CrossMark

An Implementable Splitting Algorithm for the ℓ_1-norm Regularized Split Feasibility Problem

Hongjin He[1] · Chen Ling[1] · Hong-Kun Xu[1,2]

Abstract The *split feasibility problem* (SFP) captures a wide range of inverse problems, such as signal processing, image reconstruction, and so on. Recently, applications of ℓ_1-norm regularization to linear inverse problems, a special case of SFP, have been received a considerable amount of attention in the signal/image processing and statistical learning communities. However, the study of the ℓ_1-norm regularized SFP still deserves attention, especially in terms of algorithmic issues. In this paper, we shall propose an algorithm for solving the ℓ_1-norm regularized SFP. More specifically, we first formulate the ℓ_1-norm regularized SFP as a separable convex minimization problem with linear constraints, and then introduce our splitting method, which takes advantage of the separable structure and gives rise to subproblems with closed-form solutions. We prove global convergence of the proposed algorithm under certain mild conditions. Moreover, numerical experiments on an image deblurring problem verify the efficiency of our algorithm.

✉ Hong-Kun Xu
xuhk@hdu.edu.cn; xuhk@math.nsysu.edu.tw

Hongjin He
hehjmath@hdu.edu.cn

Chen Ling
cling_zufe@sina.com

[1] Department of Mathematics, School of Science, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China

[2] Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung 80424, Taiwan

1 Introduction

The *split feasibility problem* (SFP), which was first introduced by Censor and Elfving [12] to model phase retrieval problems, models a variety of inverse problems. Its recent applications to the medical treatment of intensity-modulated radiation therapy and medical imaging reconstruction [7,11,13,15–17] have inspired more studies theoretically and practically.

The mathematical formulation of SFP is indeed a feasibility problem that amounts to finding a point $x^* \in \mathbb{R}^n$ with the property

$$x^* \in \mathcal{C} \quad \text{and} \quad Ax^* \in \mathcal{Q}, \tag{1.1}$$

where $\mathcal{C}$ and $\mathcal{Q}$ are nonempty, closed and convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator.

It is known that SFP (1.1) includes a variety of real-world inverse problems as special cases. For instance, when $\mathcal{Q} = \{b\}$ is singleton, SFP (1.1) is reduced to the classical convexly constrained linear inverse problem, that is, finding a point $x^* \in \mathbb{R}^n$ such that

$$x^* \in \mathcal{C} \quad \text{and} \quad Ax^* = b. \tag{1.2}$$

This problem has extensively been studied in the literature (see [29,30,45,48,51] and the references therein). Due possibly to the unclear structure of the general constraint set $\mathcal{Q}$, SFP (1.1) has however been received much less attention than its special case (1.2). As a matter of fact, certain algorithms that work for (1.2) seemingly have no straightforward counterparts for SFP (1.1). For instance, it still remains unclear whether the dual approach for (1.2) proposed in [45] (see also [48]) can be extended to SFP (1.1).

The SFP (1.1) can be solved by optimization algorithms since it is equivalent to the constrained minimization problem:

$$\min_{x \in C} \ f(x) := \frac{1}{2} \|Ax - P_{\mathcal{Q}} Ax\|^2. \tag{1.3}$$

[Here $P_{\mathcal{Q}}$ is the projection onto $\mathcal{Q}$ (see (2.1)).] Among the existing algorithms for SFP (1.1), the CQ algorithm introduced by Byrne [7,8] is the most popular; it generates a sequence $\{x^k\}$ through the recursion process:

$$x^{k+1} = P_{\mathcal{C}} \left(x^k - A^\top \left(Ax^k - P_{\mathcal{Q}} Ax^k \right) \right).$$

This is indeed the gradient-projection algorithm applied to the minimization problem (1.3) (it is also a special case of the proximal forward-backward splitting method [24]). Recent improvements of the CQ algorithm with relaxed stepsizes and projections can be found in [26,46,55,57,60].

It is often the case that SFP (1.1) and (1.2) are ill-posed, due to possible ill-conditionedness of A. As a result, regularization must be reinforced. In fact, for the past decades, regularization techniques have successfully been used for solving (1.2), in particular, the unconstrained case of $\mathcal{C} = \mathbb{R}^n$; see [30,32,39,49] and the references therein. Regularization methods for SFP (1.1) have however been quite a recent topic; see [10,14,55,58] where the classical Tikhonov regularization [51] is used to promote stability of solutions. In contrast with Tikhonov's regularization, the ℓ_1-norm regularization for (1.2) has been paid much more attention recently, due to the fact that the ℓ_1-norm induces sparsity of solutions [6,22,27,31,56]. A typical example is the lasso [50] (also known as basis pursuit denoising [22]) which is the minimization problem:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2 + \sigma \|x\|_1, \tag{1.4}$$

where $b \in \mathbb{R}^m$ and $\sigma > 0$ is a regularization parameter.

The ℓ_1-norm regularization of SFP (1.1) is the following constrained minimization problem:

$$\min_{x \in \mathbb{R}^n} \{ \|x\|_1 \mid x \in \mathcal{C} \text{ and } Ax \in \mathcal{Q} \}, \tag{1.5}$$

which can also be recast as

$$\min_{x \in \mathcal{C}} \frac{1}{2} \|Ax - P_{\mathcal{Q}}(Ax)\|^2 + \sigma \|x\|_1. \tag{1.6}$$

Notice that the lasso (1.4), the *basis pursuit* (BP) and *basis pursuit denoising* (BPDN) of Chen et al. [22] (see also [9,27,31,53,56]) are special cases of (1.5) and (1.6).

Compared with (1.4), the ℓ_1-regularized SFP (1.6) has been received much less attention. In fact, as far as we have noticed, (1.6) has been considered only in [38], in which certain properties of the solution set of (1.6) were discussed, leaving algorithmic approaches unattached.

The main purpose of this paper is to introduce an implementable splitting algorithm for solving the ℓ_1-regularized SFP (1.6) via the technique of *alternating direction method of multipliers* (ADMM) [33] in the framework of variational inequalities. Towards this we use the trick of variable splitting to rewrite (1.6) as

$$\begin{aligned} \min_{x,z} \ & f(x) + \sigma \|z\|_1 \\ \text{s.t. } & x = z, \ x \in \mathcal{C}, \ z \in \mathbb{R}^n, \end{aligned} \tag{1.7}$$

where $f(x)$ is defined as in (1.3). Now ADMM, which has been widely used in the areas of variational inequalities, compressive sensing, image processing, and statistical learning (cf. [2,6,19,21,35,36,40,56,59]), works for (1.7) as follows: Given the k-th iterate $\mathbf{w}^k := (x^k, z^k, \lambda^k) \in \mathcal{C} \times \mathbb{R}^n \times \mathbb{R}^n$; the $(k + 1)$-th iterate $\mathbf{w}^{k+1} := (x^{k+1}, z^{k+1}, \lambda^{k+1})$ is obtained by solving three subproblems

$$\begin{cases} z^{k+1} = \arg\min_{z \in \mathbb{R}^n} \left\{ \|z\|_1 + \frac{\gamma}{2} \left\| x^k - z - \frac{\lambda^k}{\gamma} \right\|^2 \right\}, & (1.8) \\[2ex] x^{k+1} = \arg\min_{x \in \mathcal{C}} \left\{ f(x) + \frac{\gamma}{2} \left\| x - z^{k+1} - \frac{\lambda^k}{\gamma} \right\|^2 \right\}, & (1.9) \\[2ex] \lambda^{k+1} = \lambda^k - \gamma(x^{k+1} - z^{k+1}), & (1.10) \end{cases}$$

where $\{\lambda^k\}$ is a sequence of Lagrangian multipliers and $\gamma > 0$ is a penalty parameter for violation of the linear constraints. The advantage of ADMM lies in its full exploiture of the separability structure of (1.7). Nevertheless, the subproblem (1.9) may not be easy enough to get solved due to the existence of the constraint set $\mathcal{C}$. To overcome this difficulty we will combine the techniques of linearization of ADMM and the proximal point algorithm [41] to propose a new algorithm, Algorithm 1 in Sect. 3. The main contribution of this paper is to prove the global convergence of this algorithm and apply it to solve image deblurring problems.

