x \rangle \qquad \forall x' \in \mathbb{R}^n.$$

The set of all subgradients of f at x is denoted by $\partial f(x)$. The operator ∂f, which maps each x to $\partial f(x)$, is called the *subdifferential* map associated with f.

It can be seen immediately from the definition that x^* is a global minimizer of f in $\mathbb{R}^n$ if and only if $0 \in \partial f(x^*)$. If f is differentiable at x, then $\partial f(x)$ is the singleton set $\{\nabla f(x)\}$.

The subdifferential mapping of a convex function f has the following *monotonicity* property: for any x, x', v and $v' \in \mathbb{R}^n$ such that $v \in \partial f(x)$ and $v' \in \partial f(x')$, it follows that

$$\langle x - x', v - v' \rangle \geq 0. \tag{2}$$

In addition, if f is a proper closed convex function, then ∂f is a *maximal monotone* operator [23]. This is to say that if $x, v \in \mathbb{R}^n$ are such that inequality (2) holds for all $x' \in \mathbb{R}^n$ and $v' \in \partial f(x')$, then $x \in \text{dom } f$ and $v \in \partial f(x)$.

Given $\lambda > 0$, the *resolvent mapping* (or *proximal mapping*) [20] associated with ∂f is defined as

$$(I + \lambda \partial f)^{-1}(z) := \arg \min_{x \in \mathbb{R}^n} \lambda f(x) + \frac{1}{2} \|x - z\|^2, \qquad \forall z \in \mathbb{R}^n.$$

The fact that $(I + \lambda \partial f)^{-1}$ is an everywhere well defined function, if f is proper, closed and convex, is a consequence of a fundamental result due to Minty [18]. For example, if $f(x) = \mu \|x\|_1 = \mu \sum |x_i|$ where $\mu > 0$, then

$$(I + \partial f)^{-1}(z) = \mathbf{shrink}(z, \mu),$$

where

$$\mathbf{shrink}(z, \mu)_i := \max\{|z_i| - \mu, 0\} sign(z_i). \tag{3}$$

The *Fenchel-Legendre conjugate* of a convex function f, denoted by $f^* : \mathbb{R}^n \to (-\infty, \infty]$, is defined as

$$f^*(v) := \sup_{x \in \mathbb{R}^n} \langle v, x \rangle - f(x), \qquad \forall v \in \mathbb{R}^n.$$

It is simple to see that f^* is a convex closed function. Furthermore, if f is proper, closed and convex, then f^* is a proper function (see for instance [22, Theorem 12.2] and [1, Proposition 1.10]).

Definition 2 Given any convex function $f : \mathbb{R}^n \to (-\infty, \infty]$ and $\epsilon \geq 0$, a vector $v \in \mathbb{R}^n$ is called an ϵ-*subgradient* of f at $x \in \mathbb{R}^n$ if

$$f(x') \geq f(x) + \langle v, x' - x \rangle - \epsilon \qquad \forall x' \in \mathbb{R}^n.$$

The set of all ϵ-subgradients of f at x is denoted by $\partial_\epsilon f(x)$, and $\partial_\epsilon f$ is called the ϵ-*subdifferential* mapping.

It is trivial to verify that $\partial_0 f(x) = \partial f(x)$, and $\partial f(x) \subseteq \partial_\epsilon f(x)$ for every $x \in \mathbb{R}^n$ and $\epsilon \geq 0$. The proposition below lists some useful properties of the ϵ-subdifferential that will be needed in our presentation.

Proposition 2.1 *If* $f : \mathbb{R}^n \to (-\infty, \infty]$ *is a proper closed convex function,* $g : \mathbb{R}^n \to \mathbb{R}$ *is a convex differentiable function in* $\mathbb{R}^n$, *and* $M : \mathbb{R}^m \to \mathbb{R}^n$ *is a linear transformation, then the following statements hold:*

(a) $v \in \partial_\epsilon f(x)$ *if and only if* $x \in \partial_\epsilon f^*(v)$ *for all* $\epsilon \geq 0$;
(b) $\partial(f + g)(x) = \partial f(x) + \nabla g(x)$ *for all* $x \in \mathbb{R}^n$;
(c) $\partial(f \circ M)(x) \supseteq M^*\partial f(Mx)$ *for all* $x \in \mathbb{R}^m$. *In addition, if* $ri(dom f) \cap range M \neq \emptyset$, *then* $\partial(f \circ M)(x) = M^*\partial f(Mx)$ *for every* $x \in \mathbb{R}^m$;
(d) *if* $x_i, v_i \in \mathbb{R}^n$ *and* $\epsilon_i, \alpha_i \in \mathbb{R}_+$, *for* $i = 1, \ldots, k$, *are such that*

$$v_i \in \partial_{\epsilon_i} f(x_i), \quad i = 1, \ldots, k, \qquad \sum_{i=1}^{k} \alpha_i = 1,$$

and we define

$$\overline{x} := \sum_{i=1}^{k} \alpha_i x_i, \qquad \overline{v} := \sum_{i=1}^{k} \alpha_i v_i, \qquad \overline{\epsilon} := \sum_{i=1}^{k} \alpha_i (\epsilon_i + \langle x_i - \overline{x}, v_i \rangle);$$

then, we have $\overline{\epsilon} \geq 0$ and $\overline{v} \in \partial_{\overline{\epsilon}} f(\overline{x})$.

Proof Statements (a)–(c) are classical results which can be found, for example, in [16,22]. For a proof of item (d) see [2] and references therein. □

2.2 Lagrangian duality

The *Lagrangian function* $L : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \mathbb{R}^n \to (-\infty, \infty]$ for problem (1) is defined as

$$L(u, v, z) := f(u) + g(v) + \langle Mu + Cv - d, z \rangle. \tag{4}$$

The *dual function* is the concave function $\varphi : \mathbb{R}^n \to [-\infty, \infty)$ defined by

$$\varphi(z) := \inf_{(u,v) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}} L(u, v, z),$$

and the *dual problem* to (1) is

$$\max_{z \in \mathbb{R}^n} \varphi(z). \tag{5}$$

Problem (1) will be called the *primal problem*. Straightforward calculations show that weak duality holds, i.e. $\varphi^* \leq p^*$, where p^* and φ^* are the optimal values of (1) and (5), respectively.

A vector (u^*, v^*, z^*) such that $L(u^*, v^*, z^*)$ is finite and it satisfies

$$\min_{(u,v) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}} L(u, v, z^*) = L(u^*, v^*, z^*) = \max_{z \in \mathbb{R}^n} L(u^*, v^*, z) \tag{6}$$

is called a *saddle point* of the Lagrangian function L. If (u^*, v^*, z^*) is a saddle point, then (u^*, v^*) is an optimal primal solution and z^* is an optimal dual solution [22, Theorem 28.3]. Furthermore, if a saddle point of L exists, then $p^* = \varphi^*$, i.e. there is no duality gap [22].

Notice that, if (u^*, v^*, z^*) is a saddle point, from the definition of L in (4) and equalities (6) we deduce that

$$f(u) + g(v) + \langle Mu + Cv - d, z^* \rangle \geq L(u^*, v^*, z^*)$$
$$\geq f(u^*) + g(v^*) + \langle Mu^* + Cv^* - d, z \rangle$$

for all $u \in \mathbb{R}^{m_1}$, $v \in \mathbb{R}^{m_2}$, $z \in \mathbb{R}^n$. From these relations we can directly derive the *Karush–Kuhn–Tucker* (KKT) conditions

$$0 = Mu^* + Cv^* - d,$$
$$0 \in \partial f(u^*) + M^*z^*, \tag{7}$$
$$0 \in \partial g(v^*) + C^*z^*,$$

which describe an optimal solution of problem (1). Observe that the equality in (7) implies that the primal variables (u^*, v^*) must be feasible. The inclusions in (7) are known as the dual feasibility conditions. We also have that the KKT conditions hold if and only if (u^*, v^*, z^*) is a saddle point of L.

Observe that the dual function φ can be written in terms of the Fenchel-Legendre conjugates of the functions f and g. Specifically,

$$\varphi(z) = \inf_{(u,v) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}} f(u) + g(v) + \langle Mu + Cv - d, z \rangle$$
$$= \inf_{u \in \mathbb{R}^{m_1}} f(u) + \langle Mu, z \rangle + \inf_{v \in \mathbb{R}^{m_2}} g(v) + \langle Cv, z \rangle - \langle d, z \rangle$$
$$= -f^*(-M^*z) - g^*(-C^*z) - \langle d, z \rangle.$$

Hence, if we define the functions $h_1(z) = (f^* \circ -M^*)(z)$ and $h_2(z) = (g^* \circ -C^*)$ $(z) + \langle d, z \rangle$, we have that the dual problem (5) is equivalent to minimizing $h_1 + h_2$ over $\mathbb{R}^n$. Furthermore, since f^* and g^* are convex and closed, and M^* and C^* are linear operators, it follows that h_1 and h_2 are convex closed functions [22]. Therefore, z^* is a solution of (5) if and only if

$$0 \in \partial(h_1 + h_2)(z^*). \tag{8}$$

Throughout this work, we assume that

> (A.1) there exists (u^*, v^*, z^*) a saddle point of L.

Since condition A.1 implies that the KKT conditions hold, we have from the first inclusion in (7) and Proposition 2.1(a), (c) that $z^* \in \text{dom}(f^* \circ -M^*)$, which implies that h_1 is a proper function. A similar argument shows that h_2 is also a proper function. Therefore, under hypothesis A.1, we have that the subdifferentials ∂h_1 and ∂h_2 are maximal monotone operators.

3 The projective method of multipliers

Our proposal in this work is to apply the splitting-projective methods developed in [9], by Eckstein and Svaiter, to find a solution of problem

$$0 \in \partial h_1(z) + \partial h_2(z),$$

and as a consequence a solution of the dual problem (5), since the following inclusion holds

$$\partial h_1(z) + \partial h_2(z) \subseteq \partial(h_1 + h_2)(z) \qquad \forall z \in \mathbb{R}^n$$

(see Eq. (8) and the comments above).

The framework presented in [9] reformulates the problem of finding a zero of the sum of two maximal monotone operators in terms of a convex feasibility problem, which is defined by a certain closed convex "extended" solution set. To solve the feasibility problem, the authors introduced successive projection algorithms that use, on each iteration, independent calculations involving each operator.

Specifically, if we consider the subdifferential mappings ∂h_1 and ∂h_2, then the associated extended solution set, defined as in [9], is

$$S_e(\partial h_1, \partial h_2) := \left\{ (z, w) \in \mathbb{R}^n \times \mathbb{R}^n \ : \ -w \in \partial h_1(z), \ w \in \partial h_2(z) \right\}. \tag{9}$$

Since ∂h_1 and ∂h_2 are maximal monotone operators it can be proven that $S_e(\partial h_1, \partial h_2)$ is a closed convex set in $\mathbb{R}^n \times \mathbb{R}^n$, see [9, Lemma 2]. It is also easy to verify that if (z^*, w^*) is a point in $S_e(\partial h_1, \partial h_2)$ then z^* satisfies inclusion (8) and consequently it is a solution of the dual problem. Furthermore, the following lemma holds.

Lemma 3.1 *If (u^*, v^*, z^*) is a saddle point of L, then*

$$(z^*, d - Cv^*) \in S_e(\partial h_1, \partial h_2).$$

Moreover, if we assume the following conditions

(A.2) $ri\,(dom\,f^*) \cap range\,M^* \neq \emptyset$;
(A.3) $ri\,(dom\,g^*) \cap range\,C^* \neq \emptyset$.

Then, for all $(z^*, w^*) \in S_e(\partial h_1, \partial h_2)$ there exist $u^*, v^* \in \mathbb{R}^n$ such that $w^* = d - Cv^*$, $w^* = Mu^*$ and (u^*, v^*, z^*) is a saddle point of the Lagrangian function L.

Proof If (u^*, v^*, z^*) is a saddle point of the Lagrangian function, then the KKT optimality conditions hold, and the inclusions in (7), together with Proposition 2.1(a), imply that

$$u^* \in \partial f^*(-M^*z^*) \quad \text{and} \quad v^* \in \partial g^*(-C^*z^*).$$

Thus, we have

$$-Mu^* \in -M\partial f^*(-M^*z^*) \subseteq \partial(f^* \circ -M^*)(z^*) = \partial h_1(z^*) \tag{10}$$

and

$$-Cv^* \in -C\partial g^*(-C^*z^*) \subseteq \partial(g^* \circ -C^*)(z^*); \tag{11}$$

where the second inclusions in (10) and (11) follow from Proposition 2.1(c). Adding d to both sides of (11) and using the definition of h_2 and Proposition 2.1(b) we have $d - Cv^* \in \partial h_2(z^*)$. Now, adding this last inclusion to (10) we conclude that

$$-Mu^* + d - Cv^* \in \partial h_1(z^*) + \partial h_2(z^*).$$

The first assertion of the lemma follows combining the relation above with the equality in (7) and the definition of $S_e(\partial h_1, \partial h_2)$.

