Application of the Strictly Contractive Peaceman-Rachford Splitting Method to Multi-block Separable Convex Programming

Bingsheng He[1], Han Liu[2], Juwei Lu[3], and Xiaoming Yuan[4]

May 17, 2014

Abstract. Recently, a strictly contractive Peaceman-Rachford splitting method (SC-PRSM) was proposed to solve a convex minimization model with linear constraints and a separable objective function which is the sum of two functions without coupled variables. We show by an example that the SC-PRSM cannot be directly extended to the case where the objective function is the sum of three or more functions. To solve such a multi-block model, if we treat its variables and functions as two groups and directly apply the SC-PRSM, then at least one of SC-PRSM's subproblems involves more than one function and variable which might not be easy to solve. One way to improve the solvability for this direct application of the SC-PRSM is to further decompose such a subproblem so as to generate easier decomposed subproblems which could potentially be easy enough to have closed-form solutions for some specific applications. The curse accompanying this improvement in solvability is that the SC-PRSM with further decomposed subproblems is not necessarily convergent, either. We will show its divergence by the same example. Our main goal is to show that the convergence can be guaranteed if the further decomposed subproblems of the direct application of the SC-PRSM are regularized by the proximal regularization. As a result, an SC-PRSM-based splitting algorithm with provable convergence and easy implementability is proposed for multi-block convex minimization models. We analyze the convergence for the derived algorithm, including proving its global convergence and establishing its worst-case convergence rate measured by the iteration complexity. The efficiency of the new algorithm is illustrated by testing some applications arising in image processing and statistical learning.

Keywords. Convex Programming, Peaceman-Rachford Splitting Method, Convergence Rate, Contraction Methods, Proximal Point Algorithm, Separable Models.

1 Introduction

We first consider a convex minimization model with linear constraints and an objective function in form of the sum of two functions without coupled variables:

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\, x \in \mathcal{X}, y \in \mathcal{Y}\}, \tag{1.1}$$

where $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, $\mathcal{X} \subset \Re^{n_1}$ and $\mathcal{Y} \subset \Re^{n_2}$ are closed convex sets, θ_1 and θ_2 are convex but not necessarily smooth functions. A typical application of (1.1) is that θ_1 refers to a data-fidelity term and θ_2 denotes a regularization term. Concrete applications of the model (1.1) arise frequently

[1]International Centre of Management Science and Engineering, School of Management and Engineering, and Department of Mathematics, Nanjing University, Nanjing, 200093, China; E-mail: `hebma@nju.edu.cn`. This author was supported by the NSFC grant 91130007 and the MOEC fund 20110091110004.

[2]Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; E-mail: `hanliu@princeton.edu`. This author was supported by NSF Grant III–1116730.

[3]Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA; E-mail: `junweil@princeton.edu`.

[4]Department of Mathematics, Hong Kong Baptist University, Hong Kong; E-mail: `xmyuan@hkbu.edu.hk`. This author was supported by a General Research Fund from Research Grants Council of Hong Kong.

in many areas such as image processing, statistical learning, computer vision, etc., where θ_1 and θ_2 could be further specified by particular physical or industrial elaboration for a given scenario.

A benchmark solver for (1.1) is the alternating direction method of multipliers (ADMM) originally proposed in [25] (see also [7, 22]). The ADMM iterative scheme for (1.1) reads as

$$\begin{cases} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\big\}, \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.2}$$

where $\lambda \in \Re^m$ is the Lagrange multiplier and $\beta > 0$ is a penalty parameter. Throughout we assume that the penalty parameter β is fixed. The ADMM scheme (1.2) can be regarded as a splitting version of the augmented Lagrangian method (ALM) in [34, 47]; and it outperforms the direct application of the ALM in that the functions θ_1 and θ_2 are treated individually and thus the splitted subproblems in (1.2) could be much easier than the original ALM subproblems. Recently, this feature has found impressive applications in a variety of areas, and it has inspired a "renaissance" of the ADMM in the literature. We refer to [5, 14, 23] for some review papers of the ADMM.

As analyzed intensively in [21, 26], the ADMM could be regarded as an application of the Douglas-Rachford splitting method (DRSM) in [13, 37] which is well known in PDE literature. Some authors (see e.g. [27, 36]) thus have investigated how to apply the Peacemen-Rachford splitting method (PRSM) in [37, 45], another equally known operator splitting method as the DRSM which has received wide attention in PDE literature already, to the separable convex minimization model (1.1). This motivation could be further convinced by the observation "PRSM is always faster than the DRSM whenever it is convergent", as commented in [3, 24, 26]. More specifically, the PRSM iterative scheme for (1.1) can be written as

$$\begin{cases} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.3}$$

which differs from the ADMM scheme (1.2) in that it updates the Lagrange multipliers twice at each iteration. The PRSM has the disadvantage of "less stable than ADMM", although it does outperform ADMM whenever it is convergent, as elaborated in [37, 24]. In [29], this disadvantage was explained as that the sequence generated by the PRSM is not necessarily strictly contractive with respective to the solution set of (1.1) (under the defaulted assumption that the solution set of (1.1) be nonempty), while the sequence of ADMM is. Note that we follow the definition of a strictly contractive sequence in [4]. In [11], a counterexample showing that the sequence generated by the PRSM could maintain a constant distance to the solution set was constructed. Thus, the PRSM (1.3) is not necessarily convergent. To reinforce the PRSM (1.3) with provable convergence, the following strictly contractive PRSM (SC-PRSM) was proposed in [29]:

$$\begin{cases} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\big\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.4}$$

where the parameter $\alpha \in (0, 1)$. It was shown in [29] that the parameter α plays the role of enforcing the sequence generated by (1.4) to be strictly contractive with respect to the solution set of (1.1). Hence, the convergence of the SC-PRSM (1.4) can be proved by standard techniques in the literature (e.g [4]). In [29], the efficiency of the SC-PRSM (1.4) was also verified numerically.

In addition to the model (1.1), we encounter numerous applications where the objective function has a higher degree of separability such that it can be expressed as the sum of more than two functions without coupled variables. To expose our main idea with easier notation, let us only focus on the case with three functions in the objective

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \mid Ax + By + Cz = b,\, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}, \tag{1.5}$$

where $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$, $C \in \Re^{m\times n_3}$, $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$ and $\mathcal{Y} \subset \Re^{n_2}$, $\mathcal{Z} \subset \Re^{n_3}$ are closed convex sets, θ_i $(i = 1, 2, 3)$ are convex functions. Throughout, the solution set of (1.5) is assumed to be nonempty. Some typical applications in form of (1.5) include the robust principal component analysis model with noisy and incomplete data in [50], the latent variable Gaussian graphical model selection in [8], the robust alignment model for linearly correlated images in [46], the quadratic discriminant analysis model in [40] and many others.

To solve (1.5), one natural idea is to directly extend the ADMM (1.2). The resulting scheme is

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y^{k+1}, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{1.6}$$

where $\mathcal{L}_\beta(x, y, z, \lambda)$ is the augmented Lagrangian function of (1.5) defined as

$$\mathcal{L}_\beta(x, y, z, \lambda) := \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b) + \frac{\beta}{2}\|Ax + By + Cz - b\|^2$$

with $\lambda \in \Re^m$ the Lagrange multiplier and $\beta > 0$ the penalty parameter. This direct extension of ADMM (1.6) (denoted by "E-ADMM") perfectly inherits the advantage of the original ADMM (1.2), and it can be obtained by simply decomposing the ALM subproblem into 3 subproblems in Gauss-Seidel manner at each iteration. Empirically, it often works very well, see e.g. [46, 50] for some applications. However, it was shown in [10] that the E-ADMM (1.6) is not necessarily convergent. We refer to [30, 31] for some methods whose main common idea is to ensure the convergence via correcting the output of (1.6) appropriately.

Similarly, for solving the multi-block convex minimization model (1.5), we may wish to consider directly extending the SC-PRSM scheme (1.4) as

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - (\lambda^k)^T(Ax + By^k + Cz^k - b) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{3}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k + Cz^k - b), \\ y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^{k+\frac{1}{3}})^T(Ax^{k+1} + By + Cz^k - b) + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+\frac{2}{3}} = \lambda^{k+\frac{1}{3}} - \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^k - b), \\ z^{k+1} = \arg\min\{\theta_3(z) - (\lambda^{k+\frac{2}{3}})^T(Ax^{k+1} + By^{k+1} + Cz - b) + \frac{\beta}{2}\|Ax^{k+1} + By^{k+1} + Cz - b\|^2 \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{2}{3}} - \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases} \tag{1.7}$$

We denote by "E-SC-PRSM" the scheme (1.7) hereafter. Our first purpose is to show that the E-SC-PRSM (1.7) is not necessarily convergent, see Section 5.2. The technique is similar as that in [10]. With the divergence, the E-SC-PRSM (1.7) thus cannot be used directly for (1.5).

Alternatively, one may wish to use the original SC-PRSM scheme (1.4) directly by regarding $\theta_2(y) + \theta_3(z)$ as the second function in (1.1) and regrouping (y, z) and (B, C) as the second variable and coefficient matrix in (1.1), respectively. The direct application of the SC-PRSM (1.4) to (1.5) results in the following iterative scheme:

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - (\lambda^k)^T(Ax + By^k + Cz^k - b) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k + Cz^k - b), \\ (y^{k+1}, z^{k+1}) = \arg\min\left\{ \begin{array}{c} \theta_2(y) + \theta_3(z) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By + Cz - b) \\ + \frac{\beta}{2}\|Ax^{k+1} + By + Cz - b\|^2 \mid y \in \mathcal{Y},\ z \in \mathcal{Z} \end{array} \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases} \tag{1.8}$$

Provided that both of the minimization subproblems in (1.8) are solved exactly, this direct application of the SC-PRSM is successful — its convergence is guaranteed automatically. This is the bless of applying the SC-PRSM directly to (1.5). For many concrete applications of (1.5) such as the mentioned ones, however, it is not wise to do so because the (y, z)-subproblem in (1.8) must treat θ_2 and θ_3 aggregately even though they both could be very simply. This is the curse accompanying the scheme (1.8). Under the assumption that each function θ_i in (1.5) is well structured or has some special properties in the sense that treating an minimization problem involving only one of them is easy (e.g., when the resolvent operator of $\partial\theta_i$ has a closed-form solution such as θ_i is the l_1-norm term), a nature idea to overcome the curse of (1.8) is to further decompose the (y, z)-subproblem in (1.8) in Jacobian style. Thus, the (y, z)-subproblem in (1.8) is solved approximately in the sense that the following two decomposed subproblems are required to solve instead:

$$\left\{\begin{array}{l} y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By + Cz^k - b) + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \;\big|\; y \in \mathcal{Y}\big\}, \\ z^{k+1} = \arg\min\big\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By^k + Cz - b) + \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 \;\big|\; z \in \mathcal{Z}\big\}. \end{array}\right. \tag{1.9}$$

These two subproblems in (1.9) are in general easier than the (y, z)-subproblem in (1.8), as each of them only involves one θ_i in its objective function. Another reason of implementing this decomposition in Jacobian style is that these two subproblems in (1.9) are eligible for parallel computation. This makes some particular senses for large scale cases of the model (1.5) arising from high dimension statistical learning problems or some image processing applications. With the further decomposition (1.9) for the (y, z)-subproblem in (1.8), the direct application of the SC-PRSM (1.4) to the model (1.5) becomes

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k + Cz^k - b) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k + Cz^k - b), \\ y^{k+1} = \arg\min\big\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By + Cz^k - b) + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 \;\big|\; y \in \mathcal{Y}\big\}, \\ z^{k+1} = \arg\min\big\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By^k + Cz - b) + \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 \;\big|\; z \in \mathcal{Z}\big\}. \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{array}\right. \tag{1.10}$$

Compare with (1.8), the scheme (1.10) is much more implementable because its subproblems are much easier. The properties of θ_i's, if any, can thus be fully exploited by the scheme (1.10). However, it is easy to understand that despite of the guaranteed convergence of (1.8), the convergence of (1.10) might not hold because the original (y, z)-subproblem in (1.8) is solved only approximately via (1.9). In Section 5.1, we will use the same example showing the divergence of the E-SC-PRSM (1.7) to show the divergence of (1.10). Thus, it is not rationale to use either the E-SC-PRSM (1.7) or the direct application of the SC-PRSM (1.10) with decomposed subproblem to solve the multi-block convex minimization model (1.5).

Our second purpose it show that the convergence of (1.10) can be guaranteed if the decomposed y- and z-subproblems in (1.10) are further regularized by quadratic proximal terms. This idea inspires us to propose the following SC-PRSM scheme with proximal regularization (SC-PRSM-PR)

$$\left\{\begin{array}{l} x^{k+1} = \arg\min\big\{\theta_1(x) - (\lambda^k)^T(Ax + By^k + Cz^k - b) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k + Cz^k - b), \\ y^{k+1} = \arg\min\left\{\begin{array}{l} \theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By + Cz^k - b) + \\ \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \end{array} \;\middle|\; y \in \mathcal{Y}\right\}, \\ z^{k+1} = \arg\min\left\{\begin{array}{l} \theta_3(z) - (\lambda^{k+\frac{1}{2}})^T(Ax^{k+1} + By^k + Cz - b) + \\ \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \end{array} \;\middle|\; z \in \mathcal{Z}\right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{array}\right. \tag{1.11}$$

where $\alpha \in (0, 1)$ and $\mu > \alpha$. Note that the added quadratic proximal terms $\frac{\mu\beta}{2}\|B(y - y^k)\|^2$ and $\frac{\mu\beta}{2}\|C(z - z^k)\|^2$ enjoy the same explanation of the original proximal point algorithm which has been well studied in the literature, see e.g. [11, 39, 48]. An intuitive illustration is that since the objective functions in (1.9) are only approximation tos the objective function in the (y, z)-subproblem in (1.8), we use the quadratic terms to control the proximity of the new iterate to the previous iterate. The requirement $\mu \geq \alpha$ is in certain sense to control such proximity. Note that the subproblems in (1.11) are of the same difficulty as that of those in (1.10); while the convergence of (1.11) can be rigorously proved (see Section 3).

