manuscript No.
(will be inserted by the editor)

A generalization of linearized alternating direction method of multipliers for solving two-block separable convex programming

Xiaokai Chang[1,2] · **Sanyang Liu**[1] · **Pengjun Zhao**[3] · **Dunjiang Song**[4]

Received: date / Accepted: date

Abstract The linearized alternating direction method of multipliers (ADMM), with indefinite proximal regularization, has been proved to be efficient for solving separable convex optimization subject to linear constraints. In this paper, we present a generalization of linearized ADMM (G-LADMM) to solve two-block separable convex minimization model, which linearizes all the subproblems by choosing a proper positive-definite or indefinite proximal term and updates the Lagrangian multiplier twice in different ways. Furthermore, the proposed G-LADMM can be expressed as a proximal point algorithm (PPA), and all the subproblems are just to estimate the proximity operator of the function in the objective. We specify the domain of the proximal parameter and stepsizes to guarantee that G-LADMM is globally convergent. It turns out that our convergence domain of the proximal parameter and stepsizes is significantly larger than other convergence domains in the literature. The numerical experiments illustrate the improvements of the proposed G-LADMM to solve LASSO and image decomposition problems.

Keywords alternating direction method of multipliers · proximal point algorithm · separable convex programming · linearization · indefinite proximal regularization · LASSO

Mathematics Subject Classification (2000) 65K10 · 90C25 · 90C30

1 Introduction.

Since wide applications of the separable convex minimization problem with linear constraints, such as in the dual of the (quadratic) semidefinite programming (SDP) with or without nonnegative constraints [1–3], the robust principal component analysis model with noisy and incomplete data [4,5], low-rank reconstruction [6] and image decomposition [7–9], it has attracted a significant level of attention. The alternating direction method of multipliers (ADMM) [10, 11] is a benchmark method for solving the separable convex minimization problem with linear constraints, and can be extended to solve the nonconvex and nonsmooth block problems, we refer to [12–16] for studies.

Recently, the linearized versions of ADMM (LADMM) [17–21] were developed by choosing the proximal terms appropriately, which can alleviate an ADMM subproblem as easy as estimating the

✉ Xiaokai Chang,
xkchang@lut.cn

1 School of Mathematics and Statistics, Xidian University, Xi'an 710071, P. R. China.
2 College of Science, Lanzhou University of Technology, Lanzhou, Gansu 730050, P. R. China.
3 School of Mathematics and Computer Application, Shangluo University, Shangluo, 726000, P. R. China.
4 Institutes of Science and Development, Chinese Academy of Sciences, Beijing, 100190, P. R. China.

proximity operator of a function in the objective for some applications. For the convergence rate of LADMM, we refer the reader to [22–24]. Generally, the proximal term matrix was assumed to be positive semi-definite and the convergence can be obtained under this condition. However, this is not beneficial for obtaining an improved numerical performance. Requiring Lipschitz continuous on the objective, a majorized ADMM with indefinite proximal terms was studied in [25]. In [18], it was shown by He et al. that the proximal terms can be indefinite by introducing a proximal parameter, and the convergence of LADMM can be obtained without any additional assumptions. In [17, 20], the symmetric ADMM and LADMM with indefinite proximal regularization were presented, and the range of the proximal parameter was explored. Since the proximal term with a smaller proximal parameter has a less weight in the objective function of the subproblems, it is thus possible to accelerate the numerical performance. Hence, it is crucial to determine a smaller value of the proximal parameter while the convergence of LADMM can be still guaranteed.

Mainly motivated by the work on the symmetric and linearized ADMM, we present a generalization of linearized ADMM (G-LADMM) with special positive-definite and indefinite proximal regularization for solving two-block separable convex problem, and improve the value of the proximal parameter. What's more, the G-LADMM can be simplified as a proximal point algorithm (PPA). All the subproblems are just to estimate the proximity operator of the function in the objective, and will be easy enough to have closed-form solutions for many applications (e.g., the examples of sparse optimization in various areas such as compressed sensing, image processing, and statistics). To theoretically guarantee the convergence of the proposed ADMM, we derive the convergence-ensured domain of the proximal parameter for the linearized subproblems and stepsizes for the dual variable, which is significantly larger than other convergence domains in the literature. Namely, a smaller value of the proximal parameter is obtained under the same stepsizes than that in the literature. Or, from another perspective, the range of proximal parameter is improved, and a larger stepsize can be obtained under the same proximal parameter. Finally, we illustrate the numerical efficiency of the proposed method on solving the image decomposition and LASSO problems generated randomly.

The rest of this paper is organized as follows. In Section 2, we state the problem under discussion and recall some algorithms that have been discussed in the literature. Section 3 provides a detail exposition of the G-LADMM. In Section 4, we interpret the proposed G-LADMM as a prediction-correction procedure. Based on this interpretation, the convergence is established in Section 5 by using the variational inequality. The implementation and numerical experiments, for solving the LASSO problems generated randomly and image low patch-rank decomposition are provided in Section 6. We conclude our paper in the final section.

2 The Model and Some Algorithms

In this paper, we consider the following classic separable convex minimization model:

$$\begin{array}{ll} \min & f(x) + g(y) \\ s.t. & Ax + By = c, \\ & x \in \mathcal{X}, \quad y \in \mathcal{Y}, \end{array} \tag{1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are closed convex sets in the finite-dimensional real Euclidean space. The functions $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{Y} \to \mathbb{R}$ are closed and proper convex (possibly nonsmooth). The operators $A : \mathcal{X} \to \mathbb{R}^m$ and $B : \mathcal{Y} \to \mathbb{R}^m$ are linear, and $c \in \mathbb{R}^m$ is a given vector. Throughout, we assume that the solution set of problem (1) is nonempty.

The augmented Lagrangian function of problem (1) has the following form:

$$\mathcal{L}_\sigma(x, y, \lambda) = f(x) + g(y) - \langle \lambda, Ax + By - c\rangle + \frac{\sigma}{2}\|Ax + By - c\|^2,$$

where $\lambda \in \mathbb{R}^m$, $\sigma > 0$ is a penalty parameter, $\langle \cdot, \cdot \rangle$ represents the standard inner product and $\| \cdot \|$ represents the l_2-norm throughout.

Given a chosen initial point, the classic ADMM to problem (1) consists of the iterations:

$$\left.\begin{aligned} x^{k+1} &= \arg\min_{x\in\mathcal{X}} \mathcal{L}_\sigma(x,\ y^k,\ \lambda^k), \\ y^{k+1} &= \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\sigma(x^{k+1},\ y,\ \lambda^k), \\ \lambda^{k+1} &= \lambda^k - \varrho\sigma\left(Ax^{k+1} + By^{k+1} - c\right), \end{aligned}\right\} \tag{2}$$

where $\varrho > 0$, e.g., $\varrho \in (0, \frac{1+\sqrt{5}}{2})$, is called a stepsize of the dual variable. In order to improve the performance of the classic ADMM (2) with the step-length $\varrho = 1$, Eckstein and Bertsekas [26] proposed the following generalized ADMM (G-ADMM) scheme:

$$\left.\begin{aligned} x^{k+1} &= \arg\min_{x\in\mathcal{X}} \mathcal{L}_\sigma(x, y^k, \lambda^k), \\ y^{k+1} &= \arg\min_{y\in\mathcal{Y}} g(y) - \langle \lambda^k, y\rangle + \tfrac{1}{2}\|\rho Ax^{k+1} + (1-\rho)(c - By^k) + By - c\|^2, \\ \lambda^{k+1} &= \lambda^k - \sigma\left(\rho Ax^{k+1} + (1-\rho)(c - By^k) + By^{k+1} - c\right), \end{aligned}\right\} \tag{3}$$

where $\rho \in (0, 2)$ is the relaxation factor. The scheme (3) can numerically accelerate the classic ADMM with some values of ρ, e.g., $\rho \in (1, 2)$. We refer to [27] for empirical studies of the acceleration performance of the G-ADMM.

In addition, He et al. [28] presented a symmetric ADMM (S-ADMM), which performs the following updating scheme:

$$\left.\begin{aligned} x^{k+1} &= \arg\min_{x\in\mathcal{X}} \mathcal{L}_\sigma(x, y^k, \lambda^k), \\ \lambda^{k+\frac{1}{2}} &= \lambda^k - \alpha\sigma\left(Ax^{k+1} + By^k - c\right), \\ y^{k+1} &= \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\sigma(x^{k+1}, y, \lambda^{k+\frac{1}{2}}), \\ \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \beta\sigma\left(Ax^{k+1} + By^{k+1} - c\right), \end{aligned}\right\} \tag{4}$$

and the stepsizes (α, β) were restricted into a domain in order to ensure its global convergence. From the numerical performance on solving the widely used basis pursuit model and the total-variational image debarring model, the S-ADMM (4) significantly outperforms the original ADMM in both the CPU time and the number of iterations. Bai et al. [29] proposed a generalized symmetric ADMM (GS-ADMM), in which the convergence domain for the stepsizes (α, β) is significantly improved and larger than the domain of S-ADMM introduced in [28]. Based on the special S-ADMM with $\beta = 1$, Gao and Ma [17] considered the following linearized symmetric ADMM with indefinite proximal regularization (ID-SADMM),

$$\left.\begin{aligned} x^{k+1} &= \arg\min_{x\in\mathcal{X}} \mathcal{L}_\sigma(x, y^k, \lambda^k), \\ \lambda^{k+\frac{1}{2}} &= \lambda^k - \alpha\sigma\left(Ax^{k+1} + By^k - c\right), \\ y^{k+1} &= \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\sigma(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) + \tfrac{1}{2}\|y - y^k\|_D^2, \\ \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \sigma\left(Ax^{k+1} + By^{k+1} - c\right), \end{aligned}\right\} \tag{5}$$

where$D = \tau r\mathcal{I} - \sigma B^*B$ with $r > \sigma\|B^*B\|$ (the superscript * denotes the adjoint), $\tau \in [\frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}, 1)$ and $\alpha \in (-1, 1)$.

However, the range of τ presented by Gao and Ma [17] is not optimal, see a special case of ID-SADMM with $\alpha = 0$ [18,20] (named IP-LADMM), in which $\tau \in (0.75, 1)$ and it is shown that any value of τ smaller than 0.75 can yield divergence of IP-LADMM. In this paper, we present a generalization of the scheme (5) and obtain an improved range of τ, see Figure 1 and (41), which covers that of the ID-SADMM [17] and IP-LADMM [20].

3 Generalization of Linearized ADMM

Firstly, we set up some notations and terminologies, and then introduce our G-LADMM for solving two-block separable convex programming (1).

3.1 Notations

We define the following notations that will be used later:

$$\gamma = \alpha + \beta, \tag{6}$$

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad \theta(u) = f(x) + g(y), \tag{7}$$

$$\mathcal{F}(w) = \begin{pmatrix} -A^*\lambda \\ -B^*\lambda \\ Ax + By - c \end{pmatrix} \quad \text{and} \quad \mathcal{W} = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m. \tag{8}$$

By making use of the notations in (6)-(8), the Lagrangian function of problem (1) can be defined as

$$L(u, \lambda) = \theta(u) - \langle \lambda, Ax + By - c \rangle, \tag{9}$$

for any $(u, \lambda) \in \mathcal{W}$. Then solving (1) is equivalent to finding a saddle point of $L(u, \lambda)$. By the convex analysis [30, Theorem 28.3], $\bar{w} = (\bar{u}, \bar{\lambda}) \in \mathcal{W}$ is a saddle point of $L(u, \lambda)$ if and only if the following variational inequality is satisfied:

$$\theta(u) - \theta(\bar{u}) + \langle w - \bar{w}, \mathcal{F}(\bar{w}) \rangle \geq 0, \quad \forall w \in \mathcal{W}. \tag{10}$$

If $\bar{w}$ is a saddle-point of the Lagrangian function (9) associated to problem (1), then $\bar{u}$ is a solution of problem (1). We denote by $\overline{\mathcal{W}}$ the solution set of the variational inequality (10), and it is nonempty because of the nonemptyness of the solution set of problem (1).

3.2 Generalization of Linearized ADMM

Combining the linearization to the quadratic term with the acceleration [26], We design a special linearized ADMM having the following scheme.

Algorithm 1 (G-LADMM for solving problem (1).)

Step 0. Let $\sigma > 0$ be a given parameter. Choose $w^0 \in \mathcal{W}$. Set $k = 0$.
Step 1. Compute subproblem with variable x, and update the Lagrange multiplier,

$$\left.\begin{aligned} x^{k+1} &= \arg\min_{x \in \mathcal{X}} \left\{ f(x) - \langle \lambda^k, Ax \rangle + \tfrac{\sigma}{2}\|Ax + By^k - c\|^2 \right\} + \tfrac{1}{2}\|x - x^k\|^2_{D_A}, \\ \lambda^{k+\frac{1}{2}} &= \lambda^k - \alpha\sigma\left(Ax^{k+1} + By^k - c\right), \end{aligned}\right\} \tag{11}$$

where α is a step size, and

$$D_A = r_A \mathcal{I} - \sigma A^* A, \quad \text{with} \quad r_A > \sigma\|A^* A\|. \tag{12}$$

Step 2. Compute subproblem with variable y, and update the Lagrange multiplier,

$$\left.\begin{array}{l} y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \left\{ \begin{array}{l} g(y) - \langle \lambda^{k+\frac{1}{2}}, By\rangle + \frac{\sigma}{2}\|\beta(Ax^{k+1} + By^k - c) \\ -By^k + By\|^2 + \frac{1}{2}\|y - y^k\|^2_{D_B} \end{array} \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \sigma\left(\beta(Ax^{k+1} + By^k - c) - By^k + By^{k+1}\right), \end{array}\right\} \tag{13}$$

where β is the relaxation factor, and

$$D_B = \tau r_B \mathcal{I} - \sigma B^*B, \quad \text{with} \quad r_B > \sigma\|B^*B\|, \quad \tau \in (0,1). \tag{14}$$

Step 3. Replace k by $k+1$, and return to Step 1.