By (9) we have that if $(z^*, w^*) \in S_e(\partial h_1, \partial h_2)$ then $w^* \in \partial h_2(z^*) = -C\partial g^*(-C^*z^*) + d$, where the equality follows from condition A.3 and Proposition 2.1(b), (c). Thus, there exists $v^* \in \partial g^*(-C^*z^*)$ such that $w^* = -Cv^* + d$, and applying Proposition 2.1(a) we obtain that $-C^*z^* \in \partial g(v^*)$.

Similarly, using $-w^* \in \partial h_1(z^*)$, hypothesis A.2 and Proposition 2.1(a), (c), we deduce that there is a u^* such that $-w^* = -Mu^*$ and $-M^*z^* \in \partial f(u^*)$. All these conditions put together imply that (u^*, v^*, z^*) is a saddle point of L. □

According to Lemma 3.1, we can attempt to find a saddle point of the Lagrangian function (4), by seeking a point in the extended solution set $S_e(\partial h_1, \partial h_2)$.

In order to solve the feasibility problem defined by $S_e(\partial h_1, \partial h_2)$, by successive orthogonal projection methods, the authors of [9] used the resolvent mappings associated with the operators to construct affine separating hyperplanes.

In our setting the family of algorithms in [9, Algorithm 2] follows the set of recursions

$$\lambda_k b_k + x_k = z_{k-1} + \lambda_k w_{k-1}, \qquad\qquad b_k \in \partial h_2(x_k); \quad (12)$$

$$\mu_k a_k + y_k = (1 - \alpha_k)z_{k-1} + \alpha_k x_k - \mu_k w_{k-1}, \qquad a_k \in \partial h_1(y_k); \quad (13)$$

$$\gamma_k = \frac{\langle z_{k-1} - x_k, b_k - w_{k-1} \rangle + \langle z_{k-1} - y_k, a_k + w_{k-1} \rangle}{\|a_k + b_k\|^2 + \|x_k - y_k\|^2}, \qquad (14)$$

$$z_k = z_{k-1} - \rho_k \gamma_k (a_k + b_k), \qquad (15)$$

$$w_k = w_{k-1} - \rho_k \gamma_k (x_k - y_k), \qquad (16)$$

where $\lambda_k, \mu_k > 0$ and $\alpha_k \in \mathbb{R}$ are such that $(\mu_k/\lambda_k - (\alpha_k/2)^2) > 0$, and $\rho_k \in (0, 2)$.

We observe that relations in (12) and the definition of the resolvent mapping yield that $x_k = (I + \lambda_k \partial h_2)^{-1}(z_{k-1} + \lambda_k w_{k-1})$ and $b_k = \frac{1}{\lambda_k}(z_{k-1} - x_k) + w_{k-1}$. Similarly, (13) implies that $y_k = (I + \mu_k \partial h_1)^{-1}((1 - \alpha_k)z_{k-1} + \alpha_k x_k - \mu_k w_{k-1}))$ and $a_k = \frac{1}{\mu_k}((1 - \alpha_k)z_{k-1} + \alpha_k x_k - y_k) - w_{k-1}$. Hence, steps (12) and (13) are evaluations of the proximal mappings.

With the view to see that iterations (12)–(16) truly are successive (relaxed) projection methods for the convex feasibility problem of finding a point in $S_e(\partial h_1, \partial h_2)$, we define, for all integer $k \geq 1$, the affine function $\phi_k(z, w) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as

$$\phi_k(z, w) := \langle z - x_k, b_k - w \rangle + \langle z - y_k, a_k + w \rangle, \qquad (17)$$

and its non-positive level set

$$H_{\phi_k} := \{(z, w) : \phi_k(z, w) \leq 0\}.$$

Thus, by the monotonicity of the subdifferential mappings we have that $S_e(\partial h_1, \partial h_2) \subseteq H_{\phi_k}$ and we can also verify that the following relations hold[1]

$$\nabla \phi_k = (a_k + b_k, x_k - y_k), \qquad (18)$$

---

[1] Actually, it can be shown that $\gamma_k \geq c_k$, where $c_k > 0$ (see [9, Proposition 3]).

$$\gamma_k = \frac{\phi_k(z_{k-1}, w_{k-1})}{\|\nabla \phi_k\|^2} \quad \text{and} \quad \gamma_k \geq 0 \tag{19}$$

for all integer $k \geq 1$. Therefore, we conclude that if $\rho_k = 1$ the point (z_k, w_k), calculated by the update rule given by (15)–(16), is the orthogonal projection of (z_{k-1}, w_{k-1}) onto H_{ϕ_k}. Besides, if $\rho_k \neq 1$ we have that (z_k, w_k) is an under relaxed projection of (z_{k-1}, w_{k-1}).

As was observed in the paragraph after (16), in order to apply algorithm (12)–(16) it is necessary to calculate the resolvent mappings associated with ∂h_1 and ∂h_2. The next result shows how we can invert operators $I + \lambda \partial h_1$ and $I + \lambda \partial h_2$ for any $\lambda > 0$.

Lemma 3.2 *Consider $c \in \mathbb{R}^n$, $\theta : \mathbb{R}^m \to (-\infty, \infty]$ a proper closed convex function and $A : \mathbb{R}^m \to \mathbb{R}^n$ a linear operator such that $\mathrm{dom}\,\theta^* \cap \mathrm{range}\,A^* \neq \emptyset$. Let $z \in \mathbb{R}^n$ and $\lambda > 0$. Then, if $\tilde{v} \in \mathbb{R}^m$ is a solution of problem*

$$\min_{v \in \mathbb{R}^m} \theta(v) + \langle z, Av - c \rangle + \frac{\lambda}{2} \|Av - c\|^2 \tag{20}$$

it holds that $c - A\tilde{v} \in \partial h(\hat{z})$ where $h(\cdot) = (\theta^* \circ -A^*)(\cdot) + \langle c, \cdot \rangle$ and $\hat{z} = z + \lambda(A\tilde{v} - c)$. Hence, $\hat{z} = (I + \lambda \partial h)^{-1}(z)$. Furthermore, the set of optimal solutions of (20) is nonempty.

Proof If $\tilde{v} \in \mathbb{R}^m$ is a solution of (20), deriving the optimality condition of this minimization problem we have

$$0 \in \partial \theta(\tilde{v}) + A^* z + \lambda A^*(A\tilde{v} - c) = \partial \theta(\tilde{v}) + A^*(z + \lambda(A\tilde{v} - c)).$$

From the definition of $\hat{z}$ and the identity above it follows that

$$0 \in \partial \theta(\tilde{v}) + A^* \hat{z}.$$

Now, by equation above and Proposition 2.1(a), (c) we have

$$-A\tilde{v} \in \partial(\theta^* \circ -A^*)(\hat{z}). \tag{21}$$

Since we are assuming that $\mathrm{dom}\,\theta^* \cap \mathrm{range}\,A^* \neq \emptyset$, the definition of h and Proposition 2.1(b), (c) yield

$$\partial h(z) = \partial(\theta^* \circ -A^*)(z) + c = -A\partial \theta^*(-A^* z) + c, \qquad \forall z \in \mathbb{R}^n. \tag{22}$$

Therefore, adding c to both sides of (21) and combining with Eq. (22) we deduce that $c - A\tilde{v} \in \partial h(\hat{z})$. The assertion that $\hat{z} = (I + \lambda \partial h)^{-1}(z)$ is a direct consequence of this last inclusion and the definition of $\hat{z}$.

Next, we notice that, since ∂h is maximal monotone, Minty's theorem [18] asserts that for all $z \in \mathbb{R}^n$ and $\lambda > 0$ there exist $\tilde{z}, w \in \mathbb{R}^n$ such that

$$\begin{cases} w \in \partial h(\tilde{z}), \\ \lambda w + \tilde{z} = z. \end{cases} \tag{23}$$

Therefore, the inclusion above, together with Eq. (22), implies that there exists $\overline{v} \in \partial\theta^*(-A^*\tilde{z})$ such that $w = -A\overline{v} + c$. This last inclusion yields $-A^*\tilde{z} \in \partial\theta(\overline{v})$, from which we deduce that

$$0 \in \partial\theta(\overline{v}) + A^*\tilde{z} = \partial\theta(\overline{v}) + A^*(z - \lambda w),$$

where the equality above follows from the equality in (23). Finally, replacing w by $c - A\overline{v}$ in the equation above, we obtain

$$0 \in \partial\theta(\overline{v}) + A^*(z + \lambda(A\overline{v} - c)),$$

from which follows that $\overline{v}$ is an optimal solution of problem (20). □

In what follows we assume that conditions A.2 and A.3 are satisfied. We can now introduce the Projective Method of Multipliers.

Algorithm (PMM) *Let* $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n$, $\lambda > 0$ *and* $\overline{\rho} \in [0, 1)$ *be given. For* $k = 1, 2, \ldots$.

1. *Compute* $v_k \in \mathbb{R}^{m_2}$ *as*

$$v_k \in \arg\min_{v \in \mathbb{R}^{m_2}} g(v) + \langle z_{k-1} + \lambda w_{k-1}, Cv - d\rangle + \frac{\lambda}{2} \|Cv - d\|^2, \qquad (24)$$

and $u_k \in \mathbb{R}^{m_1}$ *as*

$$u_k \in \arg\min_{u \in \mathbb{R}^{m_1}} f(u) + \langle z_{k-1} + \lambda(Cv_k - d), Mu\rangle + \frac{\lambda}{2} \|Mu\|^2. \qquad (25)$$

2. *If* $\|Mu_k + Cv_k - d\| + \|Mu_k - w_{k-1}\| = 0$ *stop. Otherwise, set*

$$\gamma_k = \frac{\lambda \|Cv_k - d + w_{k-1}\|^2 + \lambda \langle d - Cv_k - Mu_k, w_{k-1} - Mu_k\rangle}{\|Mu_k + Cv_k - d\|^2 + \lambda^2 \|Mu_k - w_{k-1}\|^2}.$$

3. *Choose* $\rho_k \in [1 - \overline{\rho}, 1 + \overline{\rho}]$ *and set*

$$z_k = z_{k-1} + \rho_k\gamma_k(Mu_k + Cv_k - d),$$
$$w_k = w_{k-1} - \rho_k\gamma_k\lambda(w_{k-1} - Mu_k).$$

We observe that condition $\|a_k + b_k\| + \|x_k - y_k\| = 0$ in algorithms (12)–(16) implies that $x_k = y_k$ and $b_k = -a_k$, and as a consequence we have $(x_k, b_k) \in S_e(\partial h_1, \partial h_2)$. Hence, algorithms (12)–(16) can also be alternatively stated with the stopping criterion $\|a_k + b_k\| + \|x_k - y_k\| = 0$, and the following result holds.

Proposition 3.1 *The* PMM *is a special instance of algorithms* (12)–(16) *(with the stopping criterion* $\|a_k + b_k\| + \|x_k - y_k\| = 0$*) where*

$$\lambda_k = \mu_k = \lambda, \qquad \alpha_k = 1, \qquad (26)$$

and

$$x_k = z_{k-1} + \lambda w_{k-1} + \lambda(Cv_k - d), \qquad b_k = d - Cv_k \in \partial h_2(x_k),$$
$$y_k = x_k - \lambda(w_{k-1} - Mu_k), \qquad a_k = -Mu_k \in \partial h_1(y_k), \tag{27}$$

for every integer $k \geq 1$.

Proof First we notice that (26) implies

$$\frac{\lambda_k}{\mu_k} - \left(\frac{\alpha_k}{2}\right)^2 = \frac{\lambda}{\lambda} - \left(\frac{1}{2}\right)^2 = \frac{3}{4}, \tag{28}$$

for all integer $k \geq 1$. Next, applying Lemma 3.2 with $\theta = g$, $A = C$, $c = d$, $z = z_{k-1} + \lambda w_{k-1}$ and $\tilde{v} = v_k$ we have that x_k and b_k, defined as in (27), satisfy $b_k \in \partial h_2(x_k)$ and $x_k = (I + \lambda \partial h_2)^{-1}(z_{k-1} + \lambda w_{k-1})$. Therefore, the pair (x_k, b_k) satisfies the relations in (12) with $\lambda_k = \lambda$.

Similarly, applying Lemma 3.2 with $\theta = f$, $A = M$, $c = 0$, $z = x_k - \lambda w_{k-1}$ and $\tilde{v} = u_k$ we have that the points y_k and a_k, given in (27), satisfy (13) with $\mu_k = \lambda$, $\alpha_k = 1$ and x_k defined in (27).