As a customized application of the original SC-PRSM (1.4) to the specific multi-block convex minimization model (1.5), the SC-PRSM-PR (1.11) is equally implementable as (1.4) in the sense their subproblems are of the same level of difficulty. Moreover, we will show that the SC-PRSM-PR (1.11) is also globally convergent and its worst-case convergence rate measured by the iteration complexity in both the ergodic and a nonergodic senses can be established. Thus, besides of its implementability, the SC-PRSM-PR (1.11) also fully inherits the theoretical properties of the original SC-PRSM (1.4) established in [29]. This is the main task in Sections 3 and 4. In Section 5, as mentioned, we will construct an example to show the divergence of the E-SC-PRSM (1.7) and the direct application of the SC-PRSM (1.10). As explained, the SC-PRSM-PR (1.11) is motivated by the particular consideration of some concrete applications of the abstract model (1.5). Thus, we will test its efficiency for some useful applications arising in the image processing and statistical learning domains; and report some preliminary experiment results in Section 6. Finally, some concluding remarks are made in Section 7.

2 Preliminaries

In this section, we summarize some known results in the literature which are useful for our analysis later; and define some auxiliary variables which can simplify the notation of our analysis.

2.1 The Variational Inequality Reformulation of (1.5)

We first reformulate the multi-block convex minimization model (1.5) as the variational inequality (VI): Find $w^* = (x^*, y^*, z^*, \lambda^*)$ such that

$$\mathrm{VI}(\Omega, F, \theta): \qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{2.1a}$$

where

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \qquad \theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z),$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix} \qquad \text{and} \qquad \Omega := \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m. \tag{2.1b}$$

Obviously, the mapping $F(w)$ defined in (2.1b) is affine with a skew-symmetric matrix; it is thus monotone. We denote by Ω^* the solution set of $\mathrm{VI}(\Omega, F, \theta)$, and it is not empty under the nonempty assumption of the solution set of (1.5).

Note that x^k is not required to generate the new $(k+1)$-th iteration in all the ADMM- or PRSM-based schemes mentioned previously, see (1.2), (1.3), (1.4), (1.6), (1.7), (1.8), (1.10) and (1.11). That is, such a scheme only requires (y^k, z^k, λ^k) to generate the next new iterate. Thus, as mentioned in [5], x is an **intermediate** variable in all the mentioned ADMM- or PRSM-based schemes. For this reason, in the following analysis, we use the notations $v^k = (y^k, z^k, \lambda^k)$ and $\mathcal{V} = \mathcal{Y} \times \mathcal{Z} \times \Re^m$, and we let

$$\mathcal{V}^* := \{v^* = (y^*, z^*, \lambda^*) \,|\, w^* = (x^*, y^*, z^*, \lambda^*) \in \Omega^*\}.$$

2.2 Some Natation

We define some auxiliary variables in this subsection which will help us alleviate the notation in the convergence analysis and improve the presentation.

First of all, we introduce a new sequence $\{\tilde{w}^k\}$ by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{z}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ z^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \end{pmatrix}, \tag{2.2}$$

where $(x^{k+1}, y^{k+1}, z^{k+1})$ is generated by the scheme (1.11) from (y^k, z^k, λ^k). According, we have

$$\tilde{v}^k = \begin{pmatrix} \tilde{y}^k \\ \tilde{z}^k \\ \tilde{\lambda}^k \end{pmatrix}, \tag{2.3}$$

where $(\tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$ is defined in (2.2).

In fact, using the notation of $\lambda^{k+\frac{1}{2}}$ in (1.11), we have

$$\lambda^{k+1} = \lambda^k - 2\alpha\beta\big(Ax^{k+1} + \frac{1}{2}B(y^k + y^{k+1}) + \frac{1}{2}C(z^k + z^{k+1}) - b\big).$$

According to (2.2), because

$$x^{k+1} = \tilde{x}^k, \qquad y^{k+1} = \tilde{y}^k \qquad z^{k+1} = \tilde{z}^k,$$

we have

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b) \qquad \text{and} \qquad \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k). \tag{2.4}$$

By a manipulation, the update form of λ^{k+1} (1.11) can be represented as

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\beta\big[(A\tilde{x}^k + By^k + Cz^k - b) - B(y^k - \tilde{y}^k) - C(z^k - \tilde{z}^k)\big] \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\big[(\lambda^k - \tilde{\lambda}^k) - \beta B(y^k - \tilde{y}^k) - \beta C(z^k - \tilde{z}^k)\big] \\ &= \lambda^k - \big[2\alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\beta B(y^k - \tilde{y}^k) - \alpha\beta C(z^k - \tilde{z}^k)\big]. \end{aligned} \tag{2.5}$$

In the following lemma, we establish the relationship between the iterates v^k and v^{k+1} generated by the SC-PRSM-PR (1.11) and the auxiliary variable $\tilde{v}^k$ defined in (2.2).

Lemma 2.1 *Let v^{k+1} be generated by the SC-PRSM-PR (1.11) with the given v^k; and $\tilde{v}^k$ be defined in (2.3). Then, we have*

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{2.6}$$

where

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\alpha\beta B & -\alpha\beta C & 2\alpha I \end{pmatrix}. \tag{2.7}$$

Proof: Together with $y^{k+1}=\tilde{y}^k$ and $z^{k+1}=\tilde{z}^k$ and using (2.5), we have the following relationship:

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^{k} \\ z^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\alpha\beta B & -\alpha\beta C & 2\alpha I \end{pmatrix} \begin{pmatrix} y^k-\tilde{y}^k \\ z^k-\tilde{z}^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}.$$

This can be rewritten as a compact form of (2.6), where M is defined in (2.7). □

3 Global Convergence

In this section, we show that the sequence generated by the SC-PRSM-PR scheme (1.11) globally converges to a solution point of $\mathrm{VI}(\Omega, F, \theta)$. We first prove some inequalities which are crucial for establishing the strict contraction for the sequence generated by the SC-PRSM-PR (1.11). We summarize them in several lemmas.

Lemma 3.1 *Let $\{w^k=(x^k,y^k,z^k,\lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{v^k=(y^k,z^k,\lambda^k)\}$; $\{\tilde{w}^k\}$ and $\{\tilde{v}^k\}$ be defined in (2.2) and (2.3), respectively. Then we have*

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k), \quad \forall\, w\in\Omega, \tag{3.1}$$

where

$$Q=\begin{pmatrix} (1+\mu)\beta B^TB & 0 & -\alpha B^T \\ 0 & (1+\mu)\beta C^TC & -\alpha C^T \\ -B & -C & \frac{1}{\beta}I \end{pmatrix}. \tag{3.2}$$

Proof: Since $\tilde{x}^k=x^{k+1}$, deriving the first-order optimality condition of the x-subproblem in (1.11), we have

$$\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{A^T[\beta(A\tilde{x}^k+By^k+Cz^k-b)-\lambda^k]\}\ge 0,\ \forall\, x\in\mathcal{X}.$$

Substituting $\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k+Cz^k-b)$ (see (2.4)) into the above inequality, we obtain

$$\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\}\ge 0, \quad \forall\, x\in\mathcal{X}. \tag{3.3}$$

Using $\lambda^{k+\frac{1}{2}}=\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)$ (also see (2.4)), the y-minimization problem in (1.11) can be written as

$$\tilde{y}^k=\arg\min\left\{ \begin{array}{l} \theta_2(y)-[\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)]^T(A\tilde{x}^k+By+Cz^k-b) \\ +\frac{\beta}{2}\|A\tilde{x}^k+By+Cz^k-b\|^2+\frac{\mu\beta}{2}\|B(y-y^k)\|^2 \end{array} \,\middle|\, y\in\mathcal{Y}\right\},$$

and its first-order optimality condition gives us

$$\begin{aligned} &\theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\big\{-B^T[\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)] \\ &\qquad +\beta B^T(A\tilde{x}^k+B\tilde{y}^k+Cz^k-b)+\mu\beta B^TB(\tilde{y}^k-y^k)\big\}\ge 0,\ \forall\, y\in\mathcal{Y}. \end{aligned} \tag{3.4}$$

Again, using (2.4), we have

$$\begin{aligned} &-[\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)]+\beta(A\tilde{x}^k+B\tilde{y}^k+Cz^k-b)+\mu\beta B(\tilde{y}^k-y^k) \\ &\quad = -[\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)]+\beta(A\tilde{x}^k+By^k+Cz^k-b)+(1+\mu)\beta B(\tilde{y}^k-y^k) \\ &\quad = -[\lambda^k-\alpha(\lambda^k-\tilde{\lambda}^k)]+(\lambda^k-\tilde{\lambda}^k)+(1+\mu)\beta B(\tilde{y}^k-y^k) \\ &\quad = -\tilde{\lambda}^k+(1+\mu)\beta B(\tilde{y}^k-y^k)-\alpha(\tilde{\lambda}^k-\lambda^k). \end{aligned}$$

Consequently, it follows from (3.4) that

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + (1+\mu)\beta B^T B(\tilde{y}^k - y^k) - \alpha B^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \ \forall\, y \in \mathcal{Y}. \quad (3.5)$$

Analogously, from the z-minimization problem in (1.11), we get

$$\theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T\{-C^T\tilde{\lambda}^k + (1+\mu)\beta C^T C(\tilde{z}^k - z^k) - \alpha C^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \ \forall\, z \in \mathcal{Z}. \quad (3.6)$$

In addition, it follows from the last equation of (2.2) that

$$(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0,$$

which can be rewritten as

$$\tilde{\lambda}^k \in \Re^m, \ \ (\lambda - \tilde{\lambda}^k)^T\{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \ \ \forall \lambda \in \Re^m. \quad (3.7)$$

Combining (3.3), (3.5), (3.6) and (3.7) together, we get

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ z - \tilde{z}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ -C^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b \end{pmatrix} \right.$$
$$\left. + \begin{pmatrix} 0 \\ (1+\mu)\beta B^T B(\tilde{y}^k - y^k) - \alpha B^T(\tilde{\lambda}^k - \lambda^k) \\ (1+\mu)\beta C^T C(\tilde{z}^k - z^k) - \alpha C^T(\tilde{\lambda}^k - \lambda^k) \\ -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0, \qquad \forall\, w \in \Omega.$$

By using the notation $F(w)$ (see (2.1)) and matrix Q (see (3.2)), the assertion (3.1) follows from the last inequality directly and the lemma is proved. □

Before we proceed the proof, recall we have defined the matrix M in (2.7). Then, together with the matrix Q defined in (3.2), let us define a new matrix H as

$$H = QM^{-1}. \quad (3.8)$$

Some useful properties of H are summarized in the following proposition.

Proposition 3.2 *The matrix H defined in (3.8) is symmetric and it can be written as*

$$H = \begin{pmatrix} (1+\mu-\frac{1}{2}\alpha)\beta B^T B & -\frac{1}{2}\alpha\beta B^T C & -\frac{1}{2}B^T \\ -\frac{1}{2}\alpha\beta C^T B & (1+\mu-\frac{1}{2}\alpha)\beta C^T C & -\frac{1}{2}C^T \\ -\frac{1}{2}B & -\frac{1}{2}C & \frac{1}{2\alpha\beta}I \end{pmatrix}. \quad (3.9)$$

Moreover, for any fixed $\alpha \in (0,1)$ and $\mu \geq \alpha$, H is positive definite.

Proof: The proof requires some liner algebra knowledge. First, note that $H = QM^{-1}$. For the matrix M defined in (2.7), we have

$$M^{-1} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \frac{\beta}{2}B & \frac{\beta}{2}C & \frac{1}{2\alpha}I \end{pmatrix}.$$

Then, by a manipulation, we obtain

$$
\begin{aligned}
H &= \begin{pmatrix} (1+\mu)\beta B^T B & 0 & -\alpha B^T \\ 0 & (1+\mu)\beta C^T C & -\alpha C^T \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ \frac{\beta}{2} B & \frac{\beta}{2} C & \frac{1}{2\alpha} I \end{pmatrix} \\
&= \begin{pmatrix} (1+\mu-\frac{1}{2}\alpha)\beta B^T B & -\frac{1}{2}\alpha\beta B^T C & -\frac{1}{2} B^T \\ -\frac{1}{2}\alpha\beta C^T B & (1+\mu-\frac{1}{2}\alpha)\beta C^T C & -\frac{1}{2} C^T \\ -\frac{1}{2} B & -\frac{1}{2} C & \frac{1}{2\alpha\beta} I \end{pmatrix}.
\end{aligned}
$$

This is just the form of (3.9) and H is symmetric. The first part is proved.

To show the positive definiteness of matrix H, we need only to observe the following 3×3 matrix

$$
\begin{pmatrix} (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2}\alpha & -\frac{1}{2} \\ -\frac{1}{2}\alpha & (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2\alpha} \end{pmatrix}.
$$

Since $\alpha \in (0,1)$ and $\mu \geq \alpha$, we have

$$
1+\mu-\frac{1}{2}\alpha > 0 \qquad \text{and} \qquad \begin{vmatrix} (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2}\alpha \\ -\frac{1}{2}\alpha & (1+\mu-\frac{1}{2}\alpha) \end{vmatrix} > 0.
$$

Note that

$$
\begin{aligned}
& \begin{vmatrix} (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2}\alpha & -\frac{1}{2} \\ -\frac{1}{2}\alpha & (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2\alpha} \end{vmatrix} \\
= & -\frac{1}{2} \begin{vmatrix} -\frac{1}{2}\alpha & -\frac{1}{2} \\ (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2} \end{vmatrix} + \frac{1}{2} \begin{vmatrix} (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2} \\ -\frac{1}{2}\alpha & -\frac{1}{2} \end{vmatrix} + \frac{1}{2\alpha} \begin{vmatrix} (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2}\alpha \\ -\frac{1}{2}\alpha & (1+\mu-\frac{1}{2}\alpha) \end{vmatrix} \\
= & \begin{vmatrix} (1+\mu-\frac{1}{2}\alpha) & -\frac{1}{2} \\ -\frac{1}{2}\alpha & -\frac{1}{2} \end{vmatrix} + \frac{1}{2\alpha}\big((1+\mu-\frac{1}{2}\alpha)^2 - (\frac{1}{2}\alpha)^2\big) \\
= & \begin{vmatrix} (1+\mu) & 0 \\ -\frac{1}{2}\alpha & -\frac{1}{2} \end{vmatrix} + \frac{1}{2\alpha}(1+\mu)(1+\mu-\alpha) \\
= & \frac{1}{2\alpha}(1+\mu)(1+\mu-2\alpha).
\end{aligned}
$$

Since $\alpha \in (0,1)$ and $\mu \geq \alpha$, H is positive definite. □

The assertion in Proposition 3.2 helps us present the convergence analysis in succinct notation. Now, with the defined matrices M, Q and H, we can further analyze the conclusion proved in Lemma 3.1. More specifically, let us first observe the right-hand side of the inequality (3.1) and rewrite it as the sum of some quadratic terms under certain matrix norms. This is done in the following lemma.