Remark 1 *The convergence of G-LADMM is only related to the sum $\gamma = \alpha + \beta$ from our following analysis, namely $\beta = 0$ can be accepted and then its iteration scheme reads as*

$$\left.\begin{array}{l} \lambda^{k+\frac{0}{2}} = \lambda^k - \sigma\left(Ax^k + By^k - c\right), \\ x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \left\{ f(x) + \frac{r_A}{2}\left\| x - x^k - \frac{A^*(\lambda^{k+\frac{0}{2}})}{r_A} \right\|^2 \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\sigma\left(Ax^{k+1} + By^k - c\right), \\ y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \left\{ g(y) + \frac{\tau r_B}{2}\left\| y - y^k - \frac{B^*(\lambda^{k+\frac{1}{2}})}{\tau r_B} \right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \sigma\left(By^{k+1} - By^k\right), \end{array}\right\} \tag{15}$$

for any $\alpha \in (0,2)$. Here, we abuse slightly the notation "$k + \frac{0}{2}$" in (15), it is not k by calculation. That is, $\lambda^{k+\frac{0}{2}}$ and λ^k are different.

From Remark 1, all the subproblems in G-LADMM are just to estimate the proximity operator of the function in the objective, without computing $(\partial f + \sigma A^*A)^{-1}$ and $(\partial g + \sigma B^*B)^{-1}$, which may be expensive. Theoretically, we can understand the third updating scheme of the Lagrange multiplier as a proximal point regularization, where the proximity to the previous one is controlled by the difference of the just-obtained primal variable y^{k+1}.

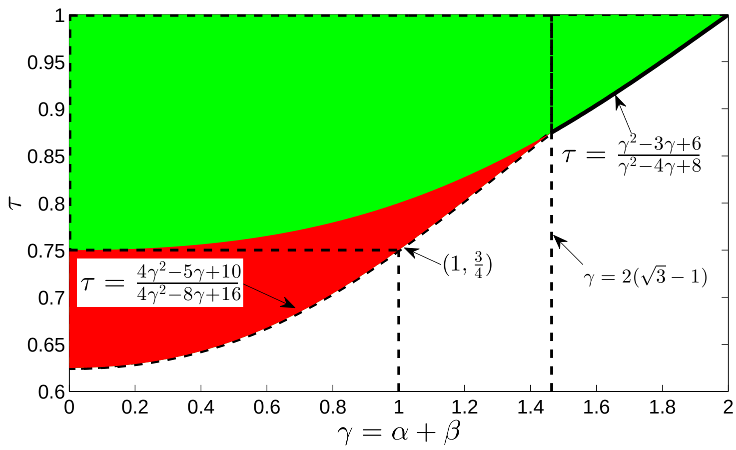


Fig. 1 The improved domain of the parameters (τ, γ).

Obviously, the proposed G-LADMM recovers ID-SADMM scheme (5) developed by Gao and Ma [17] when $D_A = 0$ in (12) and $\beta = 1$. Note that ID-SADMM is based on the symmetric ADMM, and it is a generalization of many existing linearized ADMM schemes, linearizing the quadratic term of the second subproblem, for instance with positive-definite proximal regularization as in [19,21] and with indefinite proximal regularization as in [18,20]. In addition, the convergence domain for the parameters (τ,γ), shown in Figure 1, is significantly larger than the domain given in [17]. The improved domain is labeled as red. For the detail to explain, see Sections 5.2 and 5.3. Under the convergence of G-LADMM can be theoretically still ensured, we obtain a smaller value of the proximal parameter when the same sum γ is used, which can be arbitrarily close to 5/8 when the sum γ is close to 0. Furthermore, from [20] the improved domain recovers the optimal proximal parameter $\tau > \frac{3}{4}$ when $\gamma = 1$.

4 A Prediction-Correction Interpretation

Following a similar approach in [29], we interpret G-LADMM as a prediction-correction procedure. First, we define

$$\left.\begin{array}{l} \widetilde{\lambda}^k = \lambda^k - \sigma\left(Ax^{k+1} + By^k - c\right), \\ u^k = \begin{pmatrix} x^k \\ y^k \end{pmatrix} \ , \widetilde{u}^k = \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix}, \quad w^k = \begin{pmatrix} u^k \\ \lambda^k \end{pmatrix}, \quad \widetilde{w}^k = \begin{pmatrix} \widetilde{u}^k \\ \widetilde{\lambda}^k \end{pmatrix}, \end{array}\right\} \tag{16}$$

and $D_0 = r_B\mathcal{I} - \sigma B^*B$. Then, the optimality conditions of the iterations (11) and (13) can be summarized as:

$$f(x) - f(x^{k+1}) + \left\langle x - x^{k+1}, -A^*(\lambda^k - \sigma r^k) + D_A(x^{k+1} - x^k)\right\rangle \geq 0, \forall x \in \mathcal{X}. \tag{17}$$

$$\begin{aligned} g(y) - g(y^{k+1}) + \Big\langle y - y^{k+1}, -B^*[\lambda^{k+\frac{1}{2}} - \sigma(\beta r^k - By^k + By^{k+1})]\Big\rangle \\ + \left\langle y - y^{k+1}, D_B(y^{k+1} - y^k)\right\rangle \geq 0, \forall y \in \mathcal{Y}. \end{aligned} \tag{18}$$

where $r^k = Ax^{k+1} + By^k - c$.

Combining with the definition of D_B in (14), (18) can be simplified as

$$g(y) - g(y^{k+1}) + \Big\langle y - y^{k+1}, -B^*[\lambda^{k+\frac{1}{2}} - \sigma\beta r^k] + \tau r_B(y^{k+1} - y^k)\Big\rangle \geq 0. \tag{19}$$

From the definition of $\widetilde{\lambda}^k$ in (16), we have

$$\lambda^{k+\frac{1}{2}} = \widetilde{\lambda}^k + (\alpha - 1)(\widetilde{\lambda}^k - \lambda^k) = \lambda^k - \alpha(\lambda^k - \widetilde{\lambda}^k). \tag{20}$$

Thus by $\widetilde{\lambda}^k = \lambda^k - \sigma r^k$, (17) and (19) can be written as

$$\begin{cases} f(x) - f(x^{k+1}) + \Big\langle x - x^{k+1}, -A^*\widetilde{\lambda}^k + D_A(x^{k+1} - x^k)\Big\rangle \geq 0, \\ g(y) - g(y^{k+1}) + \Big\langle y - y^{k+1}, -B^*[\widetilde{\lambda}^k + (\alpha + \beta - 1)(\widetilde{\lambda}^k - \lambda^k)] + \tau r_B(y^{k+1} - y^k)\Big\rangle \geq 0. \end{cases} \tag{21}$$

Lemma 1 ***(Prediction step)*** *Let $\{w^k\}$ be generated by G-LADMM and $\{\widetilde{w}^k\}$ be defined in (16), then for any $w \in \mathcal{W}$, we have*

$$\theta(u) - \theta(\widetilde{u}^k) + \langle w - \widetilde{w}^k, \mathcal{F}(\widetilde{w}^k)\rangle \geq \langle w - \widetilde{w}^k, \mathcal{Q}(w^k - \widetilde{w}^k)\rangle,$$

where

$$\mathcal{Q} = \left(\begin{array}{c:c:c} D_A & 0 & 0 \\ \hdashline 0 & \tau r_B\mathcal{I} & (1-\gamma)B^* \\ \hdashline 0 & -B & \frac{1}{\sigma}\mathcal{I} \end{array}\right).$$

Proof Notice that from the definition of $\widetilde{\lambda}^k$ in (16), we have

$$Ax^{k+1} + By^{k+1} - c - (By^{k+1} - By^k) + \frac{1}{\sigma}(\widetilde{\lambda}^k - \lambda^k) = 0,$$

namely,

$$\left\langle \lambda - \widetilde{\lambda}^k, \quad Ax^{k+1} + By^{k+1} - c - (By^{k+1} - By^k) + \frac{1}{\sigma}(\widetilde{\lambda}^k - \lambda^k) \right\rangle = 0, \quad \forall \lambda \in \mathbb{R}^m.$$

Together with (21) and the definition of $\mathcal{F}(w)$ in (8), we can get the result easily. □

By (16) and (20), the updating of λ^{k+1} can be represented as

$$\begin{aligned}\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \beta\sigma\left(Ax^{k+1} + By^k - c\right) + \sigma B(y^k - y^{k+1}) \\ &= \lambda^k - \alpha(\lambda^k - \widetilde{\lambda}^k) - \beta(\lambda^k - \widetilde{\lambda}^k) + \sigma B(y^k - y^{k+1}) \\ &= \lambda^k - (\alpha + \beta)(\lambda^k - \widetilde{\lambda}^k) + \sigma B(y^k - y^{k+1}).\end{aligned}$$

Thus, we have following results.

Lemma 2 ***(Correction step)*** *Let $\{w^k\}$ be generated by G-LADMM and $\{\widetilde{w}^k\}$ be defined in (16), the following equality holds*

$$w^{k+1} = w^k - \mathcal{M}(w^k - \widetilde{w}^k), \tag{22}$$

where

$$\mathcal{M} = \begin{pmatrix} \mathcal{I} & 0 & 0 \\ 0 & \mathcal{I} & 0 \\ 0 & -\sigma B & \gamma\mathcal{I} \end{pmatrix}. \tag{23}$$

Lemmas 1 and 2 show that G-LADMM can be interpreted as a prediction-correction framework. The prediction step is to compute $\widetilde{w}^k$, then w^{k+1} is obtained by the correction step (22). In the next section, we analyse the convergence of G-LADMM by this prediction-correction interpretation.

5 Convergence Analysis

For a special case of G-LADMM with $D_A = 0$, $\beta = 1$ and $\alpha \in (-1, 1)$, Gao and Ma [17] introduced sufficient conditions to obtain its convergence and provided a convergence rate. However, a smaller proximal parameter τ is particularly meaningful for generating larger step sizes for solving the subproblems. Thus, in this section, we are concentrating on a smaller proximal parameter τ under the same stepsizes, but the convergence can be established. In the following, we slightly abuse the notation $\|v\|^2_{\mathcal{M}} := \langle v, \mathcal{M}v \rangle$ when $\mathcal{M}$ is not positive definite.

5.1 Basic Properties

Lemma 3 *Let $\mathcal{H} = \mathcal{Q}\mathcal{M}^{-1}$, then the operator $\mathcal{H}$ is symmetric positive definite, if $\tau > \gamma - 1$ with $\gamma \in (0, 2)$.*

Proof By simple calculations, we can obtain

$$\mathcal{H} = \begin{pmatrix} D_A & 0 \\ 0 & \mathcal{H}_0 \end{pmatrix} \quad \text{and} \quad \mathcal{H}_0 = \frac{1}{\gamma}\left(\begin{array}{c:c} \gamma\tau r_B \mathcal{I} + \sigma(1-\gamma)B^*B & (1-\gamma)B^* \\ \hdashline (1-\gamma)B & \frac{1}{\sigma}\mathcal{I} \end{array}\right). \tag{24}$$

If $\gamma = 1$, then $\tau > 0$ and $\mathcal{H}_0$ is symmetric positive definite from $r_B > 0$ and $\sigma > 0$. Note that from the positive-definition of D_A, the result is right for $\gamma = 1$. Next, we testify the case of $\gamma \neq 1$.

