RESEARCH **Open Access**

CrossMark

Modified hybrid decomposition of the augmented Lagrangian method with larger step size for three-block separable convex programming

Min Sun[1,2*] and Yiju Wang[2]

*Correspondence:
ziyouxiaodou@163.com
[1] School of Mathematics and
Statistics, Zaozhuang University,
Zaozhuang, P.R. China
[2] School of Management, Qufu
Normal University, Qufu, P.R. China

Abstract

The Jacobian decomposition and the Gauss–Seidel decomposition of augmented Lagrangian method (ALM) are two popular methods for separable convex programming. However, their convergence is not guaranteed for three-block separable convex programming. In this paper, we present a modified hybrid decomposition of ALM (MHD-ALM) for three-block separable convex programming, which first updates all variables by a hybrid decomposition of ALM, and then corrects the output by a correction step with constant step size $\alpha \in (0, 2 - \sqrt{2})$ which is much less restricted than the step sizes in similar methods. Furthermore, we show that $2 - \sqrt{2}$ is the optimal upper bound of the constant step size α. The rationality of MHD-ALM is testified by theoretical analysis, including global convergence, ergodic convergence rate, nonergodic convergence rate, and refined ergodic convergence rate. MHD-ALM is applied to solve video background extraction problem, and numerical results indicate that it is numerically reliable and requires less computation.

Keywords: The augmented Lagrangian method; Three-block separable convex programming; Step size; Global convergence

1 Introduction

Many problems encountered in applied mathematics area can be formulated as separable convex programming, such as basis pursuit (BP) problem [1–3], video background extraction problem [4–7], image decomposition [8–10], and so on. Thus the solving of separable convex programming plays a fundamental role in applied mathematics and has drawn persistent attention. In the existing literature, several forms of separable convex programming have been investigated [11–15], in which the following three-block separable convex programming rouses more interest:

$$\min \left\{ \sum_{i=1}^{3} \theta_i(x_i) \Big| \sum_{i=1}^{3} A_i x_i = b, x_i \in \mathcal{X}_i, i = 1, 2, 3 \right\}, \tag{1}$$

where $\theta_i : \mathcal{R}^{n_i} \mapsto (-\infty, +\infty]$ $(i = 1, 2, 3)$ are lower semicontinuous proper convex functions, $A_i \in \mathcal{R}^{l \times n_i}$ $(i = 1, 2, 3)$ and $b \in \mathcal{R}^l$, $\mathcal{X}_i$ $(i = 1, 2, 3)$ are nonempty closed convex sets in

$\mathcal{R}^{n_i}$ ($i = 1, 2, 3$). Throughout this paper, we assume that the solution set of problem (1) is nonempty.

The Lagrangian and augmented Lagrangian functions of problem (1) are defined, respectively, as

$$\mathcal{L}(x_1, x_2, x_3, \lambda) = \sum_{i=1}^{3} \theta_i(x_i) - \left\langle \lambda, \sum_{i=1}^{3} A_i x_i - b \right\rangle, \tag{2}$$

$$\mathcal{L}_\beta(x_1, x_2, x_3, \lambda) = \mathcal{L}(x_1, x_2, x_3, \lambda) + \frac{\beta}{2} \left\| \sum_{i=1}^{3} A_i x_i - b \right\|^2, \tag{3}$$

where $\lambda \in \mathcal{R}^l$ is the Lagrange multiplier associated with the linear constraints in (1), and $\beta > 0$ is a penalty parameter. Applying the augmented Lagrangian method (ALM) [16] to problem (1), we can obtain the following iterative scheme:

$$\begin{cases} (x_1^{k+1}, x_2^{k+1}, x_3^{k+1}) = \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2, x_3, \lambda^k) | x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, x_3 \in \mathcal{X}_3\}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b). \end{cases} \tag{4}$$

Obviously, three variables x_1, x_2, x_3 are all involved in the minimization problem of (4), which makes the method often hard to implement. One technique to handle this is to split the subproblem into several small scale subproblems. Based on this, if we split it in a Gauss–Seidel manner and adopt the famous alternating direction method of multiplier (ADMM) [11], we obtain the following iterative scheme:

$$\begin{cases} x_1^{k+1} = \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \lambda^k) | x_1 \in \mathcal{X}_1\}, \\ x_2^{k+1} = \operatorname{argmin}\{\mathcal{L}_\beta(x_1^{k+1}, x_2, x_3^k, \lambda^k) | x_2 \in \mathcal{X}_2\}, \\ x_3^{k+1} = \operatorname{argmin}\{\mathcal{L}_\beta(x_1^{k+1}, x_2^{k+1}, x_3, \lambda^k) | x_3 \in \mathcal{X}_3\}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b). \end{cases} \tag{5}$$

On the other hand, if we split it in a Jacobian manner, we get the following full parallel iterative scheme:

$$\begin{cases} x_1^{k+1} = \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \lambda^k) | x_1 \in \mathcal{X}_1\}, \\ x_2^{k+1} = \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, x_2, x_3^k, \lambda^k) | x_2 \in \mathcal{X}_2\}, \\ x_3^{k+1} = \operatorname{argmin}\{\mathcal{L}_\beta(x_1^k, x_2^k, x_3, \lambda^k) | x_3 \in \mathcal{X}_3\}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b). \end{cases} \tag{6}$$

Compared with the minimization problem in (4), the scale of the minimization procedures in (5) and (6) is decreased, and they fully utilize the separable property of the objective function of (1), thus the new iterative schemes (5) and (6) gain some solvability. However, their convergence cannot be guaranteed under milder conditions as shown in [12, 17]. To overcome this drawback, several new techniques, such as the regularization method with large proximal parameter [18–23], the prediction-correction method with shrunk step size [12, 13, 24–26], etc., have been developed.

Compared with the regularization method, the prediction-correction method has attracted extensive interest, and during the past decades many scholars have performed studies in this direction. For example, He *et al.* [24] proposed an ADMM-based contraction type method for solving multi-block separable convex programming, which first generates a temporal iterate by (5), and then corrects it with a Gaussian back substitution procedure. Later, He *et al.* [12] developed a full Jacobian decomposition of the augmented Lagrangian method for solving multi-block separable convex programming, which first generates a temporal iterate by (6), and then corrects it with a constant step size or varying step size. Different from the above, Han *et al.* [13] proposed a partial splitting augmented Lagrangian method for solving three-block separable convex programming, which first updates the primal variables x_1, x_2, x_3 in a partially-parallel manner, and then corrects x_3, λ with a constant step size. Later, Wang *et al.* [25] presented a proximal partially-parallel splitting method for solving multi-block separable convex programming, which first updates all primal variables in a partially-parallel manner, and then corrects the output with a constant step size or varying step size. Quite recently, Chang *et al.* [26] proposed a convergent prediction-correction-based ADMM in which more minimization problems are involved. In conclusion, the above iteration schemes first generate a temporal iterate by (5) or (6) or their variants, and then generate the new iterate by correcting the temporal iterate with varying step size or a constant step size.

Varying step size needs to be dynamically updated at each iteration, which might be computationally demanding for large-scale (1). Hence in this paper, we consider the prediction-correction method with constant step size for solving problem (1). To the best of our knowledge, He *et al.* [12] first proposed a prediction-correction method with constant step size for solving (1), and they proved that the upper bound of the constant step size is 0.2679. By taking a hybrid splitting of (4) as the prediction step, Wang *et al.* [25] relaxed the upper bound of the constant step size to 0.3670 and Han *et al.* [13] further relaxed it to 0.3820. In practice, to enhance the numerical efficiency of the corresponding iteration method, larger values of the step size are preferred as long as the convergence is still guaranteed [26]. In this paper, based on the methods in [12, 13, 25], we propose a modified hybrid decomposition of the augmented Lagrangian method with constant step size, whose upper bound is relaxed to 0.5858.