The rest of this paper is structured as follows. In Sect. 2, we summarize some basic notation, definitions and properties that will be useful for our subsequent analysis. In Sect. 3, we analyze the subproblems of ADMM for (1.7) and propose an implementable algorithm by incorporating the ideas of linearization and the proximal point algorithm into the new algorithm. Then, we prove global convergence of the proposed algorithm under some mild conditions. In Sect. 4, numerical experiments on image deblurring problems are carried out to

verify the efficiency and reliability of our algorithm. Finally, concluding remarks are included in Sect. 5.

2 Preliminaries

In this section, we summarize some basic notation, concepts and well known results that will be useful for further discussions in subsequent sections.

Let $\mathbb{R}^n$ be an n-dimensional Euclidean space. For any two vectors $x, y \in \mathbb{R}^n$, $\langle x, y \rangle$ denotes the standard inner product. Furthermore, for a given symmetric and positive definite matrix M, we denote by $\|x\|_M = \sqrt{\langle x, Mx \rangle}$ the M-norm of x. For any matrix A, we denote $\|A\|$ as its matrix 2-norm.

Throughout this paper, we denote by ι_Ω the indicator function of a nonempty closed convex set Ω, i.e.,

$$\iota_\Omega(x) := \begin{cases} 0, & \text{if } x \in \Omega, \\ +\infty, & \text{if } x \notin \Omega. \end{cases}$$

The projection operator $P_{\Omega,M}$ from $\mathbb{R}^n$ onto the nonempty closed convex set Ω under M-norm is defined by

$$P_{\Omega,M}[x] := \arg\min \{ \|y - x\|_M \mid y \in \Omega \}, \quad x \in \mathbb{R}^n. \tag{2.1}$$

For simplicity, let P_Ω be the projection operator under Euclidean norm. It is well-known that the projection $P_{\Omega,M}$ (in Ω) can be characterized by the relation

$$\langle u - P_{\Omega,M}[u], M\left(v - P_{\Omega,M}[u]\right) \rangle \leq 0, \quad u \in \mathbb{R}^n, \quad v \in \Omega.$$

(The proof is available in any standard optimization textbook, such as [5, p. 211].) Next we introduce Moreau's notion [42] of proximity operators which extends the concept of projections.

Definition 2.1 Let $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous and convex function. The proximity operator of φ, denoted by $\mathbf{prox}_\varphi$, is defined as an operator from $\mathbb{R}^n$ into itself and given by

$$\mathbf{prox}_\varphi(x) = \arg\min_{y \in \mathbb{R}^n} \left\{ \varphi(y) + \frac{1}{2} \|x - y\|^2 \right\}, \quad x \in \mathbb{R}^n. \tag{2.2}$$

It is clear that $P_\Omega = \mathbf{prox}_\varphi$ when $\varphi = \iota_\Omega$. Moreover, if we denote by $\mathbf{I}$ the identity operator on $\mathbb{R}^n$, then $\mathbf{prox}_\varphi$ and $(\mathbf{I} - \mathbf{prox}_\varphi)$ share the same nonexpansivity of P_Ω and $(\mathbf{I} - P_\Omega)$. The reader is referred to [23,24,42] for more details.

Definition 2.2 Let $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous and convex function. The subdifferential of φ is the set-valued operator $\partial\varphi : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, given by, for $x \in \mathbb{R}^n$,

$$\partial\varphi(x) := \left\{ \zeta \in \mathbb{R}^n \mid \varphi(y) \geq \varphi(x) + \langle y - x, \zeta \rangle, \quad y \in \mathbb{R}^n \right\}.$$

According to the above two definitions, the minimizer in the definition (2.2) of $\mathbf{prox}_\varphi(x)$ is characterized by the inclusion

$$0 \in \partial\varphi(\mathbf{prox}_\varphi(x)) + \mathbf{prox}_\varphi(x) - x.$$

Definition 2.3 A set-valued operator T from $\mathbb{R}^n$ to $2^{\mathbb{R}^n}$ is said to be monotone if

$$\langle x_1 - x_2, \ x_1^* - x_2^* \rangle \geq 0, \quad x_1, \ x_2 \in \mathbb{R}^n, \quad x_1^* \in T(x_1), \ x_2^* \in T(x_2).$$

A monotone operator is said to be maximal monotone if its graph is not properly contained in the graph of any other monotone operator.

It is well-known [47] that the subdifferential operator $\partial\varphi$ of a proper, lower semicontinuous and convex function φ is maximal monotone.

Definition 2.4 Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex and differentiable function. We say that its gradient ∇g is Lipschitz continuous with constant $L > 0$ if

$$\|\nabla g(x_1) - \nabla g(x_2)\| \leq L\|x_1 - x_2\|, \quad x_1, \ x_2 \in \mathbb{R}^n.$$

Notice that the gradient of the function f defined in (1.6) is $\nabla f = A^\top(\mathbf{I} - P_\Omega)A$. It is easily seen that ∇f is monotone and Lipschitz continuous with constant $\|A\|^2$, i.e.,

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leq \|A\|^2\|x_1 - x_2\|, \quad x_1, \ x_2 \in \mathbb{R}^n. \tag{2.3}$$

3 Algorithm and Its Convergence Analysis

We are aimed in this section to invent a new algorithm, Algorithm 1 below, and prove its global convergence under certain mild conditions.

3.1 Algorithm

We begin with more analysis on the algorithm (1.8)–(1.9). The subproblem (1.9) is a constrained optimization problem which might not be easy enough to be solved efficiently. To make ADMM implementable for some constrained models, it is suggested directly linearizing the constrained subproblem [i.e., (1.9)] [18,20,52]. Clearly, an immediate benefit is that the solutions to the subproblems can be represented explicitly. It is noteworthy that direct linearization on ADMM must often be imposed strong conditions (e.g., strict convexity) for global convergence. Notice that the linearization on ADMM yields an approximate function involving a proximal term. Since the *proximal point algorithm* (PPA) [41] is one of the most powerful solvers in optimization, which not only provides a unified algorithmic framework including ADMM as its special case [28], but also, as Parikh and Boyd [44] emphasized, works under extremely general conditions, we combine the ideas of linearization on ADMM and PPA to propose an implementable splitting algorithm which has easier subproblems with closed-form solutions and shares similar convergence results to the PPA's. More concretely, our idea is that we first add a proximal term to the subproblem (1.8), and the resulting subproblem has a closed-form solution via the shrinkage operator. We then linearize the function f at the latest iterate x^k in such a way that the subproblem (1.9) is simple enough to possess a closed-form expression. Following these two steps, we get an intermediate Lagrangian multiplier via (1.10).

Taking a revisit on the linearization procedure, we see that the resulting subproblem is essentially a one-step approximation of ADMM, which might not be the best output for the next iterate. Accordingly, we add a correction step with lower computational costs to update the output of the *linearized ADMM* (LADMM) for the purpose of compensating the approximation errors caused by linearization. A notable feature is that the correction-step can improve the numerical performance in terms of taking fewer iterations than LADMM, in

addition to relaxing the conditions for global convergence and inducing a strong convergence result as [4].