From $r_B > \sigma\|B^*B\|$, for any $\gamma \in (0,2)$ and $\gamma \neq 1$ we have $\mathcal{H}_0 \succ \widetilde{\mathcal{H}}$ and

$$\widetilde{\mathcal{H}} = \frac{1-\gamma}{\gamma}\begin{pmatrix} \frac{\gamma\tau+1-\gamma}{1-\gamma}\sigma B^*B & B^* \\ B & \frac{\mathcal{I}}{\sigma(1-\gamma)}\mathcal{I} \end{pmatrix}.$$

Since the operator $\widetilde{\mathcal{H}}$ can be decomposed as $\widetilde{\mathcal{H}} = \widetilde{\mathcal{D}}^*\widetilde{\mathcal{H}}_0\widetilde{\mathcal{D}}$, where

$$\widetilde{\mathcal{H}}_0 = \frac{1-\gamma}{\gamma}\begin{pmatrix} \frac{\gamma\tau+1-\gamma}{1-\gamma}\mathcal{I} & \mathcal{I} \\ \mathcal{I} & \frac{1}{1-\gamma}\mathcal{I} \end{pmatrix} \text{ and } \widetilde{\mathcal{D}} = \begin{pmatrix} \sigma^{\frac{1}{2}}B & 0 \\ 0 & \sigma^{-\frac{1}{2}}\mathcal{I} \end{pmatrix}.$$

According to the fact that

$$\begin{pmatrix} \mathcal{I} & (\gamma-1)\mathcal{I} \\ 0 & \mathcal{I} \end{pmatrix}\widetilde{\mathcal{H}}_0\begin{pmatrix} \mathcal{I} & (\gamma-1)\mathcal{I} \\ 0 & \mathcal{I} \end{pmatrix}^* = \begin{pmatrix} \tau\mathcal{I}+(1-\gamma)\mathcal{I} & 0 \\ 0 & \frac{1}{\gamma}\mathcal{I} \end{pmatrix},$$

and the positive-definition of D_A, we deduce $\mathcal{H}$ is positive definite if and only if $\mathcal{H}_0$ is positive definite. Invoking $\tau > \gamma - 1$ and $\gamma \in (0,2)$, $\tau\mathcal{I}+(1-\gamma)\mathcal{I}$ is positive definite and $\gamma > 0$, then $\widetilde{\mathcal{H}}$ is positive definite or semidefinite, and $\mathcal{H}_0$ is positive definite. This completes the proof. □

Now, we define

$$\mathcal{G} := \mathcal{Q} + \mathcal{Q}^* - \mathcal{M}^*\mathcal{H}\mathcal{M} = \begin{pmatrix} D_A & 0 & 0 \\ 0 & D_B & 0 \\ 0 & 0 & \frac{2-\gamma}{\sigma}\mathcal{I} \end{pmatrix}, \tag{25}$$

by which the properties of the sequences $\{w^k\}$ and $\{\widetilde{w}^k\}$ can be stated.

Lemma 4 *For the sequences $\{w^k\}$ generated by G-LADMM and $\{\widetilde{w}^k\}$ defined in (16), we have*

$$\theta(u) - \theta(\widetilde{u}^k) + \langle w - \widetilde{w}^k, \mathcal{F}(\widetilde{w}^k)\rangle \geq \frac{1}{2}(\|w-w^{k+1}\|_{\mathcal{H}}^2 - \|w-w^k\|_{\mathcal{H}}^2) + \frac{1}{2}\|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2$$

for any $w \in \mathcal{W}$, where $\mathcal{H} = \mathcal{Q}\mathcal{M}^{-1}$ and $\mathcal{G}$ is defined in (25).

Proof From Lemmas 1, 2 and $\mathcal{Q} = \mathcal{H}\mathcal{M}$, for any $w \in \mathcal{W}$ we have

$$\begin{aligned}\theta(u) - \theta(\widetilde{u}^k) + \langle w - \widetilde{w}^k, \mathcal{F}(\widetilde{w}^k)\rangle &\geq \langle w - \widetilde{w}^k, \mathcal{H}\mathcal{M}(w^k - \widetilde{w}^k)\rangle \\ &= \langle w - \widetilde{w}^k, \mathcal{H}(w^k - w^{k+1})\rangle,\end{aligned} \tag{26}$$

where the equality follows (22). Since $\mathcal{H}$ is symmetric, we have

$$\begin{aligned}\langle w - \widetilde{w}^k, \mathcal{H}(w^k - w^{k+1})\rangle = &\frac{1}{2}(\|w-w^{k+1}\|_{\mathcal{H}}^2 - \|w-w^k\|_{\mathcal{H}}^2) \\ &+\frac{1}{2}(\|w^k-\widetilde{w}^k\|_{\mathcal{H}}^2 - \|w^{k+1}-\widetilde{w}^k\|_{\mathcal{H}}^2).\end{aligned} \tag{27}$$

In addition, using the definition of $\mathcal{G}$ in (25), we have

$$\|w^k - \widetilde{w}^k\|_{\mathcal{H}}^2 - \|w^{k+1} - \widetilde{w}^k\|_{\mathcal{H}}^2 = \|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2. \tag{28}$$

Substituting (27) and (28) into (26), we get the result. □

Lemma 5 *For the sequences $\{w^k\}$ generated by G-LADMM and $\{\widetilde{w}^k\}$ defined in (16), we have*

$$\|w^{k+1} - \overline{w}\|_{\mathcal{H}}^2 \leq \|w^k - \overline{w}\|_{\mathcal{H}}^2 - \|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2, \quad \forall \overline{w} \in \overline{\mathcal{W}}.$$

Proof Letting $w = \bar{w} \in \overline{\mathcal{W}}$ in Lemma 4, noticing that $\mathcal{F}$ is skew-symmetric, namely,

$$\langle w - \widetilde{w}, \mathcal{F}(w) - \mathcal{F}(\widetilde{w})\rangle = 0, \forall\, w, \widetilde{w} \in \mathcal{W},$$

then by (10), the result holds. □

However, the operator D_B is not necessarily positive definite when $\tau \in (0, 1)$, so $\mathcal{G}$ is not necessarily positive definite and Lemma 5 does not imply the convergence of the sequence $\{w^k\}$. We thus need further investigate the term $\|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2$ to explore the convergence of G-LADMM.

5.2 Investigating $\|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2$

As mentioned above, the key point for proving the convergence of G-LADMM is to analyze the term $\|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2$, while its desirable properties is related to the range of τ. In this section, we focus on investigating this term and aim at finding a sufficient condition on the parameter τ such that

$$\|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2 \geq \varphi(w^k, w^{k+1}) - \varphi(w^{k-1}, w^k) + \phi(w^k, w^{k+1}),$$

where $\varphi(\cdot, \cdot)$ and $\phi(\cdot, \cdot)$ are both non-negative functions, by which the convergence of G-LADMM can be established.

Lemma 6 *Let $\{w^k\}$ be generated by G-LADMM and $\{\widetilde{w}^k\}$ be defined in (16), then we have*

$$\begin{aligned}\|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2 = \|x^k - x^{k+1}\|_{D_A}^2 + \tau\|y^k - y^{k+1}\|_{D_0}^2 + \tfrac{2-\gamma}{\sigma\gamma^2}\|\lambda^k - \lambda^{k+1}\|^2 \\ + \tfrac{2-\gamma-(1-\tau)\gamma^2}{\gamma^2}\sigma\|B(y^k - y^{k+1})\|^2 + \tfrac{2(2-\gamma)}{\gamma^2}\langle\lambda^k - \lambda^{k+1}, B(y^k - y^{k+1})\rangle.\end{aligned}$$

Proof Invoking the definitions of D_0, $\widetilde{v}$ and $\mathcal{G}$, we have

$$\begin{aligned}\|w^k - \widetilde{w}^k\|_{\mathcal{G}}^2 = \|x^k - x^{k+1}\|_{D_A}^2 + \tau\|y^k - y^{k+1}\|_{D_0}^2 \\ -(1-\tau)\sigma\|B(y^k - y^{k+1})\|^2 + \frac{(2-\gamma)}{\sigma}\|\lambda^k - \widetilde{\lambda}^k\|^2.\end{aligned} \tag{29}$$

On the other hand, it follows from (13) and (16) that,

$$\lambda^k - \widetilde{\lambda}^k = \frac{1}{\gamma}(\lambda^k - \lambda^{k+1}) + \frac{\sigma}{\gamma}B(y^k - y^{k+1}),$$

and thus we get

$$\begin{aligned}\|\lambda^k - \widetilde{\lambda}^k\|^2 = \frac{1}{\gamma^2}\|\lambda^k - \lambda^{k+1}\|^2 + \frac{\sigma^2}{\gamma^2}\|B(y^k - y^{k+1})\|^2 \\ +2\frac{\sigma}{\gamma^2}\langle\lambda^k - \lambda^{k+1}, B(y^k - y^{k+1})\rangle.\end{aligned} \tag{30}$$

Substituting (30) into (29), we obtain the result. □

Lemma 7 *Let $\{w^k\}$ be generated by G-LADMM and $\{\widetilde{w}^k\}$ be defined in (16). Then, we have*

$$\begin{aligned}\langle\lambda^k - \lambda^{k+1}, B(y^k - y^{k+1})\rangle \geq \tfrac{\tau}{2}\|y^k - y^{k+1}\|_{D_0}^2 - \tfrac{\tau}{2}\|y^k - y^{k-1}\|_{D_0}^2 \\ - \tfrac{3(1-\tau)}{2}\sigma\|B(y^k - y^{k+1})\|^2 - \tfrac{(1-\tau)}{2}\sigma\|B(y^{k-1} - y^k)\|^2.\end{aligned}$$

Proof From the optimality condition of the y-subproblems in Step 2, we have

$$g(y) - g(y^{k+1}) + \langle y - y^{k+1},\ \ - B^*\lambda^{k+1} + D_B(y^{k+1} - y^k)\rangle \geq 0. \tag{31}$$

Similarly, for the previous iterate, we get

$$g(y) - g(y^k) + \langle y - y^k,\ \ - B^*\lambda^k + D_B(y^k - y^{k-1})\rangle \geq 0. \tag{32}$$

Setting $y = y^k$ and $y = y^{k+1}$ in (31) and (32) respectively, and then their addition gives

$$\langle y^k - y^{k+1},\ \ - B^*(\lambda^{k+1} - \lambda^k) + D_B[(y^{k+1} - y^k) - (y^k - y^{k-1})]\rangle \geq 0.$$

Using notations D_B and D_0, we have

$$\begin{aligned}
&2\langle \lambda^k - \lambda^{k+1},\ B(y^k - y^{k+1})\rangle \\
\geq\ & 2\tau\|y^k - y^{k+1}\|^2_{D_0} - 2\tau\langle y^k - y^{k+1},\ D_0(y^k - y^{k-1})\rangle \\
&-2(1-\tau)\sigma\|B(y^k - y^{k+1})\|^2 + 2(1-\tau)\sigma\langle y^k - y^{k+1},\ B^*B(y^k - y^{k-1})\rangle \\
\geq\ & \tau\|y^k - y^{k+1}\|^2_{D_0} - \tau\|y^k - y^{k-1}\|^2_{D_0} - 3(1-\tau)\sigma\|B(y^k - y^{k+1})\|^2 \\
&-(1-\tau)\sigma\|B(y^k - y^{k-1})\|^2.
\end{aligned}$$

The last inequality follows the Cauchy-Schwarz inequality. This completes the proof. □

Now, define two functions

$$\varphi(w^k, w^{k+1}) := \frac{2-\gamma}{\gamma^2}[\tau\|y^k - y^{k+1}\|^2_{D_0} + (1-\tau)\sigma\|B(y^k - y^{k+1})\|^2], \tag{33}$$

$$\phi(w^k, w^{k+1}) := \|x^k - x^{k+1}\|^2_{D_A} + \tau\|y^k - y^{k+1}\|^2_{D_0} + \frac{2-\gamma}{\sigma\gamma^2}\|\lambda^k - \lambda^{k+1}\|^2, \tag{34}$$

then they are non-negative for any $\tau \in (0,1)$ and $\gamma \in (0,2)$, by which a lower bound of the term $\|w^k - \widetilde{w}^k\|^2_{\mathcal{G}}$ can be deduced.

Lemma 8 *Let $\{w^k\}$ be generated by G-LADMM and $\{\widetilde{w}^k\}$ be defined in (16), if $\tau \geq \frac{\gamma^2 - 3\gamma + 6}{\gamma^2 - 4\gamma + 8}$ for any $\gamma \in (0,2)$, then*

$$\|w^k - \widetilde{w}^k\|^2_{\mathcal{G}} \geq \varphi(w^k, w^{k+1}) - \varphi(w^{k-1}, w^k) + \phi(w^k, w^{k+1}),$$

where the functions φ and ϕ are defined as in (33) and (34).

Proof By Lemmas 6 and 7 and the definitions of φ and ϕ, we obtain

$$\begin{aligned}
&\|w^k - \widetilde{w}^k\|^2_{\mathcal{G}} \\
\geq\ & \varphi(w^k, w^{k+1}) - \varphi(w^{k-1}, w^k) + \phi(w^k, w^{k+1}) \\
&+ \left(\frac{2-\gamma-(1-\tau)\gamma^2}{\gamma^2} - \frac{(2-\gamma)(1-\tau)}{\gamma^2} - 3(1-\tau)\frac{(2-\gamma)}{\gamma^2}\right)\sigma\|B(y^k - y^{k+1})\|^2 \\
=\ & \varphi(w^k, w^{k+1}) - \varphi(w^{k-1}, w^k) + \phi(w^k, w^{k+1}) \\
&+ \frac{2-\gamma-(1-\tau)(\gamma^2 - 4\gamma + 8)}{\gamma^2}\sigma\|B(y^k - y^{k+1})\|^2.
\end{aligned}$$

Notice that from $\tau \geq \frac{\gamma^2 - 3\gamma + 6}{\gamma^2 - 4\gamma + 8}$ with $\gamma \in (0,2)$, we have $2 - \gamma - (1-\tau)(\gamma^2 - 4\gamma + 8) \geq 0$, which implies the non-negativity of the last term above. This completes the proof. □

Setting

$$H_0 = \left\{(\tau, \gamma) \mid \tau \in \left[\frac{\gamma^2 - 3\gamma + 6}{\gamma^2 - 4\gamma + 8}, 1\right), \gamma \in (0,2)\right\}, \tag{35}$$

it is equivalent to the domain $\left\{(\tau, 1+\alpha) \mid \tau \in [\frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}, 1),\ \alpha \in (-1,1)\right\}$ for ID-SADMM proposed by Gao and Ma [17]. For a special case of G-LADMM with $\alpha = 0$ and $\beta = 1$ (i.e., $\gamma = 1$), the value of τ is not smaller than 0.8. However, we know that the optimal choice for τ is 0.75 for this case from [20]. Therefore, we will improve the range of the proximal parameter τ in the following section.