Lemma 3.3 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{v^k = (y^k, z^k, \lambda^k)\}$; $\{\tilde{w}^k\}$ and $\{\tilde{v}^k\}$ be defined in (2.2) and (2.3), respectively. Then we have*

$$
(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \tag{3.10}
$$

with

$$G = Q^T + Q - M^T HM, \tag{3.11}$$

where the matrices M, Q and H are defined in (2.7), (3.2) and (3.9), respectively.

Proof: The proof only requires some elementary manipulations. More specifically, using the fact $Q = HM$ (see (3.8)) and the relation (2.6), the right-hand side of (3.1) can be written as

$$(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = (v - \tilde{v}^k)^T H(v^k - v^{k+1}). \tag{3.12}$$

Applying the identity

$$(a-b)^T H(c-d) = \frac{1}{2}\{\|a-d\|_H^2 - \|a-c\|_H^2\} + \frac{1}{2}\{\|c-b\|_H^2 - \|d-b\|_H^2\},$$

to the right-hand side of (3.12) with

$$a = v, \quad b = \tilde{v}^k, \quad c = v^k, \quad \text{and} \quad d = v^{k+1},$$

we thus obtain

$$(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = \frac{1}{2}\big(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2\big) + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \tag{3.13}$$

For the last term of (3.13), we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\quad \overset{(2.6)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&\quad = \quad 2(v^k - \tilde{v}^k) HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k) M^T HM(v^k - \tilde{v}^k) \\
&\quad \overset{(3.8)}{=} \quad (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k).
\end{aligned} \tag{3.14}$$

By using (3.13), (3.14) and (3.11), the assertion of Lemma 3.3 is proved. □

In Lemma 3.3, a new matrix G is introduced in order to improve the inequality (3.1) in Lemma 3.1. Let us hold on the proof temporarily and take a closer look at the matrix G just defined in (3.11). Some properties of this matrix is summarized in the following proposition.

Proposition 3.4 *The symmetric matrix G defined in (3.11) can be rewritten as*

$$G = \begin{pmatrix} (1+\mu-\alpha)\beta B^T B & -\alpha\beta B^T C & -(1-\alpha)B^T \\ -\alpha\beta C^T B & (1+\mu-\alpha)\beta C^T C & -(1-\alpha)C^T \\ -(1-\alpha)B & -(1-\alpha)C & \frac{2-2\alpha}{\beta} I \end{pmatrix}. \tag{3.15}$$

Moreover, for fixed $\alpha \in (0,1)$ and any $\mu \geq \alpha$ (resp. $\mu > \alpha$), G is positive semi-definite (Resp., positive definite).

Proof: For the matrix G defined in (3.11), since $Q = HM$ (see (3.8)), we have

$$G = Q^T + Q - M^T HM = Q^T + Q - M^T Q.$$

By using the matrices M and Q (see (2.7) and (3.2)), we obtain

$$\begin{aligned} G &= (Q^T+Q) - \begin{pmatrix} I & 0 & -\alpha\beta B^T \\ 0 & I & -\alpha\beta C^T \\ 0 & 0 & 2\alpha I \end{pmatrix} \begin{pmatrix} (1+\mu)\beta B^T B & 0 & -\alpha B^T \\ 0 & (1+\mu)\beta C^T C & -\alpha C^T \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} \\ &= \begin{pmatrix} (2+2\mu)\beta B^T B & 0 & -(1+\alpha)B^T \\ 0 & (2+2\mu)\beta C^T C & -(1+\alpha)C^T \\ -(1+\alpha)B & -(1+\alpha)C & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} (1+\mu+\alpha)\beta B^T B & \alpha\beta B^T C & -2\alpha B^T \\ -\alpha\beta C^T B & (1+\mu+\alpha)\beta C^T C & -2\alpha C^T \\ -2\alpha B & -2\alpha C & \frac{2\alpha}{\beta} I \end{pmatrix} \\ &= \begin{pmatrix} (1+\mu-\alpha)\beta B^T B & -\alpha\beta B^T C & -(1-\alpha)B^T \\ \alpha\beta C^T B & (1+\mu-\alpha)\beta C^T C & -(1-\alpha)C^T \\ -(1-\alpha)B & -(1-\alpha)C & \frac{2-2\alpha}{\beta} I \end{pmatrix}. \end{aligned}$$

This is just the form of G (see (3.15)). The first part is completed.

To show the positive semi-definiteness (Resp., positive definiteness) of G, we need only to observe the following 3×3 matrix

$$\begin{pmatrix} (1+\mu-\alpha) & -\alpha & -(1-\alpha) \\ -\alpha & (1+\mu-\alpha) & -(1-\alpha) \\ -(1-\alpha) & -(1-\alpha) & 2(1-\alpha) \end{pmatrix}$$

Since $\alpha \in (0,1)$ and $\mu \geq \alpha$, we have

$$1+\mu-\alpha > 0 \qquad \text{and} \qquad \begin{vmatrix} (1+\mu-\alpha) & -\alpha \\ -\alpha & (1+\mu-\alpha) \end{vmatrix} > 0.$$

Note that

$$\begin{aligned} & \begin{vmatrix} (1+\mu-\alpha) & -\alpha & -(1-\alpha) \\ -\alpha & (1+\mu-\alpha) & -(1-\alpha) \\ -(1-\alpha) & -(1-\alpha) & 2(1-\alpha) \end{vmatrix} \\ = & -(1-\alpha) \begin{vmatrix} -\alpha & -(1-\alpha) \\ (1+\mu-\alpha) & -(1-\alpha) \end{vmatrix} + (1-\alpha) \begin{vmatrix} (1+\mu-\alpha) & -(1-\alpha) \\ -\alpha & -(1-\alpha) \end{vmatrix} \\ & +2(1-\alpha) \begin{vmatrix} (1+\mu-\alpha) & -\alpha \\ -\alpha & (1+\mu-\alpha) \end{vmatrix} \\ = & \; 2(1-\alpha) \begin{vmatrix} (1+\mu-\alpha) & -(1-\alpha) \\ -\alpha & -(1-\alpha) \end{vmatrix} + 2(1-\alpha) \begin{vmatrix} (1+\mu-\alpha) & -\alpha \\ -\alpha & (1+\mu-\alpha) \end{vmatrix} \\ = & \; 2(1-\alpha) \begin{vmatrix} (1+\mu) & 0 \\ -\alpha & -(1-\alpha) \end{vmatrix} + 2(1-\alpha)\big((1+\mu-\alpha)^2 - \alpha^2\big) \\ = & -2(1-\alpha)^2(1+\mu) + 2(1-\alpha)(1+\mu)(1+\mu-2\alpha) \\ = & \; 2(1-\alpha)(1+\mu)(\mu-\alpha). \end{aligned}$$

Thus, for fixed $\alpha \in (0,1)$ and any $\mu \geq \alpha$ (resp. $\mu > \alpha$), G is positive semi-definite (resp. positive definite). The proof is complete. $\square$

Now, with the proved propositions and lemmas, the inequality (3.1) in Lemma 3.1 can be significantly polished. We summarize it as a theorem and will use it later.

Theorem 3.5 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{v^k = (y^k, z^k, \lambda^k)\}$; $\{\tilde{w}^k\}$ and $\{\tilde{v}^k\}$ be defined in (2.2) and (2.3), respectively. Then, we have*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \ \ \forall w \in \Omega, \quad (3.16)$$

where H and G are defined in (3.9) and (3.11), respectively.

Proof: It is trivial by combining the assertions (3.1) and (3.10). $\square$

Now we are ready to show that the sequence $\{v^k\}$ generated by the SC-PRSM-PR (1.11) with $\alpha \in (0,1)$ and $\mu > \alpha$ is strictly contractive with respect to the solution of the VI (2.1a). Note that for this case the matrix G defined in (3.11) is positive definite as proved in Proposition 3.4.

Theorem 3.6 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{v^k = (y^k, z^k, \lambda^k)\}$; $\{\tilde{w}^k\}$ and $\{\tilde{v}^k\}$ be defined in (2.2) and (2.3), respectively. Then we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*, \quad (3.17)$$

where H and G are defined in (3.9) and (3.11), respectively.

Proof: Setting $w = w^*$ in (3.16), we get

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \quad (3.18)$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \geq 0$$

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_G^2. \quad (3.19)$$

The assertion (3.17) follows directly. $\square$

The assertion (3.17) thus implies the strict contraction of the sequence $\{v^k\}$ generated by the SC-PRSM-PR (1.11). We can thus easily prove the convergence based on this assertion, as stated in the following theorem. This assertion is also the basis for establishing the worst-case convergence rate in a nonergodic sense in Section 4.1.

Theorem 3.7 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11). The sequence $\{v^k = (y^k, z^k, \lambda^k)\}$ converges to some v^∞ which belongs to $\mathcal{V}^*$.*

Proof: According to (3.17), it holds that $\{v^k\}$ is bounded and

$$\lim_{k\to\infty} \|v^k - \tilde{v}^k\| = 0. \quad (3.20)$$

So, $\{\tilde{v}^k\}$ is also bounded. Let v^∞ be a cluster point of $\{\tilde{v}^k\}$ and $\{\tilde{v}^{k_j}\}$ is a subsequence which converges to v^∞. Let $\{\tilde{w}^k\}$ and $\{\tilde{w}^{k_j}\}$ be the induced sequences by $\{\tilde{v}^k\}$ and $\{\tilde{v}^{k_j}\}$, respectively. It follows from (3.1) that

$$\tilde{w}^{k_j} \in \Omega, \quad \theta(u) - \theta(u^\infty) + (w - \tilde{w}^{k_j})^T F(\tilde{w}^{k_j}) \geq (v - v^{k_j})^T Q(v^{k_j} - \tilde{v}^{k_j}), \quad \forall\, w \in \Omega.$$

Since the matrix Q is not singular, it follows from the continuity of $\theta(u)$ and $F(w)$ that

$$w^\infty \in \Omega, \quad \theta(u) - \theta(u^\infty) + (w - w^\infty)^T F(w^\infty) \geq 0, \quad \forall\, w \in \Omega.$$

The above variational inequality indicates that w^∞ is a solution of $\mathrm{VI}(\Omega, F)$. By using (3.20) and $\lim_{j\to\infty} v^{k_j} = v^\infty$, the subsequence $\{v^{k_j}\}$ converges to v^∞. Due to (3.17), we have

$$\|v^{k+1} - v^\infty\|_H \leq \|v^k - v^\infty\|_H$$

and thus $\{v^k\}$ converges to v^∞. The proof is complete. □

4 Worst-case Convergence Rate

In this section, we establish the worst-case $O(1/t)$ convergence rate measured by the iteration complexity in both the ergodic and a nonergodic senses for the SC-PRSM-PR (1.11), where t is the iteration counter. Note that as the work [41, 43] and many others, a worst-case $O(1/t)$ convergence rate measured by the iteration complexity means an approximate solution with an accuracy of $O(1/t)$ can be found based on t iterations of an iterative scheme; or equivalently, it requires at most $O(1/\epsilon)$ iterations to find an approximate solution with an accuracy of ϵ.

4.1 Worse-case Convergence Rate in a Nonergodic Sense

We first establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the SC-PRSM-PR (1.11). The starting point for the analysis is the assertion (3.17) in Theorem 3.6, and the analytic framework follows from the work in [33] for the ADMM scheme (1.2).

First, recall the matrices M, H and G defined respectively in (2.7), (3.9) and (3.11). Since both matrices G and $M^T HM$ are positive definite, there exists a constant $c > 0$ such that

$$\|M(v^k - \tilde{v}^k)\|_H^2 \leq c\|v^k - \tilde{v}^k\|_G^2.$$

Substituting it into (3.17) and using (2.6), it follows that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{1}{c}\|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{4.1}$$

In the following, we will show that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ is monotonically non-increasing. That is , we have

$$\|v^{k+1} - v^{k+2}\|_H^2 \leq \|v^k - v^{k+1}\|_H^2, \quad \forall\, k \geq 0.$$

The following lemma establishes an important inequality for this purpose.

Lemma 4.1 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{v^k = (y^k, z^k, \lambda^k)\}$; $\{\tilde{w}^k\}$ and $\{\tilde{v}^k\}$ be defined in (2.2) and (2.3), respectively Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq 0, \tag{4.2}$$

where the matrix Q is defined in (3.2).

Proof: Set $w = \tilde{w}^{k+1}$ in (3.1), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{4.3}$$

Note that (3.1) is also true for $k := k+1$ and thus

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}), \quad \forall\, w \in \Omega.$$

Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{4.4}$$

Adding (4.3) and (4.4) and using the monotonicity of F, we get (4.2) immediately. □

One more inequality is needed; we summarize it in the following lemma.

Lemma 4.2 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{v^k = (y^k, z^k, \lambda^k)\}$; $\{\tilde{w}^k\}$ and $\{\tilde{v}^k\}$ be defined in (2.2) and (2.3), respectively. Then, we have*

$$(v^k - \tilde{v}^k)^T M^T HM\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{4.5}$$

where the matrices M, H and Q are defined in (2.7), (3.9) and (3.2), *respectively.*

Proof: Adding the equation

$$\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\}^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} = \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}$$

to the both sides of (4.2), we get

$$(v^k - v^{k+1})^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{4.6}$$

By using (see (2.6) and (3.8))

$$v^k - v^{k+1} = M(v^k - \tilde{v}^k) \qquad \text{and} \qquad Q = HM,$$

to the term $v^k - v^{k+1}$ in the left hand side of (4.6), we obtain

$$(v^k - \tilde{v}^k)^T M^T HM\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}.$$

and the lemma is proved. □

Now, we are ready to show that the sequence $\{\|M(v^k - \tilde{v}^k)\|^2_H\}$ is non-increasing.

Theorem 4.3 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{v^k = (y^k, z^k, \lambda^k)\}$; $\{\tilde{w}^k\}$ and $\{\tilde{v}^k\}$ be defined in (2.2) and (2.3), respectively. Then, we have*

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|^2_H \leq \|M(v^k - \tilde{v}^k)\|^2_H, \tag{4.7}$$

where the matrices M and H defined in (2.7) and (3.9), respectively.