Basing upon the above analysis, we will design an algorithm that inherits the advantages of ADMM and PPA. Our algorithm is a two-stage method, consisting of a prediction step and a correction step. Below, we mainly state the details of the prediction step. First, we generate the predictor $\widetilde{z}^k$ via

$$\widetilde{z}^k = \arg\min_{z \in \mathbb{R}^n} \left\{ \sigma \|z\|_1 + \frac{\gamma}{2} \left\| x^k - z - \frac{\lambda^k}{\gamma} \right\|^2 + \frac{\rho_2}{2} \left\| z - z^k \right\|^2 \right\}, \qquad (3.1)$$

where ρ_2 is a positive constant. The reason of imposing an proximal term $\frac{\rho_2}{2} \left\| z - z^k \right\|^2$ is to regularize the z-related subproblem. Besides, we can further derive a stronger convergence result theoretically with the help of such proximal term. Obviously, (3.1) can be rewritten as

$$\widetilde{z}^k = \arg\min_{z \in \mathbb{R}^n} \left\{ \sigma \|z\|_1 + \frac{\gamma + \rho_2}{2} \left\| z - \frac{1}{\gamma + \rho_2} \left(\gamma x^k - \lambda^k + \rho_2 z^k \right) \right\|^2 \right\}.$$

Consequently, using the shrinkage operator [24,56], we can solve the last minimization and express $\widetilde{z}^k$ in closed-form as

$$\widetilde{z}^k = \text{shrink} \left(\omega^k, \frac{\sigma}{\gamma + \rho_2} \right) \quad \text{with} \quad \omega^k := \frac{1}{\gamma + \rho_2} \left(\gamma x^k - \lambda^k + \rho_2 z^k \right). \qquad (3.2)$$

Recall that the shrinkage operator 'shrink$(\cdot, \cdot)$' is defined by

$$\text{shrink}(w, \mu) := \text{sign}(w) \odot \max\{0, |w| - \mu\}, \qquad w \in \mathbb{R}^n, \ \mu > 0, \qquad (3.3)$$

where the 'sign', 'max', and the absolute value function '$|\cdot|$' are component-wise, and '$\odot$' denotes the component-wise product.

For the x-related subproblem, the minimization subproblem (1.9) is not simple enough to have closed-form solution due to the nonlinearity of the function f. We may treat the nonlinearity of f by linearization. As a matter of fact, given $(x^k, \widetilde{z}^k, \lambda^k)$, we linearize f at x^k to get an approximate subproblem of (1.9), that is,

$$\begin{aligned} \widetilde{x}^k &= \arg\min_{x \in \mathcal{C}} \left\{ \left\langle \nabla f\left(x^k\right), \ x - x^k \right\rangle + \frac{\rho_1}{2} \left\| x - x^k \right\|^2 + \frac{\gamma}{2} \left\| x - \widetilde{z}^k - \frac{\lambda^k}{\gamma} \right\|^2 \right\} \\ &= P_{\mathcal{C}} \left[\frac{1}{\rho_1 + \gamma} \left(\rho_1 x^k + \gamma \widetilde{z}^k + \lambda^k - \nabla f\left(x^k\right) \right) \right], \end{aligned} \qquad (3.4)$$

where ρ_1 is a suitable constant for approximation. With a pair of predictors $(\widetilde{x}^k, \widetilde{z}^k)$, we generate the last predictor of λ via

$$\widetilde{\lambda}^k = \lambda^k - \gamma \left(\widetilde{x}^k - \widetilde{z}^k \right). \qquad (3.5)$$

Finally, with the predictor $\widetilde{\mathbf{w}}^k = \left(\widetilde{x}^k, \widetilde{z}^k, \widetilde{\lambda}^k \right)$, we update the next iterate with a correction step to compensate the approximation errors caused by the linearization of ADMM. This correction step together with the proximal regularization further relaxes the convergence requirements and improves the numerical performance (see Sect. 4).

Before presenting the splitting algorithm, for the sake of convenience, we denote $\eta_1(x, \widetilde{x}) := \nabla f(x) - \nabla f(\widetilde{x})$, I_n the n-dimensional identity matrix, and

$$\mathbf{w} := \begin{pmatrix} x \\ z \\ \lambda \end{pmatrix}, \quad M := \begin{pmatrix} (\rho_1 + \gamma)I_n & 0 & 0 \\ 0 & \rho_2 I_n & 0 \\ 0 & 0 & \frac{1}{\gamma}I_n \end{pmatrix} \quad \text{and} \quad \eta(\mathbf{w}, \widetilde{\mathbf{w}}) := \begin{pmatrix} \eta_1(x, \widetilde{x}) \\ 0 \\ 0 \end{pmatrix}. \tag{3.6}$$

Then, the corresponding algorithm is described formally in Algorithm 1.

Algorithm 1 (Implementable Splitting Method)

1: Choose $\rho_1 \geq \|A\|^2$, $\rho_2 > 0$, $\gamma > 0$ and $\tau \in (0, 2)$. Set $\mathbf{w}^0 := (x^0, z^0, \lambda^0) \in \mathcal{C} \times \mathbb{R}^n \times \mathbb{R}^n$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3:     Generate a predictor $\widetilde{\mathbf{w}}^k = \left(\widetilde{x}^k, \widetilde{z}^k, \widetilde{\lambda}^k\right)$ via (3.2), (3.4) and (3.5), respectively.
4:     Update the next iterate $\mathbf{w}^{k+1} = \left(x^{k+1}, z^{k+1}, \lambda^{k+1}\right)$ via

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \tau\alpha_k \mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right), \tag{3.7}$$

$$\text{where} \begin{cases} \alpha_k = \dfrac{\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)}{\|\mathbf{d}(\mathbf{w}^k, \widetilde{\mathbf{w}}^k)\|_M^2}, & (3.8) \\[2mm] \phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) = \left\langle \mathbf{w}^k - \widetilde{\mathbf{w}}^k, \ M\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\right\rangle + \left\langle \widetilde{\lambda}^k - \lambda^k, \ \widetilde{x}^k - x^k \right\rangle, & (3.9) \\[2mm] \mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) = \left(\mathbf{w}^k - \widetilde{\mathbf{w}}^k\right) - M^{-1}\eta\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right). & (3.10) \end{cases}$$

5: **end for**

Remark 3.1 It is worth pointing out that Algorithm 1 is not limited to solve the ℓ_1-norm regularized SFP. When we use a general regularizer $\varphi(z)$ instead of $\sigma\|z\|_1$ in subproblem (3.1), then (3.1) amounts to

$$\widetilde{z}^k = \arg\min_{z \in \mathbb{R}^n} \left\{ \varphi(z) + \frac{\gamma + \rho_2}{2} \left\| z - \frac{1}{\gamma + \rho_2}\left(\gamma x^k - \lambda^k + \rho_2 z^k\right)\right\|^2\right\}.$$

By invoking the definition of the proximity operator given in (2.2), the above scheme can be rewritten as

$$\widetilde{z}^k = \mathbf{prox}_{\frac{1}{\gamma+\rho_2}\varphi}\left(\omega^k\right) \quad \text{with} \quad \omega^k := \frac{1}{\gamma + \rho_2}\left(\gamma x^k - \lambda^k + \rho_2 z^k\right).$$

Therefore, Algorithm 1 is easily implementable as long as the above proximity operator has explicit representation. We refer the reader to [23] for more details of proximity operators.

Remark 3.2 The direction $\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)$ defined by (3.10) apparently involves the matrix inverse of M. This is however not a computational worry since M is a block scalar matrix and its inverse is trivially obtained.

Remark 3.3 Even if we deal with the special case of $\mathcal{Q} := \{b\}$ in the ADMM algorithm (1.8)–(1.10), the second subproblem (1.9) is still not easy to solve under the assumption $\mathcal{C} \neq \mathbb{R}^n$ or $A \neq I_n$. When we further consider special cases $\mathcal{C} \subset \mathbb{R}_+^n$, an interesting work [61] showed that the resulting problem can be recast as

$$\min_{x,z} \sigma\left(z^\top e\right) + \frac{1}{2}\|Ax - b\|^2 \\ \text{s.t.} \ x = z, \ x \in \mathbb{R}^n, \ z \in \mathcal{C}, \tag{3.11}$$

where $e := (1, 1, \ldots, 1)^{\top}$. Accordingly, applying ADMM to (3.11) yields an implementable algorithm such that all subproblems have closed-form solutions. However, for general settings on $\mathcal{C}$ and $\mathcal{Q}$, we can not directly formulate $\|z\|_1$ as $(z^{\top}e)$ and the nonlinearity of $f(x) = \frac{1}{2}\|Ax - P_{\mathcal{Q}}(Ax)\|^2$ also results in a subproblem without explicit representation. Comparatively speaking, our formulation for (1.6) and linearization are more suitable for general convex sets $\mathcal{C}$ and $\mathcal{Q}$.