Moreover, identities in (27) yield

$$b_k + a_k = d - Cv_k - Mu_k, \qquad x_k - y_k = \lambda(w_{k-1} - Mu_k), \tag{29}$$

and

$$z_{k-1} - y_k = \lambda(d - Mu_k - Cv_k). \tag{30}$$

Using (29), (30) and the definitions of x_k, b_k, y_k and a_k in (27), we can rewrite γ_k in step 2 of the PMM as in (19), which is exactly Eq. (14). Also, the identities in (29) imply that we can restate the stopping criterion in step 2 of the PMM as $\|a_k + b_k\| + \|x_k - y_k\| = 0$. Finally, (29) and the update rule in step 3 of the PMM imply that

$$z_k = z_{k-1} - \rho_k \gamma_k (a_k + b_k),$$
$$w_k = w_{k-1} - \rho_k \gamma_k (x_k - y_k).$$

Thus, the proposition is proven. $\qquad\square$

From Proposition 3.1 it follows that if for some k the stopping criterion in step 2 of the PMM holds, then

$$Mu_k + Cv_k - d = 0 \quad \text{and} \quad x_k - y_k = 0. \tag{31}$$

Furthermore, by the definitions of x_k and y_k in (27), and the optimality conditions of problems (24) and (25), we have

$$0 \in \partial g(v_k) + C^* x_k \quad \text{and} \quad 0 \in \partial f(u_k) + M^* y_k, \tag{32}$$

for all integer $k \geq 1$. Combining (31) with (32) we may conclude that if the PMM stops in step 2, then (u_k, v_k, x_k) satisfies the KKT conditions, and consequently it is a saddle point of L.

Otherwise, if the PMM generates an infinite sequence, in view of Proposition 3.1, we are able to establish its global convergence using the convergence results presented in [9].

Theorem 3.1 *Consider the sequences* $\{(u_k, v_k)\}$, $\{(z_k, w_k)\}$, $\{\gamma_k\}$ *and* $\{\rho_k\}$ *generated by the* PMM. *Consider also the sequences* $\{x_k\}$, $\{b_k\}$, $\{y_k\}$ *and* $\{a_k\}$ *defined in* (27). *Then, the following statements hold.*

(a) *There exist* z^* *a solution of the dual problem* (5) *and* $w^* \in \mathbb{R}^n$ *such that* $-w^* \in \partial h_1(z^*)$, $w^* \in \partial h_2(z^*)$, *and* $(x_k, b_k) \rightarrow (z^*, w^*)$, $(y_k, -a_k) \rightarrow (z^*, w^*)$ *and* $(z_k, w_k) \rightarrow (z^*, w^*)$.
(b) $Mu_k + Cv_k - d \rightarrow 0$ *and* $x_k - y_k \rightarrow 0$.
(c) $\lim_{k\to\infty} f(u_k) + g(v_k) = p^*$.

Proof (a) According to Proposition 3.1 the PMM is an instance of [9, Algorithm 2] applied to the subdifferential operators ∂h_1 and ∂h_2, and with generated sequences $\{(z_k, w_k)\}$, calculated by step 3 of the PMM, and $\{(x_k, b_k)\}$, $\{(y_k, a_k)\}$, which are defined in (27). From assumption A.1 and Eq. (28) it follows that the hypotheses of [9, Proposition 3] are satisfied. Thus, invoking this proposition we have that there exists $(z^*, w^*) \in S_e(\partial h_1, \partial h_2)$ such that

$$(z_k, w_k) \rightarrow (z^*, w^*), \qquad (x_k, b_k) \rightarrow (z^*, w^*) \qquad \text{and} \qquad (y_k, -a_k) \rightarrow (z^*, w^*). \tag{33}$$

Moreover, since $(z^*, w^*) \in S_e(\partial h_1, \partial h_2)$ we have that $-w^* \in \partial h_1(z^*)$, $w^* \in \partial h_2(z^*)$ and z^* is a solution of the dual problem (5).

(b) By (33) it trivially follows that $x_k - y_k \rightarrow 0$ and $a_k + b_k \rightarrow 0$. Hence, using the definition of a_k and b_k we deduce that $Mu_k + Cv_k - d \rightarrow 0$.

(c) Let (u^*, v^*, z^*) be a KKT point of L, which exists from hypothesis A.1, then from the first equality in (6) we have

$$L(u^*, v^*, z^*) \leq L(u_k, v_k, z^*), \qquad \text{for } k = 1, 2, \ldots.$$

From equation above, the definition of the Lagrangian function in (4) and the KKT conditions (7) it follows that

$$f(u^*) + g(v^*) \leq f(u_k) + g(v_k) + \langle Mu_k + Cv_k - d, z^* \rangle.$$

Since $p^* = f(u^*) + g(v^*)$, combining inequality above with item (b) we deduce that

$$p^* \leq \liminf_{k\to\infty} f(u_k) + g(v_k). \tag{34}$$

Now, we observe that the first inclusion in (32), together with Definition 1, implies

$$g(v^*) \geq g(v_k) - \langle C^* x_k, v^* - v_k \rangle.$$

Equivalently, from the second inclusion in (32) and Definition 1 it follows that

$$f(u^*) \geq f(u_k) - \langle M^* y_k, u^* - u_k \rangle.$$

Adding the two equations above we obtain

$$
\begin{aligned}
p^* &\geq f(u_k) + g(v_k) - \langle C^* x_k, v^* - v_k \rangle - \langle M^* y_k, u^* - u_k \rangle \\
&= f(u_k) + g(v_k) - \langle x_k, Cv^* - Cv_k \rangle - \langle y_k, Mu^* - Mu_k \rangle \\
&= f(u_k) + g(v_k) - \langle x_k - y_k, Cv^* - Cv_k \rangle - \langle y_k, d - Mu_k - Cv_k \rangle,
\end{aligned}
$$

where the last equality follows from a simple manipulation and the equality in (7). Since $\{b_k = d - Cv_k\}$ and $\{y_k\}$ are convergent sequences, therefore bounded sequences, equation above, together with item (b), yields

$$p^* \geq \limsup_{k \to \infty} f(u_k) + g(v_k).$$

Combining inequality above with (34) we conclude the proof.

$\square$

4 Complexity results

Our goal in this section is to study the iteration complexity of the PMM for solving problem (1). In order to develop global convergence bounds for the method we will examine how well its iterates satisfy the KKT conditions. Observe that the inclusions in (32) indicate that the quantities $\|Mu_k + Cv_k - d\|$ and $\|x_k - y_k\|$ can be used to measure the accuracy of an iterate (u_k, v_k, x_k) to a saddle point of the Lagrangian function. More specifically, if we define the primal and dual residuals, associated with (u_k, v_k, x_k), by

$$
\begin{aligned}
r_k^p &:= Mu_k + Cv_k - d, \\
r_k^d &:= x_k - y_k;
\end{aligned}
$$

then, from the inclusions in (32) and the KKT conditions it follows that when $\|r_k^p\| = \|r_k^d\| = 0$, the triplet (u_k, v_k, x_k) is a saddle point of L. Therefore, the size of these residuals indicates how far the iterates are from a saddle point, and it can be viewed as an error measurement of the PMM. It is thus reasonable to seek upper bounds for these quantities for the purpose of investigating the convergence rate of the PMM.

The theorem below estimates the quality of the best iterate among $(u_1, v_1, x_1), \ldots, (u_k, v_k, x_k)$, in terms of the error measurement given by the primal and dual residuals. We refer to these estimates as *pointwise* complexity bounds for the PMM.

Theorem 4.1 *Consider the sequences $\{(u_k, v_k)\}$, $\{(z_k, w_k)\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ generated by the PMM. Consider also the sequences $\{x_k\}$, $\{b_k\}$, $\{y_k\}$ and $\{a_k\}$ defined in (27). If d_0 is the distance of (z_0, w_0) to the set $S_e(\partial h_1, \partial h_2)$, then for all $k = 1, 2, \ldots$, we have*

$$0 \in \partial g(v_k) + C^* x_k, \qquad 0 \in \partial f(u_k) + M^* y_k, \qquad (35)$$

and there exists and index $1 \leq i \leq k$ such that

$$\|Mu_i + Cv_i - d\| \leq \frac{2d_0}{(1 - \overline{\rho})\tau\sqrt{k}}, \qquad \|x_i - y_i\| \leq \frac{2d_0}{(1 - \overline{\rho})\tau\sqrt{k}}; \qquad (36)$$

where $\tau := \min\left\{\lambda, \dfrac{1}{\lambda}\right\}$.

Proof Inclusions (35) were established in (32). Therefore, what is left is to show the bounds in (36). Since the point $(z_k, w_k) = (z_{k-1}, w_{k-1}) + \rho_k((z_{k-1}^*, w_{k-1}^*) - (z_{k-1}, w_{k-1}))$, where (z_{k-1}^*, w_{k-1}^*) is the orthogonal projection of (z_{k-1}, w_{k-1}) onto H_{ϕ_k}, and $S_e(\partial h_1, \partial h_2) \subseteq H_{\phi_k}$, we take an arbitrary $(z^*, w^*) \in S_e(\partial h_1, \partial h_2)$ and observe that for $k = 1, 2, \ldots,$

$$
\begin{aligned}
\left\|(z_k, w_k) - (z^*, w^*)\right\|^2 &= \left\|(z_{k-1}, w_{k-1}) - (z^*, w^*)\right\|^2 \\
&+ \left\|\rho_k((z_{k-1}^*, w_{k-1}^*) - (z_{k-1}, w_{k-1}))\right\|^2 \\
&+ 2\rho_k \left\langle (z_{k-1}^*, w_{k-1}^*) - (z_{k-1}, w_{k-1}), (z_{k-1}, w_{k-1}) - (z^*, w^*)\right\rangle \\
&\leq \left\|(z_{k-1}, w_{k-1}) - (z^*, w^*)\right\|^2 \\
&+ \left(1 - \frac{2}{\rho_k}\right)\|(z_k, w_k) - (z_{k-1}, w_{k-1})\|^2 \\
&= \left\|(z_{k-1}, w_{k-1}) - (z^*, w^*)\right\|^2 \\
&- \rho_k(2 - \rho_k)\gamma_k^2 \left\|(Mu_k + Cv_k - d, \lambda(w_{k-1} - Mu_k))\right\|^2,
\end{aligned}
$$

where the inequality above follows from a well-known property of the orthogonal projection and the fact that $\rho_k((z_{k-1}^*, w_{k-1}^*) - (z_{k-1}, w_{k-1})) = (z_k, w_k) - (z_{k-1}, w_{k-1})$. Thus, applying the inequality above recursively, we have

$$
\begin{aligned}
\left\|(z_k, w_k) - (z^*, w^*)\right\|^2 &\leq \left\|(z_0, w_0) - (z^*, w^*)\right\|^2 \\
&- \sum_{j=1}^{k} \rho_j(2 - \rho_j)\gamma_j^2 \left\|(Mu_j + Cv_j - d, \lambda(w_{j-1} - Mu_j))\right\|^2.
\end{aligned}
\qquad (37)
$$

We rearrange terms in the equation above and notice that $\lambda(w_{j-1} - Mu_j) = x_j - y_j$, which yields

$$
\begin{aligned}
\sum_{j=1}^{k} \rho_j(2 - \rho_j)\gamma_j^2 &\left\|(Mu_j + Cv_j - d, x_j - y_j)\right\|^2 \\
&\leq \left\|(z_0, w_0) - (z^*, w^*)\right\|^2 - \left\|(z_k, w_k) - (z^*, w^*)\right\|^2 \\
&\leq \left\|(z_0, w_0) - (z^*, w^*)\right\|^2.
\end{aligned}
\qquad (38)
$$

Taking (z^*, w^*) to be the orthogonal projection of (z_0, w_0) onto $S_e(\partial h_1, \partial h_2)$ in inequality (38), we obtain

$$\sum_{j=1}^{k} \rho_j (2 - \rho_j) \gamma_j^2 \left\| (Mu_j + Cv_j - d, x_j - y_j) \right\|^2 \le d_0^2. \tag{39}$$

Now, for i such that

$$i \in \arg \min_{j=1,\dots,k} \left(\left\| (Mu_j + Cv_j - d, x_j - y_j) \right\|^2 \right),$$

we use inequality (39) and the fact that $\rho_j \in [1 - \overline{\rho}, 1 + \overline{\rho}]$ to conclude that

$$\| Mu_i + Cv_i - d \|^2 + \| x_i - y_i \|^2 \le \frac{d_0^2}{(1 - \overline{\rho})^2 \sum_{j=1}^{k} \gamma_j^2}. \tag{40}$$