Proof: Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity

$$\|a\|^2_H - \|b\|^2_H = 2a^T H(a - b) - \|a - b\|^2_H,$$

we obtain

$$\begin{aligned} &\|M(v^k - \tilde{v}^k)\|^2_H - \|M(v^{k+1} - \tilde{v}^{k+1})\|^2_H \\ &= \ 2(v^k - \tilde{v}^k)^T M^T HM\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} - \|M(v^k - \tilde{v}^k) - M(v^{k+1} - \tilde{v}^{k+1})\|^2_H. \end{aligned}$$

Inserting (4.5) into the first term of the right-hand side of the last equality, we obtain

$$\|M(v^k - \tilde{v}^k)\|^2_H - \|M(v^{k+1} - \tilde{v}^{k+1})\|^2_H \geq \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q-M^T HM)}.$$

The assertion (4.7) follows from the above inequality and Lemma (3.4) immediately. □

Now, according to (4.7) and (2.6), we have

$$\|v^{k+1} - v^{k+2}\|^2_H \leq \|v^k - v^{k+1}\|^2_H. \tag{4.8}$$

That is, the monotonicity of the sequence $\{\|v^k - v^{k+1}\|^2_H\}$ is proved. Then, with (4.1) and (4.8), we can prove a worst-case $O(1/t)$ convergence rate in a nonergodic sense for the SC-PRSM-PR (1.11) with $\alpha \in [0, 1)$. We summarize the result in the following theorem.

Theorem 4.4 *Let $\{w^k = (x^k, y^k, z^k, \lambda^k)\}$ be the sequence generated by the SC-PRSM-PR (1.11) and $\{v^k = (y^k, z^k, \lambda^k)\}$. Then we have*

$$\|v^k - v^{k+1}\|_H^2 \le \frac{c}{(k+1)}\|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*, \tag{4.9}$$

where the matrix H is defined in (3.9).

Proof: First, it follows from (3.17) that

$$\frac{1}{c}\sum_{t=0}^{\infty}\|v^t - v^{t+1}\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{4.10}$$

According to Theorem 4.3, the sequence $\{\|v^t - v^{t+1}\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(k+1)\|v^k - v^{k+1}\|_H^2 \le \sum_{i=0}^{k}\|v^i - v^{i+1}\|_H^2. \tag{4.11}$$

The assertion (4.9) follows from (4.10) and (4.11) immediately. □

Notice that $\mathcal{V}^*$ is convex and closed. Let $d := \inf\{\|v^0 - v^*\|_H \,|\, v^* \in \Omega^*\}$. Then, for any given $\epsilon > 0$, Theorem 4.4 shows that the scheme (1.11) needs at most $\lfloor d^2/\epsilon \rfloor$ iterations to ensure that $\|v^k - v^{k+1}\|_H^2 \le \epsilon$. Recall that w^{k+1} is a solution of $\mathrm{VI}(\Omega, F, \theta)$ if $\|v^k - v^{k+1}\|_H^2 = 0$. A worst-case $O(1/t)$ convergence rate in a nonergodic sense is thus established for the SC-PRSM-PR (1.11) in Theorem 4.4.

4.2 Worse-case Convergence Rate in the Ergodic Sense

In this subsection, we establish a worst-case $O(1/t)$ convergence rate in the ergodic sense for the SC-PRSM-PR (1.11). For this purpose, we only need the positive semi-definiteness of the matrix G defined in (3.11). Thus, as asserted in Proposition 3.4, we just choose $\alpha \in (0, 1)$ and $\mu \ge \alpha$ for the SC-PRSM-PR (1.11). For the analysis, the starting point is Theorem 3.5, and its follows the work in [32] for the ADMM (1.2).

Let us first recall Theorem 2.3.5 in [16], which provides us an insightful characterization for the solution set of a generic VI. In the following theorem, we specific the Theorem 2.3.5 in [16] for the particular $\mathrm{VI}(\Omega, F, \theta)$ under consideration.

Theorem 4.5 *The solution set of VI(Ω, F, θ) is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w\in\Omega}\big\{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \ge 0\big\}. \tag{4.12}$$

Proof: The proof is an incremental extension of Theorem 2.3.5 in [16], or see the the proof of Theorem 2.1 in [32]. □

Theorem 4.5 thus implies that $\tilde{w} \in \Omega$ is an approximate solution of $\mathrm{VI}(\Omega, F, \theta)$ with the accuracy $\epsilon > 0$ if it satisfies

$$\theta(u) - \theta(\tilde{u}) + F(w)^T(w - \tilde{w}) \ge -\epsilon, \;\; \forall w \in \Omega \cap \mathcal{D}(\tilde{u}),$$

where

$$\mathcal{D}(\tilde{u}) = \{u \,|\, \|u - \tilde{u}\| \le 1\}.$$

In the rest, our purpose is to show that based on t iterations of the SC-PRSM-PR (1.11), we can find $\tilde{w} \in \Omega$ such that

$$\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \le \epsilon, \;\; \forall w \in \Omega \cap \mathcal{D}(\tilde{u}), \tag{4.13}$$

with $\epsilon = O(1/t)$. That is, an approximate solution of VI(Ω, F, θ) with an accuracy of $O(1/t)$ can be found based on t iterations of the SC-PRSM-PR (1.11).

For coming analysis, let us slightly refine the assertion (3.16) in Theorem 3.5. Using the fact (see the notation of $F(w)$ in (2.1b))

$$(w - \tilde{w}^k)^T F(w) = (w - \tilde{w}^k)^T F(\tilde{w}^k),$$

then it follows from (3.16) that

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \geq \frac{1}{2}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{4.14}$$

Then, we summarize the worst-case convergence $O(1/t)$ convergence rate in the ergodic sense for the SC-PRSM-PR (1.11) in the following theorem.

Theorem 4.6 *Let $\{w^k\}$ be the sequence generated by the SC-PRSM-PR (1.11); $\{\tilde{w}^k\}$ be defined in (2.2); and t be an integer. Let $\tilde{w}_t$ be defined as the average of $\tilde{w}^k$ for $k = 1, 2, \cdots t$, i.e.,*

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k. \tag{4.15}$$

Then, we have $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega, \tag{4.16}$$

where H is defined by (3.9).

Proof: First, because of (2.2), it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (4.15) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (4.14) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\Big)^T F(w) + \frac{1}{2}\|v - v^0\|_H^2 \geq 0, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1} \sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{4.17}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \leq \frac{1}{t+1} \sum_{k=0}^{t} \theta(\tilde{u}^k).$$

Substituting it in (4.17), the assertion of this theorem follows directly. $\square$

Recall (4.13). The conclusion (4.16) thus indicates that based on the t iteration of the SC-PRSM-PR (1.11), we can find $\tilde{w}_t$ defined in (4.15) which is an approximate solution of (1.5) with an accuracy of $O(1/t)$. That is, a worst-case $O(1/t)$ convergence rate in the ergodic sense is established for the SC-PRSM-PR (1.11) in Theorem 4.6.

5 A Divergence Example

It has been shown in the last section that the SC-PRSM-PR in (1.11) is convergent when $\alpha \in (0,1)$ and $\mu > \alpha$. In this section, we give an example showing that both the direct extension of the SC-PRSM (1.7) and the direct application of the SC-PRSM (1.10) are not necessarily convergent. Thus, the motivation of considering the SC-PRSM-PR (1.11) for the multi-block convex minimization model (1.5) is justified.

The example is inspired by the counter example in [10], showing the divergence of the E-ADMM (1.6). More specifically, we consider the following system of linear equations:

$$Ax + By + Cz = 0, \tag{5.1}$$

where $A, B, C \in \Re^4$ are linearly independent such that the matrix (A, B, C) is full rank and x, y, z are all in $\Re$. This is a special case of the model (1.5) with $\theta_1 = \theta_2 = \theta_3 = 0$, $m = 4$, $n_1 = n_2 = n_3 = 1$, $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \Re$; and the coefficients matrices are A, B and C, respectively. Obviously, the system of linear equation (5.1) has the unique solution $x^* = y^* = z^* = 0$. In particular, we consider

$$(A, B, C) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \end{pmatrix}. \tag{5.2}$$

5.1 The Divergence of the Direct Application of SC-PRSM (1.10)

First, we show that the direct application of the SC-PRSM (1.10) is not necessarily convergent.

Applying the scheme (1.10) to the homogeneous system of linear equation (5.1), the resulting scheme can be written as

$$\begin{cases} A^T[\beta(Ax^{k+1} + By^k + Cz^k) - \lambda^k] = 0 \\ \alpha\beta(Ax^{k+1} + By^k + Cz^k) + \lambda^{k+\frac{1}{2}} - \lambda^k = 0 \\ B^T[\beta(Ax^{k+1} + By^{k+1} + Cz^k) - \lambda^{k+\frac{1}{2}}] = 0 \\ C^T[\beta(Ax^{k+1} + By^k + Cz^{k+1}) - \lambda^{k+\frac{1}{2}}] = 0 \\ \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^{k+1}) + \lambda^{k+1} - \lambda^{k+\frac{1}{2}} = 0. \end{cases} \tag{5.3}$$

It follows from the first equation in (5.3) that

$$x^{k+1} = \frac{1}{A^T A}(-A^T By^k - A^T Cz + A^T \lambda^k/\beta). \tag{5.4}$$

For ease of notation, let us denote $\mu^k = \lambda^k/\beta$. Then, plugging (5.4) into the rest equations in (5.3), we derive that

$$\begin{pmatrix} B^T B & 0 & 0 \\ 0 & C^T C & 0 \\ \alpha B & \alpha C & I \end{pmatrix} \begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \mu^{k+1} \end{pmatrix}$$

$$= \left[\begin{pmatrix} -\alpha B^T B & -(\alpha+1)B^T C & B^T \\ -(\alpha+1)C^T B & -\alpha C^T C & C^T \\ -\alpha B & -\alpha C & I \end{pmatrix} - \frac{1}{A^T A} \begin{pmatrix} (\alpha+1)B^T A \\ (\alpha+1)C^T A \\ 2\alpha A \end{pmatrix} (-A^T B, -A^T C, A^T) \right] \begin{pmatrix} y^k \\ z^k \\ \mu^k \end{pmatrix}.$$

Let

$$L_1 = \begin{pmatrix} B^T B & 0 & 0 \\ 0 & C^T C & 0 \\ \alpha B & \alpha C & I \end{pmatrix},$$

$$
R_1 = \begin{pmatrix} -\alpha B^T B & -(\alpha+1)B^T C & B^T \\ -(\alpha+1)C^T B & -\alpha C^T C & C^T \\ -\alpha B & -\alpha C & I \end{pmatrix} - \frac{1}{A^T A} \begin{pmatrix} (\alpha+1)B^T A \\ (\alpha+1)C^T A \\ 2\alpha A \end{pmatrix} \left(-A^T B, -A^T C, A^T \right),
$$

and denote

$$
M_1 = L_1^{-1} R_1.
$$

Then, the scheme (5.3) can be written compactly as

$$
\begin{pmatrix} y^k \\ z^k \\ \mu^k \end{pmatrix} = M_1^k \begin{pmatrix} y^0 \\ z^0 \\ \mu^0 \end{pmatrix}.
$$

Obviously, if the the spectral radius of M_1, denoted by $\rho(M_1) := |\lambda_{\max}(M_1)|$ (the largest eigenvalue of M_1), is not smaller than 1, then the sequence generated by the scheme above is not possible to converge to the solution point $(x^*, y^*, z^*) = (0, 0, 0)$ of the system (5.1) for any starting point.

Consider the example where the coefficients (A, B, C) for (5.1) are given in (5.2). Then, with trivial manipulation, we know that

$$
L_1 = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 13 & 0 & 0 & 0 & 0 \\ \alpha & \alpha & 1 & 0 & 0 & 0 \\ \alpha & 2\alpha & 0 & 1 & 0 & 0 \\ 2\alpha & 2\alpha & 0 & 0 & 1 & 0 \\ 2\alpha & 2\alpha & 0 & 0 & 0 & 1 \end{pmatrix}
$$

and

$$
R_1 = \frac{1}{4} \begin{pmatrix} 36-4\alpha & -2\alpha-2 & -6\alpha-2 & -6\alpha-2 & 2-6\alpha & 2-6\alpha \\ -2\alpha-2 & 49-3\alpha & -7\alpha-3 & 1-7\alpha & 1-7\alpha & 1-7\alpha \\ 8\alpha & 10\alpha & 4-2\alpha & -2\alpha & -2\alpha & -2\alpha \\ 8\alpha & 6\alpha & -2\alpha & 4-2\alpha & -2\alpha & -2\alpha \\ 4\alpha & 6\alpha & -2\alpha & -2\alpha & 4-2\alpha & -2\alpha \\ 4\alpha & 6\alpha & -2\alpha & -2\alpha & -2\alpha & 4-2\alpha \end{pmatrix}.
$$

In Figure 1, we plot the values of $\rho(M_1)$ for different choices of α varying from 0 to 1 with the equal distance of 0.02. It is obvious that $\rho(M_1) \geq 1$ for all tested cases, and it is monotonically increasing with respect to $\alpha \in (0, 1)$. Therefore, the sequence generated by the scheme above is not convergent to the solution point of the system (5.1) for any starting point. It is thus illustrated that the direct application of the SC-PRSM (1.10) is not necessarily convergent.

5.2 The Divergence of the E-SC-PRSM (1.7)

Now we turn to show the divergence of the E-SC-PRSM (1.7) when it is applied to solve the same example (5.1).