3.2 Convergence Analysis

In this section we prove the global convergence of Algorithm 1. We begin with stating a fundamental fact derived from the first-order optimality condition, that is, problem (1.7) is equivalent to the *variational inequality* (VI) of finding a vector $\mathbf{w}^* \in \Omega$ with the property

$$\langle \mathbf{w} - \mathbf{w}^*, \; F(\mathbf{w}^*) \rangle \geq 0, \quad \mathbf{w} \in \Omega, \tag{3.12a}$$

with

$$\mathbf{w} := \begin{pmatrix} x \\ z \\ \lambda \end{pmatrix}, \quad F(\mathbf{w}) := \begin{pmatrix} \nabla f(x) - \lambda \\ \zeta + \lambda \\ x - z \end{pmatrix} \quad \text{and} \quad \Omega := \mathcal{C} \times \mathbb{R}^n \times \mathbb{R}^n, \tag{3.12b}$$

where $\zeta \in \partial(\sigma \| \cdot \|_1)(z)$. The global convergence of Algorithm 1 will be established within the VI framework.

Lemma 3.1 *Suppose that $\mathbf{w}^*$ is a solution of (3.12). Then, we have*

$$\langle \mathbf{w}^k - \mathbf{w}^*, \; M\mathbf{d}(\mathbf{w}^k, \widetilde{\mathbf{w}}^k) \rangle \geq \phi(\mathbf{w}^k, \widetilde{\mathbf{w}}^k). \tag{3.13}$$

Proof First, the iterative scheme (3.4) can be easily recast into a VI, that is,

$$\langle x - \widetilde{x}^k, \; \nabla f(x^k) + (\rho_1 + \gamma)(\widetilde{x}^k - x^k) + \gamma x^k - \gamma \widetilde{z}^k - \lambda^k \rangle \geq 0, \quad x \in \mathcal{C}.$$

Using (3.5) and rearranging terms of the above inequality yields

$$\langle x - \widetilde{x}^k, \; \nabla f(\widetilde{x}^k) - \widetilde{\lambda}^k + (\rho_1 + \gamma)(\widetilde{x}^k - x^k) + \eta_1(x^k, \widetilde{x}^k) \rangle \geq \gamma \langle x - \widetilde{x}^k, \; \widetilde{x}^k - x^k \rangle, \tag{3.14}$$

where $\eta_1(x^k, \widetilde{x}^k)$ is given in (3.6).

Similarly, it follows from (3.1) and, respectively, (3.5) that

$$\begin{cases} \langle z - \widetilde{z}^k, \; \zeta^k + \widetilde{\lambda}^k + \rho_2(\widetilde{z}^k - z^k) \rangle \geq \gamma \langle \widetilde{z}^k - z, \; \widetilde{x}^k - x^k \rangle, \quad z \in \mathbb{R}^n, & (3.15) \\ \langle \lambda - \widetilde{\lambda}^k, \; \widetilde{x}^k - \widetilde{z}^k + \dfrac{1}{\gamma}(\widetilde{\lambda}^k - \lambda^k) \rangle = 0, \quad \lambda \in \mathbb{R}^n, & (3.16) \end{cases}$$

where $\zeta^k \in \partial(\sigma \| \cdot \|_1)(z^k)$. Upon summing up (3.14)–(3.16), we can rewrite it into a compact form as follows

$$\langle \mathbf{w} - \widetilde{\mathbf{w}}^k, \; F(\widetilde{\mathbf{w}}^k) - M\mathbf{d}(\mathbf{w}^k, \widetilde{\mathbf{w}}^k) \rangle \geq \gamma \langle \widetilde{z}^k - \widetilde{x}^k + x - z, \; \widetilde{x}^k - x^k \rangle, \quad \mathbf{w} \in \Omega. \tag{3.17}$$

Setting now $\mathbf{w} := \mathbf{w}^*$ in the above inequality and noticing the fact that $x^* = z^*$, we get

$$\langle \mathbf{w}^* - \widetilde{\mathbf{w}}^k, \; F(\widetilde{\mathbf{w}}^k) - M\mathbf{d}(\mathbf{w}^k, \widetilde{\mathbf{w}}^k) \rangle \geq \langle \widetilde{\lambda}^k - \lambda^k, \; \widetilde{x}^k - x^k \rangle.$$

Equivalently,

$$\left\langle \widetilde{\mathbf{w}}^k - \mathbf{w}^*, \ M\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\right\rangle \geq \left\langle \widetilde{\mathbf{w}}^k - \mathbf{w}^*, \ F\left(\widetilde{\mathbf{w}}^k\right)\right\rangle + \left\langle \widetilde{\lambda}^k - \lambda^k, \ \widetilde{x}^k - x^k\right\rangle. \tag{3.18}$$

Since f and $\sigma\| \cdot \|_1$ are convex functions, it is easy to derive that the function F defined in (3.12) is monotone. It turns out that

$$\left\langle \widetilde{\mathbf{w}}^k - \mathbf{w}^*, \ F\left(\widetilde{\mathbf{w}}^k\right)\right\rangle \geq \left\langle \widetilde{\mathbf{w}}^k - \mathbf{w}^*, \ F\left(\mathbf{w}^*\right)\right\rangle \geq 0.$$

This together with (3.18) further implies that

$$\left\langle \widetilde{\mathbf{w}}^k - \mathbf{w}^*, \ M\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\right\rangle \geq \left\langle \widetilde{\lambda}^k - \lambda^k, \ \widetilde{x}^k - x^k\right\rangle.$$

The desired inequality (3.13) now immediately follows from the above inequality and the definition of $\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)$ given by (3.9).   □

Indeed, the above result implies that $-\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)$ is a descent search direction of the unknown distance function $\frac{1}{2}\|\mathbf{w} - \mathbf{w}^*\|_M^2$ at the point $\mathbf{w} = \mathbf{w}^k$. Below, we show that the stepsize defined by (3.8) is reasonable. To see this we define two functions

$$\mathbf{w}^{k+1}(\alpha) := \mathbf{w}^k - \alpha\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) \tag{3.19}$$

and

$$\Theta(\alpha) := \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - \|\mathbf{w}^{k+1}(\alpha) - \mathbf{w}^*\|_M^2. \tag{3.20}$$

Note that the function $\Theta(\alpha)$ may be viewed as the progress-function to measure the improvement obtained at the k-th iteration of Algorithm 1. We need to maximize $\Theta(\alpha)$ for seeking an optimal improvement.

Lemma 3.2 *Suppose that $\mathbf{w}^*$ is a solution of (3.12). Then we have*

$$\Theta(\alpha) \geq q(\alpha), \tag{3.21}$$

where

$$q(\alpha) = 2\alpha\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) - \alpha^2\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2.$$

Proof It follows from (3.19) and (3.20) that

$$\begin{aligned}
\Theta(\alpha) &= \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - \|\mathbf{w}^{k+1}(\alpha) - \mathbf{w}^*\|_M^2 \\
&= \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - \|\mathbf{w}^k - \alpha\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) - \mathbf{w}^*\|_M^2 \\
&= 2\alpha\left\langle \mathbf{w}^k - \mathbf{w}^*, \ M\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\right\rangle - \alpha^2\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2 \\
&\geq 2\alpha\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) - \alpha^2\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2,
\end{aligned}$$

where the last inequality follows from Lemma 3.1.   □

Since $q(\alpha)$ is a quadratic function of α, it is easy to find that $q(\alpha)$ attains its maximum value at the point

$$\alpha_k := \frac{\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)}{\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2}. \tag{3.22}$$

It then turns out that, for any relaxation factor $\tau > 0$,

$$\Theta(\tau\alpha_k) \geq 2\tau\alpha_k\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) - \tau^2\alpha_k^2\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2$$
$$= \tau(2-\tau)\alpha_k\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right).$$

To ensure that an improvement can be obtained at each iteration, we should limit $\tau \in (0, 2)$ so that the right-hand side in the above relation is positive. However, we suggest to choose $\tau \in [1, 2)$ for fast convergence in practice (see Sect. 4).