The rest of this paper is organized as follows. Section 2 lists some notations and basic results. In Sect. 3, we present a modified hybrid decomposition of the augmented Lagrangian method with larger step size for problem (1) and establish its global convergence and refined convergence rate. Furthermore, a simple example is given to illustrate that $2 - \sqrt{2} \cong 0.5858$ is the optimal upper bound of the constant step size in MHD-ALM. In Sect. 4, some numerical results are given to demonstrate the numerical advantage of larger step size. Finally, a brief conclusion including some possible future works is drawn in Sect. 5.

2 Preliminaries

In this section, we give some notations and basic results about the minimization problem (1), which will be used in the forthcoming discussions.

Throughout this paper, we define the following notations:

$$x = (x_1, x_2, x_3), \qquad v = (x_2, x_3, \lambda), \qquad w = (x_1, x_2, x_3, \lambda),$$

$$\theta(x) = \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3)$$

and

$$\mathcal{A} = (A_1, A_2, A_3), \qquad \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3, \qquad \mathcal{V} = \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{R}^l, \qquad \mathcal{W} = \mathcal{X} \times \mathcal{R}^l.$$

Definition 2.1 A tuple $(x^*, \lambda^*) \in \mathcal{W}$ is called a saddle point of the Lagrangian function (2) if it satisfies the inequalities

$$\mathcal{L}_{\lambda \in \mathcal{R}^l}(x^*, \lambda) \le \mathcal{L}(x^*, \lambda^*) \le \mathcal{L}_{x \in \mathcal{X}}(x, \lambda^*). \tag{7}$$

Solving problem (1) is equivalent to finding a saddle point of $\mathcal{L}(x, \lambda)$ [26, 27]. Therefore, to solve (1), we only need to solve the two inequalities in (7), which can be written as the following mixed variational inequality:

$$\theta(x) - \theta(x^*) + (w - w^*)^\top F(w^*) \ge 0, \quad \forall w \in \mathcal{W}, \tag{8}$$

where

$$F(w) = \begin{pmatrix} -A_1^\top \lambda \\ -A_2^\top \lambda \\ -A_3^\top \lambda \\ \sum_{i=1}^{3} A_i x_i - b \end{pmatrix} = \begin{pmatrix} -\mathcal{A}^\top \lambda \\ \mathcal{A}x - b \end{pmatrix} = \begin{pmatrix} 0 & -\mathcal{A}^\top \\ \mathcal{A} & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} - \begin{pmatrix} 0 \\ b \end{pmatrix}. \tag{9}$$

Because $F(w)$ is a linear mapping with skew-symmetric coefficient matrix, it satisfies the following property:

$$(w' - w)^\top F(w') = (w' - w)^\top F(w), \quad \forall w', w \in \mathcal{W}. \tag{10}$$

The mixed variational inequality (8) is denoted by MVI($\mathcal{W}, F, \theta$), whose solution set is denoted by $\mathcal{W}^*$, which is nonempty from the assumption on problem (1).

To solve MVI($\mathcal{W}, F, \theta$), He *et al.* [28] presented the following prototype algorithm:

A prototype algorithm for MVI($\mathcal{W}, \boldsymbol{F}, \theta$), denoted by ProAlo:

Prediction: For given v^k, find $\hat{w}^k \in \mathcal{W}$ and Q satisfying

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(\hat{w}^k) \ge (v - \hat{v}^k)^\top Q(v^k - \hat{v}^k), \quad \forall w \in \mathcal{W}, \tag{11}$$

where the matrix Q has the property: $(Q + Q^\top)$ is positive definite.

Correction: Determine a nonsingular matrix M, a scalar $\alpha > 0$, and generate the new iterate v^{k+1} via

$$v^{k+1} = v^k - \alpha M(v^k - \hat{v}^k). \tag{12}$$

Condition 2.1 The matrices Q, M in the ProAlo satisfy that the three matrices $Q + Q^\top$, $H := QM^{-1}$, $G(\alpha) := Q + Q^\top - \alpha M^\top HM$ are positive definite.

Under Condition 2.1, He *et al.* [28] established the convergence results of ProAlo, including the global convergence, the worst-case $\mathcal{O}(1/t)$ convergence rate in ergodic or nonergodic sense, where t is the iteration counter. See Theorems 3.3, 4.2, 4.5 in [28].

To end this section, we give the following lemma which will be used in the subsequent section.

Lemma 2.1 ([27]) *Let $\mathcal{X} \subseteq \mathcal{R}^n$ be a closed nonempty convex set, $\theta(x)$ and $f(x)$ be two convex functions. If the function $\theta(x)$ is nondifferentiable, the function $f(x)$ is differentiable, and the solution set of the problem $\min\{\theta(x) + f(x)|x \in \mathcal{X}\}$ is nonempty, then*

$$x^* \in \operatorname{argmin}\{\theta(x) + f(x)|x \in \mathcal{X}\}$$

if and only if

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^\top \nabla f(x^*) \ge 0, \quad \forall x \in \mathcal{X}.$$

3 Algorithm and its convergence

In this section, we give the process of the modified hybrid decomposition of the augmented Lagrangian method (MHD-ALM) for three-block separable convex programming (1) and establish its convergence results, including global convergence, ergodic convergence rate, nonergodic convergence rate, and refined ergodic convergence rate.

Algorithm: MHD-ALM

Step 0. Let parameters $\alpha \in (0, 2 - \sqrt{2})$, $\beta > 0$, tolerance error $\varepsilon > 0$. Choose an initial point $v^0 = (x_2^0, x_3^0, \lambda^0) \in \mathcal{V}$. Set $k = 0$.

Step 1. Compute the prediction iterate $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \tilde{x}_3^k, \tilde{\lambda}^k)$ via

$$\begin{cases} \tilde{x}_1^k = \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \lambda^k)|x_1 \in \mathcal{X}_1\}, \\ \tilde{x}_2^k = \operatorname{argmin}\{\mathcal{L}_\beta(\tilde{x}_1^k, x_2, x_3^k, \lambda^k)|x_2 \in \mathcal{X}_2\}, \\ \tilde{x}_3^k = \operatorname{argmin}\{\mathcal{L}_\beta(\tilde{x}_1^k, x_2^k, x_3, \lambda^k)|x_3 \in \mathcal{X}_3\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A_1\tilde{x}_1^k + A_2\tilde{x}_2^k + A_3\tilde{x}_3^k - b). \end{cases} \quad (13)$$

Step 2. If $\max\{\|A_2 x_2^k - A_2\tilde{x}_2^k\|, \|A_3 x_3^k - A_3\tilde{x}_3^k\|, \|\lambda^k - \tilde{\lambda}^k\|\} \le \varepsilon$, then stop; otherwise, go to Step 3.

Step 3. Generate the new iterate $w^{k+1} = (x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ by

$$\begin{cases} x_1^{k+1} = \tilde{x}_1^k, \\ x_2^{k+1} = x_2^k - \alpha(x_2^k - \tilde{x}_2^k), \\ x_3^{k+1} = x_3^k - \alpha(x_3^k - \tilde{x}_3^k), \\ \lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k). \end{cases} \quad (14)$$

Replace $k + 1$ by k, and go to Step 1.

Remark 3.1 Different from the iterative schemes (5) and (6), the iterative scheme (13) first updates the primal variable x_1 and then updates the primal variables x_2, x_3 in a parallel

manner. Furthermore, the feasible set of the step size α in MHD-ALM is extended from $(0, 0.2679)$ in [12], $(0, 0.3820)$ in [13], $(0, 0.3670)$ in [25] to $(0, 0.5858)$.