Next, we notice that Proposition 3.1, together with the equality in (19), implies

$$\gamma_j = \frac{\phi_j(z_{j-1}, w_{j-1})}{\left\| \nabla \phi_j \right\|^2}, \qquad \text{for } j = 1, \dots, k, \tag{41}$$

where ϕ_j is the affine function given in (17) associated with x_j, y_j, b_j and a_j defined in (27). Moreover, combining Eqs. (17), (27), (29) and (30) we have

$$\begin{aligned}
\phi_j(z_{j-1}, w_{j-1}) &= \lambda \left\| Cv_j - d + w_{j-1} \right\|^2 + \lambda \langle d - Cv_j - Mu_j, w_{j-1} - Mu_j \rangle \\
&= \frac{\lambda}{2} \left\| Cv_j - d + w_{j-1} \right\|^2 \\
&\quad + \frac{\lambda}{2} \left(\left\| d - Cv_j - Mu_j \right\|^2 + \left\| w_{j-1} - Mu_j \right\|^2 \right).
\end{aligned}$$

Hence, we substitute the relation above into (41) to obtain

$$\begin{aligned}
\gamma_j &= \frac{\lambda \left\| Cv_j - d + w_{j-1} \right\|^2}{2 \left\| \nabla \phi_j \right\|^2} + \frac{\lambda \left\| d - Cv_j - Mu_j \right\|^2 + \lambda \left\| w_{j-1} - Mu_j \right\|^2}{2 \left\| \nabla \phi_j \right\|^2} \\
&\ge \frac{\lambda \left\| d - Cv_j - Mu_j \right\|^2 + \lambda \left\| w_{j-1} - Mu_j \right\|^2}{2 \left\| \nabla \phi_j \right\|^2}.
\end{aligned} \tag{42}$$

Now, we use the following estimate

$$\begin{aligned}
&\lambda \left\| d - Cv_j - Mu_j \right\|^2 + \lambda \left\| w_{j-1} - Mu_j \right\|^2 \\
&= \lambda \left\| d - Cv_j - Mu_j \right\|^2 + \frac{1}{\lambda} \lambda^2 \left\| w_{j-1} - Mu_j \right\|^2
\end{aligned}$$

$$\geq \tau(\|d - Cv_j - Mu_j\|^2 + \lambda^2 \|w_{j-1} - Mu_j\|^2)$$
$$= \tau \|\nabla\phi_j\|^2,$$

and the inequality in (42) to deduce that

$$\gamma_j \geq \frac{\tau}{2}, \qquad \text{for } j = 1, \ldots, k. \tag{43}$$

This last inequality, together with (40), implies

$$\|Mu_i + Cv_i - d\|^2 + \|x_i - y_i\|^2 \leq \frac{4d_0^2}{(1 - \overline{\rho})^2 \tau^2 k},$$

from which the theorem follows. □

We now develop alternative complexity bounds for the PMM, which we call *ergodic* complexity bounds. We define a sequence of ergodic iterates as weighted averages of the iterates and derive a convergence rate for the PMM, which as before, is obtained from estimates of the residuals for the KKT conditions associated with these ergodic sequences.

The idea of considering averages of the iterates in the analysis of the convergence rate for methods for solving problem (1) has been already used in other works. For instance, in [15,19] it was shown a worst-case $\mathcal{O}(1/k)$ convergence rate for the ADMM in the ergodic sense.

The sequences of ergodic means $\{\overline{u}_k\}$, $\{\overline{v}_k\}$, $\{\overline{x}_k\}$ and $\{\overline{y}_k\}$ associated with $\{u_k\}$, $\{v_k\}$, $\{x_k\}$ and $\{y_k\}$, respectively, are defined as

$$\overline{u}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j u_j, \qquad \overline{v}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j v_j,$$

$$\overline{x}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j x_j, \qquad \overline{y}_k := \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j y_j, \qquad \text{where } \Gamma_k := \sum_{j=1}^{k}\rho_j\gamma_j. \tag{44}$$

Lemma 4.1 *For all integer $k \geq 1$ define*

$$\overline{\epsilon}_k^u := \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j \langle u_j - \overline{u}_k, -M^*y_j\rangle,$$

$$\overline{\epsilon}_k^v := \frac{1}{\Gamma_k}\sum_{j=1}^{k}\rho_j\gamma_j \langle v_j - \overline{v}_k, -C^*x_j\rangle. \tag{45}$$

Then, $\overline{\epsilon}_k^v \geq 0$, $\overline{\epsilon}_k^u \geq 0$ and

$$0 \in \partial_{\overline{\epsilon}_k^v} g(\overline{v}_k) + C^*\overline{x}_k, \qquad 0 \in \partial_{\overline{\epsilon}_k^u} f(\overline{u}_k) + M^*\overline{y}_k. \tag{46}$$

Proof From inclusions in (35) we have

$$-C^* x_k \in \partial g(v_k) \qquad \text{and} \qquad -M^* y_k \in \partial f(u_k).$$

Thus, the assertion that $\bar{\epsilon}_k^v \geq 0$ and the first inclusion in (46) are a direct consequence of the first inclusion in the equation above, the definitions of $\bar{x}_k$, $\bar{v}_k$ and $\bar{\epsilon}_k^v$, the fact that C^* is a linear operator and Proposition 2.1(d).

Similarly, the second inclusion in (46) and the fact that $\bar{\epsilon}_k^u \geq 0$ follow from the definitions of $\bar{y}_k$, $\bar{u}_k$ and $\bar{\epsilon}_k^u$, linearity of the M^* operator, the second inclusion in the relation above and Proposition 2.1(d). □

According to Lemma 4.1, if $\|\bar{r}_k^p\| = \|\bar{r}_k^d\| = 0$ and $\bar{\epsilon}_k^u = \bar{\epsilon}_k^v = 0$, where $\bar{r}_k^p = M\bar{u}_k + C\bar{v}_k - d$ and $\bar{r}_k^d = \bar{x}_k - \bar{y}_k$; then it follows that $(\bar{u}_k, \bar{v}_k, \bar{x}_k)$ satisfies the KKT conditions and, consequently, it is a saddle point of the Lagrangian function. Thus, we have computable residuals for the sequence of ergodic means, i.e. the residual vector $(\bar{r}_k^p, \bar{r}_k^d, \bar{\epsilon}_k^u, \bar{\epsilon}_k^v)$, and we can attempt to construct bounds on its size.

For this purpose, we first prove the following technical result. It establishes an estimate for the quantity $\bar{\epsilon}_k^u + \bar{\epsilon}_k^v$.

Lemma 4.2 *Let $\{u_k\}$, $\{v_k\}$, $\{z_k\}$, $\{w_k\}$, $\{\gamma_k\}$ and $\{\rho_k\}$ be the sequences generated by the PMM and $\{x_k\}$, $\{y_k\}$ be defined in (27). Define also the sequences of ergodic iterates $\{\bar{u}_k\}$, $\{\bar{v}_k\}$, $\{\bar{x}_k\}$, $\{\bar{y}_k\}$, $\{\bar{\epsilon}_k^u\}$ and $\{\bar{\epsilon}_k^v\}$ as in (44) and (45). Then, for every integer $k \geq 1$, we have*

$$\bar{\epsilon}_k^u + \bar{\epsilon}_k^v$$
$$\leq \frac{1}{\Gamma_k} \left[\frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \left(\lambda^2 \|Mu_j + Cv_j - d\|^2 + \|d - Cv_j - w_{j-1}\|^2 \right) \right.$$
$$\left. + 4d_0^2 \right]. \tag{47}$$

Proof We first show that

$$\bar{\epsilon}_k^u + \bar{\epsilon}_k^v = -\frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \phi_j(\bar{y}_k, d - C\bar{v}_k). \tag{48}$$

By the definitions of ϕ_j, b_j and a_j we have

$$\phi_j(\bar{y}_k, d - C\bar{v}_k) = \langle \bar{y}_k - x_j, C\bar{v}_k - Cv_j \rangle + \langle \bar{y}_k - y_j, d - C\bar{v}_k - Mu_j \rangle$$
$$= -\langle \bar{y}_k, Cv_j \rangle - \langle x_j, C\bar{v}_k - Cv_j \rangle$$
$$+ \langle \bar{y}_k - y_j, d \rangle + \langle y_j, C\bar{v}_k \rangle - \langle \bar{y}_k - y_j, Mu_j \rangle. \tag{49}$$

We use the definitions of $\overline{y}_k$, $\overline{v}_k$, Γ_k and the fact that C is a linear map, to obtain

$$\frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \left(\langle \overline{y}_k - y_j, d \rangle - \langle \overline{y}_k, Cv_j \rangle + \langle y_j, C\overline{v}_k \rangle \right)$$
$$= \langle \overline{y}_k - \overline{y}_k, d \rangle - \langle \overline{y}_k, C\overline{v}_k \rangle + \langle \overline{y}_k, C\overline{v}_k \rangle = 0.$$

Now, multiplying (49) by $\rho_j \gamma_j / \Gamma_k$, adding from $j = 1$ to k and combining with the relation above, we conclude that

$$\frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \phi_j (\overline{y}_k, d - C\overline{v}_k)$$
$$= -\frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \left(\langle x_j, C\overline{v}_k - Cv_j \rangle + \langle \overline{y}_k - y_j, Mu_j \rangle \right). \tag{50}$$

Next, we observe that

$$\overline{\epsilon}_k^u + \overline{\epsilon}_k^v = \frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \left(\langle M\overline{u}_k - Mu_j, y_j \rangle + \langle C\overline{v}_k - Cv_j, x_j \rangle \right)$$
$$= \frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \left(\langle Mu_j, \overline{y}_k - y_j \rangle + \langle C\overline{v}_k - Cv_j, x_j \rangle \right),$$

where the last equality above is a consequence of the definitions of $\overline{y}_k$ and $M\overline{u}_k$. We deduce formula (48) combining the equation above with (50).

For an arbitrary $(z, w) \in \mathbb{R}^n \times \mathbb{R}^n$ and all integer $j \geq 1$ we have

$$\frac{1}{2} \left\| (z, w) - (z_{j-1}, w_{j-1}) \right\|^2$$
$$= \frac{1}{2} \left\| (z, w) - (z_j, w_j) \right\|^2 + \langle (z, w) - (z_j, w_j), (z_j, w_j) - (z_{j-1}, w_{j-1}) \rangle$$
$$+ \frac{1}{2} \left\| (z_j, w_j) - (z_{j-1}, w_{j-1}) \right\|^2$$
$$= \frac{1}{2} \left\| (z, w) - (z_j, w_j) \right\|^2 - \rho_j \gamma_j \langle (z, w) - (z_j, w_j), \nabla \phi_j \rangle + \frac{1}{2} \rho_j^2 \gamma_j^2 \left\| \nabla \phi_j \right\|^2, \tag{51}$$

where the second equality follows from the identity $(z_j, w_j) = (z_{j-1}, w_{j-1}) - \rho_j \gamma_j \nabla \phi_j$, which is a consequence of step 3 in the PMM, (29) and (18). Now, we notice that

$$\langle (z, w) - (z_j, w_j), \nabla \phi_j \rangle$$
$$= \langle (z, w) - (y_j, d - Cv_j), \nabla \phi_j \rangle + \langle (y_j, d - Cv_j) - (z_j, w_j), \nabla \phi_j \rangle$$

$$
\begin{aligned}
&= \phi_j(z, w) - \phi_j(z_j, w_j) \\
&= \phi_j(z, w) - \phi_j((z_{j-1}, w_{j-1}) - \rho_j \gamma_j \nabla \phi_j) \\
&= \phi_j(z, w) - \phi_j(z_{j-1}, w_{j-1}) + \rho_j \gamma_j \left\| \nabla \phi_j \right\|^2,
\end{aligned}
$$

where the second and fourth equalities are due to (17), (18) and (27). Substituting the equation above into (51) yields

$$
\begin{aligned}
&\frac{1}{2} \left\| (z, w) - (z_{j-1}, w_{j-1}) \right\|^2 \\
&= \frac{1}{2} \left\| (z, w) - (z_j, w_j) \right\|^2 - \rho_j \gamma_j \phi_j(z, w) + \rho_j \gamma_j \phi_j(z_{j-1}, w_{j-1}) \\
&\quad - \rho_j^2 \gamma_j^2 \left\| \nabla \phi_j \right\|^2 + \frac{1}{2} \rho_j^2 \gamma_j^2 \left\| \nabla \phi_j \right\|^2 \\
&= \frac{1}{2} \left\| (z, w) - (z_j, w_j) \right\|^2 - \rho_j \gamma_j \phi_j(z, w) + \rho_j \gamma_j^2 \left\| \nabla \phi_j \right\|^2 - \frac{1}{2} \rho_j^2 \gamma_j^2 \left\| \nabla \phi_j \right\|^2,
\end{aligned}
$$

where formula (19) is used for obtaining the last equality. Rearranging terms in the equation above and adding from $j = 1$ to k, we obtain

$$
\begin{aligned}
-\sum_{j=1}^{k} \rho_j \gamma_j \phi_j(z, w) = &\frac{1}{2} \left\| (z, w) - (z_0, w_0) \right\|^2 - \frac{1}{2} \left\| (z, w) - (z_k, w_k) \right\|^2 \\
&- \sum_{j=1}^{k} \frac{1}{2} \rho_j (2 - \rho_j) \gamma_j^2 \left\| \nabla \phi_j \right\|^2.
\end{aligned}
$$