Lemma 3.3 *Suppose that $\rho_1 \geq \|A\|^2$. Then, the step size α_k defined by (3.8) is bounded away below from zero; that is, $\inf_{k\geq 1} \alpha_k \geq \alpha_{\min} > 0$ for some positive constant $\alpha_{\min}$.*

Proof An appropriate application of the Cauchy–Schwartz inequality

$$2\langle\mathbf{a}, \mathbf{b}\rangle \leq \kappa\|\mathbf{a}\|^2 + \frac{1}{\kappa}\|\mathbf{b}\|^2, \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^n, \ \kappa > 0$$

implies that

$$\left\langle\widetilde{\lambda}^k - \lambda^k, \ \widetilde{x}^k - x^k\right\rangle \leq \frac{1}{2\gamma}\|\widetilde{\lambda}^k - \lambda^k\|^2 + \frac{\gamma}{2}\|\widetilde{x}^k - x^k\|^2.$$

Consequently, it follows from (3.9) and (3.10) that

$$\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) = \left\langle\mathbf{w}^k - \widetilde{\mathbf{w}}^k, \ M\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\right\rangle + \left\langle\widetilde{\lambda}^k - \lambda^k, \ \widetilde{x}^k - x^k\right\rangle$$
$$\geq \left(\rho_1 + \frac{\gamma}{2}\right)\|\widetilde{x}^k - x^k\|^2 + \rho_2\|\widetilde{z}^k - z^k\|^2$$
$$+ \frac{1}{2\gamma}\|\widetilde{\lambda}^k - \lambda^k\|^2 - \left\langle x^k - \widetilde{x}^k, \ \eta_1\left(x^k, \widetilde{x}^k\right)\right\rangle$$
$$\geq \left(\rho_1 - \|A\|^2 + \frac{\gamma}{2}\right)\|\widetilde{x}^k - x^k\|^2 + \rho_2\|\widetilde{z}^k - z^k\|^2 + \frac{1}{2\gamma}\|\widetilde{\lambda}^k - \lambda^k\|^2$$
$$\geq C_{\min}\|\widetilde{\mathbf{w}}^k - \mathbf{w}^k\|^2, \tag{3.23}$$

where the second inequality follows from the Cauchy–Schwartz inequality and the Lipschitz continuity of ∇f, and the last inequality from the introduction of the constant $C_{\min} := \min\left\{\frac{\gamma}{2}, \ \rho_2, \ \frac{1}{2\gamma}\right\}$.

On the other hand, it follows again from the Cauchy–Schwartz inequality and monotonicity of ∇f that

$$\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2 = \left\langle\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right), \ M\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\right\rangle$$
$$= \|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|_M^2 - 2\left\langle x^k - \widetilde{x}^k, \ \eta_1\left(x^k, \widetilde{x}^k\right)\right\rangle + \frac{1}{\rho_1 + \gamma}\|\eta_1\left(x^k, \widetilde{x}^k\right)\|^2$$
$$\leq \left(\rho_1 + \gamma + \frac{\|A\|^4}{\rho_1 + \gamma}\right)\|x^k - \widetilde{x}^k\|^2 + \rho_2\|z^k - \widetilde{z}^k\|^2 + \frac{1}{\gamma}\|\lambda^k - \widetilde{\lambda}^k\|^2$$
$$\leq C_{\max}\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|^2, \tag{3.24}$$

where $C_{\max} := \max\left\{\frac{(\rho_1+\gamma)^2+\|A\|^4}{\rho_1+\gamma}, \ \rho_2, \ \frac{1}{\gamma}\right\}$. Thus, combining (3.23) and (3.24) immediately yields

$$\alpha_k \geq \frac{C_{\min}}{C_{\max}} =: \alpha_{\min} > 0.$$

The assertion of this lemma is proved.   □

Theorem 3.1 *Suppose that $\rho_1 \geq \|A\|^2$ and let $\mathbf{w}^*$ be an arbitrary solution of (3.12). Then the sequence $\{\mathbf{w}^k\}$ generated by Algorithm 1 satisfies the following property*

$$\|\mathbf{w}^{k+1} - \mathbf{w}^*\|_M^2 \leq \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - \tau(2-\tau)\alpha_{\min}C_{\min}\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|^2. \tag{3.25}$$

Proof From (3.7), (3.22) and (3.23) together with Lemmas 3.1 and 3.3, it follows that

$$
\begin{aligned}
\|\mathbf{w}^{k+1} - \mathbf{w}^*\|_M^2 &= \|\mathbf{w}^k - \tau\alpha_k\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) - \mathbf{w}^*\|_M^2 \\
&= \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - 2\tau\alpha_k\left\langle\mathbf{w}^k - \mathbf{w}^*, \ M\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\right\rangle + \tau^2\alpha_k^2\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2 \\
&\leq \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - 2\tau\alpha_k\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) + \tau^2\alpha_k^2\|\mathbf{d}\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right)\|_M^2 \\
&= \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - \tau(2-\tau)\alpha_k\phi\left(\mathbf{w}^k, \widetilde{\mathbf{w}}^k\right) \\
&\leq \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - \tau(2-\tau)\alpha_{\min}C_{\min}\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|^2.
\end{aligned}
$$

We obtained the desired result.   □

Theorem 3.2 *Suppose that $\rho_1 \geq \|A\|^2$. Then, the sequence $\{\mathbf{w}^k\}$ generated by Algorithm 1 is globally convergent to a solution of VI (3.12).*

Proof Let $\mathbf{w}^*$ be a solution of (3.12). It is immediately clear from (3.25) that

$$\|\mathbf{w}^{k+1} - \mathbf{w}^*\|_M^2 \leq \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 \leq \cdots \leq \|\mathbf{w}^0 - \mathbf{w}^*\|_M^2. \tag{3.26}$$

That is, the sequence $\{\|\mathbf{w}^k - \mathbf{w}^*\|_M\}$ is decreasing. In particular, the sequence $\{\mathbf{w}^k\}$ is bounded and the

$$\lim_{k\to\infty} \|\mathbf{w}^k - \mathbf{w}^*\|_M \quad \text{exists.} \tag{3.27}$$

By rewriting (3.25) as

$$\tau(2-\tau)\alpha_{\min}C_{\min}\|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\|^2 \leq \|\mathbf{w}^k - \mathbf{w}^*\|_M^2 - \|\mathbf{w}^{k+1} - \mathbf{w}^*\|_M^2,$$

we immediately conclude that

$$\lim_{k\to\infty} \|\mathbf{w}^k - \widetilde{\mathbf{w}}^k\| = 0. \tag{3.28}$$

It turns out that $\lim_{k\to\infty}\|x^k - \widetilde{x}^k\| = 0$ and that the sequences $\{\mathbf{w}^k\}$ and $\{\widetilde{\mathbf{w}}^k\}$ have the same cluster points. To prove the convergence of the sequence $\{\mathbf{w}^k\}$, let $\{\mathbf{w}^{k_j}\}$ be a subsequence of $\{\mathbf{w}^k\}$ that converges to some point $\mathbf{w}^\infty$; hence, the subsequence $\{\widetilde{\mathbf{w}}^{k_j}\}$ of $\{\widetilde{\mathbf{w}}^k\}$ also converges to $\mathbf{w}^\infty$.

Now taking the limit as $j \to \infty$ over the subsequence $\{k_j\}$ in (3.17) and using the continuity of F, we arrive at

$$\left\langle \mathbf{w} - \mathbf{w}^\infty, \ F(\mathbf{w}^\infty)\right\rangle \geq 0, \quad \mathbf{w} \in \Omega.$$

Namely, $\mathbf{w}^\infty$ solves VI (3.12).

Finally, substituting $\mathbf{w}^\infty$ for $\mathbf{w}^*$ in (3.27) and using the fact that $\|\mathbf{w}^k - \mathbf{w}^\infty\|_M$ is convergent, we conclude that

$$\lim_{k\to\infty} \|\mathbf{w}^k - \mathbf{w}^\infty\|_M = \lim_{j\to\infty} \|\mathbf{w}^{k_j} - \mathbf{w}^\infty\|_M = 0.$$

This proves that the full sequence $\{\mathbf{w}^k\}$ converges to $\mathbf{w}^\infty$.   □

4 Numerical Experiments

The development of implementable algorithms tailored for ℓ_1-norm regularized SFP [(even if for its special case of (1.2)] is in infant stage. Our main goal here is not to investigate the competitiveness of this approach with other algorithms, but simply to illustrate the reliability and the sensitivity with respect to some involved parameters of the presented method. Certainly, we also report some results to support our motivation by comparing Algorithm 1 (abbreviated as 'ISM') with ADMM and the straightforward linearized ADMM (we use "LADMM" for short) corresponding to (3.2), (3.4) and (3.5). Specifically, we demonstrate the application of the three algorithms to an image deblurring problem and report some preliminary numerical results to achieve the aforementioned goals.