The methods in [12, 13, 24–26] and MHD-ALM all fall into the algorithmic framework of prediction-correction methods. The main differences among these methods are: (i) in the prediction step, the methods in [24, 26] update all the primal variables in a sequential order; the method in [12] updates all the primal variables in a parallel manner; the methods in [13, 25] and MHD-ALM update all the primal variables in a partial parallel manner, i.e., they first update x_1 and then update x_2, x_3 in a parallel manner; (ii) in the correction step, the method in [13] updates x_3, λ; the method in [26] and MHD-ALM update x_2, x_3, λ, and the methods in [12, 24, 25] update all the variables.

The convergence analysis of MHD-ALM needs the following assumption and auxiliary sequence.

Assumption 3.1 The matrices A_2, A_3 in problem (1) are both full column rank.

Define an auxiliary sequence $\hat{w}^k = (\hat{x}_1^k, \hat{x}_2^k, \hat{x}_3^k, \hat{\lambda}^k)$ as

$$\hat{x}_i^k = \tilde{x}_i^k \quad (i = 1, 2, 3), \qquad \hat{\lambda}^k = \lambda^k - \beta\big(A_1 \tilde{x}_1^k + A_2 x_2^k + A_3 x_3^k - b\big). \tag{15}$$

To prove the convergence results of MHD-ALM, we only need cast it into the ProAlo and ensure the following two conditions hold: (i) the generated sequence satisfying (11), (12); (ii) the resulting matrices Q, M satisfying Condition 2.1 in Sect. 2. We first verify the first condition. Based on Lemma 2.1, we can derive the first order optimality conditions of the subproblems in (13), which are summarized in the following lemma.

Lemma 3.1 *Let* $\{w^k\}$ *be the sequence generated by MHD-ALM and* $\{\hat{w}^k\}$ *be defined as in* (15). *Then it holds that*

$$\theta(x) - \theta\big(\hat{x}^k\big) + \big(w - \hat{w}^k\big)^\top F\big(\hat{w}^k\big) \geq \big(v - \hat{v}^k\big)^\top Q\big(v^k - \hat{v}^k\big), \quad \forall w \in \mathcal{W}, \tag{16}$$

where the matrix Q is defined by

$$Q = \begin{pmatrix} \beta A_2^\top A_2 & 0 & 0 \\ 0 & \beta A_3^\top A_3 & 0 \\ -A_2 & -A_3 & I_l/\beta \end{pmatrix}. \tag{17}$$

Proof Based on Lemma 2.1 and using the notation of $\hat{w}^k$ in (15), the first order optimality conditions for the three minimization problems in (13) can be summarized as the following inequalities:

$$\theta_1(x_1) - \theta_1\big(\hat{x}_1^k\big) + \big(x_1 - \hat{x}_1^k\big)^\top \big(-A_1^\top \hat{\lambda}^k\big) \geq 0, \quad \forall x_1 \in \mathcal{X}_1,$$

$$\theta_2(x_2) - \theta_2\big(\hat{x}_2^k\big) + \big(x_2 - \hat{x}_2^k\big)^\top \big(-A_2^\top \hat{\lambda}^k - \beta A_2^\top A_2\big(x_2^k - \hat{x}_2^k\big)\big) \geq 0, \quad \forall x_2 \in \mathcal{X}_2,$$

$$\theta_3(x_3) - \theta_3\big(\hat{x}_3^k\big) + \big(x_3 - \hat{x}_3^k\big)^\top \big(-A_3^\top \hat{\lambda}^k - \beta A_3^\top A_3\big(x_3^k - \hat{x}_3^k\big)\big) \geq 0, \quad \forall x_3 \in \mathcal{X}_3.$$

Furthermore, the definition of the variable $\hat{\lambda}^k$ in (15) gives

$$\left(\lambda - \hat{\lambda}^k\right)^\top \left(A_1\hat{x}_1^k + A_2\hat{x}_2^k + A_3\hat{x}_3^k - b + A_2\left(x_2^k - \hat{x}_2^k\right) + A_3\left(x_3^k - \hat{x}_3^k\right) + \frac{1}{\beta}\left(\hat{\lambda}^k - \lambda^k\right)\right) = 0,$$

$$\forall \lambda \in \mathcal{R}^l.$$

Adding the above four inequalities, rearranging terms, and using the definition of the matrix Q, the function $F(w)$, we can get the result (16). This completes the proof. □

Remark 3.2 When $\max\{\|A_2x_2^k - A_2\tilde{x}_2^k\|, \|A_3x_3^k - A_3\tilde{x}_3^k\|, \|\lambda^k - \tilde{\lambda}^k\|\} = 0$, by (15), we get $A_2x_2^k = A_2\hat{x}_2^k$, $A_3x_3^k = A_3\hat{x}_3^k$, $\lambda^k = \hat{\lambda}^k$. Thus, $Q(v^k - \hat{v}^k) = 0$. This and inequality (16) indicate that

$$\theta(x) - \theta\left(\hat{x}^k\right) + \left(w - \hat{w}^k\right)^\top F\left(\hat{w}^k\right) \geq 0, \quad \forall w \in \mathcal{W}.$$

Therefore, $w^k \in \mathcal{W}^*$, and the stopping criterion of MHD-ALM is reasonable.

By the definition of $\hat{\lambda}^k$ in (15), the updating formula of $\tilde{\lambda}^k$ can be represented as

$$\begin{aligned}
\tilde{\lambda}^k &= \lambda^k - \beta\left(A_1\tilde{x}_1^k + A_2\tilde{x}_2^k + A_3\tilde{x}_3^k - b\right) \\
&= \lambda^k - \left(\lambda^k - \hat{\lambda}^k - \beta\left(A_2\left(x_2^k - \hat{x}_2^k\right) + A_3\left(x_3^k - \hat{x}_3^k\right)\right)\right) \\
&= \lambda^k - \left(-\beta A_2, -\beta A_3, I_l\right)\left(v^k - \hat{v}^k\right).
\end{aligned}$$

This together with (14), (15) gives

$$\begin{cases} x_1^{k+1} = \tilde{x}_1^k, \\ v^{k+1} = v^k - \alpha M(v^k - \hat{v}^k), \end{cases} \tag{18}$$

where the matrix M is defined as

$$M = \begin{pmatrix} I_l & 0 & 0 \\ 0 & I_l & 0 \\ -\beta A_2 & -\beta A_3 & I_l \end{pmatrix}. \tag{19}$$

Now to establish the convergence results of MHD-ALM, we only need to verify that the matrices Q, M satisfy Condition 2.1 in Sect. 2.

Lemma 3.2 *Let the matrices Q, M be defined as in (17) and (19). If $\alpha \in (0, 0.5858)$ and Assumption 3.1 hold, then we have*
  (i) *the symmetric matrix $Q + Q^\top$ is positive definite;*
  (ii) *the matrix $H = QM^{-1}$ is symmetric and positive definite;*
  (iii) *the matrix $G(\alpha) = Q + Q^\top - \alpha M^\top HM$ is symmetric and positive definite.*

Proof (i) From the definition of Q, we have

$$Q + Q^\top = \begin{pmatrix} 2\beta A_2^\top A_2 & 0 & -A_2^\top \\ 0 & 2\beta A_3^\top A_3 & -A_3^\top \\ -A_2 & -A_3 & 2I_l/\beta \end{pmatrix}.$$

Therefore, for any $v = (x_2, x_3, \lambda) \neq 0$, we have

$$v^\top (Q + Q^\top) v = 2\beta \|A_2 x_2\|^2 + 2\beta \|A_3 x_3\|^2 + 2\lambda^\top (A_2 x_2 + A_3 x_3) + \frac{2}{\beta} \|\lambda\|^2. \tag{20}$$

- If $(x_2, x_3) = 0$, then $\lambda \neq 0$, so by (20), we get $v^\top (Q + Q^\top) v = \frac{2}{\beta} \|\lambda\|^2 > 0$.
- If $(x_2, x_3) \neq 0$, then from (20) we get

$$v^\top (Q + Q^\top) v \geq \beta \|A_2 x_2\|^2 + \beta \|A_3 x_3\|^2 > 0,$$

where the first inequality follows from the inequality $2x^\top y \geq -(\beta \|x\|^2 + \|y\|^2/\beta)$, and the second inequality comes from Assumption 3.1.