5.3 Improving the Domain H_0

Fix $\kappa \in [0,1)$ and $\tau \in (\kappa, 1)$, define

$$\delta := \frac{\tau - \kappa}{2(1-\kappa)},$$

then $\delta \in (0, 1/2)$ and $\frac{1}{4(1-\delta)} < \frac{1}{4} + \frac{1}{2}\delta$. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\langle \lambda^k - \lambda^{k+1}, B(y^k - y^{k+1})\rangle \\
\geq &-\frac{1}{4(1-\delta)}\sigma\|B(y^k - y^{k+1})\|^2 - (1-\delta)\frac{1}{\sigma}\|\lambda^k - \lambda^{k+1}\|^2 \\
\geq &-\left(\frac{1}{4} + \frac{1}{2}\delta\right)\sigma\|B(y^k - y^{k+1})\|^2 - (1-\delta)\frac{1}{\sigma}\|\lambda^k - \lambda^{k+1}\|^2 \\
= &\frac{2\kappa - \tau - 1}{4(1-\kappa)}\sigma\|B(y^k - y^{k+1})\|^2 - \frac{2-\tau-\kappa}{2(1-\kappa)}\frac{1}{\sigma}\|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned}$$

Combining it with Lemma 7, we derive

$$\begin{aligned}
&2\langle \lambda^k - \lambda^{k+1},\ B(y^k - y^{k+1})\rangle \\
\geq &\frac{\tau}{2}\|y^k - y^{k+1}\|^2_{D_0} - \frac{\tau}{2}\|y^k - y^{k-1}\|^2_{D_0} + \frac{8\kappa + 5\tau - 7 - 6\kappa\tau}{4(1-\kappa)}\sigma\|B(y^k - y^{k+1})\|^2 \\
&-\frac{(1-\tau)}{2}\sigma\|B(y^k - y^{k-1})\|^2 - \frac{2-\tau-\kappa}{2(1-\kappa)}\frac{1}{\sigma}\|\lambda^k - \lambda^{k+1}\|^2,
\end{aligned} \tag{36}$$

with which we can obtain an improved lower bound of $\|w^k - \widetilde{w}^k\|^2_{\mathcal{G}}$.

Lemma 9 *Fix $\gamma \in (0,2)$, $\kappa \in [0,1)$ and $\tau \in (\kappa, 1)$, define*

$$\psi(\tau, \kappa, \gamma) := 4(1-\kappa)(\tau - 1)\gamma^2 - (6\kappa + 7\tau - 5 - 8\kappa\tau)\gamma + (12\kappa + 14\tau - 10 - 16\kappa\tau), \tag{37}$$

suppose $\psi(\tau, \kappa, \gamma) \geq 0$. Let $\{w^k\}$ be generated by G-LADMM and $\{\widetilde{w}^k\}$ be defined in (16), then we have

$$\|w^k - \widetilde{w}^k\|^2_{\mathcal{G}} \geq \frac{1}{2}\varphi(w^k, w^{k+1}) - \frac{1}{2}\varphi(w^{k-1}, w^k) + \widetilde{\phi}(w^k, w^{k+1}),$$

where the function φ is defined as in (33) and

$$\widetilde{\phi}(w^k, w^{k+1}) := \|x^k - x^{k+1}\|^2_{D_A} + \tau\|y^k - y^{k+1}\|^2_{D_0} + \frac{\delta(2-\gamma)}{\sigma\gamma^2}\|\lambda^k - \lambda^{k+1}\|^2, \tag{38}$$

with $\delta = \frac{\tau-\kappa}{2(1-\kappa)} \in (0, 1/2)$.

Proof By Lemma 6 and (36), we have

$$
\begin{aligned}
&\|w^k-\widetilde{w}^k\|_{\mathcal{G}}^2\\
&\geq \widetilde{\phi}(w^k,w^{k+1})+\frac{1}{2}\varphi(w^k,w^{k+1})-\frac{1}{2}\varphi(w^{k-1},w^k)+\frac{2-\gamma-(1-\tau)\gamma^2}{\gamma^2}\sigma\|B(y^k-y^{k+1})\|^2\\
&\quad+\left(\frac{(2-\gamma)(8\kappa+5\tau-7-6\kappa\tau)}{4\gamma^2(1-\kappa)}-\frac{(2-\gamma)(1-\tau)}{2\gamma^2}\right)\sigma\|B(y^k-y^{k+1})\|^2\\
&=\widetilde{\phi}(w^k,w^{k+1})+\frac{1}{2}(\varphi(w^k,w^{k+1})-\varphi(w^{k-1},w^k))\\
&\quad+\frac{4(1-\kappa)(2-\gamma-\gamma^2+\tau\gamma^2)+(2-\gamma)(10\kappa+7\tau-9-8\kappa\tau)}{4\gamma^2(1-\kappa)}\sigma\|B(y^k-y^{k+1})\|^2.
\end{aligned}
$$

Invoking $\kappa<\tau<1$ and $\gamma>0$, we deduce $4\gamma^2(1-\kappa)>0$, and

$$
\begin{aligned}
&4(1-\kappa)(2-\gamma-\gamma^2+\tau\gamma^2)+(2-\gamma)(10\kappa+7\tau-9-8\kappa\tau)\\
&=[4(1-\kappa)\gamma^2+(8\kappa-7)\gamma+14-16\kappa]\tau-[4(1-\kappa)\gamma^2+(6\kappa-5)\gamma+10-12\kappa]\\
&=\psi(\tau,\kappa,\gamma)\geq 0.
\end{aligned}
$$

This completes the proof. □

Lemma 9 gives a new lower bound of $\|w^k-\widetilde{w}^k\|_{\mathcal{G}}^2$, but it is based on the condition $\psi(\tau,\kappa,\gamma)\geq 0$. In sequel, we will investigate the range of (τ,γ) from the condition $\psi(\tau,\kappa,\gamma)\geq 0$.

For simplicity to discuss, we use the following notations:

$$
\left.
\begin{array}{l}
\theta_1(\kappa,\gamma):=4(1-\kappa)\gamma^2+(8\kappa-7)\gamma+14-16\kappa,\\
\theta_2(\kappa,\gamma):=4(1-\kappa)\gamma^2+(6\kappa-5)\gamma+10-12\kappa,\\
\xi(\kappa,\gamma):=\frac{\theta_2(\kappa,\gamma)}{\theta_1(\kappa,\gamma)},\\
\eta(\kappa):=\frac{7-8\kappa+\sqrt{(7-8\kappa)(24\kappa-25)}}{8-8\kappa}.
\end{array}
\right\}
\tag{39}
$$

By these notations in (39) and the definition of $\psi(\tau,\kappa,\gamma)$ in (37), we have

$$
\psi(\tau,\kappa,\gamma)=\tau\theta_1(\kappa,\gamma)-\theta_2(\kappa,\gamma).
$$

Note that $\theta_1(\kappa,\gamma)$ is a quadratic function with variable γ, since $4(1-\kappa)>0$ and the symmetry axis is $\gamma_0=\frac{7-8\kappa}{8-8\kappa}$, the minimization of $\theta_1(\kappa,\gamma)$ is $\theta_1(\kappa,\gamma_0)=\frac{(8\kappa-7)(24\kappa-25)}{8-8\kappa}$ for any $\kappa\in[0,1)$. Now, we classify different situations of discussing the solution of $\psi(\tau,\kappa,\gamma)\geq 0$:

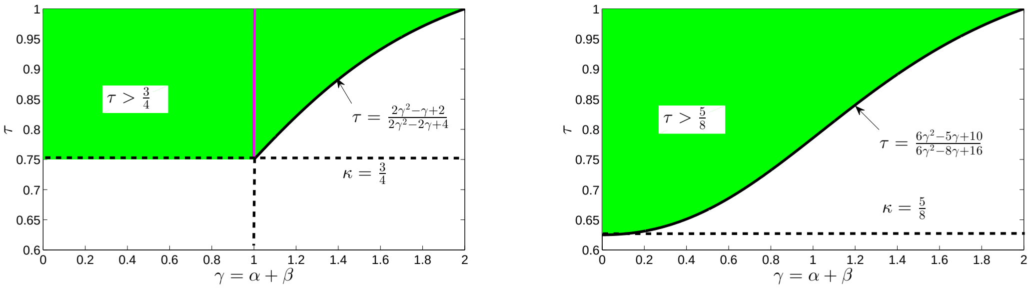
Case 1. If $\kappa\in[0,\frac{7}{8})$, then $\gamma_0>0$ and $\theta_1(\kappa,\gamma)\geq\theta_1(\kappa,\gamma_0)>0$ for any $\gamma\in(0,2)$;
Case 2. If $\kappa=\frac{7}{8}$, then $\theta_1(\kappa,\gamma)=\frac{1}{2}\gamma^2>0$ for any $\gamma\in(0,2)$;
Case 3. If $\kappa\in(\frac{7}{8},1)$ and $2>\gamma>\eta(\kappa)$, we derive $\theta_1(\kappa,\gamma)>0$;
Case 4. If $\kappa\in(\frac{7}{8},1)$ and $0<\gamma<\eta(\kappa)$, we derive $\theta_1(\kappa,\gamma)<0$ but $\xi(\kappa,\gamma)>1$, thus we deduce $\psi(\tau,\kappa,\gamma)<0$ for any $\kappa<\tau<1$;
Case 5. If $\kappa\in(\frac{7}{8},1)$ and $\gamma=\eta(\kappa)$, then we have $\theta_1(\kappa,\gamma)=0$ and $\theta_2(\kappa,\gamma)>0$, which implies $\psi(\tau,\kappa,\gamma)<0$.

Taking five cases above with $\theta_1(\kappa,\gamma)>0$, $\theta_1(\kappa,\gamma)=0$ or $\theta_1(\kappa,\gamma)<0$, a sufficient condition for ensuring $\psi(\tau,\kappa,\gamma)\geq 0$ can be observed from the first three cases, and expressed as:

$$
\tau\geq\xi(\kappa,\gamma)\text{ and }\tau>\kappa,\quad\text{if }\kappa\in[0,\frac{7}{8}],\gamma\in(0,2)\text{ or }\kappa\in(\frac{7}{8},1),\gamma\in(\eta(\kappa),2).
\tag{40}
$$

From (40), the relation of τ and κ can be observed

$$
\begin{cases}
\tau\geq\frac{10-12\kappa}{14-16\kappa}, & \kappa\in[0,\frac{5}{8}),\\
\tau>\kappa, & \kappa\in[\frac{5}{8},1).
\end{cases}
$$



(a) $\kappa = \frac{3}{4}$ by He et al. with $\alpha = 0$ and $\beta = 1$. (b) $\kappa = \frac{5}{8}$.

Fig. 2 The comparison of domains H by OLADMM with $\kappa = \frac{3}{4}$ and G-LADMM with $\kappa = \frac{5}{8}$.

Thus, we obtain a new lower bound $\frac{5}{8}$ of τ when $\kappa = \frac{5}{8}$. In this case, we have $\tau \geq \frac{6\gamma^2-5\gamma+10}{6\gamma^2-8\gamma+16}$ from $\tau \geq \xi(\kappa, \gamma)$, see Figure 2(b).

Actually, we can obtain many different regions for each $\kappa \in [0, 1)$, for instance, the region $(\tau \geq \frac{2\gamma^2-\gamma+2}{2\gamma^2-2\gamma+4})$ with $\kappa = \frac{3}{4}$ shown in Figure 2(a), which recovers the optimal condition $\tau \in (0.75, 1)$ for the OLADMM ($\gamma = 1$) proposed by He et al [20].

Upon removing the auxiliary variable κ, a sufficient condition relating τ and γ can be obtained to establish $\psi(\tau, \kappa, \gamma) \geq 0$, labeled as H_1:

$$\begin{aligned} H_1 &= \bigcup_{\kappa \in [0,1)} \{(\tau, \gamma) \mid \psi(\tau, \kappa, \gamma) \geq 0, \tau \in (\kappa, 1), \gamma \in (0, 2)\} \\ &= \left\{ (\tau, \gamma) \;\middle|\; \tau \in \left(\frac{4\gamma^2 - 5\gamma + 10}{4\gamma^2 - 8\gamma + 16}, \;\; 1 \right), \gamma \in (0, 2) \right\}. \end{aligned}$$

Figure 3 shows the domain H_1 and its sub-domain with some fixed κ. Considering the condition $\gamma - 1 < \tau < 1$ in Lemma 3, we can obtain an improved domain H defined as

$$\begin{aligned} H &= H_0 \cup H_1 \\ &= \left\{ (\tau, \gamma) \;\middle|\; \begin{array}{ll} 1 > \tau > \frac{4\gamma^2-5\gamma+10}{4\gamma^2-8\gamma+16}, & \gamma \in (0, 2(\sqrt{3}-1)] \\ 1 > \tau \geq \frac{\gamma^2-3\gamma+6}{\gamma^2-4\gamma+8}, & \gamma \in (2(\sqrt{3}-1), 2) \end{array} \right\} \end{aligned} \tag{41}$$

to guarantee the convergence of G-LADMM, where H_0 is defined in (35).