When the scheme (1.7) is applied to solve the special model (5.1), the iterative scheme can be specified as

$$
\begin{cases} A^T[\beta(Ax^{k+1} + By^k + Cz^k) - \lambda^k] = 0 \\ \alpha\beta(Ax^{k+1} + By^k + Cz^k) + \lambda^{k+\frac{1}{3}} - \lambda^k = 0 \\ B^T[\beta(Ax^{k+1} + By^{k+1} + Cz^k) - \lambda^{k+\frac{1}{3}}] = 0 \\ \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^k) + \lambda^{k+\frac{2}{3}} - \lambda^{k+\frac{1}{3}} = 0 \\ C^T[\beta(Ax^{k+1} + By^{k+1} + Cz^{k+1}) - \lambda^{k+\frac{2}{3}}] = 0 \\ \alpha\beta(Ax^{k+1} + By^{k+1} + Cz^{k+1}) + \lambda^{k+1} - \lambda^{k+\frac{2}{3}} = 0. \end{cases} \tag{5.5}
$$

Similarly as the last subsection, we can solve x^{k+1} first based on the first equation in (5.5) and then substitute it into the other equations. This leads to the following equation:

$$\begin{pmatrix} B^TB & 0 & 0 \\ (1+\alpha)C^TB & C^TC & 0 \\ 2\alpha B & \alpha C & I \end{pmatrix} \begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \mu^{k+1} \end{pmatrix}$$
$$= \left[\begin{pmatrix} -\alpha B^TB & -(\alpha+1)B^TC & B^T \\ -\alpha C^TB & -2\alpha C^TC & C^T \\ -\alpha B & -2\alpha C & I \end{pmatrix} - \frac{1}{A^TA} \begin{pmatrix} (\alpha+1)B^TA \\ (2\alpha+1)C^TA \\ 3\alpha A \end{pmatrix} (-A^TB, -A^TC, A^T) \right] \begin{pmatrix} y^k \\ z^k \\ \mu^k \end{pmatrix}.$$

Then, we denote

$$L_2 = \begin{pmatrix} B^TB & 0 & 0 \\ (1+\alpha)C^TB & C^TC & 0 \\ 2\alpha B & \alpha C & I \end{pmatrix},$$

$$R_2 = \begin{pmatrix} -\alpha B^TB & -(\alpha+1)B^TC & B^T \\ -\alpha C^TB & -2\alpha C^TC & C^T \\ -\alpha B & -2\alpha C & I \end{pmatrix} - \frac{1}{A^TA} \begin{pmatrix} (\alpha+1)B^TA \\ (2\alpha+1)C^TA \\ 3\alpha A \end{pmatrix} (-A^TB, -A^TC, A^T).$$

and

$$M_2 = L_2^{-1}R_2.$$

With the specific definitions of (A, B, C) in (5.2), we can easily show that

$$L_2 = \begin{pmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 11\alpha+11 & 13 & 0 & 0 & 0 & 0 \\ 2\alpha & \alpha & 1 & 0 & 0 & 0 \\ 2\alpha & 2\alpha & 0 & 1 & 0 & 0 \\ 4\alpha & 2\alpha & 0 & 0 & 1 & 0 \\ 4\alpha & 2\alpha & 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$R_2 = \frac{1}{4} \begin{pmatrix} 36-4\alpha & -2\alpha-2 & -6\alpha-2 & -6\alpha-2 & 2-6\alpha & 2-6\alpha \\ 40\alpha+42 & 49-6\alpha & -14\alpha-3 & 1-14\alpha & 1-14\alpha & 1-14\alpha \\ 14\alpha & 13\alpha & 4-3\alpha & -3\alpha & -3\alpha & -3\alpha \\ 14\alpha & 5\alpha & -3\alpha & 4-3\alpha & -3\alpha & -3\alpha \\ 10\alpha & 5\alpha & -3\alpha & -3\alpha & 4-3\alpha & -3\alpha \\ 10\alpha & 5\alpha & -3\alpha & -3\alpha & -3\alpha & 4-3\alpha \end{pmatrix}.$$

Hence, to check the spectral radius of $\rho(M_2) := \lambda_{\max}(M_2)$ (the largest eigenvalue of M_2), we take different values of α varying from 0 to 1 with the equal distance of 0.02 and plot the values of $\rho(M_2)$. In Figure 1, we show that $\rho(M_2) \geq 1$ for all the tested cases, and it is monotonically increasing with respect to $\alpha \in (0, 1)$. Therefore, the sequence generated by the scheme above is not convergent to the solution point of the system (5.1). It is thus illustrated that the E-SC-PRSM (1.7) is not necessarily convergent.

6 Numerical Results

In this section, we test the proposed SC-PRSM-PR (1.11) for some applications arising in the image processing and statistical learning domains, and report the numerical results. Since the SC-PRSM-PR (1.11) can be regarded as a customized application of the original SC-PRSM (1.4) to the specific multi-block convex minimization model (1.5) and it is an operator splitting type method, we will mainly compare it with some methods of the same kind. In particular, as well justified in the

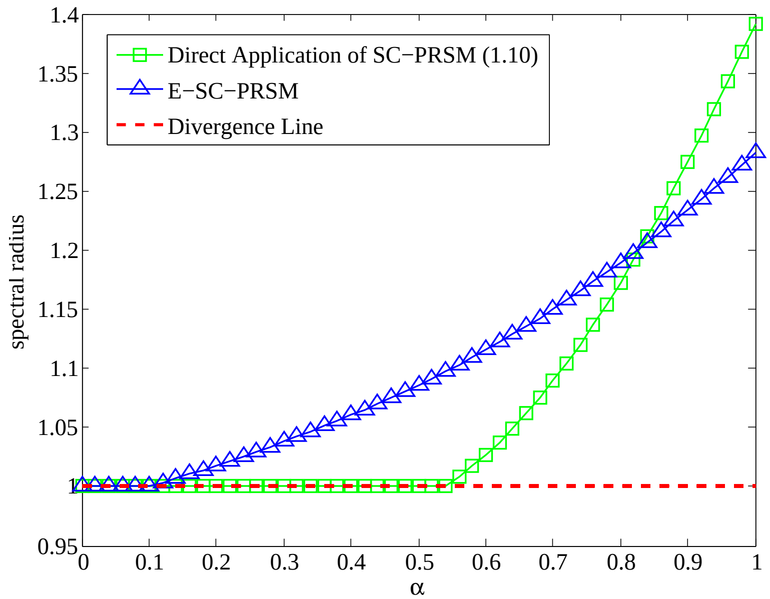


Figure 1: The magnitude of the spectral radius for M_1 and M_2 with respect to $\alpha \in (0, 1)$

literature (e.g.,[46, 50]), the E-ADMM (1.6) often performs well despite of its lack of convergence. We will thus compare the SC-PRSM-PR with it. Moreover, because of the same reason as (1.6), it is interesting to verify the empirical efficiency of the E-SC-PRSM (1.7), even though its divergence has been just shown. In fact, as we shall report, for the tested examples, the E-SC-PRSM does perform well too.

Our code was written by Matlab 2010b and all numerical experiments were performed on a Dell(R) laptop computer with 1.5GHz AMD(TM) A8 processor and a 4GB memory. Since our numerical experiments were conducted on an ordinary laptop without parallel processors, for the y- and z-subproblems in (1.11) at each iteration which are eligible for parallel computation, we only count the longer time.

6.1 Image Restoration with Mixed Impulsive and Gaussian Noises

Let $u \in \Re^n$ represent a digital image with $n = n_1 \times n_2$. Note that a two-dimensional image can be represented by vectorizing it as a one-dimensional vector in certain order, e.g. the lexicographic order. Suppose the clean image u is corrupted by both blur and noise. We consider the case where the noise is the mixture of an additive Gaussian white noise and an impulse noise. The corrupted (also observed) image is denoted by u^0. Image restoration is to recover the clean image u from the observed image u^0.

We consider the following image restoration model for mixed noise removal which was proposed in [35]:

$$\min_{u,f} \left\{ \tau \|u\|_{\mathrm{TV}} + \frac{\rho}{2}\|u - f\|^2 + \|P_{\mathcal{A}}(Hf - u^0)\|_1 \right\}. \tag{6.1}$$

In (6.1), $\|\cdot\|$ and $\cdot\|_1$ denote the l_2- and l_1 norms, respectively; $\|\cdot\|_{\mathrm{TV}}$ is the discrete total variation defined by

$$\|u\|_{\mathrm{TV}} = \sum_{1 \le i,j \le n} \sqrt{\left|(\nabla_1 u)_{j,k}\right|^2 + \left|(\nabla_2 u)_{j,k}\right|^2},$$

where $\nabla_1 : \Re^n \to \Re^n$ and $\nabla_2 : \Re^n \to \Re^n$ are the discrete horizontal and vertical partial derivatives, respectively; and we denote $\nabla = (\nabla_1, \nabla_2)$, see [49]; thus $\|u\|_{\mathrm{TV}}$ can be written as $\|\nabla u\|_1$; H is the matrix representation (convolution operator) of a spatially-invariant blur; $\mathcal{A}$ represents the set of pixels which are corrupted by the impulsive noise (all the pixels outside $\mathcal{A}$ are corrupted by the Gaussian noise); $P_{\mathcal{A}}$ is the characteristic function of the set $\mathcal{A}$, i.e., $P_{\mathcal{A}}(u)$ has the value 1 for any pixel within $\mathcal{A}$ and 0 for any pixel outside $\mathcal{A}$; u^0 is the corrupted image with blurry and mixed noise; and τ and ρ are positive constants. To identify the set $\mathcal{A}$, in [35] it was proposed to apply the adaptive median filter (AMF) first to remove most of the impulsive noise within $\mathcal{A}$.

We first show that the model (6.1) can be reformulated as a special case of (1.5). In fact, by introducing the auxiliary variables w, v and z, we can reformulate (6.1) as

$$\begin{array}{ll} \min & \tau||w||_1 + \frac{\rho}{2}\|v\|^2 + \|P_{\mathcal{A}}(z)\|_1 \\ \text{s.t.} & w = \nabla u \\ & v = u - f \\ & z = Hf - u^0, \end{array} \tag{6.2}$$

which is a special case of the abstract model (1.5) with the following specification:

- $x := u$, $y := f$ and $z := (w, v, z)$; $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are the whole real spaces in appropriate dimensionality;
- $\theta_1(x) := 0$, $\theta_2(y) := 0$ and $\theta_3(z) := \theta_3(w, v, z) = \tau||w||_1 + \frac{\rho}{2}\|v\|^2 + \|P_{\mathcal{A}}(z)\|_1$;
- and

$$A := \begin{bmatrix} \nabla \\ I \\ 0 \end{bmatrix}, \; B := \begin{bmatrix} 0 \\ -I \\ H \end{bmatrix}, \; C := \begin{bmatrix} -I & 0 & 0 \\ 0 & -I & 0 \\ 0 & 0 & -I \end{bmatrix}, \; b := \begin{bmatrix} 0 \\ 0 \\ u^0 \end{bmatrix}. \tag{6.3}$$

Thus, the methods "E-ADMM", "E-SC-PRSM" and "SC-PRSM-PR" are all applicable to the model (6.2). Below we elaborate on the minimization subproblems arising in the SC-PRSM-PR (1.11) and show that they all have closed-form solutions. We skip the elaboration on the subproblems of E-ADMM and E-SC-PRSM which can be easily found in the literature or similar as those of the SC-PRSM-PR.

When the SC-PRSM-PR (1.11) is implemented to (6.2), the first subproblem (i.e., the u-subproblem) is

$$u^{k+1} = \arg\min_{u \in \Re^{n\times n}} \left\{ \|\nabla u - w^k - \frac{\lambda_1^k}{\beta}\|^2 + \|u - f^k - v^k - \frac{\lambda_2^k}{\beta}\|^2 \right\},$$

whose solution is given by

$$\beta(\nabla^T\nabla + I)u^{k+1} = \lambda_2^k + \beta(f^k + v^k) + \nabla^T(\lambda_1^k + \beta w^k),$$

which can be solved efficiently by the fast Fourier transform (FFT) or the discrete cosine transform (DCT) (see e.g., [28] for details). In fact, applying the FFT to diagonalize ∇ such that $\nabla = \mathcal{F}^{-1}D\mathcal{F}$, where $\mathcal{F}$ is the Fourier transformation matrix and D is a diagonal matrix, we can rewrite the equation above as

$$\beta(D^TD + I)\mathcal{F}u^{k+1} = \mathcal{F}\lambda_2^k + \beta(\mathcal{F}f^k + \mathcal{F}v^k) + D^T(\mathcal{F}\lambda_1^k + \beta\mathcal{F}w^k),$$

where $\mathcal{F}u^{k+1}$ can be obtained by an FFT and then u^{k+1} is recovered by an inverse FFT.

After updating the Lagrange multiplier $\lambda_j^{k+1/2}$ for $j = 1, 2, 3$ according to (1.11), the second subproblem (i.e., the f-subproblem) reads as

$$f^{k+1} = \arg\min_{f \in \Re^{n\times n}} \left\{ \|u^k - f - v^k - \lambda_2^{k+1/2}/\beta\|^2 + \|Hf - z^k - u^0 - \lambda_3^{k+1/2}/\beta\|^2 + \mu\|f - f^k\|^2 \right\},$$

whose solution is given by

$$\beta(H^T H + (\mu+1))f^{k+1} = H^T(\lambda_3^{k+1/2} + \beta(z^k + u^0)) + \beta(u^k - v^k) - \lambda_2^{k+1/2} + \beta\mu f^k.$$

For this system of equations, since the matrix H is a spatially convolution operator which can be diagonalized by the FFT as well, by a similar strategy for the u-subproblem, we can solve f^{k+1} efficiently.

For the third subproblem (i.e., the (w, v, z)-subproblem) in the SC-PRSM-PR (1.11), it can be separated into three smaller subproblems as follows:

$$\begin{array}{rcl} w^{k+1} & = & \arg\min_w \left\{ \tau||w||_1 + \frac{\beta}{2}\|\nabla u^{k+1} - w - \frac{\lambda_1^{k+1/2}}{\beta}\|^2 + \frac{\mu\beta}{2}\|w - w^k\|^2 \right\} \\ & = & \mathrm{Shrink}_{\frac{\tau}{\beta(1+\mu)}} \left(\frac{1}{\mu+1}(\nabla u^{k+1} - \lambda_1^{k+1/2}/\beta + \mu w^k) \right); \\ v^{k+1} & = & \arg\min_v \left\{ \rho\|v\|^2 + \beta\|u^{k+1} - f^k - v - \lambda_2^{k+1/2}/\beta\|^2 + \mu\beta\|v - v^k\|^2 \right\} \\ & = & (\beta(u^{k+1} - f^k) - \lambda_2^{k+1/2} + \mu\beta v^k)/(\rho + (1+\mu)\beta) \\ z^{k+1} & = & \arg\min_z \left\{ \|P_{\mathcal{A}}(z)\|_1 + \frac{\beta}{2}\|Hf^k - u^0 - \lambda_2^k/\beta\|^2 + \frac{\mu\beta}{2}\|z - z^k\|^2 \right\}, \end{array}$$

with

$$(z^{k+1})_i = \begin{cases} \mathrm{Shrink}_{1/((1+\mu)\beta)} \left(\frac{1}{\mu+1}(Hf^k - u^0 - \lambda_3^{k+1/2}/\beta + \mu z^k) \right), & \text{if } i \in \mathcal{A}; \\ \frac{1}{\mu+1}(Hf^k - u^0 - \lambda_3^{k+1/2}/\beta + \mu z^k), & \text{otherwise.} \end{cases}$$

and $\mathrm{shrink}_\sigma(\cdot)$ denotes the shrinkage operator (see e.g. [12]). That is:

$$\mathrm{Shrink}_\sigma(a) = sign(a) \circ \max\{|a| - \sigma, 0\}, \forall a \in \mathcal{R}^n,$$

with $\sigma > 0$, where $\mathrm{sign}(\cdot)$ is the sign function and the operator "$\circ$" stands for the componentwise scalar multiplication.