(ii) From the definition of Q, M, we have

$$H = \begin{pmatrix} \beta A_2^\top A_2 & 0 & 0 \\ 0 & \beta A_3^\top A_3 & 0 \\ 0 & 0 & I_l/\beta \end{pmatrix},$$

which is obviously positive definite by Assumption 3.1.

(iii) Similarly, from the definition of Q, M, we have

$$G(\alpha) = \begin{pmatrix} \beta(2 - 2\alpha)A_2^\top A_2 & -\alpha\beta A_2^\top A_3 & -(1-\alpha)A_2^\top \\ -\alpha\beta A_3^\top A_2 & \beta(2 - 2\alpha)A_3^\top A_3 & -(1-\alpha)A_3^\top \\ -(1-\alpha)A_2 & -(1-\alpha)A_3 & (2-\alpha)I_l/\beta \end{pmatrix}$$

$$= L^\top R(\alpha) L,$$

where

$$L = \begin{pmatrix} \sqrt{\beta}A_2 & 0 & 0 \\ 0 & \sqrt{\beta}A_3 & 0 \\ 0 & 0 & I_l/\sqrt{\beta} \end{pmatrix}, \qquad R(\alpha) = \begin{pmatrix} (2 - 2\alpha)I_l & -\alpha I_l & -(1-\alpha)I_l \\ -\alpha I_l & (2 - 2\alpha)I_l & -(1-\alpha)I_l \\ -(1-\alpha)I_l & -(1-\alpha)I_l & (2-\alpha)I_l \end{pmatrix}.$$

This together with Assumption 3.1 implies that we only need to prove the matrix $R(\alpha)$ is positive definite. In fact, it can be written as

$$R(\alpha) = \begin{pmatrix} 2 - 2\alpha & -\alpha & -(1-\alpha) \\ -\alpha & 2 - 2\alpha & -(1-\alpha) \\ -(1-\alpha) & -(1-\alpha) & 2 - \alpha \end{pmatrix} \otimes I_l,$$

where $\otimes$ denotes the matrix Kronecker product. Thus, we only need to prove the 3 order matrix

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

is positive definite, whose three eigenvalues are $\lambda_1 = 2 - \alpha$, $\lambda_2 = 2 - 2\alpha - \sqrt{3\alpha^2 - 4\alpha + 2}$, $\lambda_3 = 2 - 2\alpha + \sqrt{3\alpha^2 - 4\alpha + 2}$. Then, solving the following three inequalities simultaneously

$$
\begin{cases}
2 - \alpha > 0, \\
2 - 2\alpha - \sqrt{3\alpha^2 - 4\alpha + 2} > 0, \\
2 - 2\alpha + \sqrt{3\alpha^2 - 4\alpha + 2} > 0,
\end{cases}
$$

we get $0 < \alpha < 2 - \sqrt{2} \cong 0.5858$. Therefore, the matrix G is positive definite for any $\alpha \in (0, 0.5858)$. This completes the proof. $\qquad\square$

Lemma 3.2 indicates that the matrices Q, M defined as in (17) and (19) satisfy Condition 2.1 in Sect. 2, and thus we get the following convergence results of MHD-ALM based on Theorems 3.3, 4.2, 4.5 in [28].

Theorem 3.1 (Global convergence) *Let $\{w^k\}$ be the sequence generated by MHD-ALM. Then it converges to a vector w^∞, which belongs to $\mathcal{W}^*$.*

Theorem 3.2 (Ergodic convergence rate) *Let $\{w^k\}$ be the sequence generated by MHD-ALM, $\{\hat{w}^k\}$ be the corresponding sequence defined in (15). Set*

$$
\bar{w}^t = \frac{1}{t} \sum_{k=0}^{t} \hat{w}^k.
$$

Then, for any integer $t \geq 1$, we have

$$
\theta(\bar{x}^t) - \theta(x) + (\bar{w}^t - w)^\top F(w) \leq \frac{1}{2\alpha t} \| v - v^0 \|_H^2, \quad \forall w \in \mathcal{W}. \tag{21}
$$

Theorem 3.3 (Nonergodic convergence rate) *Let $\{w^k\}$ be the sequence generated by MHD-ALM. Then, for any $w^* \in \mathcal{W}^*$ and integer $t \geq 1$, we have*

$$
\| M(v^t - \hat{v}^t) \|_H^2 \leq \frac{1}{c_0 t} \| v^0 - v^* \|_H^2,
$$

where $c_0 > 0$ is a constant.

The term

$$
\frac{1}{2\alpha t} \| v - v^0 \|_H^2
$$

on the right-hand side of (21) is used to measure the ergodic convergence rate of MHD-ALM. However, it is not only independent of the distance between the initial iterate w^0 and the solution set $\mathcal{W}^*$ but also hard to estimate due to the variable v. Therefore, inequality (21) is not a reasonable criterion to measure the nonergodic convergence rate of MHD-ALM. In the following, we shall give a refined result from the objective function and constraint condition of problem (1), which is more reasonable, accurate, and intuitive.

Lemma 3.3 *Let* $\{w^k\}$ *be the sequence generated by MHD-ALM. Then, for any* $w \in \mathcal{W}$, *we have*

$$\alpha\big(\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(\hat{w}^k)\big) \geq \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{\alpha}{2}\|v^k - \hat{v}^k\|_H^2. \quad (22)$$

Proof The proof is similar to that of Lemma 3.1 in [28] and is omitted for brevity of this paper. This completes the proof. □

Theorem 3.4 (Refined ergodic convergence rate) *Let* $\{w^k\}$ *be the sequence generated by MHD-ALM,* $\{\hat{w}^k\}$ *be the sequence defined in* (15). *Set*

$$\bar{w}^t = \frac{1}{t}\sum_{k=0}^{t-1}\hat{w}^k.$$

Then, for any integer $t \geq 1$, *there exists a constant* $c > 0$ *such that*

$$\begin{cases} |\theta(\bar{x}^t) - \theta(x^*)| \leq \frac{c}{2\alpha t}, \\ \|\mathcal{A}\bar{x}^t - b\| \leq \frac{c}{2\alpha t}. \end{cases}$$

Proof Choose $w^* = (x^*, \lambda^*) \in \mathcal{W}^*$. Then, for any $\lambda \in \mathcal{R}^l$, we have $\tilde{w}^* := (x^*, \lambda) \in \mathcal{W}$. From the definition of $F(w)$ in (9), we have

$$\begin{aligned} (\tilde{w}^* &- \hat{w}^k)^\top F(\hat{w}^k) \\ &= (\tilde{w}^* - \hat{w}^k)^\top F(\tilde{w}^*) \\ &= \begin{pmatrix} x^* - \hat{x}^k \\ \lambda - \hat{\lambda}^k \end{pmatrix}^\top \begin{pmatrix} -\mathcal{A}^\top \lambda \\ \mathcal{A}x^* - b \end{pmatrix} \\ &= -\lambda^\top(\mathcal{A}x^* - \mathcal{A}\hat{x}^k) \\ &= \lambda^\top(\mathcal{A}\hat{x}^k - b), \end{aligned}$$

where the first equation follows from (10). Setting $w = \tilde{w}^*$ in (22), we get

$$\alpha\big(\theta(\hat{x}^k) - \theta(x^*) - (\tilde{w}^* - \hat{w}^k)^\top F(\hat{w}^k)\big) \leq \frac{1}{2}\big(\|\tilde{v}^* - v^k\|_H^2 - \|\tilde{v}^* - v^{k+1}\|_H^2\big) - \frac{\alpha}{2}\|v^k - \hat{v}^k\|_H^2.$$