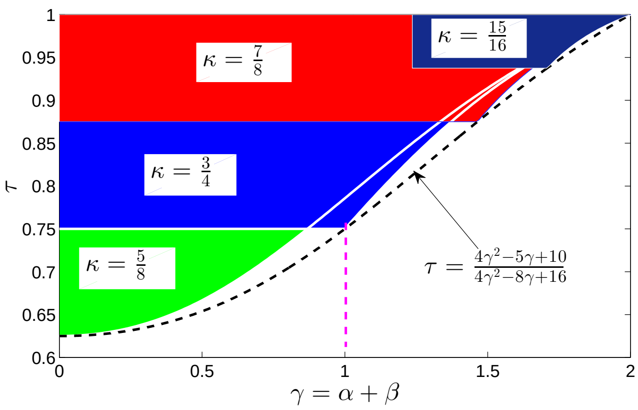


Fig. 3 The domain H_1 and its sub-domain with different fixed κ.

5.4 Convergence

In this section, we prove the convergence of G-LADMM with $(\tau, \gamma) \in H_1$, since the convergence with $(\tau, \gamma) \in H_0$ can be found in [17]. Based on Lemmas 5 and 9, we now establish the following result:

Lemma 10 *Let the sequence $\{w^k\}$ be generated by G-LADMM. Then if $(\tau, \gamma) \in H_1$, we have*

$$\|w^{k+1} - \bar{w}\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^k, w^{k+1}) \le \|w^k - \bar{w}\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^{k-1}, w^k) - \widetilde{\phi}(w^k, w^{k+1}), \quad \forall \bar{w} \in \overline{\mathcal{W}},$$

where φ and $\widetilde{\phi}$ are defined as in (33) and (38).

Since D_A and D_0 are positive definite, $\widetilde{\phi}(w^k, w^{k+1}) \ge 0$ for any $(\tau, \gamma) \in H_1$. By Lemma 10, the iterative contraction of the sequence $\{\|w^{k+1} - \bar{w}\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^k, w^{k+1})\}$ can be obtained for any $(\tau, \gamma) \in H_1$.

Theorem 1 *Suppose the sequence $\{w^k\}$ is generated by G-LADMM with $(\tau, \gamma) \in H_1$, then it converges to a point w^∞ in $\overline{\mathcal{W}}$.*

Proof First, it follows from Lemma 10 and (38) that

$$\begin{aligned}\widetilde{\phi}(w^k, w^{k+1}) &= \|x^k - x^{k+1}\|_{D_A}^2 + \tau\|y^k - y^{k+1}\|_{D_0}^2 + \frac{\delta(2-\gamma)}{\sigma\gamma^2}\|\lambda^k - \lambda^{k+1}\|^2 \\ &\le \|w^k - \bar{w}\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^{k-1}, w^k) - \|w^{k+1} - \bar{w}\|_{\mathcal{H}}^2 - \frac{1}{2}\varphi(w^k, w^{k+1}).\end{aligned}$$

Summing from $k = 1$ to ∞ gives

$$\begin{aligned}&\sum_{k=1}^{\infty}(\|x^k - x^{k+1}\|_{D_A}^2 + \tau\|y^k - y^{k+1}\|_{D_0}^2 + \tfrac{\delta(2-\gamma)}{\sigma\gamma^2}\|\lambda^k - \lambda^{k+1}\|^2) \\ &\le \|w^1 - \bar{w}\|_{\mathcal{H}}^2 + \tfrac{1}{2}\varphi(w^0, w^1).\end{aligned}$$

Recall D_A and D_0 are positive definite, and from $\frac{\delta(2-\gamma)}{\sigma\gamma^2} > 0$, we thus deduce

$$\lim_{k\to\infty}\|w^k - w^{k+1}\| = 0. \tag{42}$$

In addition, letting $\bar{w} \in \overline{\mathcal{W}}$ be an arbitrarily fixed point, then

$$\|w^{k+1} - \bar{w}\|_{\mathcal{H}}^2 \le \|w^k - \bar{w}\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^{k-1}, w^k) \le \|w^1 - \bar{w}\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^0, w^1).$$

Thus, the sequence $\{w^k\}$ must be bounded and has at least one cluster point $w^\infty = (x^\infty, y^\infty, \lambda^\infty)$. Namely, there exists a subsequence $\{k_j\}_{j=1}^{\infty}$ such that $\lim_{j\to\infty} w^{k_j} = w^\infty$.

Note that the operator $\mathcal{M}$ defined in (23) is nonsingular and $\widetilde{w}^k = w^k + \mathcal{M}^{-1}(w^{k+1} - w^k)$, then the sequence $\{\widetilde{w}^k\}$ is also bounded and it follows from (42) that

$$\lim_{j\to\infty}\widetilde{w}^{k_j} = \lim_{j\to\infty}(w^{k_j} + \mathcal{M}^{-1}(w^{k_j+1} - w^{k_j})) = w^\infty.$$

Thus for any $w \in \mathcal{W}$, by Lemma 1 we have

$$\theta(u) - \theta(\widetilde{u}^{k_j}) + \langle w - \widetilde{w}^{k_j}, \mathcal{F}(\widetilde{w}^{k_j})\rangle \ge \langle w - \widetilde{w}^{k_j}, \mathcal{Q}(w^{k_j} - \widetilde{w}^{k_j})\rangle, \quad \widetilde{w}^{k_j} \in \mathcal{W}.$$

Note that the continuity of $\theta(u)$ and $\mathcal{F}(w)$, it holds that

$$w^\infty \in \mathcal{W}, \quad \theta(u) - \theta(u^\infty) + \langle w - w^\infty, \mathcal{F}(w^\infty)\rangle \ge 0,$$

where $u^\infty = \begin{pmatrix} x^\infty \\ y^\infty \end{pmatrix}$. The above variational inequality indicates that w^∞ is a solution point of the variational inequality (10), i.e. $w^\infty \in \overline{\mathcal{W}}$. By Lemma 10, for all $l > k_j$ we have

$$\|w^l - w^\infty\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^{l-1}, w^l) \le \|w^{k_j} - w^\infty\|_{\mathcal{H}}^2 + \frac{1}{2}\varphi(w^{k_j-1}, w^{k_j})).$$

This together with (42) and the positive definiteness of $\mathcal{H}$ illustrate $\lim_{l\to\infty} w^l = w^\infty$. Therefore, the whole sequence $\{w^k\}$ converges to the solution $w^\infty \in \overline{\mathcal{W}}$. This completes the proof. □

Remark 2 *If the linear operator A has full column rank, and the operator $(\partial f + \sigma A^*A)^{-1}$ is easy to deal with, the linearization in the x-subproblem will be unnecessary, G-LADMM can be written as*

$$\left.\begin{array}{l} x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \left\{ f(x) + \frac{\sigma}{2}\left\| Ax + By^k - c - \frac{\lambda^k}{\sigma}\right\|^2 \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\sigma\left(Ax^{k+1} + By^k - c\right), \\ y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \left\{ g(y) + \frac{\tau r_B}{2}\left\| y - y^k - \frac{B^*(\lambda^{k+\frac{1}{2}})}{\tau r_B}\right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \sigma\left(By^{k+1} - By^k\right), \end{array}\right\} \tag{43}$$

for any $\alpha \in (0,2)$. In this case, we have $D_A = 0$ but the convergence of G-LADMM can be guaranteed, by using the fact that the linear operator A has full column rank,

$$Ax^{k+1} = A\widetilde{x}^k = \frac{\lambda^k - \widetilde{\lambda}^k}{\sigma} - (By^k - c),$$

and the similar process as in the proof of Theorem 1.

Using Theorem 1 and the similar idea as in [17, Section 6], we can also obtain that G-LADMM has the worst-case $\mathcal{O}(\frac{1}{t})$ nonasymptotic convergence rate, which is omitted here for conciseness.

6 Numerical Experiments.

We present numerical results to demonstrate the computational performance of G-LADMM, and compare it numerically with some other efficient methods for solving LASSO model and image decomposition. We denote the random number generator by *seed* for generating data again in MATLAB R2013b. All experiments are performed on an Intel(R) Core(TM) i5-4590 CPU@ 3.30 GHz PC with 8GB of RAM running on 64-bit Windows operating system.

6.1 LASSO Model.

In statistics and machine learning, least absolute shrinkage and selection operator (LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. The so-called Lagrangian form of the LASSO model [31,32] can be written as:

$$\min_{y\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Ay - b\|^2 + \eta\|y\|_1 \right\} \tag{44}$$

where $\|y\|_1 := \sum_{i=1}^n |y_i|$, $A \in \mathbb{R}^{m\times n}$ is a design matrix usually with $m \ll n$, m is the number of data points, n is the number of features, $b \in \mathbb{R}^m$ is the response vector and $\eta > 0$ is a regularization parameter. The LASSO model provides a sparse estimation of y when there are more features than data points. It can also be explained as a model for finding a sparse solution of the under-determined system of linear equations $Ay = b$.

As in [17], by introducing an auxiliary variable x problem (44) can be reformulated as a 2-block separable convex problem:

$$\min \left\{ \frac{1}{2}\|x-b\|^2 + \eta\|y\|_1 \;\middle|\; x - Ay = 0,\; x \in \mathbb{R}^m,\; y \in \mathbb{R}^n \right\}, \tag{45}$$

then we can apply the proposed G-LADMM to it. Since the x-subproblem has a closed-form solution for given the quadratic term, we do not implement the linearization. By (43), for any $\alpha \in (0,2)$ the iterative scheme is

$$\left.\begin{array}{l} x^{k+1} = \arg\min\limits_{x\in\mathbb{R}^m} \left\{ \frac{1}{2}\|x-b\|^2 + \frac{\sigma}{2}\|x - Ay^k - \frac{\lambda^k}{\sigma}\|^2 \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\sigma\left(x^{k+1} - Ay^k\right) \\ y^{k+1} = \arg\min\limits_{y\in\mathbb{R}^n} \left\{ \eta\|y\|_1 + \frac{\tau r_A}{2}\left\| y - y^k - \frac{A^T(\lambda^{k+\frac{1}{2}})}{\tau r_A} \right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} + \sigma A\left(y^{k+1} - y^k\right), \end{array}\right\}$$

where $r_A > \sigma\|A^T A\|$.

We generate the data as in [17, section 6.1], the matlab code is :

```
p=100/n;
x0 = sprandn(n,1,p);
A = randn(m,n);
A = A*spdiags(1./sqrt(sum(A.^2))',0,n,n);
b = A*x0 + sqrt(0.001)*randn(m,1);
```

We report the numerical results obtained by G-LADMM, and compare the performance with the linearized SADMM with positive-definite proximal regularization (PD-SADMM) and ID-SADMM [17].

We test five cases of the dimension of A ranging from 900×3000 to 1500×5000 with $\eta = 0.1$, and set $r_A = \zeta\|A^T A\|$ with $\zeta = 1.001$. Note that from the definition of $\mathcal{H}$ in (24), G-LADMM with $\beta = 1$ is equivalent to ID-SADMM when $\gamma > 2(\sqrt{3}-1) \approx 1.4641$, the stepsize $\gamma = \alpha$ is thus set to 0.5, 0.7, 0.9 and 1.1 in this section. Since small proximal term can allow for larger steps and $\gamma < 2(\sqrt{3}-1)$, here we always let $\tau = \zeta\frac{4\gamma^2-5\gamma+10}{4\gamma^2-8\gamma+16}$. The termination tolerance is defined as in [17,31] by

$$\|x^{k+1} - Ay^{k+1}\|^2 \le \epsilon^{pri} \quad \text{and} \quad \|\sigma A(y^{k+1} - y^k)\|^2 \le \epsilon^{dual},$$

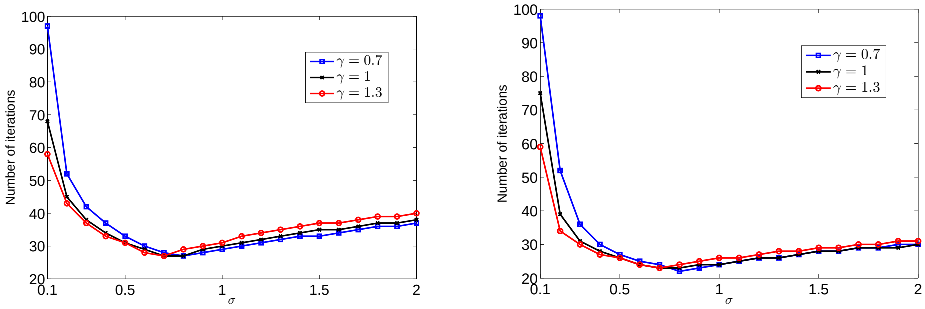
where

$$\epsilon^{pri} = \sqrt{n}\epsilon^{abs} + \epsilon^{rel}\max\{\|x^{k+1}\|^2, \|Ay^{k+1}\|^2\} \quad \text{and} \quad \epsilon^{dual} = \sqrt{n}\epsilon^{abs} + \epsilon^{rel}\|y^{k+1}\|^2,$$

with $\epsilon^{abs} = 10^{-4}$ and $\epsilon^{rel} = 10^{-2}$ for all the methods. The initial points (y^0, λ^0) are chosen to be zero, and the maximal number of iterations is set as 1000.