All codes were written by MATLAB 2010b and run on a Lenovo desktop computer with Intel Pentium Dual-Core processor 2.33GHz and 2Gb main memory.

The problem under test here is a wavelet-based image deblurring problem, which is a special case of (1.5) with a box constraint $\mathcal{C}$ and $\mathcal{Q} := \{b\}$. More concretely, the model for this problem can be formulated as:

$$\min_x \left\{ \sigma \|Wx\|_1 + \frac{1}{2}\|Ax - b\|^2 \mid x \in \mathcal{C} \right\}, \tag{4.1}$$

where b represents the (vectorized) observed image, the matrix $A := BW$ consists of a blur operator B and an inverse of a three stage Haar wavelet transform W, and the set $\mathcal{C}$ is a box area in $\mathbb{R}^n$. Indeed, if the image is considered with double precision entries, then the pixel values lie in $\mathcal{C} := [0, 1]$, and an 8-bit gray-scale image corresponds to $\mathcal{C} := [0, 255]$. Notice that the unconstrained alternative of (4.1), i.e., $\mathcal{C} := \mathbb{R}^n$, is more popular in the literature (see, for instance, [1,3,37]). However, as shown in [18,20,34,43], more precise restoration, especially for binary images, may occur by absorbing an additional box constraint $\mathcal{C}$ into the model. In this section, we are concerned with the constrained model (4.1) and investigate the behavior of Algorithm 1 for this problem.

In these experiments, we adopt the reflective boundary condition and set the regularization parameter σ as 2e−5. We test eight different images, which were scaled into the range between 0 and 1, i.e., $\mathcal{C} := [0, 1]$, and corrupted in the same way used in [3]. More specifically, each image was degraded by a Gaussian blur of size 9×9 and standard deviation 4 and corrupted by adding an additive zero-mean white Gaussian noise with standard deviation 10^{-3}. The original and degraded eight images are listed in Figs. 1 and 2, respectively.

We first define the usual signal-to-noise ratio (SNR) in decibel (dB) to measure the quality of the recovered image, i.e.,

$$\text{SNR} := 20 \log_{10} \frac{\|x\|}{\|\tilde{x} - x\|},$$

where x is a clean image and $\tilde{x}$ is a restored image. Clearly, a larger SNR value means that we restored a better image. Without loss of generality, we use

$$\frac{\|x^{k+1} - x^k\|}{\|x^k\|} \leq \text{tol} \tag{4.2}$$

to be the stopping criterion throughout the experiments. Throughout, we report the number of iterations (Iter.); computing time in seconds (Time), and SNR values (SNR).

We first investigate the numerical behavior of our ISM for (4.1). Notice that the global convergence of our ISM is built up under the assumption $\rho_1 \geq \|A\|^2$. We thus set $\rho_1 = 1.5\|A\|^2$ throughout the experiments. Below, we test 'chart.tiff' image and investigate the
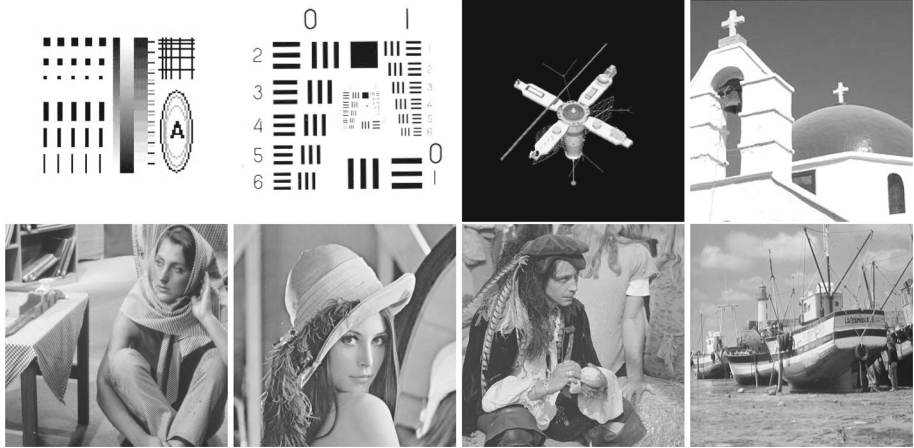
Fig. 1 Original images. *Left* to *right*, *top* to *bottom* shape.jpg (128×128); chart.tiff (256×256); satellite (256×256); church (256×256); barbara (512×512); lena (512×512); man (512×512); boat (512×512)

Fig. 2 Observed images

performance of the rest of parameters involved in our method. For the relaxation parameter τ, we set $\gamma = 0.02$, $\rho_2 = 0.01$, and test three cases of $\tau = \{1.0, 1.4, 1.8\}$. For the penalty parameter γ, we take $\tau = 1.8$, $\rho_2 = 0.01$ and then consider four cases of $\gamma = \{0.1, 0.02, 0.001, 0.0001\}$. Finally, we set $\tau = 1.8$ and $\gamma = 0.02$ to investigate the performance of ρ_2 with $\rho_2 = \{1, 0.05, 0.1, 0.01\}$. For each case, we plot the evolution of SNR with respect to iterations in Fig. 3.

It can be easily seen from the first plot in Fig. 3 that larger τ performs better than the smaller ones, which also supports the theoretical analysis in Sect. 3 [see Lemma 3.2 and (3.22)]. From the last two plots in Fig. 3, however, smaller γ and ρ_2 bring better performance relatively. Indeed, we can choose $\rho_2 = 0$ due to $\eta(\mathbf{w}, \widetilde{\mathbf{w}})$ defined in (3.6) can be simplified as $\eta_1(x, \widetilde{x})$. Of course a suitable $\rho_2 > 0$ may have benefits for stable performance in certain applications, because $\frac{\rho_2}{2}\|z - z^k\|^2$ serves a similar role as Tikhonov regularization. Though
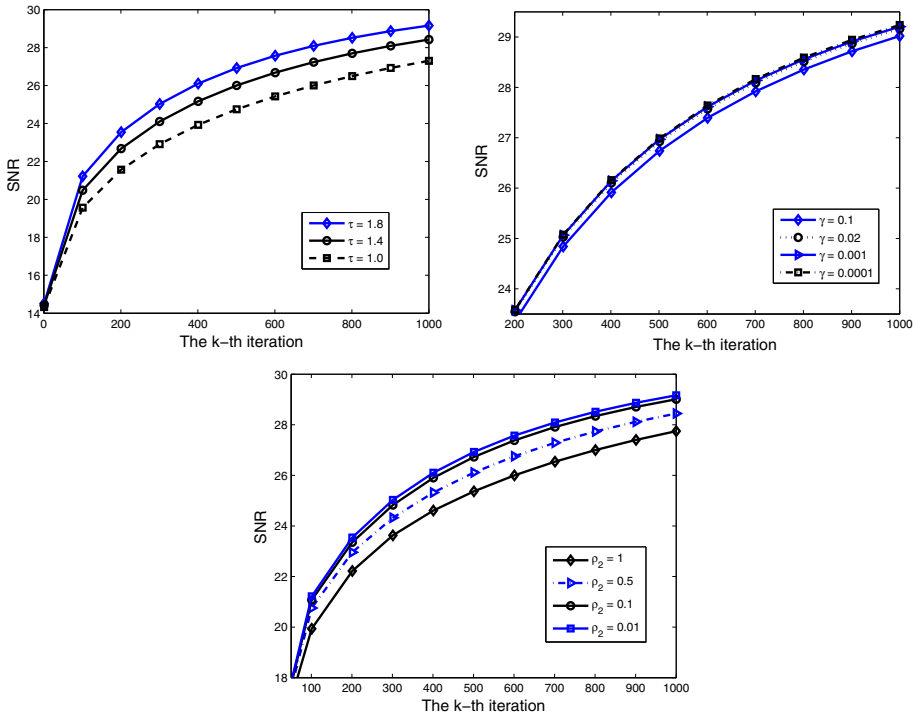
Fig. 3 Numerical performance of our ISM with different parameters