Combining the above two inequalities gives

$$\alpha\big(\theta(\hat{x}^k) - \theta(x^*) - \lambda^\top(\mathcal{A}\hat{x}^k - b)\big) \leq \frac{1}{2}\big(\|\tilde{w}^* - w^k\|_H^2 - \|\tilde{v}^* - v^{k+1}\|_H^2\big) - \frac{\alpha}{2}\|v^k - \hat{v}^k\|_H^2.$$

Summing the above inequality from $k = 0$ to $t - 1$ yields

$$\sum_{k=0}^{t-1}\theta(\hat{x}^k) - t\theta(x^*) - \lambda^\top\left(\mathcal{A}\sum_{k=0}^{t-1}\hat{x}^k - tb\right) \leq \frac{1}{2\alpha}\|\tilde{v}^* - v^0\|_H^2.$$

Dividing both sides of the above inequality by t, we get

$$\frac{1}{t}\sum_{k=0}^{t-1}\theta\left(\hat{x}^k\right)-\theta\left(x^*\right)-\lambda^\top\left(\mathcal{A}\bar{x}^t-b\right)\leq\frac{1}{2\alpha t}\left\|\tilde{v}^*-v^0\right\|_H^2.$$

Then it follows from the convexity of θ_i ($i=1,2,3$) that

$$\theta\left(\bar{x}^t\right)-\theta\left(x^*\right)-\lambda^\top\left(\mathcal{A}\bar{x}^t-b\right)\leq\frac{1}{2\alpha t}\left\|\begin{pmatrix}y^0-y^*\\\lambda^0-\lambda\end{pmatrix}\right\|_H^2, \tag{23}$$

where $y^0=(x_2^0,x_3^0)$, $y^*=(x_2^*,x_3^*)$. Since (23) holds for any λ, we can set

$$\lambda=-\frac{\mathcal{A}\bar{x}^t-b}{\|\mathcal{A}\bar{x}^t-b\|},$$

and consequently,

$$\theta\left(\bar{x}^t\right)-\theta\left(x^*\right)+\|\mathcal{A}\bar{x}^t-b\|\leq\frac{1}{2\alpha t}\sup_{\|\lambda\|\leq1}\left\|\begin{pmatrix}y^0-y^*\\\lambda^0-\lambda\end{pmatrix}\right\|_H^2.$$

Set

$$c=\sup_{\|\lambda\|\leq1}\left\|\begin{pmatrix}y^0-y^*\\\lambda^0-\lambda\end{pmatrix}\right\|_H^2,$$

and we thus get

$$\theta\left(\bar{x}^t\right)-\theta\left(x^*\right)+\left\|\mathcal{A}\bar{x}^t-b\right\|\leq\frac{c}{2\alpha t}.$$

Since $x^*\in\mathcal{X}^*$ (here $\mathcal{X}^*$ denotes the solution set of problem (1)), we have

$$\theta\left(\bar{x}^t\right)-\theta\left(x^*\right)\geq0.$$

Combining the above two inequalities gives

$$\begin{cases}|\theta(\bar{x}^t)-\theta(x^*)|\leq\frac{c}{2\alpha t},\\\|\mathcal{A}\bar{x}^t-b\|\leq\frac{c}{2\alpha t},\end{cases}$$

which completes the proof. □

As mentioned in Sect. 1, He *et al.* [12] used a simple example to show that the iterative scheme (6) may diverge for two-block separable convex programming. If we set $\theta_1=0$, $A_1=0$ in (1) and MHD-ALM, then MHD-ALM reduces to the method in [12]. In this case, the feasible set of α in [12] is (0,0.3670), the same as that of the method in [25] for three-block separable convex programming. Now we use this example given in [12] to show that: (i) larger values of $\alpha\in(0,2-\sqrt{2})$ can enhance the performance of MHD-ALM; (ii) MHD-ALM with $\alpha\geq2-\sqrt{2}\cong0.5858$ may diverge.

Example 3.1 Consider the linear equation

$$x_2 + x_3 = 0. \tag{24}$$

Obviously, the linear equation (24) is a special case of problem (1) with the specifications: $\theta_1 = \theta_2 = \theta_3 = 0$, $A_1 = 0$, $A_2 = A_3 = 1$, $b = 0$, $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3 = \mathcal{R}$. Due to $\theta_1 = 0$, $A_1 = 0$, in the following we do not consider the variable x_1. The solution set of the corresponding mixed variation inequalities is

$$\mathcal{W}^* = \left\{ \left(x_2^*, x_3^*, \lambda^* \right) | x_2^* + x_3^* = 0, \lambda^* = 0 \right\}.$$

For MHD-ALM, we set $\beta = 1$, the initial point $x_2^0 = x_3^0 = 0$, $\lambda^0 = 1$, and choose

$$\alpha \in \{0.20, 0.21, 0.22, \dots, 0.55\}.$$

The stopping criterion is set as

$$\max\left\{ \left| x_2^k + x_3^k \right|, \left| \lambda^k \right| \right\} \le 10^{-5},$$

or the number of iterations exceeds 10,000.

The numerical results are graphically shown in Fig. 1, which illustrates that when $\alpha \le 0.5$, the number of iterations is descending with respect to α, while when $\alpha \in (0.5, 0.55)$, the number of iterations increases quickly. Therefore, $\alpha = 0.5$ is optimal for this problem, and larger values of α in its feasible set indeed can enhance the numerical performance of MHD-ALM. Of course, some extreme values, such as the values near the upper bound 0.5858, are not appropriate choices.

Figure 1 Sensitivity test on the step size α
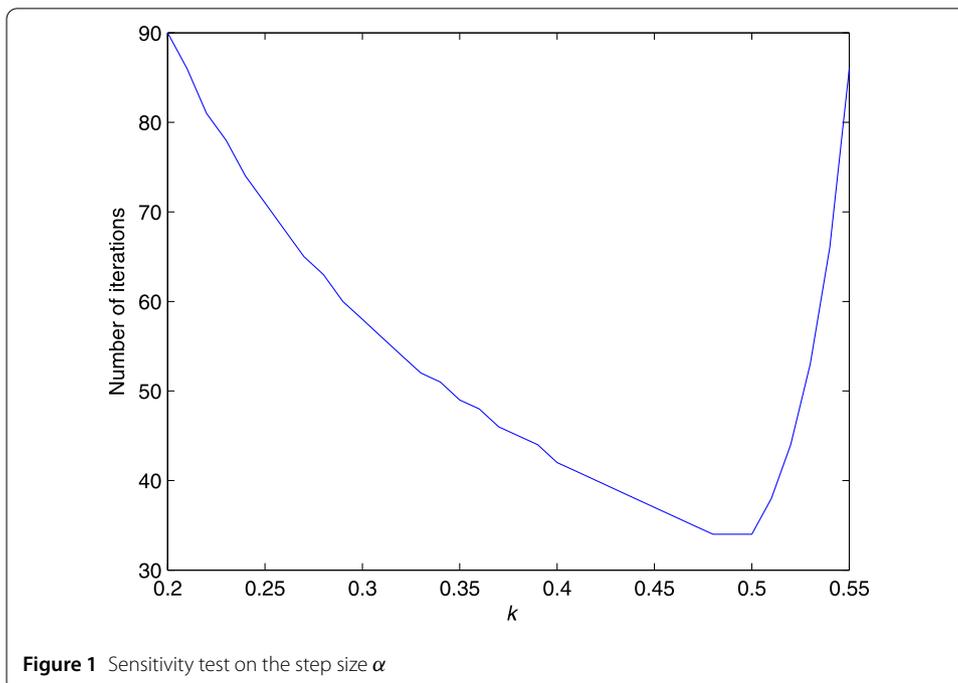
Now, we show that MHD-ALM may diverge for $\alpha \geq 2 - \sqrt{2}$. By some simple manipulations, the iterative scheme of (13) and (14) for problem (24) can be written in the following compact form:

$$\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} \end{pmatrix} = P(\alpha) \begin{pmatrix} x_2^k \\ x_3^k \\ \lambda^k \end{pmatrix},$$ (25)

where

$$P(\alpha) = \begin{pmatrix} 1 - \alpha & -\alpha & \alpha \\ -\alpha & 1 - \alpha & \alpha \\ \alpha & \alpha & 1 - 2\alpha \end{pmatrix}.$$

Three eigenvalues of the matrix $P(\alpha)$ are

$$\lambda_1 = 1, \qquad \lambda_2 = 1 - 2\alpha + \sqrt{2}\alpha, \qquad \lambda_3 = 1 - 2\alpha - \sqrt{2}\alpha.$$

Now let us consider the following two cases:

(1) For any $\alpha > 2 - \sqrt{2}$, we have

$$\lambda_3 = 1 - (2 + \sqrt{2})\alpha < 1 - (2 + \sqrt{2})(2 - \sqrt{2}) = -1.$$

Then $\rho(P(\alpha)) > 1$ for $\alpha > 2 - \sqrt{2}$, where $\rho(P(\alpha))$ is the spectral radius of $P(\alpha)$. Hence, the iterative scheme (25) with $\alpha > 2 - \sqrt{2}$ is divergent for this problem.

(2) For $\alpha = 2 - \sqrt{2}$, by eigenvalue decomposition, the matrix $P(2 - \sqrt{2})$ can be decomposed as

$$P(2 - \sqrt{2}) = VDV^\top,$$

where

$$V = \begin{pmatrix} -1/2 & -\sqrt{2}/2 & 1/2 \\ -1/2 & \sqrt{2}/2 & 1/2 \\ \sqrt{2}/2 & 0 & \sqrt{2}/2 \end{pmatrix}, \qquad D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4\sqrt{2} - 5 \end{pmatrix}.$$

Thus, by (25), we get

$$\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} -\frac{\sqrt{2}((-1)^k - (4\sqrt{2} - 5)^k)}{4} \\ -\frac{\sqrt{2}((-1)^k - (4\sqrt{2} - 5)^k)}{4} \\ \frac{(-1)^k}{2} + \frac{(4\sqrt{2} - 5)^k}{2} \end{pmatrix},$$

from which it holds that

$$\begin{pmatrix} x_2^{2k} \\ x_3^{2k} \\ \lambda^{2k} \end{pmatrix} \to \begin{pmatrix} -\frac{\sqrt{2}}{4} \\ -\frac{\sqrt{2}}{4} \\ \frac{1}{2} \end{pmatrix}, \qquad \begin{pmatrix} x_2^{2k+1} \\ x_3^{2k+1} \\ \lambda^{2k+1} \end{pmatrix} \to \begin{pmatrix} \frac{\sqrt{2}}{4} \\ \frac{\sqrt{2}}{4} \\ -\frac{1}{2} \end{pmatrix}.$$

Hence, the iterative scheme (25) with $\alpha = 2 - \sqrt{2}$ is also divergent for this problem. Overall, $2 - \sqrt{2}$ is the optimal upper bound of the step size α in MHD-ALM.

Now let us consider some special cases and extensions of MHD-ALM:

(1) Problem (1) with $\theta_1 = 0, A_1 = 0$ reduces to two-block separable convex programming, which can be solved by MHD-ALM as follows:

$$
\begin{cases}
\tilde{x}_2^k = \operatorname{argmin}\{\mathcal{L}_\beta(x_2, x_3^k, \lambda^k) | x_2 \in \mathcal{X}_2\}, \\
\tilde{x}_3^k = \operatorname{argmin}\{\mathcal{L}_\beta(x_2^k, x_3, \lambda^k) | x_3 \in \mathcal{X}_3\}, \\
\tilde{\lambda}^k = \lambda^k - \beta(A_2 \tilde{x}_2^k + A_3 \tilde{x}_3^k - b),
\end{cases}
\tag{26}
$$

and

$$
\begin{cases}
x_2^{k+1} = x_2^k - \alpha(x_2^k - \tilde{x}_2^k), \\
x_3^{k+1} = x_3^k - \alpha(x_3^k - \tilde{x}_3^k), \\
\lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k).
\end{cases}
\tag{27}
$$

Since the iterative scheme (26), (27) is a special case of MHD-ALM, it is convergent for any $\alpha \in (0, 2 - \sqrt{2})$. Furthermore, by Example 3.1, $2 - \sqrt{2}$ is the optimal upper bound of the constant step size α in (26), (27).

(2) Similarly, problem (1) with $\theta_3 = 0, A_3 = 0$ also reduces to two-block separable convex programming, which can be solved by MHD-ALM as follows:

$$
\begin{cases}
\tilde{x}_1^k = \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, \lambda^k) | x_1 \in \mathcal{X}_1\}, \\
\tilde{x}_2^k = \operatorname{argmin}\{\mathcal{L}_\beta(\tilde{x}_1^k, x_2, \lambda^k) | x_2 \in \mathcal{X}_2\}, \\
\tilde{\lambda}^k = \lambda^k - \beta(A_1 \tilde{x}_1^k + A_2 \tilde{x}_2^k - b),
\end{cases}
\tag{28}
$$

and

$$
\begin{cases}
x_1^{k+1} = \tilde{x}_1^k, \\
x_2^{k+1} = x_2^k - \alpha(x_2^k - \tilde{x}_2^k), \\
\lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k).
\end{cases}
\tag{29}
$$

Following a similar analysis procedure, we can prove that the iterative scheme (28), (29) is convergent for any $\alpha \in (0, 1)$.

(3) Extending MHD-ALM to solve four-block separable convex programming:

$$
\min\left\{ \sum_{i=1}^4 \theta_i(x_i) \,\Big|\, \sum_{i=1}^4 A_i x_i = b, x_i \in \mathcal{X}_i, i = 1, 2, 3, 4 \right\},
$$

we can get the following iterative scheme:

$$
\begin{cases}
\tilde{x}_1^k = \operatorname{argmin}\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, x_4^k, \lambda^k) | x_1 \in \mathcal{X}_1\}, \\
\tilde{x}_2^k = \operatorname{argmin}\{\mathcal{L}_\beta(\tilde{x}_1^k, x_2, x_3^k, x_4^k, \lambda^k) | x_2 \in \mathcal{X}_2\}, \\
\tilde{x}_3^k = \operatorname{argmin}\{\mathcal{L}_\beta(\tilde{x}_1^k, x_2^k, x_3, x_4^k, \lambda^k) | x_3 \in \mathcal{X}_3\}, \\
\tilde{x}_4^k = \operatorname{argmin}\{\mathcal{L}_\beta(\tilde{x}_1^k, x_2^k, x_3^k, x_4, \lambda^k) | x_4 \in \mathcal{X}_4\}, \\
\tilde{\lambda}^k = \lambda^k - \beta(\mathcal{A}_1 \tilde{x}^k - b),
\end{cases}
\tag{30}
$$

and

$$
\begin{cases}
x_1^{k+1} = \tilde{x}_1^k, \\
x_2^{k+1} = x_2^k - \alpha(x_2^k - \tilde{x}_2^k), \\
x_3^{k+1} = x_3^k - \alpha(x_3^k - \tilde{x}_3^k), \\
x_4^{k+1} = x_4^k - \alpha(x_4^k - \tilde{x}_4^k), \\
\lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k).
\end{cases}
\tag{31}
$$

With similar reasoning, we can prove that the iterative scheme (30), (31) is convergent for any $\alpha \in (0, 2 - \sqrt{3})$.