Finally, we update λ^{k+1} based on $\lambda^{k+1/2}$ and the just-solved u^{k+1}, f^{k+1} and $(w^{k+1}, v^{k+1}, z^{k+1})$.

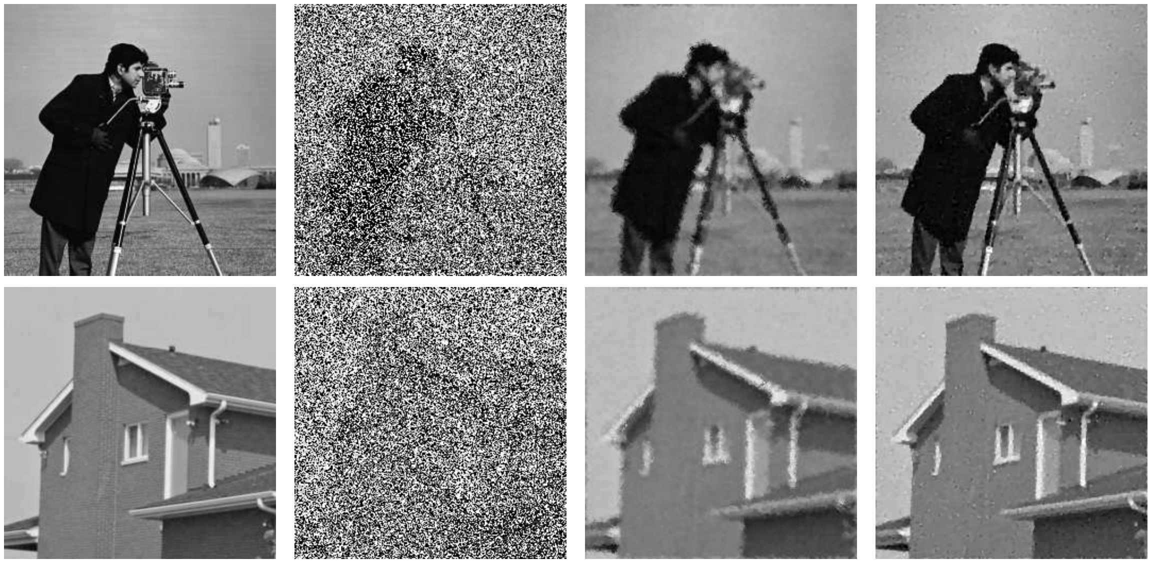
For the model (6.2), we test two images: Cameraman.png and House.png. Both are of the size 256×256. These two images are first convoluted by a blurring kernel with radius 3 and then corrupted by impulsive noise with the intensity 0.7 and Gaussian white noise with the variance 0.01. We first apply the AMF (see e.g. [35]) with window size 19 to identify the corrupted index set $\mathcal{A}$ and remove the impulsive noise and get the filtered images. The original, degraded and filtered images are shown in Figure 2.

We now numerically compare the SC-PRSM-PR with the E-ADMM and E-SC-PRSM. We take the parameters $\tau = 0.02$ and $\rho = 1$ in the model (6.2). For a fair comparison, we choose the same values for common parameters of different methods: $\beta = 6$ for all and $\alpha = 0.15$ for the SC-PRSM-PR and E-SC-PRSM. For the additional parameter μ of the SC-PRSM-PR, we take it as $\mu = 0.16$. We use the stopping criterion (6.20) with $\epsilon = 4 \times 10^{-5}$ and the maximum iteration number is set to be 1000. The initial iterates for all methods are the degraded images. We use the signal-to-noise ratio (SNR) in the unit dB as the measure of the performance of the restored images from all methods. The SNR is defined as

$$\mathrm{SNR} = 10 \log_{10} \frac{\|u\|^2}{\|\hat{u} - u\|^2},$$

where u is the original image and $\hat{u}$ is the restored image. In the experiment, we use the stopping criterion, which is popularly adopted in the literature of image processing:

$$\mathrm{Tol} := \frac{\|f^{k+1} - f^k\|_F}{1 + \|f^k\|_F} < 4 \times 10^{-5}. \tag{6.4}$$



original degraded filtered restored

Figure 2: The original images(first column), the degraded images (second column), the filtered images by AMF (third column) and the restored images by SC-PRSM-PR.

The images restored by the SC-PRSM-PR are shown in Figure 2. In Figure 3, we plot the evolution curves of the SNR values with respect to the computing time in seconds for the tested images. Table 6.1 reports some statistics of the comparison among these methods, including the number of iterations ("Iter."), computing time in second ("CPU(s)") and the SNR value in unit of dB ("SNR(dB)") of the restored image. This set of experiment shows: 1) the E-SC-PRSM (1.7), as we have expected, does work well for the tested examples empirically even though its lack of convergence has been demonstrated in Section 5.2; and 2) the proposed SC-PRSM-PR with proved convergence is very competitive with, and sometimes is even faster than, the E-ADMM and E-SC-PRSM whose convergence is lacked. Note that the SC-PRSM-PR (1.11) usually requires more iteration numbers; but two subproblems of it at each iteration can be solved in parallel. This helps it save computing time.

	Cameraman.png			House.png		
Algorithm	Iter.	CPU (s)	SNR (dB)	Iter.	CPU (s)	SNR (dB)
E-ADMM	901	31.73	19.20	814	28.49	23.84
E-SC-PRSM	767	31.47	19.19	695	27.81	23.84
SC-PRSM-PR	**935**	**29.76**	**19.20**	**845**	**25.77**	**23.85**

Table 1: Numerical Comparison for the image restoration model (6.1).

6.2 Robust Principal Component Analysis with Missing and Noisy Data

It worths mentioning that the model (1.5) and the previous analysis are for vector spaces. But without any difficulty, they can be trivially extended to the case where matrix variables are considered and some linear mappings rather than matrices are accompanying the variables in the constraints. In the upcoming subsections, we test two cases of the model (1.5) with matrix variables, both arising in statistical learning.

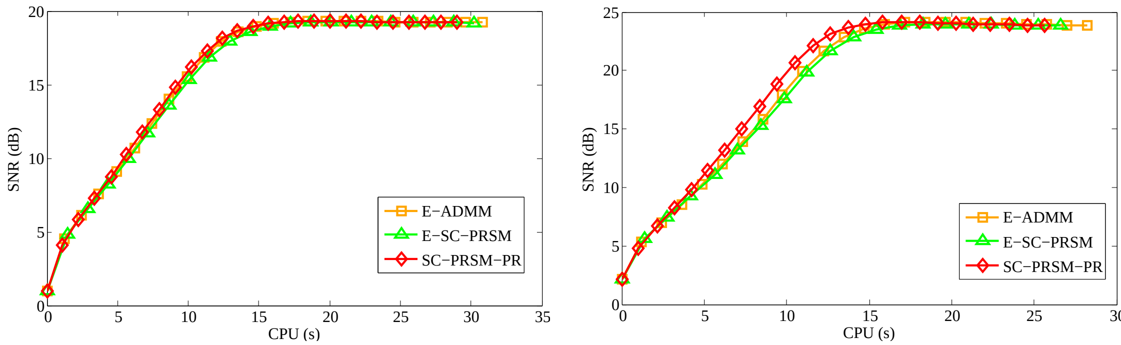


Figure 3: Image restoration from the mixture of noises (left: Cameraman.png and right: House.png on the right): evolutions of the SNR (dB) with respect to the CPU time.

We first test the model of robust principal component analysis (RPCA) with missing and noisy data. This model aims at decomposing a matrix $M \in \Re^{m\times n}$ as the sum of a low-rank matrix $R \in \Re^{m\times n}$ and a sparse matrix $S \in \Re^{m\times n}$; but not all the entries of M are known and M is corrupted by a noise matrix. More specifically, we focus on the unconstrained the model studied in [50]:

$$\min \quad ||R||_* + \gamma||S||_1 + \tfrac{\nu}{2}\|P_\Omega(M-R-S)\|_F^2, \tag{6.5}$$

where $||\cdot||_*$ is the nuclear norm defined as the sum of all singular values of a matrix, $||\cdot||_1$ is the sum of absolute values of all entries of a matrix, $||\cdot||_F$ is the Frobenius norm which is the square sum of all entries of a matrix; $\gamma > 0$ is a constant balancing the low-rank and sparsity and $\nu > 0$ is a constant reflecting the Gaussian noise level; Ω is a subset of the index set of entries $\{1,2,\cdots,m\}\times\{1,2,\cdots,n\}$, and we assume that only those entries $\{C_{ij},(i,j)\in\Omega\}$ can be observed; the incomplete observation information is summarized by the operator $P_\Omega : \Re^{m\times n}\to\Re^{m\times n}$, which is the orthogonal projection onto the span of matrices vanishing outside of Ω so that the ij-th entry of $P_\Omega(X)$ is X_{ij} if $(i,j)\in\Omega$ and zero otherwise. Note that the model (6.5) is a generalization of the model of matrix decomposition in [9] and the model of robust principal component analysis in [6].

Introducing an auxiliary variable $Z \in \Re^{m\times n}$, we can reformulate (6.5) as

$$\begin{array}{ll} \min & \gamma||S||_1 + ||R||_* + \frac{\nu}{2}\|P_\Omega(Z)\|_F^2, \\ s.t. & \qquad S+R+Z=M. \end{array} \tag{6.6}$$

which is a special case of (1.5) with $x=S$, $y=R$, $z=Z$; A, B and C are all identity mappings; $b=M$; $\theta_1(x)=\gamma||S||_1$, $\theta_2(y)=||R||_*$, $\theta_3(z)=\frac{\nu}{2}\|P_\Omega(Z)\|_F^2$, $\mathcal{X}=\mathcal{Y}=\mathcal{Z}=\Re^{m\times n}$. Then, applying the SC-PRSM-PR (1.11) to the model (6.6) and again omitting some details, we can see that all the resulting subproblems have closed-form solutions. More specifically, the resulting SC-PRSM-PR scheme for (6.6) can be explicitly expressed as

$$\left\{\begin{array}{lcl} S^{k+1} &=& \mathcal{S}_{\frac{1}{\beta}}(-R^k - Z^k + M + \frac{1}{\beta}\lambda^k) \\ \lambda^{k+\frac{1}{2}} &=& \lambda^k - \alpha\beta(R^{k+1}+S^k+Z^k-M) \\ R^{k+1} &=& \mathcal{D}_{\frac{\gamma}{\beta(\mu+1)}}\left(\frac{\mu}{\mu+1}S^k - \frac{1}{\mu+1}(S^{k+1}+Z^k-M-\frac{1}{\beta}\lambda^{k+\frac{1}{2}})\right) \\ Z^{k+1} &=& \tilde{Z}^k \\ \lambda^{k+1} &=& \lambda^{k+\frac{1}{2}} - \alpha\beta(R^{k+1}+S^{k+1}+Z^{k+1}-M), \end{array}\right. \tag{6.7}$$

where $\tilde{Z}^k$ is given by

$$\tilde{Z}^k_{ij} = \left\{\begin{array}{ll} \frac{\mu\beta}{\nu+(\mu+1)\beta}Z^k - \frac{\beta}{\nu+(\mu+1)\beta}(S^{k+1}+R^{k+1}-M-\frac{1}{\beta}\lambda^{k+\frac{1}{2}}), & \text{if } (i,j)\in\Omega; \\ \frac{\mu}{\mu+1}Z^k - \frac{1}{\mu+1}(S^{k+1}+R^{k+1}-M-\frac{1}{\beta}\lambda^{k+\frac{1}{2}}), & \text{otherwise}. \end{array}\right. .$$

Algorithm	Iter.	CPU(s)	rank($\hat{R}$)	$\lvert\text{supp}(\hat{S})\rvert$
E-ADMM	44	48.73	7	243,991
E-SC-PRSM	46	79.28	14	579,658
SC-PRSM-PR	**33**	**40.97**	**9**	**301,387**

Table 2: Numerical Comparison for the RPCA model (6.6)

Note that in (6.7), $\mathcal{S}_{\frac{1}{\beta}}$ is the matrix version of the shrinkage operator defined before, with the detail of

$$(S_{\frac{1}{\beta}})_{ij} = (1 - \frac{1}{\beta}/|X_{ij}|)_+ \cdot X_{ij}, \quad 1 \leq i \leq m, 1 \leq j \leq n, \tag{6.8}$$

for $\beta > 0$. In addition, $\mathcal{D}_\tau(X)$ with $\tau > 0$ is the singular value soft-thresholding operator defined as follows. If a matrix X with rank r has the singular value decomposition (SVD)

$$X = U\Lambda V^*, \quad \Lambda = \text{diag}(\{\sigma_i\}_{1\leq i\leq r}), \tag{6.9}$$

then we define

$$\mathcal{D}_\tau(X) = U\mathcal{D}_\tau(\Lambda)V^*, \quad \mathcal{D}_\tau(\Lambda) = \text{diag}(\{(\sigma_i - \tau)_+\}_{1\leq i\leq r}). \tag{6.10}$$

As tested in [6, 46, 50], one application of the RPCA model (6.6) is to extract the background from a surveillance video. For this application, the low-rank and sparse components, R and S, represent the background and foreground of the video M, respectively. We test the surveillance video at the hall of an airport, which is available at `http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html`. The video consists of 200 frames in size of 114×176. Thus, the video data can be realigned as a matrix $M \in \Re^{m\times n}$ with $m = 25,344, n = 200$. We obtain the observed subset Ω by randomly generating 80% of the entries of D and add the Gaussian noise with mean zero and variance 10^{-3} to D to obtain the observed matrix M. According to [6], we set the regularization parameters $\gamma = 1/\sqrt{m}$ and $\nu = 100$. The 10th, 100th and 200th frames of the original and the corrupted video are displayed at the first and second rows in Figure 4, respectively.

To implement the E-ADMM, E-SC-PRSM and SC-PRSM-PR, we set $\beta = 0.005|\Omega|/\|M\|_1$ for all. We choose $\alpha = 0.25$ for the E-SC-PRSM and SC-PRSM-PR. For the additional parameter μ of SC-PRSM-PR, we take $\mu = 0.26$. The initial iterates all start at zero matrices and the stopping criterion of all methods are taken as

$$\max\left\{\frac{\|R^{k+1} - R^k\|_F}{1 + \|R^k\|_F}, \frac{\|S^{k+1} - S^k\|_F}{1 + \|S^k\|_F}\right\} < 10^{-2}.$$

Some frames of the foreground recovered by the SC-PRSM-PR is shown at the third row of Figure 4. In Figure 5, we plot the respectively evolutions of the primal and dual residuals for all the methods under comparison with respect to both the computing time and number of iterations. Table 2 reports some quantitative comparison among these methods, including the number of iterations ("Iter"), computing time in seconds ("CPU(s)"), and rank of the recovery video foreground ("rank($\hat{R}$)") and the number of nonzero entries of the video background ("$|\text{supp}(\hat{S})|$"). The statistics in Table 2 and curves in Figure 5 demonstrate that the SC-PRSM-PR outperforms other two methods. This set of experiment further verifies the efficiency of the proposed SC-PRSM-PR (1.11).