Table 1 Numerical performance of ISM under different tolerances 'tol'

Image	tol = 10^{-3}			tol = 5×10^{-4}			tol = 10^{-4}			tol = 5×10^{-5}		
	Iter.	Time	SNR	Iter.	Time	SNR	Iter.	Time	SNR	Iter.	Time	SNR
Shape	31	0.91	14.57	91	2.17	15.82	499	12.00	19.59	848	19.91	20.99
Chart	45	6.11	18.97	91	12.58	20.89	329	44.95	25.36	532	73.56	27.14
Satellite	50	7.02	12.93	104	14.36	13.62	458	62.89	15.19	788	108.80	15.64
Church	22	3.00	21.73	43	5.80	22.59	196	26.69	24.92	336	45.53	25.67
Barbara	16	9.78	17.87	31	18.86	18.08	206	126.55	18.88	377	232.00	19.11
Lena	16	9.72	23.32	28	16.95	23.76	150	92.69	25.30	276	170.63	25.79
Man	18	11.11	20.63	39	23.97	21.24	196	120.45	22.70	356	219.58	23.15
Boat	19	11.78	20.60	45	27.69	21.48	213	132.39	23.40	377	233.78	24.00

we can not demonstrate (possibly impossible, indeed) the best choices of the parameters for all problems, the results in Fig. 3 provide a direction for us to choose them empirically.

According to the above data, we now take $\tau = 1.8$, $\gamma = 0.02$, $\rho_2 = 0.01$ and report the numerical results of our ISM under different stopping tolerances 'tol'. More specifically, we set four different tolerances tol = $\{10^{-3}, 5 \times 10^{-4}, 10^{-4}, 5 \cdot 10^{-5}\}$ in (4.2) and summarize the corresponding results in Table 1.

Table 2 Numerical comparison (tol $= 10^{-5}$)

Image	ISM			LADMM			ADMM		
	Iter.	Time	SNR	Iter.	Time	SNR	Iter. (InIt.)	Time	SNR
Shape	2404	55.78	23.09	2990	39.14	22.50	427 (8249)	76.34	24.09
Chart	1383	192.38	29.95	1786	134.56	29.16	185 (3759)	206.59	31.31
Satellite	2370	328.52	16.10	2905	218.52	16.02	388 (9735)	531.83	16.17
Church	1154	161.91	26.87	1307	98.02	26.52	354 (6906)	380.69	27.31
Barbara	1403	870.09	19.47	1569	771.97	19.37	478 (5223)	1619.06	19.83
Lena	1024	632.80	26.50	1126	554.08	26.30	297 (3292)	1013.45	26.78
Man	1260	780.11	23.78	1426	702.05	23.62	327 (3590)	1108.84	24.05
Boat	1300	798.52	24.90	1493	731.98	24.65	332 (3960)	1212.42	25.35

Fig. 4 Recovered images by the three methods. From *top* to *bottom*: ISM, LADMM and ADMM

The data in Table 1 show that the number of iterations is increasing significantly when we set a smaller tolerance 'tol'. However, our method is still efficient for the case of tol $\geq 5 \times 10^{-4}$. More importantly, our ISM is easily implementable due to its simple subproblems.

Hereafter, we turn to compare ISM with LADMM and ADMM showing the necessity of linearization and correction step. For the parameters involved in LADMM and ADMM, we follow the same settings of ISM, that is, $\rho_1 = 1.5\|A\|^2$ for LADMM and $\gamma = 0.02$ for both LADMM and ADMM. Here, we set the tolerance as tol $= 10^{-5}$ and reported the corresponding results in Table 2. Note that the subproblem (1.9) of ADMM does not have closed-form solution. We thus implemented the *projected Barzilai–Borwein method* in [25]

to pursuit an approximate solution to this problem and allow a maximal number of 30 for the inner loop. Accordingly, we also report the inner iteration (denoted by "InIt.") of ADMM in Table 2. The recovered images are listed in Fig. 4.

As shown in Table 2, we can see that ADMM requires fewest iterations to recover a satisfactory image. However, ADMM needs a time-consuming inner loop to pursuit an approximate solution such that ADMM takes more computing time than ISM and LADMM, which can be clearly seen from the number of inner iterations. Thus, the linearization is necessary for the problem under consideration. Comparing ISM and LADMM, we can see that ISM takes fewer iterations to get higher SNR than LADMM, but more computing time. This supports that our correction step can improve the performance of ISM in terms of iterations. Since the correction step requires to compute $\eta_1(x^k, \tilde{x}^k)$, ISM needs one more *discrete cosine transform* and its inverse, which thus takes more computing time than LADMM. From the recovered images in Fig. 4, the three methods can recover desirable images. In summary, all experiments in this section verify the efficiency and reliability of our ISM.

5 Conclusions

We have considered the ℓ_1-norm regularized SFP, which is an extension of the timely ℓ_1-norm regularized linear inverse problems, but has been received much less attention. Due to the nondifferentiability of ℓ_1-norm regularizer, many traditional iterative methods in the literature of SFP can not be directly applied to the ℓ_1-regularized model. Therefore, we have proposed a splitting algorithm, which exploits the structure of the problem such that its subproblems are simple enough to have closed-form solutions under the assumptions that projections onto the sets $\mathcal{C}$ and $\mathcal{Q}$ are readily computed. Our preliminary numerical results have supported the idea of the application of ℓ_1-norm regularization to SFP, and verified the reliability and efficiency of Algorithm 1. Recently, an extension of SFP, i.e., *multiple-set split feasibility problem* (MSFP), has been received considerable attention, e.g., see [13,54,62,63]. Therefore, to investigate the application of ℓ_1-norm regularization to MSFP and the applicability of the proposed algorithm to the regularized MSFP is our future work. On the other hand, as stated in [38], larger regularization parameter σ in (1.6) will yields sparser solutions, and it is difficult to choose σ that fits well for all problems. Thus, how to choose a suitable σ automatically in algorithmic implementation is also our further investigation.

References

1. Afonso, M., Bioucas-Dias, J., Figueiredo, M.: Fast image recovery using variable splitting and constrained optimization. IEEE Trans. Image Process. **19**, 2345–2356 (2010)
2. Bai, Z., Chen, M., Yuan, X.: Applications of the alternating direction method of multipliers to the semi-definite inverse quadratic eigenvalue problems. Inverse Probl. **29**, 075,011 (2013)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)
4. Beck, A., Teboulle, M.: Gradient-based algorithms with applications to signal recovery problems. In: Palomar, D.P., Eldar, Y.C. (eds.) Convex Optimization in Signal Processing and Communications, pp. 33–88. Cambridge University Press, New York (2010)