Consequently, we have

$$
-\sum_{j=1}^{k} \rho_j \gamma_j \phi_j(z, w) \le \frac{1}{2} \left\| (z, w) - (z_0, w_0) \right\|^2, \qquad \forall (z, w) \in \mathbb{R}^n \times \mathbb{R}^n.
$$

Now, we use inequality above with $(z, w) = (\overline{y}_k, d - C\overline{v}_k)$, and combine with (48), to obtain

$$
\begin{aligned}
\overline{\epsilon}_k^u + \overline{\epsilon}_k^v &\le \frac{1}{2\Gamma_k} \left\| (\overline{y}_k, d - C\overline{v}_k) - (z_0, w_0) \right\|^2 \\
&\le \frac{1}{2\Gamma_k} \sum_{j=1}^{k} \frac{\rho_j \gamma_j}{\Gamma_k} \left\| (y_j, d - Cv_j) - (z_0, w_0) \right\|^2 \\
&\le \frac{1}{\Gamma_k} \sum_{j=1}^{k} \frac{\rho_j \gamma_j}{\Gamma_k} \left(\left\| (y_j, d - Cv_j) - (z_{j-1}, w_{j-1}) \right\|^2 \right. \\
&\quad \left. + \left\| (z_{j-1}, w_{j-1}) - (z_0, w_0) \right\|^2 \right),
\end{aligned}
\qquad (52)
$$

where the second inequality above is due to the definitions of $\overline{y}_k$, $\overline{v}_k$, the fact that C is a linear operator and the convexity of $\|\cdot\|^2$. Further, the third inequality in equation above is obtained using the triangle inequality for norms.

Next, we notice that inequality (37) implies

$$\left\| (z_j, w_j) - (z^*, w^*) \right\| \leq \left\| (z^*, w^*) - (z_0, w_0) \right\|,$$

for all integers $j \geq 0$ and all $(z^*, w^*) \in S_e(\partial h_1, \partial h_2)$. Taking (z^*, w^*) to be the orthogonal projection of (z_0, w_0) onto $S_e(\partial h_1, \partial h_2)$ in the relation above and using the triangle inequality, we deduce that

$$\left\| (z_j, w_j) - (z_0, w_0) \right\| \leq \left\| (z_j, w_j) - (z^*, w^*) \right\| + \left\| (z^*, w^*) - (z_0, w_0) \right\| \leq 2d_0. \tag{53}$$

Combining (53) with (52) we have

$$\overline{\epsilon}_k^u + \overline{\epsilon}_k^v \leq \frac{1}{\Gamma_k} \left[\frac{1}{\Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j \left\| (y_j, d - Cv_j) - (z_{j-1}, w_{j-1}) \right\|^2 + 4d_0^2 \right].$$

To end the proof we substitute the identity $y_j - z_{j-1} = \lambda(Mu_j + Cv_j - d)$, which follows from the definition of y_j in (27), into the above inequality. $\qquad\square$

The following theorem provides estimates for the quality of the measure of the ergodic means $\overline{u}_k$, $\overline{v}_k$, $\overline{x}_k$ and $\overline{y}_k$. More specifically, we show that the residuals associated with the ergodic sequences are $\mathcal{O}(1/k)$.

Theorem 4.2 *Assume the hypotheses of* Theorem 4.1. *Consider also the sequences* $\{\overline{u}_k\}$, $\{\overline{v}_k\}$, $\{\overline{x}_k\}$ *and* $\{\overline{y}_k\}$ *given in* (44), *and* $\{\overline{\epsilon}_k^u\}$, $\{\overline{\epsilon}_k^v\}$ *defined in* (45). *Then, for all integer* $k \geq 1$, *we have*

$$0 \in \partial_{\overline{\epsilon}_k^v} g(\overline{v}_k) + C^* \overline{x}_k, \qquad 0 \in \partial_{\overline{\epsilon}_k^u} f(\overline{u}_k) + M^* \overline{y}_k, \tag{54}$$

and

$$\| M\overline{u}_k + C\overline{v}_k - d \| \leq \frac{4d_0}{k(1 - \overline{\rho})\tau}, \qquad \left\| \overline{x}_k - \overline{y}_k \right\| \leq \frac{4d_0}{k(1 - \overline{\rho})\tau}, \tag{55}$$

$$\overline{\epsilon}_k^u + \overline{\epsilon}_k^v \leq \frac{8d_0^2 \vartheta}{k(1 - \overline{\rho})\tau}; \tag{56}$$

where $\vartheta := \dfrac{1}{\tau^2(1 - \overline{\rho})^2} + 1.$

Proof The inclusions in (54) were proven in Lemma 4.1. To prove the estimates in (55) we first observe that, since

$$x_k - y_k = \lambda(w_{k-1} - Mu_k) \qquad \text{for } k = 1, 2, \ldots,$$

by the update rule in step 3 of the PMM we have

$$(z_k, w_k) = (z_{k-1}, w_{k-1}) - \rho_k \gamma_k (d - Cv_k - Mu_k, x_k - y_k)$$

$$= (z_0, w_0) - \sum_{j=1}^{k} \rho_j \gamma_j (d - Cv_j - Mu_j, x_j - y_j)$$

$$= (z_0, w_0) - \Gamma_k (d - C\overline{v}_k - M\overline{u}_k, \overline{x}_k - \overline{y}_k), \tag{57}$$

where the last equality above follows from the definitions of $\Gamma_k, \overline{v}_k, \overline{u}_k, \overline{x}_k$ and $\overline{y}_k$ in (44), and the fact that M and C are linear operators. Therefore, from (57) we deduce that

$$\left\| (d - C\overline{v}_k - M\overline{u}_k, \overline{x}_k - \overline{y}_k) \right\| = \frac{1}{\Gamma_k} \left\| (z_0, w_0) - (z_k, w_k) \right\|,$$

and combining the identity above with estimate (53) we obtain

$$\left\| (d - C\overline{v}_k - M\overline{u}_k, \overline{x}_k - \overline{y}_k) \right\| \leq \frac{2d_0}{\Gamma_k}. \tag{58}$$

Next, we notice that Eq. (43) and the fact that $\rho_j \in [1 - \overline{\rho}, 1 + \overline{\rho}]$ imply

$$\Gamma_k = \sum_{j=1}^{k} \rho_j \gamma_j \geq \sum_{j=1}^{k} (1 - \overline{\rho}) \frac{\tau}{2} = (1 - \overline{\rho}) \frac{\tau}{2} k. \tag{59}$$

The inequality above, together with (58), yields

$$\left\| (d - C\overline{v}_k - M\overline{u}_k, \overline{x}_k - \overline{y}_k) \right\| \leq \frac{4d_0}{(1 - \overline{\rho})\tau k},$$

from which the bounds in (55) follow directly.

Now, using the equality in (42) we have

$$\gamma_j \geq \frac{\lambda \left\| Cv_j - d + w_{j-1} \right\|^2}{2 \left\| \nabla \phi_j \right\|^2} + \frac{\lambda \left\| d - Cv_j - Mu_j \right\|^2}{2 \left\| \nabla \phi_j \right\|^2},$$

and as a consequence we obtain

$$\left\| \nabla \phi_j \right\|^2 \gamma_j$$
$$\geq \frac{\tau}{2} \left(\left\| Cv_j - d + w_{j-1} \right\|^2 + \lambda^2 \left\| d - Cv_j - Mu_j \right\|^2 \right), \quad \text{for } j = 1, \dots, k.$$

Multiplying the inequality above by $\rho_j \gamma_j 2/\tau$, adding from $j = 1$ to k and using (47), we have

$$\overline{\epsilon}_k^u + \overline{\epsilon}_k^v \le \frac{1}{\Gamma_k} \left[\frac{2}{\tau \Gamma_k} \sum_{j=1}^{k} \rho_j \gamma_j^2 \left\| \nabla \phi_j \right\|^2 + 4d_0^2 \right]$$

$$= \frac{1}{\Gamma_k} \left[\frac{2}{\tau \Gamma_k} \sum_{j=1}^{k} \frac{1}{2 - \rho_j} \rho_j (2 - \rho_j) \gamma_j^2 \left\| \nabla \phi_j \right\|^2 + 4d_0^2 \right].$$

Finally, relation above, together with (39) and the fact that $\rho_j \in [1 - \overline{\rho}, 1 + \overline{\rho}]$, yields

$$\overline{\epsilon}_k^u + \overline{\epsilon}_k^v \le \frac{1}{\Gamma_k} \left[\frac{2}{\tau \Gamma_k (1 - \overline{\rho})} d_0^2 + 4d_0^2 \right].$$

The bound in (56) is achieved using this last inequality and (59). □

5 Applications

This section describes some simple, preliminary computational testing of the PMM. These experiments are not meant to be exhaustive, but only to provide some preliminary indication of the performance of the PMM.

Specifically, we discuss the specialization of the PMM to two common test problems. First, we consider the total variation model for image denoising (TV denoising). Then, we consider a compressed sensing problem for Magnetic Resonance Imaging, and we exhibit some numerical experiments to illustrate the performance of the PMM when solving these problems.

5.1 TV denoising

Total variation (TV) or ROF model is a common image model developed by Rudin et al. [24] for the problem of removing noise from an image. If $b \in \mathbb{R}^{m \times n}$ is an observed noisy image, the TV problem for image denoising estimates the unknown original image $u \in \mathbb{R}^{m \times n}$ by solving the minimization problem

$$\min_{u \in \mathbb{R}^{m \times n}} \zeta TV(u) + \frac{1}{2} \|u - b\|_F^2 , \tag{60}$$

where TV is the total variation norm defined as

$$TV(u) := \|\nabla_1 u\|_1 + \|\nabla_2 u\|_1 . \tag{61}$$

Here $\nabla_1 : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ and $\nabla_2 : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ are the discrete forward gradients in the first and second direction, respectively, given by

$$(\nabla_1 u)_{ij} = u_{i+1,j} - u_{i,j}, \qquad (\nabla_2 u)_{ij} = u_{i,j+1} - u_{i,j},$$
$$i = 1, \ldots, m, \ j = 1, \ldots, n, \ u \in \mathbb{R}^{m \times n};$$

and we assume standard reflexive boundary conditions

$$u_{m+1,j} - u_{m,j} = 0, \quad j = 1, \ldots, n \quad \text{and} \quad u_{i,n+1} - u_{i,n} = 0, \quad i = 1, \ldots, m.$$

The regularization parameter $\zeta > 0$ controls the tradeoff between fidelity to measurements and the smoothness term given by the total variation.

To solve the TV problem using the PMM we first have to sate it in the form of a linearly constrained minimization problem (1). If we define $\Omega := \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n}$, and the linear map $\nabla : \mathbb{R}^{m \times n} \to \Omega$ by

$$\nabla u := (\nabla_1 u, \nabla_2 u);$$

then, taking $v = \nabla u \in \Omega$, we have that (60) is equivalent to the optimization problem

$$\min_{(u,v) \in \mathbb{R}^{m \times n} \times \Omega} \left\{ \zeta \|v\|_1 + \frac{1}{2} \|u - b\|_F^2 : \nabla u - v = 0 \right\}. \tag{62}$$

Now, we solve (62) by applying the PMM with $f(u) = \frac{1}{2} \|u - b\|_F^2$, $g(v) = \zeta \|v\|_1$, $M = \nabla$, $C = -I$ and $d = 0$.