4 Numerical results

In this section, we demonstrate the practical efficiency of MHD-ALM by applying it to recover low-rank and sparse components of matrices from incomplete and noisy observation. Furthermore, to give more insight into the behavior of MHD-ALM, we compare it with the full Jacobian decomposition of the augmented Lagrangian method (FJD-ALM) [12] and the proximal partially parallel splitting method with constant step size (PPPSM) [25]. All experiments are performed on a Pentium(R) Dual-Core CPU T4400@2.2 GHz PC with 4 GB of RAM running on 64-bit Windows operating system.

The mathematical model of recovering low-rank and sparse components of matrices from incomplete and noisy observation is [20]

$$
\min_{L,S,U} \left\{ \|L\|_* + \tau \|S\|_1 + \frac{1}{2\mu} \|P_\Omega(U)\|^2 \right\},
$$

$$
\text{s.t. } L + S + U = D,
\tag{32}
$$

where $D \in \mathcal{R}^{p \times q}$ is a given matrix, $\tau > 0$ is a balancing parameter, $\mu > 0$ is a penalty parameter, $\Omega \subseteq \{1, 2, \ldots, p\} \times \{1, 2, \ldots, q\}$ is the index set of the observable entries of D, and $P_\Omega : \mathcal{R}^{p \times q} \to \mathcal{R}^{p \times q}$ is the projection operator defined by

$$
[P_\Omega(X)]_{ij} = \begin{cases} X_{ij}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{otherwise,} \end{cases} \quad 1 \le i \le p, 1 \le j \le q.
$$

Problem (32) is a concrete model of the generic problem (1), and MHD-ALM is applicable. For this problem, the three minimization problems in (13) all admit closed-form solutions, which can be found in [20].

4.1 Simulation example

We generate the synthetic data of (32) in the same way as [5, 20]. Specifically, let L^*, S^* be the low-rank matrix, the sparse matrix, respectively, and rr, spr, and sr represent the ratios of the low-rank ratio of L^* (i.e., r/p), the number of nonzero entries of S^* (i.e., $\|S^*\|_0/(pq)$), and the observed entries (i.e., $|\Omega|/(pq)$), respectively. The observed part of the matrix D is generated by the following Matlab scripts, in which b is the vectorization of D:

```
X = randn(m, rr * m) * randn(rr * m, n);    Omega = randperm(m * n);

p = round(sr * m * n);    Omega = Omega(1 : p);    Omega = Omega';
```

$$Y = \text{zeros}(m, n); \quad L = \text{round}\big(\min(spr, sr) * m * n\big); \quad nzp = \text{Omega}(1 : L);$$

$$Y(nzp) = \big(\text{rand}(L, 1) * 2 - 1\big) * 500; b = X + Y;$$

$$b = b(\text{Omega}); \quad sigma = 0.001; \quad b = b + sigma * \text{randn}(p, 1);$$

In this experiment, we set $\tau = 1/\sqrt{p}$, $\mu = \sqrt{p + \sqrt{8p\sigma}}/10$, $\beta = \frac{0.06|\Omega|}{\|P_\Omega(D)\|_1}$, the initial iterate $(L^0, S^0, U^0, \lambda^0) = (0, 0, 0, 0)$, and use the stopping criterion

$$\max\left\{ \frac{\|\tilde{L}^k - L^k\|}{1 + \|L^k\|}, \frac{\|\tilde{S}^k - S^k\|}{1 + \|S^k\|} \right\} < 10^{-4}, \tag{33}$$

or the number of iterations exceeds 500.

The parameters in the three tested methods are listed as follows:

FJD-ALM: $\alpha = 0.38$.

PPPSM: $S_i = 0$, $H = \beta I$, $Q = I$, $\alpha = 0.36$.

MHD-ALM: $\alpha = 0.5$.

In Tables 1 and 2, we report the numerical results of three tested methods, in which the number of iterations (denoted by 'Iter.'), the elapsed CPU time in seconds (denoted by 'Time'), the relative error of the recovered low-rank matrix, and the relative error of the recovered sparse matrix are reported when the stopping criterion (33) is satisfied.

Numerical results in Tables 1 and 2 indicate that: (i) all methods successfully solved all the tested cases; (ii) both MHD-ALM and PPPSM perform better than FJD-ALM, and MHD-ALM performed the best. The reason maybe that FJD-ALM updates all the primal variables in a parallel manner, while PPPSM and MHD-ALM update x_2, x_3 based on the newest updated x_1 to accelerate the convergence speed. Furthermore, the step size α of MHD-ALM is larger than that of PPPSM, and the latter is larger than that of FJD-ALM. Therefore, larger values of α can enhance the efficiency of the corresponding method.

Table 1 Numerical comparisons between different algorithms for $p = q = 500$

rr	spr	sr	Method	Iter.	Time	$\frac{\|L^k - L^*\|}{\|L^*\|}$	$\frac{\|S^k - S^*\|}{\|S^*\|}$
0.05	0.05	0.9	FJD-ALM	97	7.42	5.7325e−04	9.7830e−05
			PPPSM	133	14.70	2.4034e−04	3.6358e−05
			MHD-ALM	44	5.17	7.5910e−04	1.5025e−04
0.05	0.05	0.6	FJD-ALM	120	11.03	1.6528e−03	1.3102e−04
			PPPSM	181	17.62	5.4329e−04	5.1513e−05
			MHD-ALM	75	6.76	1.7148e−03	1.1343e−04
0.1	0.1	0.9	FJD-ALM	127	11.39	2.2566e−03	2.2374e−04
			PPPSM	175	17.79	5.2215e−04	6.1755e−05
			MHD-ALM	80	6.86	1.5097e−03	1.5994e−04
0.1	0.1	0.8	FJD-ALM	188	17.45	4.2439e−03	2.7536e−04
			PPPSM	222	20.78	1.2697e−03	9.5339e−05
			MHD-ALM	119	10.15	2.8338e−03	1.9242e−04
0.1	0.15	0.9	FJD-ALM	209	20.18	3.9031e−03	2.3423e−04
			PPPSM	222	25.95	1.3444e−03	9.5133e−05
			MHD-ALM	120	12.37	3.2420e−03	2.0704e−04

Table 2 Numerical comparisons between different algorithms for $p = q = 1000$

rr	spr	sr	Method	Iter.	Time	$\frac{\|L^k - L^*\|}{\|L^*\|}$	$\frac{\|S^k - S^*\|}{\|S^*\|}$
0.05	0.05	0.9	FJD-ALM	105	93.16	2.6808e−04	6.1154e−05
			PPPSM	129	131.08	9.9216e−05	2.3683e−05
			MHD-ALM	46	46.82	3.5053e−04	9.5144e−05
0.05	0.05	0.6	FJD-ALM	139	117.67	7.1439e−04	6.8845e−05
			PPPSM	147	138.08	2.1260e−04	3.4427e−05
			MHD-ALM	57	53.31	5.9883e−04	9.8047e−05
0.1	0.1	0.9	FJD-ALM	119	100.13	9.3082e−04	1.6381e−04
			PPPSM	188	179.91	2.3296e−04	4.2820e−05
			MHD-ALM	61	57.04	8.9591e−04	1.6198e−04
0.1	0.1	0.8	FJD-ALM	133	114.80	1.2760e−03	1.8083e−04
			PPPSM	192	182.46	3.7423e−04	5.2624e−05
			MHD-ALM	76	67.08	1.3497e−03	1.6979e−04
0.1	0.15	0.9	FJD-ALM	139	117.89	1.9819e−03	2.3040e−04
			PPPSM	206	188.47	4.4773e−04	5.7452e−05
			MHD-ALM	84	76.34	1.6396e−03	1.8795e−04