5. Bertsekas, D., Tsitsiklis, J.: Parallel and Distributed Computation, Numerical Methods. Prentice-Hall, Englewood Cliffs, NJ (1989)
6. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–122 (2010)
7. Byrne, C.: Iterative oblique projection onto convex sets and the split feasibility problem. Inverse Probl. **18**, 441–453 (2002)
8. Byrne, C.: A unified treatment of some iterative algorithms in signal processing and image reconstruction. Inverse Probl. **20**, 103–120 (2004)
9. Cai, J., Osher, S., Shen, Z.: Linearized Bregman iterations for compressed sensing. Math. Comput. **78**, 1515–1536 (2009)
10. Ceng, L., Ansari, Q., Yao, J.: Relaxed extragradient methods for finding minimum-norm solutions of the split feasibility problem. Nonlinear Anal. **75**, 2115–2116 (2012)
11. Censor, Y., Bortfeld, T., Martin, B., Trofimov, A.: A unified approach for inversion problems in intensity-modulated radiation therapy. Phys. Med. Biol. **51**, 2353–2365 (2006)
12. Censor, Y., Elfving, T.: A multiprojection algorithm using Bregman projections in product space. Numer. Algorithms **8**, 221–239 (1994)
13. Censor, Y., Elfving, T., Kopf, N., Bortfeld, T.: The multiple-sets split feasibility problem and its applications for inverse problems. Inverse Probl. **21**, 2071–2084 (2005)
14. Censor, Y., Gibali, A., Reich, S.: Algorithms for the split variational inequality problem. Numer. Algorithms **59**, 301–323 (2012)
15. Censor, Y., Jiang, M., Wang, G.: Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning, and Inverse Problems. Medical Physics Publishing Madison, Wisconsin (2010)
16. Censor, Y., Motova, A., Segal, A.: Perturbed projections and subgradient projections for the multiple-sets split feasibility problem. J. Math. Anal. Appl. **327**, 1244–1256 (2007)
17. Censor, Y., Segal, A.: Iterative projection methods in biomedical inverse problems. In: Censor, Y., Jiang, M., Louis, A. (eds.) Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT), pp. 65–96. Edizioni della Normale, Pisa (2008)
18. Chan, R., Tao, M., Yuan, X.: Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers. SIAM J. Imaging Sci. **6**, 680–697 (2013)
19. Chan, R., Yang, J., Yuan, X.: Alternating direction method for image inpainting in wavelet domain. SIAM J. Imaging Sci. **4**, 807–826 (2011)
20. Chan, R.H., Tao, M., Yuan, X.: Linearized alternating direction method of multipliers for constrained linear least squares problems. East Asian J. Appl. Math. **2**, 326–341 (2012)
21. Chen, C., He, B., Yuan, X.: Matrix completion via alternating direction method. IMA J. Numer. Anal. **32**, 227–245 (2012)
22. Chen, S., Donoho, D., Saunders, M.: Automatic decomposition by basis pursuit. SIAM Rev. **43**, 129–159 (2001)
23. Combettes, P., Pesquet, J.: Proximal splitting methods in signal processing. In: Bauschke, H., Burachik, R., Combettes, P., Elser, V., Luke, D., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and Its Applications, vol. 49, pp. 185–212. Springer, New York (2011)
24. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Simul. **4**, 1168–1200 (2005)
25. Dai, Y., Fletcher, R.: Projected Barzilai–Borwein methods for large-scale box-constrained quadratic programming. Numerische Mathematik **100**, 21–47 (2005)
26. Dang, Y., Gao, Y.: The strong convergence of a KM-CQ-like algorithm for a split feasibility problem. Inverse Probl. **27**, 015007 (2011)
27. Donoho, D.: Compressed sensing. IEEE Trans. Inform. Theory **52**, 1289–1306 (2006)
28. Eckstein, J., Bertsekas, D.: On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
29. Eicke, B.: Iteration methods for convexly constrained ill-posed problems in Hilbert space. Numer. Funct. Anal. Optim. **13**, 413–429 (1992)
30. Engl, H., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer Academic Publishers Group, Dordrecht, The Netherlands (1996)
31. Figueiredo, M., Nowak, R., Wright, S.: Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. IEEE J. Sel. Top. Signal Process. **1**, 586–598 (2007)
32. Frick, K., Grasmair, M.: Regularization of linear ill-posed problems by the augmented Lagrangian method and variational inequalities. Inverse Probl. **28**, 104,005 (2012)
33. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Comput. Math. Appl. **2**, 16–40 (1976)

34. Han, D., He, H., Yang, H., Yuan, X.: A customized Douglas–Rachford splitting algorithm for separable convex minimization with linear constraints. Numer. Math. **127**, 167–200 (2014)
35. He, B., Liao, L., Han, D., Yang, H.: A new inexact alternating direction method for monotone variational inequalities. Math. Program. **92**, 103–118 (2002)
36. He, B., Xu, M., Yuan, X.: Solving large-scale least squares covariance matrix problems by alternating direction methods. SIAM J. Matrix Anal. Appl. **32**, 136–152 (2011)
37. He, B., Yuan, X., Zhang, W.: A customized proximal point algorithm for convex minimization with linear constraints. Comput. Optim. Appl. **56**, 559–572 (2013)
38. He, S., Zhu, W.: A note on approximating curve with 1-norm regularization method for the split feasibility problem. J. Appl. Math. **2012**, Article ID 683,890, 10 pp (2012)
39. Hochstenbach, M., Reichel, L.: Fractional Tikhonov regularization for linear discrete ill-posed problems. BIT Numer. Math. **51**, 197–215 (2011)
40. Ma, S.: Alternating direction method of multipliers for sparse principal component analysis. J. Oper. Res. Soc. China **1**, 253–274 (2013)
41. Martinet, B.: Régularization d' inéquations variationelles par approximations sucessives. Rev. Fr. Inform. Rech. Opér. **4**, 154–159 (1970)
42. Moreau, J.: Fonctions convexe duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Ser. A Math. **255**, 2897–2899 (1962)
43. Morini, S., Porcelli, M., Chan, R.: A reduced Newton method for constrained linear least squares problems. J. Comput. Appl. Math. **233**, 2200–2212 (2010)
44. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**, 123–231 (2013)
45. Potter, L., Arun, K.: A dual approach to linear inverse problems with convex constraints. SIAM J. Control Optim. **31**, 1080–1092 (1993)
46. Qu, B., Xiu, N.: A note on the CQ algorithm for the split feasibility problem. Inverse Probl. **21**, 1655–1665 (2005)
47. Rockafellar, R.: On the maximal monotonicity of subdifferential mappings. Pac. J. Math. **33**, 209–216 (1970)
48. Sabharwal, A., Potter, L.: Convexly constrained linear inverse problems: iterative least-squares and regularization. IEEE Trans. Signal Process. **46**, 2345–2352 (1998)
49. Schopfer, F., Louis, A., Schuster, T.: Nonlinear iterative methods for linear ill-posed problems in Banach spaces. Inverse Probl. **22**, 311–329 (2006)
50. Tibshirani, R.: Regression shrinkage and selection via the LASSO. J. R. Stat. Soc. Ser. B **58**, 267–288 (1996)
51. Tikhonov, A., Arsenin, V.: Solutions of Ill-Posed Problems. Winston, New York (1977)
52. Wang, X., Yuan, X.: The linearized alternating direction method of multipliers for Dantzig selector. SIAM J. Sci. Comput. **34**, A2792–A2811 (2012)
53. Wright, S., Nowak, R., Figueiredo, M.: Sparse reconstruction by separable approximation. IEEE Trans. Signal Process. **57**, 2479–2493 (2009)
54. Xu, H.: A variable Krasnosel'skii–Mann algorithm and the multiple-set split feasibility problem. Inverse Probl. **22**, 2021–2034 (2006)
55. Xu, H.: Iterative methods for the split feasibility problem in infinite dimensional Hilbert spaces. Inverse Probl. **26**, 105,018 (2010)
56. Yang, J., Zhang, Y.: Alternating direction algorithms for ℓ_1-problems in compressive sensing. SIAM J. Sci. Comput. **332**, 250–278 (2011)
57. Yang, Q.: The relaxed CQ algorithm solving the split feasibility problem. Inverse Probl. **20**, 1261–1266 (2004)
58. Yao, Y., Wu, J., Liou, Y.: Regularized methods for the split feasibility problem. Abstr. Appl. Anal. **2012**, Article ID 140,679, 13 pp (2012)
59. Yuan, X.: Alternating direction methods for covariance selection models. J. Sci. Comput. **51**, 261–273 (2012)
60. Zhang, H., Wang, Y.: A new CQ method for solving split feasibility problem. Front. Math. China **5**, 37–46 (2010)
61. Zhang, J., Morini, B.: Solving regularized linear least-squares problems by the alternating direction method with applications to image restoration. Electron. Trans. Numer. Anal. **40**, 356–372 (2013)
62. Zhang, W., Han, D., Li, Z.: A self-adaptive projection method for solving the multiple-sets split feasibility problem. Inverse Probl. **25**, 115,001 (2009)
63. Zhang, W., Han, D., Yuan, X.: An efficient simultaneous method for the constrained multiple-sets split feasibility problem. Comput. Optim. Appl. **52**, 825–843 (2012)