First, we would like to investigate the performance of G-LADMM for solving problem (45) with variance of the penalty parameter σ. The results shown in Figure 4 imply that the iterations with different γ have a similar changing pattern: decreases originally and then increases along with the decrease of the value of σ. Furthermore, the G-LADMM with $\sigma = 0.8$ performs better than that with other cases, and PD-ADMM and ID-ADMM have the same performance. Hence, in the following experiments on problem (45), we set $\sigma = 0.8$.

The detailed numerical results are reported in Table 1 and Figure 5, which compares the number of iterations and runtime in seconds for different dimensions of A. From the results in the table, we see that G-LADMM performs better than PD-SADMM and ID-SADMM. From the ratio both in the number of iterations and runtime, G-LADMM can achieve an improvement of about 20%-30% reduction over the PD-SADMM. Comparing with ID-SADMM, the improvement of G-LADMM is becoming more and more significant with the decrease of the value of γ (or α).



(a) 900×3000. (b) 1350×4500.

Fig. 4 Number of iterations with respect to variant σ in G-LADMM with $\sigma \in [0.1, 2]$.

Table 1 Comparison between the number of iterations (time in seconds) taken by G-LADMM, PD-SADMM and ID-SADMM for the Lagrangian form of the LASSO model ($seed = 1$).

Dim. of A	G-LADMM	PD-SADMM	Ratio(%)	ID-SADMM	Ratio(%)
$m \times n$	$\gamma = 0.5$	$\alpha = -0.5$		$\alpha = -0.5$	
900×3000	25(1.50)	33(1.98)	0.76(0.76)	27(1.63)	0.93(0.92)
1050×3500	25(2.07)	35(2.87)	0.71(0.72)	27(2.27)	0.93(0.91)
1200×4000	23(2.45)	33(3.52)	0.70(0.70)	25(2.67)	0.92(0.92)
1350×4500	24(3.25)	36(4.87)	0.67(0.67)	27(3.68)	0.89(0.88)
1500×5000	22(3.79)	31(5.26)	0.71(0.72)	24(4.14)	0.92(0.92)
$m \times n$	$\gamma = 0.7$	$\alpha = -0.3$		$\alpha = -0.3$	
900×3000	24(1.50)	34(2.09)	0.71(0.72)	26(1.60)	0.92(0.94)
1050×3500	23(1.95)	32(2.72)	0.72(0.72)	25(2.12)	0.92(0.92)
1200×4000	24(2.58)	34(3.69)	0.71(0.70)	26(2.81)	0.92(0.92)
1350×4500	25(3.46)	37(5.08)	0.68(0.68)	27(3.71)	0.93(0.93)
1500×5000	22(3.75)	32(5.44)	0.69(0.69)	24(4.09)	0.92(0.92)
$m \times n$	$\gamma = 0.9$	$\alpha = -0.1$		$\alpha = -0.1$	
900×3000	29(1.81)	36(2.25)	0.81(0.81)	30(1.87)	0.97(0.97)
1050×3500	25(2.13)	33(2.82)	0.76(0.76)	26(2.22)	0.96(0.96)
1200×4000	18(1.97)	25(2.75)	0.72(0.72)	19(2.08)	0.95(0.95)
1350×4500	21(2.91)	29(4.01)	0.72(0.73)	22(3.04)	0.95(0.96)
1500×5000	23(3.95)	33(5.65)	0.70(0.70)	24(4.10)	0.96(0.96)
$m \times n$	$\gamma = 1.1$	$\alpha = 0.1$		$\alpha = 0.1$	
900×3000	27(1.68)	35(2.19)	0.77(0.77)	28(1.75)	0.96(0.96)
1050×3500	26(2.22)	35(2.99)	0.74(0.74)	27(2.31)	0.96(0.96)
1200×4000	26(2.86)	34(3.74)	0.76(0.77)	27(2.96)	0.96(0.97)
1350×4500	25(3.45)	32(4.41)	0.78(0.78)	25(3.45)	1.00(1.00)
1500×5000	23(3.94)	30(5.14)	0.77(0.77)	23(3.95)	1.00(1.00)

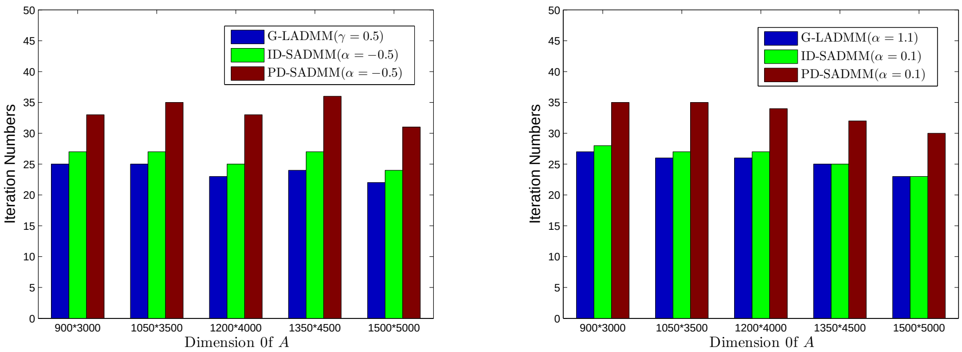
To further demonstrate the efficacy of G-LADMM, we applied G-LADMM to solving the other reformulation of problem (44):

$$\min \left\{ \frac{1}{2}\|Ax - b\|^2 + \eta\|y\|_1 \;\middle|\; x - y = 0,\ x \in \mathbb{R}^n,\ y \in \mathbb{R}^n \right\}, \tag{46}$$

and compare with the classic ADMM (cADMM) (2) with the aggressive step size of $\varrho = 1.618$ and the under-relaxation ADMM [1] (urADMM) introduced by Boyd et al. in [31] with the over-relaxation parameter being $\alpha = 1.6$. For all solvers, we set $\sigma = 0.8$ and use the same stopping criteria as in [31, Section 3.3.1], with $\epsilon^{abs} = 10^{-6}$ and $\epsilon^{rel} = 10^{-4}$.

From the performance of G-LADMM for solving reformulation (46) with variance of the stepsize parameter γ, shown in Figure 6, the iteration number decreases firstly and then increases along

[1] The MATLAB codes can be downloaded from: http://stanford.edu/~boyd/papers/admm/lasso/lasso.html



(a) $\gamma = 0.5$ (b) $\gamma = 1.1$

Fig. 5 Comparison results on $\gamma = 0.5$ and $\gamma = 1.1$ respectively for LASSO model.

with the increasing of the value of γ. The main reason for the change is that the value of the proximal parameter τ will increase when a larger stepsize is used, but generally smaller τ and larger γ can bring better numerical performance. The numerical performance of G-LADMM is sensitive to the stepsize parameter γ, while G-LADMM with $\gamma = 1.4$ can yield lower iterations. We may observe from Table 2 that G-LADMM and cADMM are clearly more efficient than urADMM, while cADMM is slightly faster than G-LADMM.

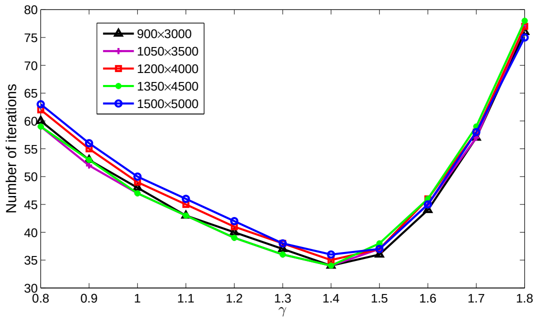


Fig. 6 Evolutions of iteration (iter) with respect to variant γ in G-LADMM for solving reformulation (46).

Table 2 Comparison between the number of iterations (time in seconds) taken by G-LADMM, nrADMM and cADMM for problem (46) ($seed = 1$).

Dim. of A $m \times n$	G-LADMM $\gamma = 1.4$	nrADMM $\alpha = 1.6$	Ratio(%)	cADMM $\varrho = 1.618$	Ratio(%)
900×3000	34(0.34)	51(0.47)	0.67(0.73)	31(0.29)	1.10(1.17)
1050×3500	34(0.57)	50(0.72)	0.68(0.79)	31(0.57)	1.10(0.99)
1200×4000	35(0.91)	51(1.07)	0.69(0.85)	32(0.74)	1.09(1.24)
1350×4500	34(1.01)	48(1.43)	0.71(0.71)	30(0.91)	1.13(1.11)
1500×5000	36(1.18)	53(1.53)	0.68(0.77)	32(1.11)	1.13(1.06)

6.2 Image Decomposition Model.

The low patch-rank decomposition is to decompose a natural image into two meaningful components. One is the structural, geometrical part and sketchy approximation of image coined as cartoon, and the other one is the oscillations and repeated patterns of image coined as texture. This problem can be mathematically modeled as:

$$\min \quad \tau_1 \||\nabla u|\|_1 + \tau_2 \|Pv\|_* + \frac{\tau_3}{2}\|K(u+v)-m\|_F^2 \tag{47}$$

where $\||\nabla \cdot |\|_1$ denotes the total variation semi-norm in order to induce the cartoon part u, $\|\cdot\|_*$ denotes the nuclear norm (defined as the sum of all singular values) to reflect the low-rank feature of the matrix Pv where v is the texture part. $P : \mathbb{R}^n \to \mathbb{R}^{r^2 \times s}$ is a patch mapping to describes the preceding process on rearranging the texture part of an image v as a matrix V, where $s = \lceil \frac{n}{r^2} \rceil$ and $\lceil \cdot \rceil$ is a operator to round a scalar as the nearest integer towards infinity. K is a linear operator corresponding to certain corruption on the target image m. For more details about P and K, we refer to [33–35].

We reformulate (47) as a 2-block separable problem:

$$\begin{aligned} \min \quad & f(x_1) + g(x_2) \\ s.t. \quad & Ax_1 + Bx_2 = 0 \end{aligned} \tag{48}$$

where

$$\begin{cases} x_1 = (u,v) \\ x_2 = (x,y,z) \end{cases}, \begin{cases} f(x_1) = 0 \\ g(x_2) = \tau_1\||\nabla x|\|_1 + \tau_2\|Py\|_* + \frac{\tau_3}{2}\|K(z)-m\|_F^2. \end{cases}$$

and

$$A = \begin{pmatrix} \nabla & 0 \\ 0 & P \\ I & I \end{pmatrix}, B = \begin{pmatrix} -I & 0 & 0 \\ 0 & -I & 0 \\ 0 & 0 & -I \end{pmatrix}.$$

Since the linear operator A and B have full column rank and computing $(\partial f + \sigma A^* A)^{-1}$ is expensive, our G-LADMM with iterative scheme (15) will be applied to problem (48).

6.2.1 Subproblems and Implement Issues.

Note from (15) and $f(x_1) = 0$ that, for $\alpha \in (0,2)$, the corresponding iterative scheme is

$$\begin{cases} \lambda^{k+\frac{0}{2}} = \lambda^k - \sigma\left(Ax_1^k + Bx_2^k\right), \\ x_1^{k+1} = x_1^k + \frac{1}{r_A} A^*(\lambda^{k+\frac{0}{2}}), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\sigma\left(Ax_1^{k+1} + Bx_2^k\right), \\ x_2^{k+1} = \arg\min_{x_2\in\mathcal{Y}} \left\{ g(x_2) + \frac{\tau r_B}{2}\left\| x_2 - x_2^k - \frac{B^*(\lambda^{k+\frac{1}{2}})}{\tau r_B}\right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \sigma\left(Bx_2^{k+1} - Bx_2^k\right). \end{cases} \tag{49}$$

To get better numerical results, we use three penalty parameter $\sigma_i (i = 1,2,3)$ for three equality constraints $\nabla u - x = 0$, $Pv - y = 0$ and $u + v - z = 0$, then we will use $r_{Ai} = \zeta\sigma_i\|A^*A\|$ and $r_{Bi} = \zeta\sigma_i\|B^*B\|$ for any $i = 1,2,3$ and $\zeta > 1$.

According to the first-order optimality condition of the $x_1 = (u,v)$-subproblem, we derive

$$u^{k+1} = u^k + \frac{1}{r_{A1}}\nabla^T(\lambda_1^k - \sigma_1(\nabla u^k - x^k)) + \frac{1}{r_{A3}}(\lambda_3^k - \sigma_3(u^k + v^k - z^k)),$$

$$v^{k+1} = v^k + \frac{1}{r_{A1}}P^T(\lambda_2^k - \sigma_2(Pv^k - y^k)) + \frac{1}{r_{A3}}(\lambda_3^k - \sigma_3(u^k + v^k - z^k)).$$

The $x_2 = (x, y, z)$-subproblem corresponds to the following optimization problem and each variable can be derived separately.

(i). The x-subproblem can be written as

$$\begin{aligned} x^{k+1} &= \arg\min \tau_1 \| |x| \|_1 + \frac{\tau r_{B1}}{2} \left\| x - x^k + \frac{\lambda_1^{k+\frac{1}{2}}}{\tau r_{B1}} \right\|^2 \\ &= \text{Shrink}\left(x^k - \frac{\lambda_1^{k+\frac{1}{2}}}{\tau r_{B1}}, \frac{\tau_1}{\tau r_{B1}} \right), \end{aligned}$$

where $\text{Shrink}(\cdot, \cdot)$ is the soft thresholding operator [38].