Given $z_{k-1}, w_{k-1} \in \Omega$, the PMM requires the solution of problems,

$$v_k = \arg \min_{v \in \Omega} \zeta \|v\|_1 - \langle z_{k-1} + \lambda w_{k-1}, v \rangle + \frac{\lambda}{2} \|v\|_F^2, \tag{63}$$

and

$$u_k = \arg \min_{u \in \mathbb{R}^{m \times n}} \frac{1}{2} \|u - b\|_F^2 + \langle z_{k-1} - \lambda v_k, \nabla u \rangle + \frac{\lambda}{2} \|\nabla u\|_F^2. \tag{64}$$

The optimality condition of problem (63), yields

$$0 \in \zeta \partial \|\cdot\|_1 (v_k) - (z_{k-1} + \lambda w_{k-1}) + \lambda v_k;$$

hence,

$$v_k = \left(I + \frac{\zeta}{\lambda} \partial \|\cdot\|_1 \right)^{-1} \left(\frac{1}{\lambda} z_{k-1} + w_{k-1} \right).$$

Therefore, the solution of problem (63) can be computed explicitly as

$$v_k = \mathbf{shrink} \left(\frac{1}{\lambda} z_{k-1} + w_{k-1}, \frac{\zeta}{\lambda} \right),$$

where the **shrink** operator is defined in (3). Deriving the optimality condition for problem (64) we have that

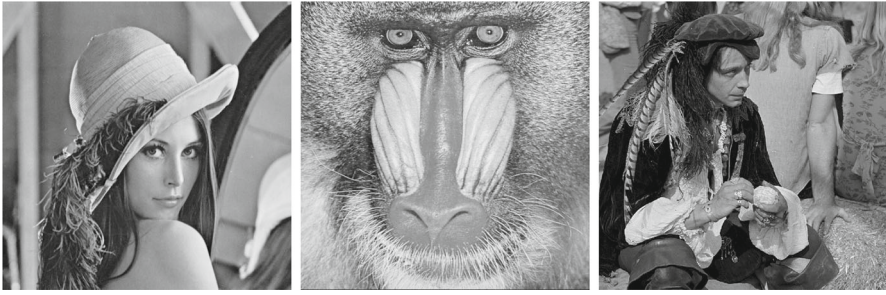$$0 = u_k - b + \nabla^* (z_{k-1} - \lambda v_k) + \lambda \nabla^* \nabla u_k,$$

Fig. 1 Test images: Lena (left), Baboon (center) and Man (right)

from which it follows that u_k has to be the solution of the system of linear equations

$$(I + \lambda \nabla^* \nabla)u_k = b - \nabla^*(z_{k-1} - \lambda v_k).$$

Thus, the PMM applied to problem (62) produces the iteration:

$$v_k = \textbf{shrink}\left(\frac{1}{\lambda}z_{k-1} + w_{k-1}, \frac{\zeta}{\lambda}\right), \tag{65}$$

$$(I + \lambda \nabla^* \nabla)u_k = b - \nabla^*(z_{k-1} - \lambda v_k), \tag{66}$$

$$\gamma_k = \frac{\lambda \|w_{k-1} - v_k\|^2 + \lambda \langle v_k - \nabla u_k, w_{k-1} - \nabla u_k \rangle}{\|\nabla u_k - v_k\|^2 + \lambda^2 \|\nabla u_k - w_{k-1}\|^2}, \tag{67}$$

$$z_k = z_{k-1} + \rho_k \gamma_k (\nabla u_k - v_k), \tag{68}$$

$$w_k = w_{k-1} - \rho_k \gamma_k \lambda (w_{k-1} - \nabla u_k). \tag{69}$$

We used three images to test the PMM in our experiments: the first was "Lena" image of size 512×512, the second was "Baboon" image of size 512×512, and the third was "Man" image of size 768×768, see Fig. 1. All images were contaminated with Gaussian noise using the Matlab function "`imnoise`" and it was used several values of variance (σ). The PMM was implemented in Matlab code and it was chosen $\lambda = 1$ in all tests, since we have found that choosing this value for λ was effective for all the experiments. The variables z_0 and w_0 were initialized to be zero.

As a way to provide a reference, we also report the results obtained with ADMM, which is actually equivalent to the Split Bregman (SB) method [10,14] for TV regularized problems. For a fair comparison, we implemented the generalized ADMM [7] with over and under relaxation factors, see also [6]. For the ADMM, λ was also set to be 1 in all iterations. Moreover, we took as initial guess the contaminated image and the multiplier was initialized to be zero. As in [14] iterations for both method were terminated when condition $\|u_k - u_{k-1}\| / \|u_k\| \leq 10^{-3}$ was met; since this stopping criterion is satisfied faster than the stopping condition given by the KKT residuals, while yielding good denoised images.

In Fig. 2 we present some denoising results. It shows the noise contaminated images and the reconstructed images with the PMM. Additionally, in Fig. 3 we report the
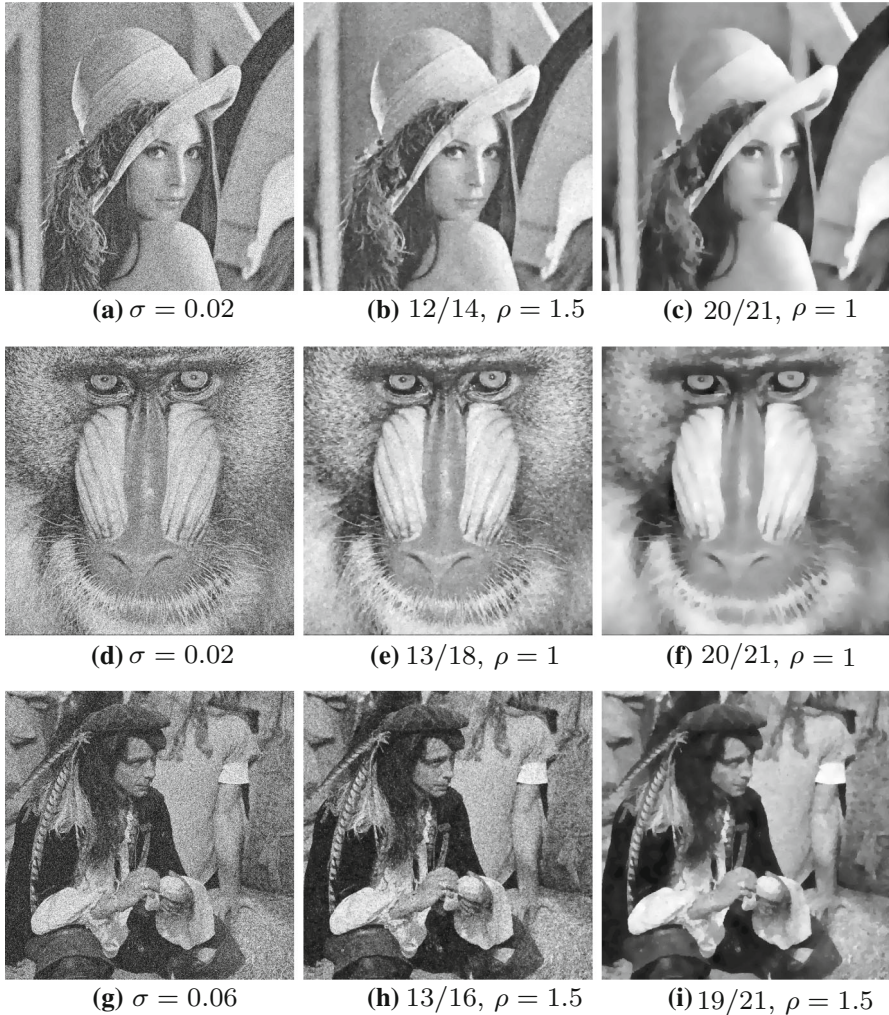
Fig. 2 Denoising results. Images **a**, **d** and **g** are contaminated with Gaussian noise, the value of variance (σ) is reported below each image. Images were denoised with $\zeta = 20$ (center) and $\zeta = 50$ (right). The value of ρ and the number of iterations required to satisfy the stopping criterion, for both the PMM/ADMM, are listed below each image

primal and dual residuals for the KKT optimality conditions for problem (62) for both methods, in some specific tests. The primal and dual residuals for the PMM were defined in Sect. 4. For the ADMM the primal residual is also defined as $\nabla u_k - v_k$, i.e. it is the residual for the equality constraint at iteration k. The dual residual for the ADMM is defined as the residual for the dual feasibility condition (see Eq. (7) and the comments below). Since the exact solution of the problems are known, we also plotted in Fig. 3 the error $\|u_k - u^*\|$ versus iteration, where u^* is the exact solution. In these experiments both methods were stopped at iteration 50. It can be observed
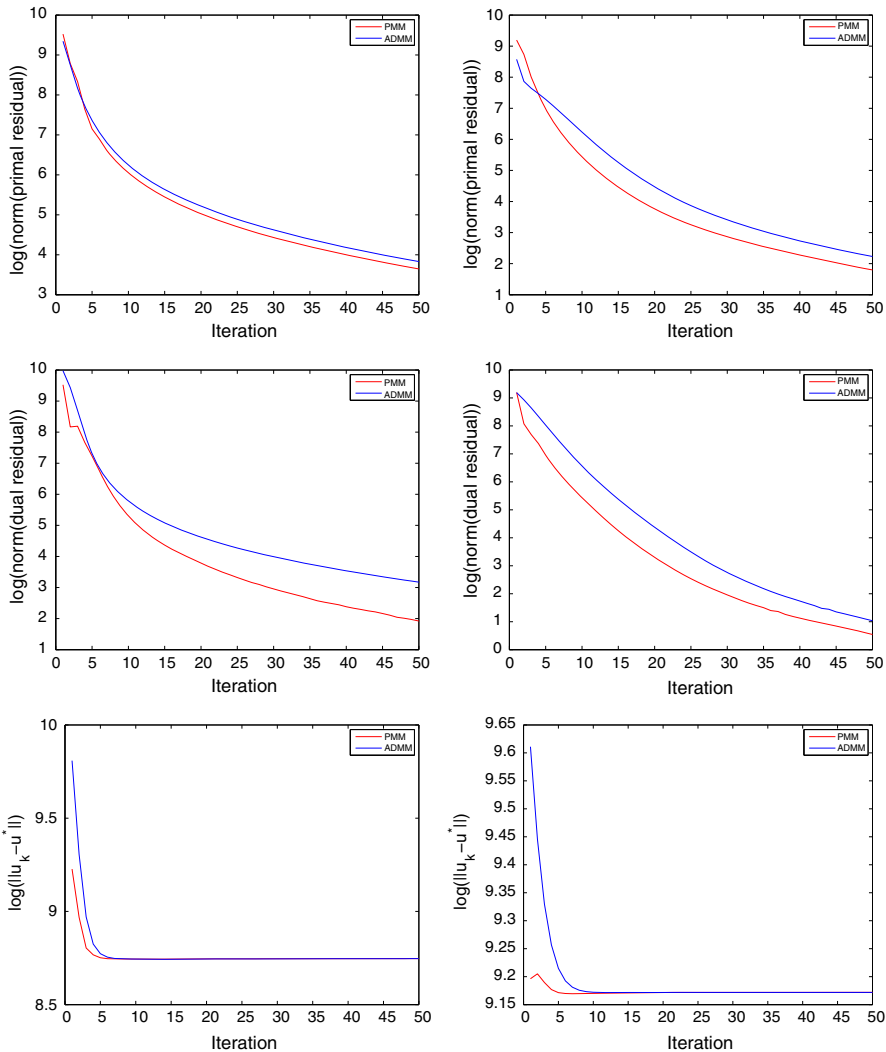
Fig. 3 Residual curves of the PMM and ADMM for the TV denoising problems. (Top) primal error $\|\nabla u_k - v_k\|$ versus iteration number k. (Center) dual error $\|x_k - y_k\|$ versus iteration number k. (Bottom) error $\|u_k - u^*\|$ versus iteration number k (u^* is the exact solution). (Left) convergence results are for the tested image Lena with $\sigma = 0.06$, $\zeta = 50$ and $\rho = 1.5$. (Right) convergence results are for the tested image Baboon with $\sigma = 0.02$, $\zeta = 20$ and $\rho = 1$

in Fig. 3 that the speed of the PMM and ADMM measured by the residual curves are very similar; however the residuals for the PMM decay faster, and this difference is more evident in the dual residual curve.

In Table 1 we present a more detailed comparison between the methods. It reports the iteration counts and total time, in seconds, required for the PMM and ADMM in specific experiments. In these tests we used $\rho_k = 1$ or $\rho_k = 1.5$ for every integer

Table 1 Iterations and computation times (seconds) in parenthesis required for the TV problem

Image	ζ	ρ	σ	PMM	ADMM	Image	ζ	ρ	σ	PMM	ADMM
Lena	20	1	0.02	13 (2.789)	17 (3.510)	Baboon	20	1	0.06	15 (2.311)	21 (2.928)
Lena	20	1.5	0.02	12 (1.803)	14 (2.050)	Baboon	20	1.5	0.06	13 (2.025)	15 (2.336)
Lena	50	1	0.06	19 (2.855)	21 (3.050)	Man	50	1	0.02	24 (7.537)	24 (7.364)
Lena	50	1.5	0.06	17 (2.642)	18 (2.656)	Man	50	1.5	0.02	21 (6.700)	22 (6.597)
Baboon	50	1	0.02	20 (2.862)	21 (2.811)	Man	20	1	0.06	16 (5.322)	21 (6.540)
Baboon	50	1.5	0.02	19 (2.920)	19 (2.789)	Man	20	1.5	0.06	13 (4.395)	16 (5.143)

$k \geq 1$, in both methods, to observe the acceleration performance. We observe that in the tests the PMM executed fewer iterations than ADMM, and the PMM was generally faster. We also observe that both methods accelerate when $\rho = 1.5$.