6.3 Quadratic Discriminant Analysis

Then we test a quadratic discriminant analysis (QDA) model. A rather new and challenging problem in the statistical learning area, the QDA aims at classifying two sets of normal distribution data (denoted by X_1 and X_2) with different but close covariance matrices generalized from the linear

Figure 4: The 10th, 100th and 200th frames of the original surveillance video (first row), the corrupted video (second row) and the sparse video recovered from SC-PRSM-PR (third row).

discriminant analysis, see e.g., [18, 19, 40] for details. An assumption for the QDA is that the data $X_1 \in \Re^{n_1 \times d}$ is generated from $\mathcal{N}(\mu_1, \Sigma_1)$ and the data $X_2 \in \Re^{n_2 \times d}$ is generated from $\mathcal{N}(\mu_2, \Sigma_2)$, where d is the dimension of data, n_1 and n_2 are respectively sample sizes. We denote by $\Sigma = \Sigma_1^{-1} - \Sigma_2^{-1}$ the difference between the inverse covariance matrices Σ_1^{-1} and Σ_1^{-1}; estimating Σ is important for classification in high dimensional statistics. In order to estimate a high-dimensional covariance, a QDA model usually assumes that Σ has some special features —- one of such features is the sparsity, see e.g. [18].

In this subsection, we propose a new model under the assumption that the matrix Σ, the difference between the inverse covariance matrices Σ_1^{-1} and Σ_1^{-1}, is represented as $\Sigma = S + R$ where S is a sparse matrix and R is a low-rank matrix. Considering the sparsity and low-rank features simultaneously is indeed important especially for some high-dimensional statistical learning problems, see e.g. [1, 17]. Let the sample covariance matrices of X_1 and X_2 be $\hat{\Sigma}_1 = X_1^T X_1 / n_1$ and $\hat{\Sigma}_2 = X_2^T X_2 / n_2$, respectively. Obviously, we have $\Sigma_1 \Omega \Sigma_2 = \Sigma_2 - \Sigma_1$. Assuming the sparsity and low-rank features of Σ simultaneously and considering the scenarios with noise on the data sets, we propose the new model to estimate $\Sigma = S + R$:

$$\begin{array}{ll} \min & \gamma ||S||_1 + ||R||_* \\ s.t. & ||\hat{\Sigma}_1 (S+R) \hat{\Sigma}_2 - (\hat{\Sigma}_2 - \hat{\Sigma}_1)||_\infty \le r, \end{array} \tag{6.11}$$

where again $\gamma > 0$ is a trade-off constant between the sparsity and low-rank features, $r > 0$ is a tolerance reflecting the noise level, $\|\cdot\|_*$ and $\|\cdot\|_1$ are defined as before in (6.5), and $\|\cdot\|_\infty := \max_{i,j} |U_{ij}|$ denotes the entry-wise maximum norm of a matrix.

Introducing an auxiliary variable U, we can reformulate (6.11) as

$$\begin{array}{ll} \min & ||R||_* + \gamma ||S||_1 \\ s.t. & U - \hat{\Sigma}_1 (R+S) \hat{\Sigma}_2 = \hat{\Sigma}_1 - \hat{\Sigma}_2, \\ & ||U||_\infty \le r, \end{array} \tag{6.12}$$

which is a special case of (1.5) with $x = U$, $y = S$, $z = R$, A is the identity mapping, B and C are the mappings defined by $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, $b = \hat{\Sigma}_1 - \hat{\Sigma}_2$, $\mathcal{X} := \{U \in \Re^{d \times d}, \mid ||U||_\infty \le r\}$, $\mathcal{Y} = \mathcal{Z} = \Re^{d \times d}$.

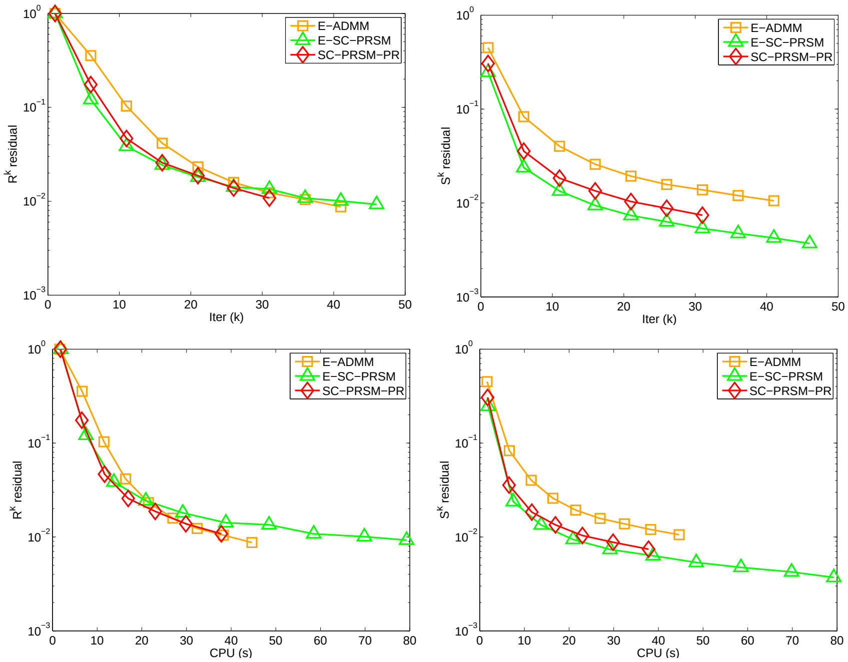


Figure 5: RPCA model (6.6): evolution of the primal and dual residuals for the E-ADMM, E-SC-PRSM and SC-PRSM-PR *w.r.t.* the iterations (first row) and computing time (second row).

To further see why the model (6.12) can be casted into (1.5), we can vectorize the matrices R and S, and then constraint in (6.12) can be written as

$$\mathbf{vec}(U) - (\hat{\Sigma}_2^T \otimes \hat{\Sigma}_1)[\mathbf{vec}(R) + \mathbf{vec}(S)] = \mathbf{vec}(\hat{\Sigma}_1 - \hat{\Sigma}_2), \tag{6.13}$$

where $\mathbf{vec}(X)$ denotes the vectorization of the matrix $X \in \Re^{n\times n}$ by stacking the columns of X into a single column vector in $\Re^{n^2}$, and $\otimes$ is a Kronecker product.

Applying the SC-PRSM-PR scheme (1.11) to the model (6.12), the scheme reads as

$$\left\{\begin{array}{lcl} U^{k+1} & = & \arg\min_{||U||_\infty < r}\{\frac{\beta}{2}||U - \hat{\Sigma}_1(R^k + S^k)\hat{\Sigma}_2 - (\hat{\Sigma}_1 - \hat{\Sigma}_2) - \frac{1}{\beta}\lambda^k||_F^2\} \\ \lambda^{k+\frac{1}{2}} & = & \lambda^k - \alpha\beta(U^{k+1} - \hat{\Sigma}_1(R^k + S^k)\hat{\Sigma}_2 - (\hat{\Sigma}_1 - \hat{\Sigma}_2)) \\ S^{k+1} & = & \arg\min\{\gamma||S||_1 + \frac{\beta}{2}||U^{k+1} - \hat{\Sigma}_1(R^k + S)\hat{\Sigma}_2 - (\hat{\Sigma}_1 - \hat{\Sigma}_2) - \frac{1}{\beta}\lambda^{k+1/2}||_F^2 + \frac{\mu\beta}{2}||\hat{\Sigma}_1(S - S^k)\hat{\Sigma}_2||^2\} \\ R^{k+1} & = & \arg\min\{||R||_* + \frac{\beta}{2}||U^{k+1} - \hat{\Sigma}_1(R + S^k)\hat{\Sigma}_2 - (\hat{\Sigma}_1 - \hat{\Sigma}_2) - \frac{1}{\beta}\lambda^{k+1/2}||_F^2 + \frac{\mu\beta}{2}||\hat{\Sigma}_1(R - R^k)\hat{\Sigma}_2||^2\} \\ \lambda^{k+1} & = & \lambda^{k+\frac{1}{2}} - \alpha\beta(U^{k+1} - \hat{\Sigma}_1(R^{k+1} + S^{k+1})\hat{\Sigma}_2 - (\hat{\Sigma}_1 - \hat{\Sigma}_2)), \end{array}\right. \tag{6.14}$$

Now, let us elaborate on the subproblems in (6.14). First, for the U-subproblem in (6.14), its closed-form solution is given by

$$U^{k+1} = \mathcal{T}_r\left(\hat{\Sigma}_1(R^k + S^k)\hat{\Sigma}_2 + (\hat{\Sigma}_1 - \hat{\Sigma}_2) + \frac{1}{\beta}\lambda^k\right), \tag{6.15}$$

where $(\mathcal{T}_r(A))_{ij}$ is defined as

$$(\mathcal{T}_r(A))_{ij} = \mathrm{sign}(A_{ij}) \cdot \max(|A_{ij}|, r).$$

For the S- and R-subproblems in (6.14), they do not have closed-form solutions and must be solved iteratively. Again, we just use the ADMM (1.2) to solve them. For example, for the S-subproblem, we can reformulate it as

$$\begin{array}{ll} \min & \gamma||S||_1 + \frac{\beta}{2}||U^{k+1} - \hat{\Sigma}_1(R^k + A)\hat{\Sigma}_2 - (\hat{\Sigma}_1 - \hat{\Sigma}_2) - \frac{1}{\beta}\lambda^{k+1/2}||_F^2 + \frac{\mu\beta}{2}||\hat{\Sigma}_1(A - S^k)\hat{\Sigma}_2||^2 \\ s.t. & A - S = 0, \end{array} \tag{6.16}$$

where A is an auxiliary variable. Applying the ADMM (1.2) with $\beta = 1$ to (6.16) leads to the scheme:

$$\left\{\begin{array}{rcl} \mathbf{vec}(A^{k+1}) & = & \big([\hat{\Sigma}_2^T\hat{\Sigma}_2]\otimes[(\mu+1)\hat{\Sigma}_1^T\hat{\Sigma}_1]+I_{d^2}\big)^{-1}\mathbf{vec}\big(\hat{\Sigma}_1^T(U^{k+1}-(\hat{\Sigma}_1-\hat{\Sigma}_2)-\frac{1}{\beta}\lambda^k)\hat{\Sigma}_2^T \\ & & \quad +\mu\hat{\Sigma}_1^T\hat{\Sigma}_1R^k\hat{\Sigma}_2\hat{\Sigma}_2^T+A^k+\frac{1}{\beta}\lambda_S^k\big) \\ S^{k+1} & = & \mathcal{S}_{\frac{\gamma}{\beta}}(A^{k+1}-\frac{1}{\beta}\lambda_S^k) \\ \lambda_S^{k+1} & = & \lambda_S^k-(A^{k+1}-S^{k+1}) \end{array}\right. \tag{6.17}$$

Similarly, we can reformulate the R-subproblem in (6.14) as

$$\begin{array}{ll} \min & ||R||_*+\frac{\beta}{2}||U^{k+1}-\hat{\Sigma}_1(A+S^k)\hat{\Sigma}_2-(\hat{\Sigma}_1-\hat{\Sigma}_2)-\frac{1}{\beta}\lambda^{k+1/2}||_F^2+\frac{\mu\beta}{2}||\hat{\Sigma}_1(A-R^k)\hat{\Sigma}_2||^2 \\ s.t. & A-R=0, \end{array} \tag{6.18}$$

where A is an auxiliary variable; and apply the ADMM (1.2) with $\beta = 1$ to (6.18). The resulting scheme is

$$\left\{\begin{array}{rcl} \mathbf{vec}(A^{k+1}) & = & \big([\hat{\Sigma}_2^T\hat{\Sigma}_2]\otimes[(\mu+1)\hat{\Sigma}_1^T\hat{\Sigma}_1]+I_{d^2}\big)^{-1}\mathbf{vec}\big(\hat{\Sigma}_1^T(U^{k+1}-(\hat{\Sigma}_1-\hat{\Sigma}_2)-\frac{1}{\beta}\lambda^k)\hat{\Sigma}_2^T \\ & & \quad +\mu\hat{\Sigma}_1^T\hat{\Sigma}_1S^k\hat{\Sigma}_2\hat{\Sigma}_2^T+A^k+\frac{1}{\beta}\lambda_R^k\big) \\ R^{k+1} & = & \mathcal{D}_{\frac{1}{\beta}}(A^{k+1}-\frac{1}{\beta}\lambda_R^k) \\ \lambda_R^{k+1} & = & \lambda_R^k-(A^{k+1}-R^{k+1}) \end{array}\right. \tag{6.19}$$

Note that the operators $\mathcal{S}_{\frac{\gamma}{\beta}}$ in (6.17) and $\mathcal{D}_{\frac{1}{\beta}}$ in (6.19) are defined in (6.8) and (6.10), respectively.

For generating the two d-dimensional normal distributions $\mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{N}(\mu_2, \Sigma_2)$, we set $d = 20$ and the mean vector as $\mu_1 = \mu_2 = (0, 0, \dots, 0)^T$. We first generate a random matrix $U \in \Re^{20\times 20}$ whose entries are i.i.d. $\mathcal{N}(0, 1)$ and a random diagonal matrix $D \in \Re^{20\times 20}$ whose diagonal elements are i.i.d uniform distribution on $[1, 2]$, then let $\Sigma_1 = U^T DU$. To obtain a low rank semi-positive definite matrix, we generate a random matrix $R_1 \in \Re^{20\times 10}$ whose entries are i.i.d. $\mathcal{N}(0, 1)$ and let $R = R_1R_1^T$. Therefore, we have $\text{rank}(R) = 10$. To obtain a sparse positive definite matrix, we first generate a sparse symmetric matrix $S_1 \in \Re^{20\times 20}$ with 50 nonzero entries and each nonzero entries are i.i.d. $\mathcal{N}(0, 1)$. In order to guarantee the positive definiteness of S, we let $S = S_1 + 2|\lambda_{\min}(S_1)|I_{20}$, where $\lambda_{\min}(S_1)$ is the smallest eigenvalue of S_1. Let $\Sigma_2 = (\Sigma_1^{-1} + S + R)^{-1}$, and we have $\Omega = \Sigma_1^{-1} - \Sigma_2^{-1} = S + R$. We generate $n = 5000$ data $X_1 \in \Re^{n\times d} \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $X_2 \in \Re^{n\times d} \sim \mathcal{N}(\mu_2, \Sigma_2)$ and obtain the sample covariance matrix $\hat{\Sigma}_1 = X_1^TX_1/d$ and $\hat{\Sigma}_2 = X_2^TX_2/d$. We set the regularization parameters $\gamma = 0.5/\sqrt{d}$ and $r = 2\sqrt{d} \approx 8.944$ in the model (6.11).