4.2 Application example

In this subsection, we apply the proposed method to solve the video background extraction problem with missing and noisy data [29]. There is a video taken in an airport, which consists of 200 grayscale frames with each frame having 144×176 pixels. We need to separate its background and foreground. Vectorizing all frames of the video, we get a matrix $D \in \mathcal{R}^{25,344 \times 50}$, and each column represents a frame. Let $L, S \in \mathcal{R}^{25,344 \times 200}$ be the matrix representations of its background and foreground (i.e., the moving objects), respectively. Then the rank of L is equal to one exactly, and S should be sparse with only a small number of nonzero elements. We consider only a fraction entries of D can be observed, whose indices are collected in the index set Ω. Then the background extraction problem with missing and noisy data can be casted as problem (32). In the experiment, the parameters in MHD-ALM are set as $\alpha = 0.5$, $\beta = \frac{0.005|\Omega|}{\|P_\Omega(D)\|_1}$, the parameters in (32) are set as $\tau = 1/\sqrt{p}$, $\mu = 0.01$, and the initial iterate $(L^0, S^0, U^0, \lambda^0) = (0, 0, 0, 0)$. We use the same stopping criterion as (33) with the tolerance 10^{-2}.

Figure 2 displays the separation results of the 10th and 125th frames of the video with sr $= 0.7$, which indicate that the proposed MHD-ALM successfully separates the background and foreground of the two frames.

5 Conclusion

In this paper, a hybrid decomposition of the augmented Lagrangian method is proposed for three-block separable convex programming, whose most important characteristic is that its correction step adopts a constant step size. We showed that the optimal upper bound of the constant step size is $2 - \sqrt{2}$. Preliminary numerical results indicate that the proposed method is more efficient than similar methods in the literature.

The following two issues deserve further researching: (i) Due to Condition 2.1 being only a sufficient condition to ensure the convergence of the ProAlo, is 1 the optimal upper bound of α in the iterative scheme (28), (29)? Similarly, is $2 - \sqrt{2}$ the optimal upper bound of α in the iterative scheme (30), (31)? (ii) If we choose different step sizes for x_2, x_3, λ in the correction step of MHD-ALM, the feasible set of these step sizes needs more discussion.

Figure 2 The 10th and 125th frames of the clean video and the corresponding corrupted frames with $\mathtt{sr} = 0.7$ (the top and third lines); the extracted background and foreground by MHD-ALM (the second and fourth lines)

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that there are no competing interests.

Authors' contributions

The first author provided the problems and gave the proof of the main results, and the second author finished the numerical experiment. All authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. **20**, 33–61 (1998)
2. Sun, M., Liu, J.: A proximal Peaceman–Rachford splitting method for compressive sensing. J. Appl. Math. Comput. **50**, 349–363 (2016)
3. Sun, M., Liu, J.: An accelerated proximal augmented Lagrangian method and its application in compressive sensing. J. Inequal. Appl. **2017**, 263 (2017)
4. Candés, E.J., Li, X.D., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM **58**(1), 1–37 (2011)
5. Tao, M., Yuan, X.M.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. **21**, 57–81 (2011)
6. Sun, M., Wang, Y.J., Liu, J.: Generalized Peaceman–Rachford splitting method for multiple-block separable convex programming with applications to robust PCA. Calcolo **54**(1), 77–94 (2017)
7. Sun, M., Sun, H.C., Wang, Y.J.: Two proximal splitting methods for multi-block separable programming with applications to stable principal component pursuit. J. Appl. Math. Comput. **56**, 411–438 (2018)
8. He, B.S., Yuan, X.M., Zhang, W.X.: A customized proximal point algorithm for convex minimization with linear constraints. Comput. Optim. Appl. **56**, 559–572 (2013)
9. He, B.S., Liu, H., Wang, Z.R., Yuan, X.M.: A strictly contractive Peaceman–Rachford splitting method for convex programming. SIAM J. Optim. **24**(3), 1011–1040 (2014)
10. He, B.S., Tao, M., Yuan, X.M.: A splitting method for separable convex programming. IMA J. Numer. Anal. **35**(1), 394–426 (2015)
11. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput. Math. Appl. **2**, 17–40 (1976)
12. He, B.S., Hou, L.S., Yuan, X.M.: On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. SIAM J. Optim. **25**(4), 2274–2312 (2015)
13. Han, D.R., Kong, W.W., Zhang, W.X.: A partial splitting augmented Lagrangian method for low patch-rank image decomposition. J. Math. Imaging Vis. **51**(1), 145–160 (2015)
14. Wang, Y.J., Zhou, G.L., Caccetta, L., Liu, W.Q.: An alternative Lagrange-dual based algorithm for sparse signal reconstruction. IEEE Trans. Signal Process. **59**, 1895–1901 (2011)
15. Wang, Y.J., Liu, W.Q., Caccetta, L., Zhou, G.L.: Parameter selection for nonnegative ℓ_1 matrix/tensor sparse decomposition. Oper. Res. Lett. **43**, 423–426 (2015)
16. Hestenes, M.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
17. Chen, C.H., He, B.S., Ye, Y.Y., Yuan, X.M.: The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. Math. Program. **155**(1), 57–79 (2016)
18. Sun, D.F., Toh, K.C., Yang, L.: A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type of constraints. SIAM J. Optim. **25**, 882–915 (2015)
19. He, B.S., Xu, H.K., Yuan, X.M.: On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM. J. Sci. Comput. **66**(3), 1204–1217 (2016)
20. Hou, L.S., He, H.J., Yang, J.F.: A partially parallel splitting method for multiple-block separable convex programming with applications to robust PCA. Comput. Optim. Appl. **63**(1), 273–303 (2016)
21. Wang, J.J., Song, W.: An algorithm twisted from generalized ADMM for multi-block separable convex minimization models. J. Comput. Appl. Math. **309**, 342–358 (2017)
22. Sun, M., Liu, J.: The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization. J. Inequal. Appl. **2017**, 19 (2017)
23. Sun, M., Sun, H.C.: Improved proximal ADMM with partially parallel splitting for multi-block separable convex programming. J. Appl. Math. Comput. **58**, 151–181 (2018)
24. He, B.S., Tao, M., Yuan, X.M.: Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J. Optim. **22**, 313–340 (2012)
25. Wang, K., Desai, J., He, H.J.: A proximal partially-parallel splitting method for separable convex programs. Optim. Methods Softw. **32**(1), 39–68 (2017)
26. Chang, X.K., Liu, S.Y., Zhao, P.J., Li, X.: Convergent prediction-correction-based ADMM for multi-block separable convex programming. J. Comput. Appl. Math. **335**, 270–288 (2018)
27. He, B.S., Ma, F., Yuan, X.M.: Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming. Optimization-online, 5569 (2016)
28. He, B.S., Yuan, X.M.: On the direct extension of ADMM for multi-block separable convex programming and beyond: from variational inequality perspective. Optimization-online, 4293 (2014)
29. Bouwmans, T.: Traditional and recent approaches in background modeling for foreground detection: an overview. Comput. Sci. Rev. **11–12**, 31–66 (2014)