However, since it is known that the generalized ADMM performs better in general when ρ is larger than 1.61 (see for instance [5,8]), we make a comparison of the two methods taking $\rho \geq 1.61$. The results of these experiments are reported in Table 2. It can be seen that for these values of ρ, although the performances of both methods are very similar in processing time and number of iterations, the PMM continues to be faster than the ADMM.

The operation of highest computational cost within each iteration of the PMM, and ADMM, for the TV problem, consists in solving problem (66). In our tests we solved this step for both algorithms using the Conjugate Gradient (CG) method with tolerance 10^{-5}. This strategy consistently yielded convergence in fewer iterations when using the PMM. Table 3 presents the total number of iteration executed by the CG method in each algorithm for some specific experiments. In the tests presented both methods were stopped at iteration 20.

However, the authors of [14] observed that the ADMM (SB method) attained optimal efficiency by executing, at each iteration of the algorithm, just a single iteration of an iterative method to solve problem (66). Motivated by this, we test the PMM performing one iteration of the CG method at each step of the algorithm. As Fig. 4 below shows, in this case, the PMM also yields good denoised images.

5.2 Compressed sensing

In many areas of applied mathematics and computer science it is often desirable to reconstruct a signal from small amount of data. Compressed sensing is a signal processing technique that allow the reconstruction of signals and images from small number of measurements, provided that they have a sparse representation. This technique has gained considerable attention in the signal processing community since the works of Candès et al. [3], and of Donoho [4], and it has had a significant impact in several applications, for example in imaging, video and medical imaging.

For testing the PMM we consider a particular application of compressed sensing in Magnetic Resonance Imaging (MRI), which is an essential medical imaging tool. MRI is based on the reconstruction of an image from a subset of measurements in the Fourier domain. This imaging problem can be modeled by the optimization problem

$$\min_u \ TV(u) + \frac{\zeta}{2} \|RFu - b\|_F^2 \,, \tag{70}$$

where TV is the total variation norm (61), F is the Discrete Fourier Transform, R is a diagonal matrix, b is the known Fourier data and u is the unknown image that we wish to reconstruct.

The matrix R has a 1 along the diagonal at entries corresponding to the Fourier coefficients that were measured, and 0 for the unknown coefficients. The second term in (70) induces the Fourier transform of the reconstructed image to be close to the measured data, while the TV term in the minimization enforces "smoothness" of the

Table 2 Iterations and computation times (seconds) in parenthesis required for the TV problem when $\rho \geq 1.61$

Image	ζ	ρ	σ	PMM	ADMM	Image	ζ	ρ	σ	PMM	ADMM
Lena	60	1.61	0.03	20 (1.407)	21 (1.437)	Baboon	60	1.7	0.03	20 (1.384)	21 (1.509)
Lena	50	1.61	0.05	17 (1.258)	18 (1.278)	Baboon	50	1.8	0.03	18 (1.282)	19 (1.383)
Lena	20	1.7	0.05	12 (0.901)	14 (0.999)	Man	30	1.61	0.06	14 (2.270)	16 (2.600)
Lena	70	1.8	0.06	21 (1.491)	22 (1.632)	Man	70	1.61	0.06	25 (3.875)	26 (4.063)
Baboon	30	1.61	0.04	13 (0.972)	15 (1.097)	Man	60	1.7	0.04	22 (3.344)	24 (3.641)
Baboon	40	1.61	0.02	16 (1.187)	17 (1.212)	Man	70	1.8	0.05	25 (3.814)	26 (3.880)

Table 3 Total number of iteration of CG method. Tests were stopped at iteration 20

Image	ζ	ρ	σ	PMM	ADMM
Lena	20	1	0.02	108	117
Lena	20	1.5	0.06	101	110
Baboon	20	1.5	0.02	102	110
Baboon	50	1	0.02	121	122
Man	20	1.5	0.06	104	112
Man	50	1	0.06	124	126

(a) $\sigma = 0.06$ **(b)** $20(1.594)$, $\zeta = 50$, $\rho = 1$ **(c)** $22(1.557)$, $\zeta = 50$, $\rho = 1$

(d) $\sigma = 0.02$ **(e)** $14(2.436)$, $\zeta = 20$, $\rho = 1.5$ **(f)** $16(2.514)$, $\zeta = 20$, $\rho = 1.5$
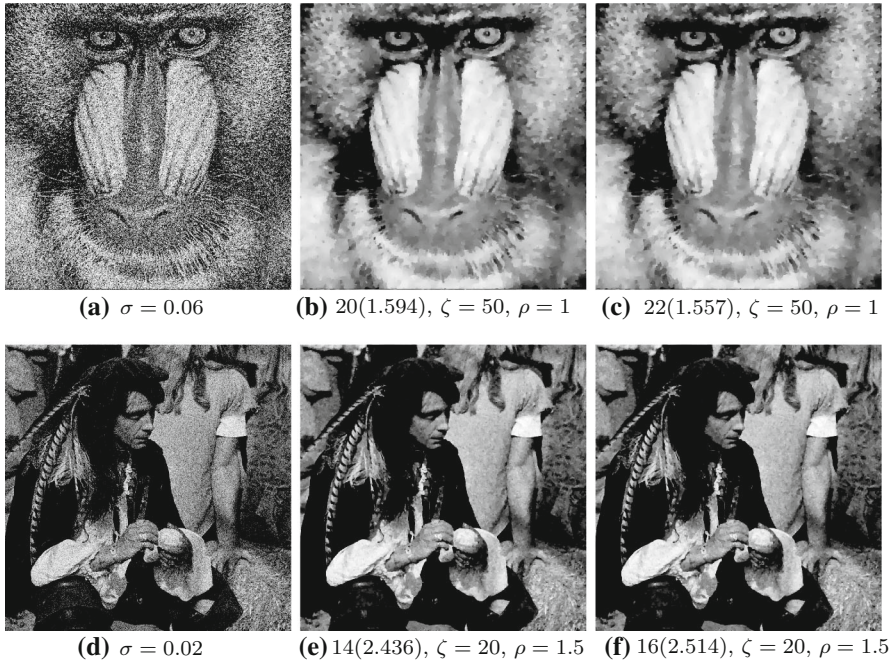
Fig. 4 Denoising with one iteration of CG method per iteration. (Left) noisy images, the value of variance is reported below each image. (Center) images denoised with PMM. (Right) images denoised with ADMM. The number of iterations, the total time in seconds (in parenthesis); as well as the used values of ζ and ρ are displayed below each image

image. The parameter $\zeta > 0$ provides a tradeoff between the fidelity term and the smoothness term.

Problem (70) can be posed as a linearly constrained minimization problem (1) in much the same manner as was done for the TV problem in the previous subsection. Therefore, to apply the PMM to (70) we take $f(u) = \frac{\zeta}{2} \|RFu - b\|_F^2$, $g(v) = \|v\|_1$, $M = \nabla$, $C = -I$ and $d = 0$. The resulting minimization problems are

$$v_k = \arg\min_v \ \|v\|_1 - \langle z_{k-1} + \lambda w_{k-1}, v \rangle + \frac{\lambda}{2} \|v\|_F^2, \tag{71}$$
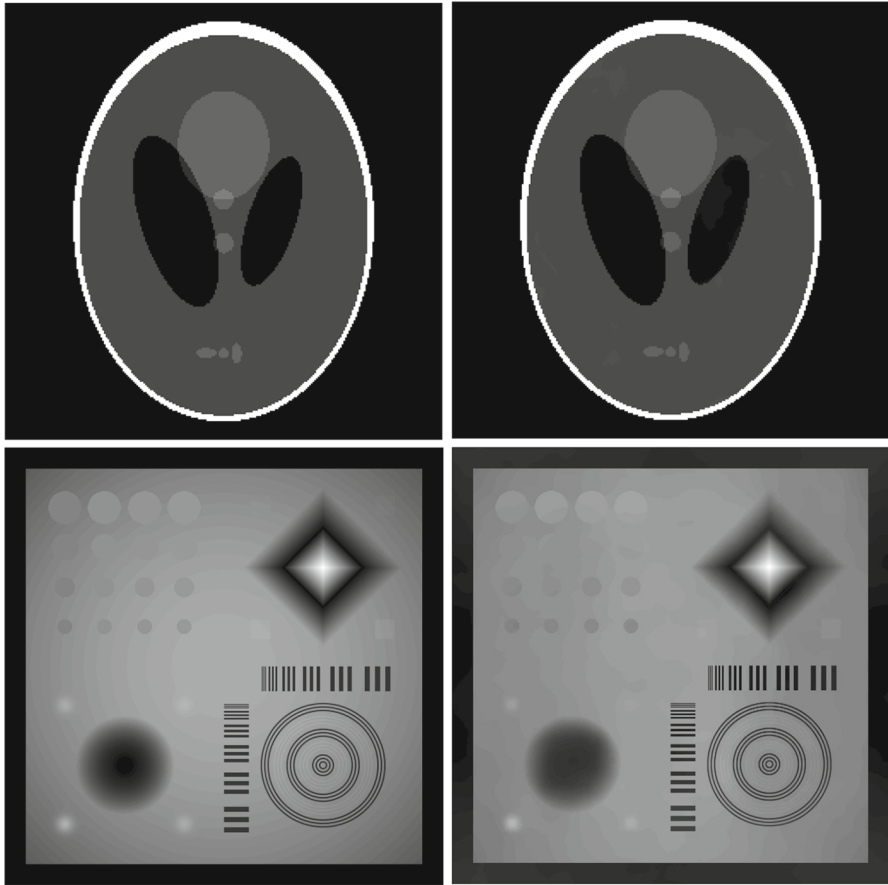
Fig. 5 (Left) images used in compressed sensing tests. (Right) the images reconstructed with the PMM. The Shepp–Logan phantom of size 256×256 (top) was recovered with 25% sampling and $\rho = 1.5$. The CS-Phantom of size 512×512 (bottom) was recovered with 50% sampling and $\rho = 1.5$

and

$$u_k = \arg\min_u \ \frac{\zeta}{2} \|RFu - b\|_F^2 + \langle z_{k-1} - \lambda v_k, \nabla u \rangle + \frac{\lambda}{2} \|\nabla u\|_F^2 . \qquad (72)$$

Problem (71) can be solved explicitly using the **shrink** operator (3). Indeed, by the optimality conditions for this problem we have

$$v_k = \mathbf{shrink}\left(\frac{1}{\lambda} z_{k-1} + w_{k-1}, \frac{1}{\lambda}\right).$$

The optimality condition for the minimization problem (72) is

$$0 = \zeta F^T R^T (RFu_k - b) + \nabla^*(z_{k-1} - \lambda v_k) + \lambda \nabla^* \nabla u_k,$$
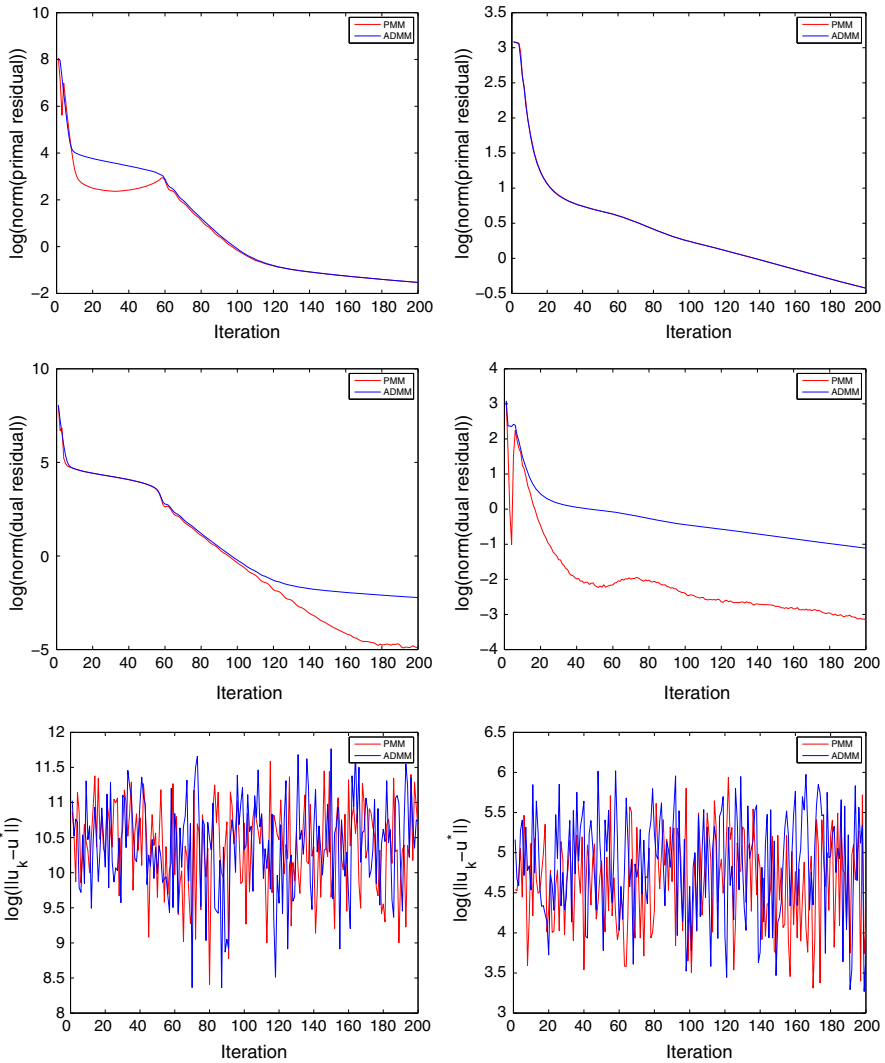
Fig. 6 Residual curves of the PMM and ADMM for the compressed sensing problems. (Top) primal error $\|\nabla u_k - v_k\|$ versus iteration number. (Center) dual error $\|x_k - y_k\|$ versus iteration number. (Bottom) error $\|u_k - u^*\|$ versus iteration number (u^* is the exact solution). (Left) convergence results are for the Shepp–Logan phantom of size 256×256 with 25% sampling and $\rho = 1.5$. (Right) convergence results are for the CS-Phantom of size 512×512 with 50% sampling and $\rho = 1.5$

or equivalently

$$(\zeta F^T R^T R F + \lambda \nabla^* \nabla) u_k = \zeta F^T R^T b - \nabla^* (z_{k-1} - \lambda v_k).$$