(ii). The y-subproblem is equivalent to

$$\begin{aligned} y^{k+1} &= \arg\min \tau_2 \| |y| \|_* + \frac{\tau r_{B2}}{2} \left\| y - y^k + \frac{\lambda_2^{k+\frac{1}{2}}}{\tau r_{B2}} \right\|^2 \\ &= \text{SVD}\left(y^k - \frac{\lambda_2^{k+\frac{1}{2}}}{\tau r_{B2}}, \frac{\tau_2}{\tau r_{B2}} \right). \end{aligned}$$

Here, for any $c > 0$ and matrix $T \in \mathbb{R}^{m\times n}$, the operator $\text{SVD} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is defined by $\text{SVD}(T, c) := U \max\{\Xi - c, 0\} V^\mathsf{T}$, where $U \Xi V^\mathsf{T}$ idenotes the singular value decomposition of matrix T.

(iii). The z-subproblem amounts to

$$\begin{aligned} z^{k+1} &= \arg\min \frac{\tau_3}{2} \|Kz - m\|^2 + \frac{\tau r_{B3}}{2} \left\| z - z^k + \frac{\lambda_3^{k+\frac{1}{2}}}{\tau r_{B3}} \right\|^2 \\ &= \left[\tau_3 K^\mathsf{T} K + \tau r_{B3} I \right]^{-1} (\tau_3 K^\mathsf{T} m + \tau r_{B3} z^k - \lambda_3^{k+\frac{1}{2}}). \end{aligned}$$

In this section, we test two different cases of (48). One is $K = I$, which is to decompose clean images without noise or blur. The other is $K = S$ in which some pixels of m are missing. We report the numerical results obtained by G-LADMM, and compare its numerical performance with PC-ADMM [36] and ADM-G [9] for solving the 3-block reformulation (see problem (50) in [36]) of (47), and the alternating proximal gradient method (APGM) in [37] for solving (48). All the tested algorithms take zero as the initial point. We set $r = 11$ for the patch mapping P, run all algorithms for 150 iterations, and plot the evolutions of objective function values with respect to iterations and computing time in seconds.

Two synthetic images and a real image listed in Figure 7 will be tested. For the synthetic images in Figure 7 (a) and (b), cartoon and texture parts are superposed with a ratio of 7:3. We adopt the periodic boundary condition for the images to be tested and thus the fast Fourier transform (FFT) will be implemented.

For synthetic images, the ground-truth of cartoons and textures are known. We thus use the following relative error to measure the quality of the cartoons and textures decomposed by the tested algorithms

$$\text{RelError}(\mu) := \frac{\|\mu - \mu^*\|}{\|\mu^*\|} \quad \text{and} \quad \text{RelError}(\nu) := \frac{\|\nu - \nu^*\|}{\|\nu^*\|},$$

where μ and ν are the decomposed cartoons and textures, respectively, and μ^* and ν^* are the corresponding ground-truths of synthetic images.

The signal-to-noise ratio (SNR) value is commonly used to measure the quality of the restored or reconstructed images in unit of dB, which is defined as

$$\text{SNR} := 20\log_{10} \frac{\|m^*\|}{\|\bar{m} - m^*\|},$$

where $\bar{m}$ is the approximation of the ground truth m^*.

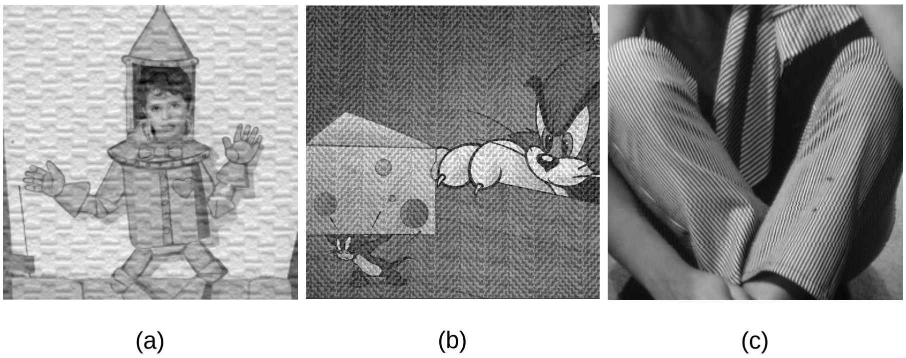


Fig. 7 Synthetic images: (a) 256 × 256 "Boy" and (b) 512 × 512 "Tomjerry". Real images: (c) 256 × 256 "Barbara".

6.2.2 The Case of K=I

For the case of $K = I$, the parameters are fixed as $(\tau_1, \tau_2, \tau_3) = (0.01, 0.005, 1)$, and $(\sigma_1, \sigma_2, \sigma_3) = (10, 10, 10)$ for PC-ADMM with $\alpha = 0.5$ and ADM-G with $\alpha = 0.9$. Since G-LADMM and APGM deal with 2-block problem (48), we set $(\sigma_1, \sigma_2, \sigma_3) = (1, 1, 1)$.

For synthetic images, we firstly investigate the performance of G-LADMM for solving (48) with variance of the stepsize α. Numerical results are reported in Figure 8, where "RelError-Mean" denotes the difference between RelError and their mean value. To balance the relative error of u and v, we set $\alpha = 1.35$ in G-LADMM (with $\gamma = 1.35$) for synthetic images.

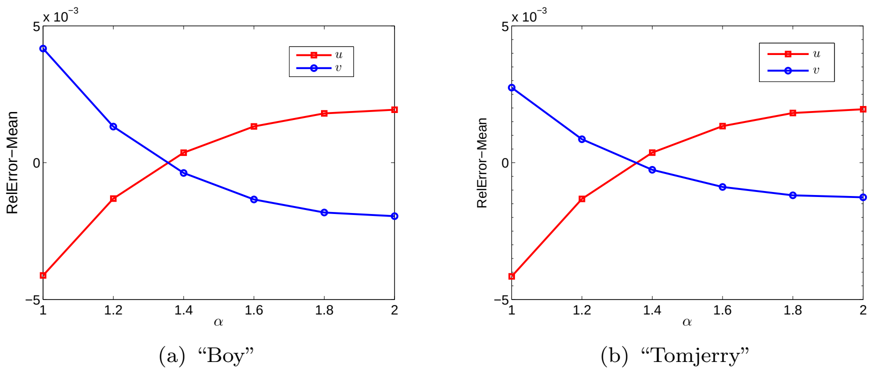


Fig. 8 Performance of G-LADMM with different values of the stepsize α.

For synthetic image "Tomjerry", the evolutions of objective function values, "RelError(u)" and "RelError(v)" with respect to iterations or computing time in seconds are plotted in Figure 9. For real image "Barbara", the evolutions of objective function values with respect to iterations and computing time in seconds are plotted in Figure 10. From the results shown in Figures, the performance of G-LADMM is comparable with APGM, and superior to PC-ADMM and ADM-G. The value of RelError(u) by ADM-G is less than G-LADMM, but its RelError(v) is greater than 0.7. We display the decomposed cartoons and textures obtained by G-LADMM and PC-ADMM in Figure 11.

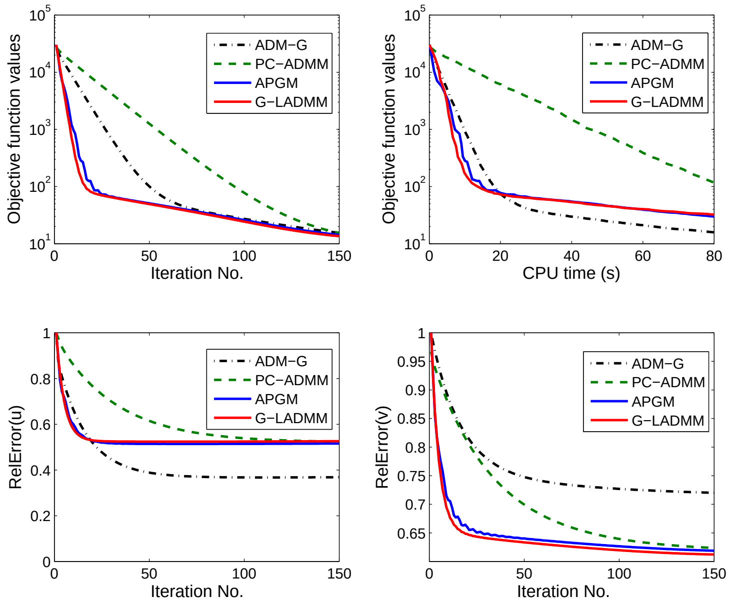


Fig. 9 Evolutions of objective function values, Error(u) and Error(v) w.r.t. iterations and computing time for "Tomjerry".

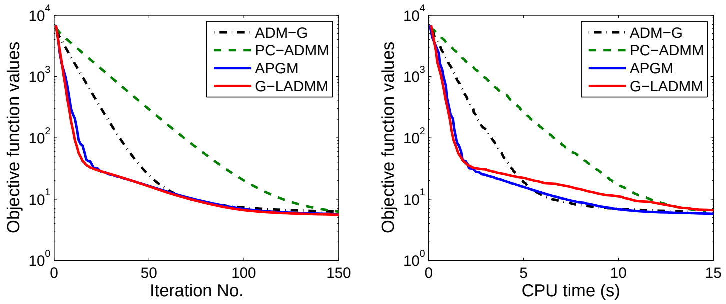


Fig. 10 Evolutions of objective function values w.r.t. iterations and computing time for "Barbara".

6.2.3 The Case of K=S

For the case of $K = S$, the parameters are fixed as $(\tau_1, \tau_2, \tau_3) = (0.02, 0.01, 1)$, and $(\sigma_1, \sigma_2, \sigma_3) = (10, 10, 0.1)$ for PC-ADMM with $\alpha = 0.5$ and ADM-G with $\alpha = 0.9$, while for G-LADMM and APGM, we set $(\tau_1, \tau_2, \tau_3) = (0.003, 0.05, 1)$, and $(\sigma_1, \sigma_2, \sigma_3) = (0.008, 0.008, 0.00008)$. The missing pixels in images are tested as in [33, 36, 39]. For the corrupted images listed in Figure 14, the SNR values are 10.39dB for "Boy", 9.06dB for "Tomjerry" and 10.88dB for "Barbara". For the tested problems with $K = S$, the G-LADMM (using iterative scheme (49)) with a larger α would yield better SNR values empirically, we thus set $\alpha = 1.9$ in this subsection.

We show the evolutions of objective function values and SNR values with respect to iterations and computing time in seconds in Figures 12 and 13. We illustrate the decomposed results of cartoons and textures obtained by G-LADMM in Figure 14 for the tested images. Visually, the corrupted images are separated as cartoons and textures in the rough. In addition, we superposed the decomposed cartoons and textures correspondingly, and then derived the reconstructed images, which listed in the right column of Figure 14.

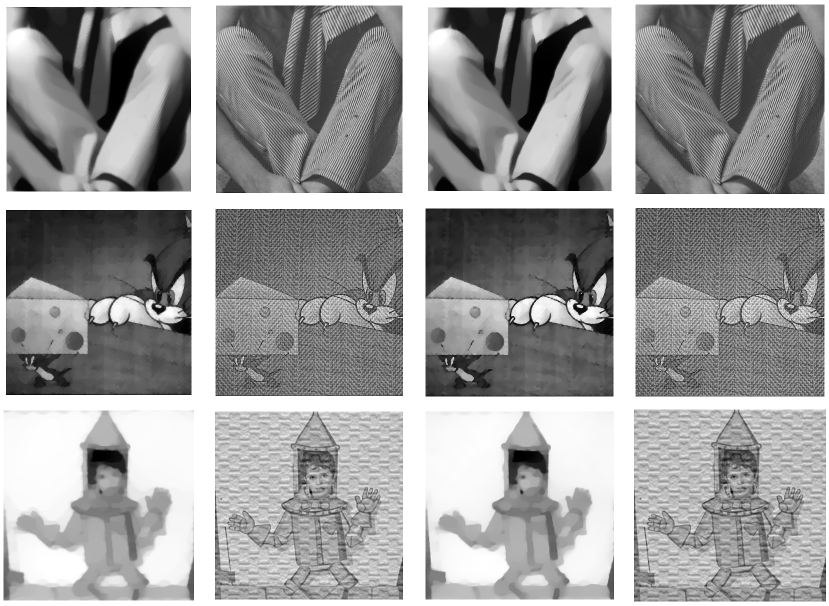


Fig. 11 Decomposed cartoons and textures of synthetic images and real image. From left column to right column: cartoons by PC-ADMM, textures by PC-ADMM, cartoons by G-LADMM and textures by G-LADMM.

To summarize our numerical experiments, we want to make some observations. Firstly, from the curves in Figures 12 and 13, all test algorithms can successfully separate the cartoon from texture for target images with corruptions, and the numerical performances of G-LADMM are comparable with APGM, and slightly better than ADM-G and PC-ADMM. The SNR of reconstructed images from G-LADMM in Figure 14 are 24.23dB for "Barbara", 22.23dB for "Tomjerry" and 32.32dB for "Boy".