To compare the SC-PRSM-PR with the E-ADMM and E-SC-PRSM, we set $\beta = 2$ for all; and $\alpha = 0.1$ for the SC-PRSM and SC-PRSM-PR; and $\mu = 0.11$ for the SC-PRSM-PR. The initial values are all zeros matrices, and the same ADMM algorithm (1.2) for $\beta = 1$ is used for solving the subproblems in all three schemes. The iteration schemes for the subproblems are similar to (6.17) and (6.19). We use the stopping criterion that

$$\text{Tol} := \max\{\beta\|R^{k+1} - R^k\|, \beta\|S^{k+1} - S^k\|, \frac{1}{\beta}\|\lambda^{k+1} - \lambda^k\|\} \le d\epsilon, \tag{6.20}$$

with $\epsilon = 1 \times 10^{-2}$ and the maximum iteration number is set as 200. Similar as [5], the quantities $\beta\|R^{k+1} - R^k\|$ and $\beta\|S^{k+1} - S^k\|$ measure the primal residual and $\frac{1}{\beta}\|\lambda^{k+1} - \lambda^k\|$ measures the dual residual of the optimality of an iterate generated by the scheme (6.14).

In Figure 6, we plot the evolution curves of the objective function values, the primal and dual residuals with respect to the iteration numbers and computing time. These curves show that the E-ADMM and E-SC-PRSM are stuck in reducing the primal and dual residuals (see the figures in the third and fourth columns in Figure 6), and the stopping criterion (6.20) is not fulfilled after running out of the maximal number of iterations set beforehand. Moreover, from the zoomed figures (see the

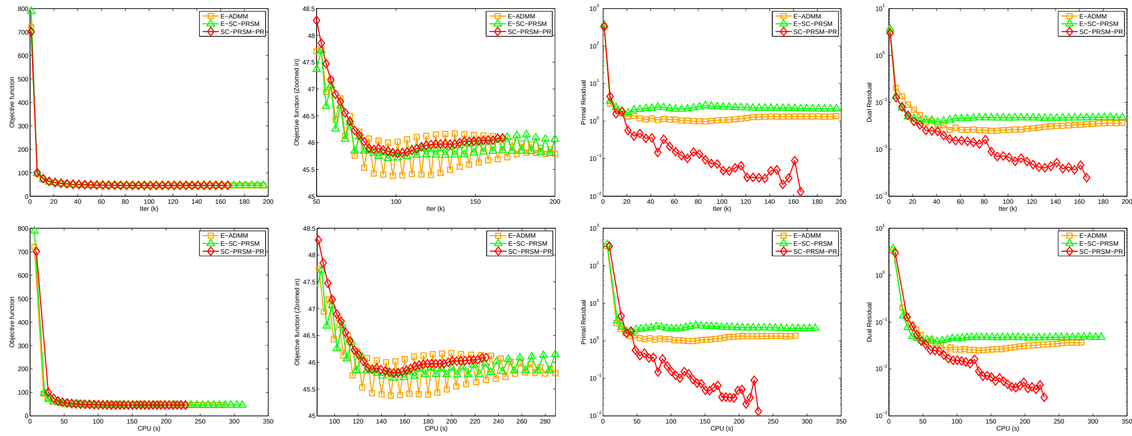


Figure 6: Quadratic discriminant analysis model (6.12): evolution of (from left to right) the objective function value, the zoomed objective value after the 50th iteration, primal residual and dual residual for E-ADMM, E-SC-PRSM, and SC-PRSM-PR $w.r.t$ the iterations (first row) and computing time (second row).

second column in Figure 6), we see that the objective function values at the iterates of the E-ADMM and E-SC-PRSM oscillate around, which corresponds to the curves showing that the reductions of the primal and dual residuals get stuck. In fact, these curves in Figure 6 also further verify the divergence of the E-ADMM and E-SC-PRSM. On the contrary, the proposed SC-RPSM-PR (1.11) converges well in reducing the objective function values, the primal and dual residuals. In Table 3, we report some statistics for the comparison of these three methods over the model (6.11), including the number of iterations ("Iter"), computing time in seconds ("CPU(s)"), rank of R ("rank($\hat{R}$)"), the number of nonzero entries of S ("$|\text{supp}(\hat{S})|$") and the violation of the constraints $\|U\|_\infty$. Recall that our simulated data set requires that $\|U\|_\infty \le r \approx 8.944$. Thus, according to Figure 6 and Table 3, although these three methods perform almost the same in recovering the rank of R and the sparsity of S, the E-SC-PRSM clearly outperforms the others in reducing the objective function values and the violation of constraints (which are exactly the measurement of optimality) for the model (6.11) with faster speed. This justifies the advantage of the proposed E-SC-PRSM.

Algorithm	Iter.	CPU(s)	rank($\hat{R}$)	$\lvert\text{supp}(\hat{S})\rvert$	$\|U\|_\infty$
E-ADMM	200	288.732	17	67	9.29
E-SC-PRSM	200	318.59	17	69	9.30
SC-PRSM-PR	**168**	**230.73**	**17**	**64**	**8.96**

Table 3: Numerical Comparison for the QDA model (6.11).

7 Conclusions

This paper further discusses the strictly contractive Peaceman-Rachford splitting method (SC-PRSM), which was recently proposed for the convex minimization model with linear constraints and a separable objective function in form of the sum of two functions without couple variables. We aim at a multi-block convex minimization model with a higher degree of separability, where the objective function is the sum of more than two functions. We show by an example that some natural ideas of directly applying or extending the original SC-PRSM to this multi-block convex minimization model do not guarantee the convergence. Our strategy is to regroup the functions and variables of the original multi-block model as two blocks, then apply the original SC-PRSM and decompose the resulting subproblems into smallers ones to gain the solvability, and finally regularize the decomposed subproblems by proximal regularization terms to ensure the convergence. The resulting scheme, called SC-PRSM with proximal regularization (SC-PRSM-PR), maintains the implementation advantage of operator-splitting type methods with easy subproblems and meanwhile has the theoretical advantage of provable convergence. We further analyze its worst-case convergence rate measured by the iteration complexity. The efficiency of the SC-PRSM-PR is verified by preliminary numerical results for solving some applications arising in the image processing and statistical learning domains.

This paper is an example of the fact that an operator splitting method with proved convergence for a convex programming with a two-block separable structure cannot be used directly for a model in the same nature but with a higher-degree of separable structure. For using such a method, some tricky treatments on its subproblems (such as our strategies of proximally regularizing the decomposed subproblems) are usually necessary to ensure the convergence. This paper can also be regarded as an example of how to design an algorithm for a more complicated model based on an available algorithm for an easier model in the context of separable convex programming. The flow path of algorithmic design might be valuable for designing customized algorithms in other contexts. For example, we can consider directly applying the original ADMM scheme (1.2) to the multi-block convex minimization model (1.5). Then, just as the SC-PRSM-PR, we can decompose the resulting (y, z)-subproblem and regularize the decomposed subproblems by proximal terms. We can also consider employing an alternating decomposition for the (y, z)-subproblem in (1.8). The philosophy of algorithmic design and the corresponding analytic framework proposed in this paper for the SC-PRSM-PR could be useful for developing such a variant and establish its convergence rigorously.

Finally, to expose the main idea with easier notation, we only focus on the case with a three-block separable structure in this paper; and the analysis can be extended to the generic case with an arbitrary number of block, which of course requires more sophisticated analysis.

References

[1] A. Agarwal and S. Negahban and M. J. Wainwright, *Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions*, Ann. Statis., 40 (2012), pp. 1171-1197.

[2] A. Andreas, E. Theodoros and P. Massimiliano, *Convex multi-task feature learning*, Machine Learning, 73(3) (2008), pp. 243–272.

[3] D. P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, 1982.

[4] E. Blum and W. Oettli, Mathematische Optimierung. Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung, Springer-Verlag, Berlin-Heidelberg-New York, 1975.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1-122.

[6] E. Candès, X. Li, Y. Ma and J. Wright. *Robust principal component analysis?*, J. ACM., 58(1) (2011), pp. 1–37.

[7] T. F. Chan and R. Glowinski, *Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations*, technical report, Stanford University, 1978.

[8] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, *Latent variable graphical model selection via convex optimization*, Ann. Stat., 40 (2012), pp. 1935-1967.

[9] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo and A. S. Willskyc, *Rank-sparsity incoherence for matrix decomposition*, SIAM J. Optim., 21 (2) (2001), pp. 572-596.

[10] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, manuscript, 2013.

[11] E. Corman and X. M. Yuan, *A generalized proximal point algorithm and its convergence rate*, SIAM J. Optim, under revision.

[12] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[13] J. Douglas and H. H. Rachford, *On the numerical solution of the heat conduction problem in* 2 *and* 3 *space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421-439.

[14] J. Eckstein, *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, RUTCOR Research Report RRR 32-2012, December 2012.

[15] J. Eckstein and D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293-318.

[16] F. Facchinei and J. S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I. Springer Series in Operations Research, Springer Verlag, New York, 2003.

[17] J. Fan, Y. Liao and M. Mincheva, *Large covariance estimation by thresholding principal orthogonal complements*, J. Royal Statis. Soc.: Series B, 75 (2013), pp. 603-680.

[18] J. Fan, T. Ke, H. Liu and L. Xia, *QUADRO: A Supervised Dimension Reduction Method via Rayleigh Quotient Optimization*, arXiv: 1311.5542, 2013.

[19] J. H. Friedman, *Regularized discriminant analysis*, J. Amer. Stat. Assoc., XX (1989), pp. 165–175.

[20] M. Fortin and R. Glowinski, Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems, Stud. Math. Appl. 15, North-Holland, Amsterdam, 1983.

[21] D. Gabay, *Applications of the method of multipliers to variational inequalities*, Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, (1983), pp. 299–331.

[22] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Comput. Math. Appli., 2 (1976), pp. 17-40.

[23] R. Glowinski, *On alternating directon methods of multipliers: a historical perspective*, Springer Proceedings of a Conference Dedicated to J. Periaux, to appear.

[24] R. Glowinski, T. Kärkkäinen and K. Majava, *On the convergence of operator-splitting methods*, in Numerical Methods for Scienfic computing, Variational Problems and Applications, edited by Y. Kuznetsov, P. Neittanmaki and O. Pironneau, Barcelona, 2003.

[25] R. Glowinski, and A. Marrocco, *Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, Revue Fr. Autom. Inform. Rech. Opér., Anal. Numér. 2 (1975), pp. 41–76.

[26] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[27] D. R. Han and X. M. Yuan, *Convergence analysis of the Peaceman-Rachford splitting method for nonsmooth convex optimization*, manuscript, 2012.

[28] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, Deblurring Images: Matrices, Spectra, and Filtering, SIAM, Philadelphia, 2006.

[29] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, *A strictly contractive Peaceman-Rachford splitting method for convex programming*, SIAM J. Optim, to appear.

[30] B. S. He, M. Tao and X. M. Yuan, *Alternating direction method with Gaussian back substitution for separable convex programming*, SIAM J. Optim, 22 (2012), pp. 313-340.

[31] B. S. He, M. Tao and X. M. Yuan, *Convergence rate and iteration complexity on the alternating direction method of multipliers with a substitution procedure for separable convex programming*, Math. Oper. Res., under revision.

[32] B. S. He and X. M. Yuan, *On the O(1/n) convergence rate of Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700-709.

[33] B. S. He and X. M. Yuan, *On nonergodic convergence rate of the Douglas-Rachford alternating direction method*, manuscript, 2012.

[34] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4(1969), pp. 303-320.

[35] Y. M. Huang, M. K. Ng and Y. W. Wen, *Fast image restoration methods for impulse and Gaussian noises removal*, IEEE Signal Proces. Lett., 16 (2009), pp. 457–460.

[36] K. C. Kiwiel, C. H. Rosa, and A. Ruszczynski, *Proximal decomposition via alternating linearization*, SIAM J. Optim., 9 (1999), pp. 668-689.

[37] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Num. Analy., 16(1979), pp. 964-979.

[38] M. Lukas, S. De Geer and P. Bühlmann, *The group lasso for logistic regression*, J. Royal Statis. Soc.: Series B, 70(1) (2008), pp. 53–71.

[39] B. Martinet, *Regularision d'inéquations variationnelles par approximations successive*, Revue Francaise d'Automatique et Informatique Recherche Opérationnelle, 126 (1970), pp. 154–159.

[40] G. McLachlan, Discriminant analysis and statistical pattern recognition, Wiley-Interscience, 2004.

[41] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, New York, 1983.

[42] Y. Nesterov, *Gradient methods for minimizing composite objective function*, Math. Prog., Ser.B, 140(1) (2013), pp. 125-161..

[43] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543-547.

[44] Y. E. Nesterov, *Smooth minimization of nonsmooth functions*, Math. Program., 103 (2005), pp.127-152.

[45] D. H. Peaceman and H. H. Rachford, *The numerical solution of parabolic elliptic differential equations*, SIAM J. Appl. Math. 3 (1955), pp. 28-41.

[46] Y. G. Peng, A. Ganesh, J. Wright, W. L. Xu and Y. Ma, *Robust alignment by sparse and low-rank decomposition for linearly correlated images*, IEEE Con. Comput. Vision Patt. Recog. (CVPR), pp. 763-770, 2010.

[47] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, 1969.

[48] R. T. Rockafellar, *Monotone operators and proximal point algorithm*, SIAM J. Con. Optim., 14 (1976), pp. 877–898.

[49] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D., 60 (1992), pp. 259–268.

[50] M. Tao and X. M. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21(1) (2011), pp. 57–81.

[51] A. Yonatan, F. Michael, S. Nathan and U. Shimon, *Uncovering shared structures in multiclass classification*, In *Proceedings of the 24th international conference on Machine learning*, (2007), pp. 17–24.

[52] W. A. Theodore, An introduction to multivariate statistical analysis, Wiley New York, 1958.

[53] M. Uuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, J. Royal Statis. Soc.: Series B, 68(1) (2006), pp. 49–67.