Table 4 Iterations and computation times (seconds) in parenthesis required for the CS problem

Image	%	ρ	PMM	ADMM
Shepp–Logan 256×256	25	1.61	379 (4.048)	380 (4.204)
Shepp–Logan 512×512	25	1.61	108 (6.695)	108 (7.154)
Shepp–Logan 768×768	50	1.8	40 (5.332)	43 (5.808)
Shepp–Logan 1024×1024	25	1.7	47 (11.58)	48 (13.35)
CS-phantom 256×256	50	1.7	559 (6.160)	559 (7.079)
CS-phantom 512×512	50	1.61	468 (29.06)	468 (32.06)
CS-phantom 768×768	75	1.61	137 (18.49)	137 (20.16)
CS-phantom 1024×1024	75	1.8	170 (40.91)	171 (45.39)

Thus, we obtain u_k, the solution of the system above, by

$$u_k = F^T(\zeta R^T R + \lambda F \nabla^* \nabla F^T)^{-1} F(\zeta F^T R^T b - \nabla^*(z_{k-1} - \lambda v_k)).$$

We tested the PMM on two synthetic phantom. The first is the digital Shepp–Logan phantom, which was created with the Matlab function "`phantom`". The second experiment was done with a CS-Phantom, which was taken from the mathworks web site. Experiments were conducted with different sizes of the phantoms and measuring at random 25%, 50% or 75% of the Fourier coefficients. As stopping condition for these problems, it was used the criterion given by the residuals for the KKT conditions. More specifically, the PMM and ADMM were stopped when the primal and dual residuals, associated with each method, were less than $10^{-6}(m * n)$, where m and n are the dimensions of the image. For all the experiments we used $\zeta = 500$ and $\lambda = 1$, since we found that these choices were effective for both methods.

Figure 5 shows the test images and their reconstructions using the PMM. The performance of the PMM and ADMM can be seen in Fig. 6, which reports the residual curves for both methods, as were the error $\|u_k - u^*\|$, where u^* is the exact solution. Observe that the primal curves for both methods are very similar along all iterations. However, the decay for the dual residual curve for the PMM is much faster than the dual residual for the ADMM.

Table 4 presents some numerical results for the compressed sensing problem. It reports the iteration count and total time required for the PMM and ADMM in the experiments. In these tests, it was used $\rho \geq 1.61$. We observe that the PMM solves all the problems faster than ADMM.

5.3 The dual residual

It was observed in our numerical experiments that, despite the overall rate of decrease for the PMM and the ADMM are very similar, the dual variable in the PMM sequence is smaller than the ADMM dual variable. This could be an advantage for the PMM, and motivates us to study the performance of the method using a stopping criterion based on the dual residual.

Table 5 Performance results using stooping criterion (73)

Problem	PMM			ADMM		
	# It	Time (s)	$relt$	# It	Time (s)	$relt$
TV (Man, $\zeta = 60$, $\rho = 1.61$, $\sigma = 0.06$)	257	24.62	0.18472	1000*	69.86	0.18574
TV (Man, $\zeta = 40$, $\rho = 1.7$, $\sigma = 0.05$)	146	14.91	0.16468	614	41.18	0.16468
TV (Man, $\zeta = 20$, $\rho = 1.8$, $\sigma = 0.03$)	68	7.077	0.14992	197	14.31	0.14992
TV (Lena, $\zeta = 50$, $\rho = 1.61$, $\sigma = 0.05$)	221	9.901	0.09670	750	23.94	0.09671
TV (Lena, $\zeta = 60$, $\rho = 1.7$, $\sigma = 0.06$)	290	12.40	0.10520	1000*	30.79	0.10521
TV (Lena, $\zeta = 30$, $\rho = 1.8$, $\sigma = 0.02$)	147	6.297	0.07453	427	13.58	0.07454
TV (Baboon, $\zeta = 50$, $\rho = 1.61$, $\sigma = 0.03$)	222	9.487	0.17651	679	21.33	0.17651
TV (Baboon, $\zeta = 30$, $\rho = 1.7$, $\sigma = 0.04$)	86	3.873	0.1576	206	6.830	0.1576
TV (Baboon, $\zeta = 70$, $\rho = 1.8$, $\sigma = 0.07$)	301	13.70	0.19101	1000*	32.82	0.19102
CS (Shepp–Logan 256 × 256, $\zeta = 500$, $\rho = 1.61$, 25%)	54	0.661	0.02701	207	2.088	0.01693
CS (Shepp–Logan 512 × 512, $\zeta = 500$, $\rho = 1.8$, 75%)	29	1.893	0.00455	31	2.051	0.00621
CS (Shepp–Logan 768 × 768, $\zeta = 500$, $\rho = 1.7$, 50%)	25	3.357	0.01943	29	4.046	0.01554
CS (Shepp–Logan 1024 × 1024, $\zeta = 500$, $\rho = 1.61$, 25%)	36	8.990	0.04263	40	10.43	0.03621
CS (Phantom 256 × 256, $\zeta = 500$, $\rho = 1.7$, 50%)	78	0.837	0.11268	466	4.751	0.10778
CS (Phantom 512 × 512, $\zeta = 500$, $\rho = 1.61$, 75%)	20	1.244	0.14151	192	11.82	0.07704
CS (Phantom 768 × 768, $\zeta = 500$, $\rho = 1.61$, 25%)	36	4.777	0.21254	179	24.77	0.20003
CS (Phantom 1024 × 1024, $\zeta = 500$, $\rho = 1.8$, 75%)	23	5.623	0.09183	115	28.3	0.07034

Fig. 7 TV problem for the test image Lena, which was contaminated with Gaussian noise with variance $\sigma = 0.02$. (Left) image denoised with PMM. (Right) image denoised with ADMM. The image was denoised using $\zeta = 30$ and $\rho = 1.8$
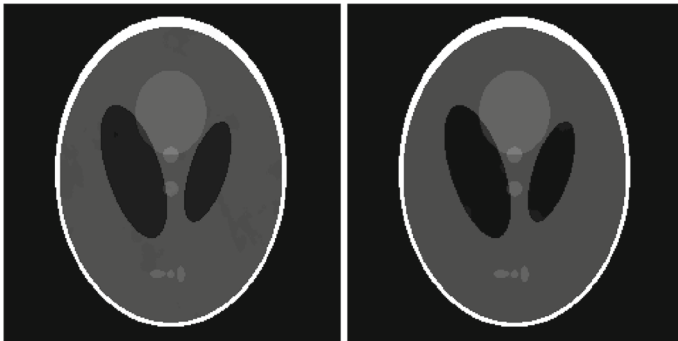

Fig. 8 Compressed sensing problem for the test image Shepp–Logan phantom of size 256×256 with 25% sampling. (Left) image recovered with the PMM. (Right) image recovered with the ADMM. In the experiments were used $\zeta = 500$ and $\rho = 1.61$

In this subsection we present some preliminary computational results considering a termination condition that only uses information from the dual residual sequences. We use as test problems the TV (60) and CS (70) problems discussed in the previous subsections. The algorithms were run until condition

$$\|d_k\| / (m * n) \leq 10^{-6} \tag{73}$$

was satisfied, where d_k is the corresponding dual residual of the sequence at iteration k, and m and n are the dimensions of the images. In all the experiments we fix $\lambda = 1$ and use $\rho \geq 1.61$.

Table 5 displays the performance of the PMM and ADMM using criterion (73). The columns mean: Problem, the considered problem (TV or CS with the corresponding image, ζ, ρ and noise or sampling); # It, iterations needed to satisfy (73); time (s), CPU time (s); and $relt$, the relative error $\|u_k - u^*\| / \|u^*\|$ at the last iterate, which can be computed since the exact solution u^* is known.

We observe that the PMM is faster than the ADMM in all the problems, and in 12 instances it is at least 2 times faster than the ADMM. We also observe that, although the high speed exhibited by the PMM, the quality of the solution found by it is very similar to that found by the ADMM, which can be seen by comparing the *relt* columns of Table 5. Additionally, we notice that in three experiments the ADMM stops because it reaches the maximum number of allowed iterations (1000)[2], therefore it does not solve the instance by the required accuracy.

Figures 7 and 8 show the image reconstruction results for some tests. It can be observed in Fig. 7 that for the TV problem both methods recover good images using (73). This is not surprising since the stopping criterion used in Sect. 5.1 is more flexible than (73), and the restoration results were satisfactory (see Sect. 5.1). It turns out that for the CS problem, although the termination condition considered in Sect. 5.2 is more restrictive than (73), the PMM and ADMM can also reconstruct images with good quality using this last stopping criterion, as can be seen in Fig. 8.

References

1. Brezis, H.: Functional Analysis. Sobolev Spaces and Partial Differential Equations. Universitext. Springer, New York (2011)
2. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F., ϵ-enlargements of maximal monotone operators: theory and applications. In: Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997), vol. 22 of Appl. Optim, pp. 25–43. Kluwer Academic Publishers, Dordrecht (1999)
3. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
4. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
5. Eckstein, J.: Parallel alternating direction multiplier decomposition of convex programs. J. Optim. Theory Appl. **80**(1), 39–62 (1994)
6. Eckstein, J.: Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results. RUTCOR Research Reports 32 (2012)
7. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. Ser. A **55**(3), 293–318 (1992)
8. Eckstein, J., Ferris, M.C.: Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. INFORMS J. Comput. **10**(2), 218–235 (1998)
9. Eckstein, J., Svaiter, B.F.: A family of projective splitting methods for the sum of two maximal monotone operators. Math. Program. Ser. B **111**(1–2), 173–199 (2008)
10. Esser, E.: Applications of lagrangian-based alternating direction methods and connections to split bregman. CAM Rep. **9**, 31 (2009)
11. Gabay, D.: Chapter ix applications of the method of multipliers to variational inequalities. Studies Math. Appl. **15**, 299–331 (1983)
12. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
13. Glowinski, R., Marrocco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér. **9**(R–2), 41–76 (1975)

---

[2] Marked with an asterisk (*).

14. Goldstein, T., Osher, S.: The split Bregman method for $L1$-regularized problems. SIAM J. Imaging Sci. **2**(2), 323–343 (2009)
15. He, B., Yuan, X.: On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J. Numer. Anal. **50**(2), 700–709 (2012)
16. Hiriart–Urruty, J.B., Lemaréchal, C.: Convex analysis and minimization algorithms II: advanced theory and bundle methods. In: Grundlehren der Mathematischen Wissenschaften, vol. 306. Springer, Berlin (1993)
17. Lions, P.-L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
18. Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**, 341–346 (1962)
19. Monteiro, R.D.C., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM J. Optim. **23**(1), 475–507 (2013)
20. Moreau, J.-J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
21. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**(2), 383–390 (1979)
22. Rockafellar, R.T.: Convex Analysis. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, NJ (1970)
23. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pac. J. Math. **33**, 209–216 (1970)
24. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms (Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991)). Phys. D **60**(1–4), 259–268 (1992)
25. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Control Optim. **29**(1), 119–138 (1991)