Secondly, G-LADMM converges to the smaller objective function value faster than other algorithms in Figure 12. However, in Figure 13, for "Tomjerry" and "Boy", G-LADMM and APGM are with the slower SNR increasing in terms of the iteration number. On the other hand, the monotone decreasing for the objective (even for $\|w^k - \bar{w}\|$ from Section 5.1) can not be established in G-LADMM, the curves from G-LADMM verify this point and fluctuate significantly in the first 40 iterations.

7 Conclusion.

In this paper, a generalization of linearized ADMM with positive-definite and indefinite proximal regularization was presented for solving 2-block separable convex programming. We provided an improved convergence domain for the stepsizes of the dual variables and the proximal parameter of the linearized subproblems, which is significantly larger than that given in the literature. However, the optimal choice of the proximal parameter remains unknown except for the G-LADMM with $\beta = 1$.

With the help of the variational inequality, we observed the convergence of the proposed algorithm. The numerical experiments demonstrated the improvements of the G-LADMM on the LASSO and image problems. Next, we will consider the application of linearized ADMM with indefinite proximal regularization to multi-block separable convex problems.

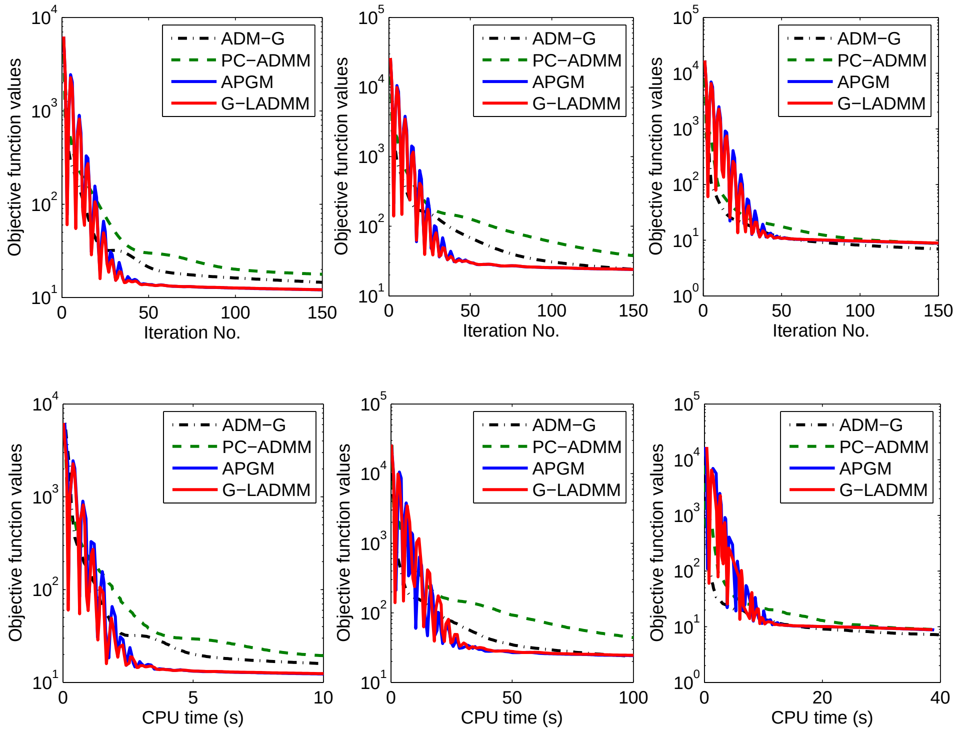


Fig. 12 Evolutions of objective function values w.r.t. iterations and computing time. From left column to right column for: "Barbara", "Tomjerry" and "Boy".

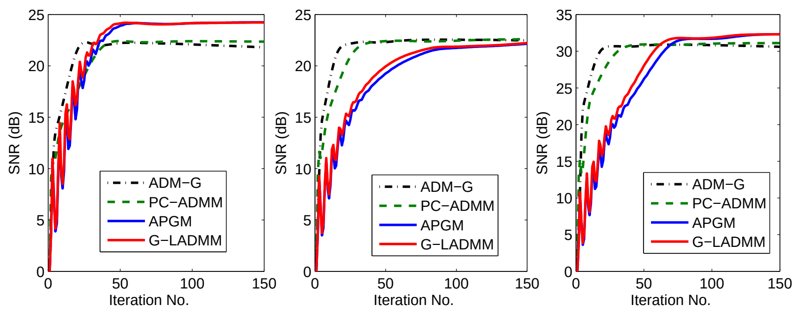


Fig. 13 Evolutions of SNR w.r.t. iterations. From left column to right column for: "Barbara", "Tomjerry" and "Boy".

Acknowledgements The authors would like to thank editor in chief and the anonymous referees for their valuable comments and suggestions which helped to improve the original version of this paper. The research of Xiaokai Chang was supported by the Fundamental Research Funds for Central Universities and the Hongliu Foundation of First-class Disciplines of Lanzhou University of Technology, China. The project was supported by the National Natural Science Foundation of China under Grant 61877046 and 41471338.

References

1. Sun, D., 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, 882–91 (2015)

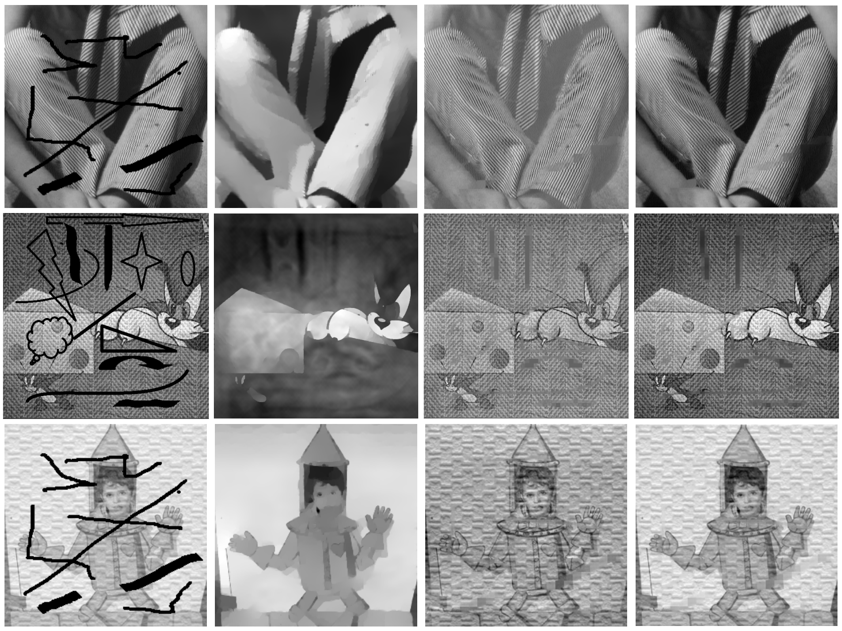


Fig. 14 Decomposed cartoons and textures on images with missing pixels by G-LADMM. From left column to right column: corrupted image, cartoon, texture and cartoon+texture.

2. Chang, X., Liu, S., Li, X.: Modified alternating direction method of multipliers for convex quadratic semidefinite programming. Neurocomputing 214, 575–586 (2016)
3. Chang, X., Liu, S., Zhao P.: A note on the sufficient initial condition ensuring the convergence of directly extended 3-block ADMM for special semidefinite programming. Optimization 67(10), 1729–1743 (2018)
4. Cand'es, E.J., Li, X., Ma, Y., Wright J.: Robust principal component analysis?. J. ACM 58, 1–37 (2011)
5. Bouwmans, T., Aybat, N., Zahzah, E.: Handbook on robust low-rank and sparse matrix decomposition: Applications in image and video processing. CRC Press, Taylor and Francis Group, May 2016.
6. Li K., Sundin M., Rojas C.R., Chatterjee S., Jansson M.: Alternating strategies with internal ADMM for low-rank matrix reconstruction. Signal Process. 121, 153–159 (2016)
7. Tao, M., Yuan, X.: Recovering low-rank and sparse components of matrices from incomplete and noisy observations. SIAM J. Optim. 21, 57–81 (2011)
8. He, B., Tao, M., Yuan, X.: Alternating direction method with gaussian back substitution for separable convex programming. SIAM J. Optim. 22, 313–340 (2012)
9. He, B., Yuan, X.: A class of ADMM-based algorithms for multi-block separable convex programming. Comput. Optim. Appl. 70(3):791–826, (2018)
10. Chan, T.F., Glowinski, R.: Finite element approximation and iterative solution of a class of mildly nonlinear elliptic equations. Stanford report STAN-CS-78-674, Computer Science Department, Stanford University, Palo Alto (1978)
11. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. 2, 17–40 (1976).
12. Hong, M., Luo, Z., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. SIAM J. Optim. 26(1), 3836–3840 (2014).
13. Glowinski, R.: ADMM and non-convex variational problems. In: Glowinski, R., Osher, S.J., Yin,W. (eds.) Splitting Methods in Communication, Imaging, Science, and Engineering, pp. 251–299. Springer, New York (2016)
14. Wang Y., Yin W., Zeng J.: Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. J. Sci. Comput. 78(1), 29-63 (2019)
15. Wang, F., Cao, W., Xu, Z.: Convergence of multi-block Bregman ADMM for nonconvex composite problems. Science China Information Sciences 61(12):122101 (2018)
16. Guo, K., Han, D., Yuan, X.: Convergence analysis of Douglas–Rachford splitting method for "strongly + weakly" convex programming. SIAM J. Numer. Anal. 55(4), 1549–1577 (2017)
17. Gao, B., Ma, F.: Symmetric alternating direction method with indefinite proximal regularization for linearly constrained convex optimization. J. Optimiz. Theory. App. 176(1), 178–204 (2018)

18. He, B., Ma, F., Yuan, X.: Linearized alternating direction method of multipliers via positive-indefinite proximal regularization for convex programming. Optimization Online (2016), http://www.optimization-online.org/DB_HTML/2016/07/5569.html.
19. Wang, X., Yuan, X.: The linearized alternating direction method of multipliers for dantzig selector. SIAM J. Sci. Comput. 34(5), A2792–A2811 (2012)
20. He, B., Ma, F., Yuan, X.: Optimal Linearized Alternating Direction Method of Multipliers for Convex Programming. Optimization Online (2017), http://www.optimization-online.org/DB_HTML/2017/09/6228.html.
21. Yang, J., Yuan, X.: Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. Math. Comput. 82 (281), 301–329 (2013)
22. Deng W., Yin W.: On the Global and Linear Convergence of the Generalized Alternating Direction Method of Multipliers. J. Sci. Comput. 66(3):889–916 (2016)
23. Sun M., Liu J.: The convergence rate of the proximal alternating direction method of multipliers with indefinite proximal regularization. J. Inequal. Appl. 2017(1):1-15, (2017)
24. Goncalves M. L. N., Melo J. G., Monteiro R. D. C.: Extending the ergodic convergence rate of the proximal ADMM. 2016. arXiv:1611.02903.
25. Li, M., Sun, D., Toh, K.C.: A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization. SIAM J. Optim. 26(2), 922–950 (2016)
26. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. 55, 293–318 (1992)
27. Cai, X., Gu, G., He, B., Yuan, X.: A proximal point algorithm revisit on alternating direction method of multipliers. Sci. China Math. 56(10), 2179–2186 (2013).
28. He, B., Ma, F., Yuan, X.: Convergence study on the symmetric version of ADMM with larger step sizes. SIAM J. Imaging Sci. 9, 1467–1501 (2016)
29. Bai, J., Li, J., Xu, F., Zhang H.: Generalized symmetric ADMM for separable convex programming. Comput. Optim. Appl. 70(1), 127–170 (2018).
30. Rockafellar R. T.: Convex Analysis. Princeton, 1970, https://www.convexoptimization.com/TOOLS/AnalysisRockafellar.pdf
31. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (2011).
32. Tibshirani, R.: Regression shrinkage and selection via the lasso: a retrospective. J. R. Stat. Soc. Ser. B. 73(3), 273–282 (2011)
33. Han, D., Kong, W., Zhang, W.: A partial splitting augmented lagrangian method for low patch-rank image decomposition. J. Math. Imaging Vis. 51(1), 145–160 (2015)
34. Schaeffer, H., Osher, S.: A low patch-rank interpretation of texture. SIAM J. Imaging Sci. 6(1), 226–262 (2013)
35. Ng, M.K., Yuan, X., Zhang, W.: A coupled variational image decomposition and restoration model for blurred cartoon-plus-texture images with missing pixels. IEEE Tran. Imaging Proc. 22, 2233–2246 (2013)
36. Chang, X., Liu, S., Zhao, P., Li, X.: Convergent prediction-ccorrection-based ADMM for multi-block separable convex programming. J. Comput. Appl. Math. 335, 270–288 (2018)
37. Ma S.: Alternating proximal gradient method for convex minimization. J. Sci. Comput. 68(2):546-572 (2016)
38. Donoho, D.L., Tsaig, Y.: Fast solution of l_1-norm minimization problems when the solution may be sparse. IEEE Trans. Inf. Theory 54(11), 4789–4812 (2008)
39. Dai, Y., Han, D., Yuan, X.: A sequential updating scheme of the Lagrange multiplier for separable convex programming. Math. Comput. 86(303),315–343 (2017)