Journal of Optimization Theory and Applications (2018) 179:163–196
https://doi.org/10.1007/s10957-018-1338-x

CrossMark

On Glowinski's Open Question on the Alternating Direction Method of Multipliers

Min Tao[1] · **Xiaoming Yuan**[2]

Received: 14 June 2017 / Accepted: 15 June 2018 / Published online: 10 July 2018

Abstract

The alternating direction method of multipliers was proposed by Glowinski and Marrocco in 1974, and it has been widely used in a broad spectrum of areas, especially in some sparsity-driven application domains. In 1982, Fortin and Glowinski suggested to enlarge the range of the dual step size for updating the multiplier from 1 to the open interval of zero to the golden ratio, and this strategy immediately accelerates the convergence of alternating direction method of multipliers for most of its applications. Meanwhile, Glowinski raised the question of whether or not the range can be further enlarged to the open interval of zero to 2; this question remains open with nearly no progress in the past decades. In this paper, we answer this question affirmatively for the case where both the functions in the objective function are quadratic. Thus, Glowinski's open question is partially answered. We further establish the global linear convergence of the alternating direction method of multipliers with this enlarged step size range for the quadratic programming under a tight condition.

Keywords Alternating direction method of multipliers · Glowinski's open question · Quadratic programming · Step size · Linear convergence

Mathematics Subject Classification 90C25 · 90C30 · 65K05 · 90C20

Communicated by Roland Glowinski.

✉ Min Tao
taom@nju.edu.cn

Xiaoming Yuan
xmyuan@hku.hk

1 Department of Mathematics, National Key Laboratory for Novel Software Technology, Nanjing University, Jiangsu, China

2 Department of Mathematics, The University of Hong Kong, Hong Kong, China

1 Introduction

The alternating direction method of multipliers (ADMM) was proposed by Glowinski and Marrocco in [1]. Recently, the ADMM has found many applications in a variety of areas [2–6] because of its simplicity in implementation and usually good numerical performance, and it has received increasing attention in the literature. We refer to, [2–4,7], for some review papers of the ADMM.

Notice that solving the subproblems related to the primal variables dominates the implementation of the ADMM scheme. It is meaningful to discuss the theory of enlarging the dual step size because it may offer an immediate accelerating on the convergence of the ADMM. In [8],[1] Fortin and Glowinski proposed a variant of the ADMM scheme (see (3)) by introducing a dual step size in the multiplier updating step, and established the global convergence for the step size range of zero to the golden ratio. Though the convergence of the ADMM scheme with this range is well studied; see, e.g., [9–15], numerically it has been observed as well that some values exceeding the golden ratio and < 2 may still perform convergence. It is natural to ask whether we can enlarge the step size range to the open interval of zero to 2 in the ADMM scheme while still guarantee the convergence. This question was raised by Glowinski in [11]. In this paper, we establish the global convergence of the ADMM scheme with the dual step size in the open interval of zero to 2 for solving quadratic programming problems, hence partially answer Glowinski's open question. Moreover, we establish the global linear convergence under a new while tight condition.

The remaining part of this paper is organized as follows. In Sect. 2, we present the problem under discussion and recall some known results in the ADMM studies. In Sect. 3, we summarize some notations and definitions to be used, present the assumptions for further discussion, and prove a number of lemmas. Then, we conduct some preparatory analysis in Sect. 4, including the specification of the matrix recursion of the ADMM scheme for the quadratic programming model problem, the KKT condition, and an example showing the divergence of the ADMM with step size 2. In Sect. 5, the convergence of the ADMM scheme with the open interval of zero to 2 is proved for the quadratic programming. This is the main result of the paper. The global linear convergence is established in Sect. 6, under a new condition different from some existing work. Both the convergence and global linear convergence are verified numerically by some examples in Sects. 5 and 6. Finally, some conclusions are drawn in Sect. 7.

2 The Model Problem and Some Known Results

We consider the following canonical convex minimization problem with linear constraints and a separable objective function without coupled variables:

$$\min_{x,y}\{\theta_1(x) + \theta_2(y) \;:\; Ax + By = b,\ x \in \mathcal{X},\ \ y \in \mathcal{Y}\}, \tag{1}$$

[1] This is a translation from its original French version in 1982.

where $A \in \mathbb{R}^{m\times n_1}$, $B \in \mathbb{R}^{m\times n_2}$, $b \in \mathbb{R}^m$, $\mathcal{X} \subset \mathbb{R}^{n_1}$ and $\mathcal{Y} \subset \mathbb{R}^{n_2}$ are closed convex sets, $\theta_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ and $\theta_2 : \mathbb{R}^{n_2} \to \mathbb{R}$ are convex (not necessarily smooth) functions. Let the augmented Lagrangian function of (1) be

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

with $z \in \mathbb{R}^m$ the Lagrange multiplier and $\beta > 0$ the penalty parameter. The iterative scheme of ADMM for (1) reads as

$$x^{k+1} = \arg\min_{x\in\mathcal{X}} \mathcal{L}_\beta(x, y^k, z^k), \tag{2a}$$
$$y^{k+1} = \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y, z^k), \tag{2b}$$
$$z^{k+1} = z^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{2c}$$

Furthermore, Fortin and Glowinski in [8] proposed the following variant of the ADMM scheme:

$$x^{k+1} = \arg\min_{x\in\mathcal{X}} \mathcal{L}_\beta(x, y^k, z^k), \tag{3a}$$
$$y^{k+1} = \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y, z^k), \tag{3b}$$
$$z^{k+1} = z^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), \tag{3c}$$

with $\gamma \in]0, \frac{1+\sqrt{5}}{2}[$. It is worthwhile to mention that the parameter γ in (3c) is different from the involved parameter in the so-called generalized ADMM that was discussed in [16,17] based on the idea in [18] (see also [19]). The convergence of (3) with $\gamma \in]0, \frac{1+\sqrt{5}}{2}[$ has been well addressed in various contexts; see, e.g., [9–15]. Numerically, it has been widely verified that a large value of γ close to $\frac{1+\sqrt{5}}{2}$ can accelerate the convergence of ADMM immediately; see, e.g., [5,6,11,20]. Glowinski raised the question in [11] (see pp. 182 therein) as: " If G is linear, it has been proved by Gabay and Mercier [1] that ALG2 converges if $0 < \rho_n = \rho < 2r$. The proof of this result is rather technical, and an open question is to decide whether it can be extended to the more general cases considered here." The function "G" in [11] corresponds to the function θ_2 in problem (1); "ALG2" refers to the ADMM scheme (3); "[1]" refers to [10] and $\rho := \gamma\beta$ in our setting. With the well studied results for $\gamma \in]0, \frac{1+\sqrt{5}}{2}[$ in (3), the gap from $\gamma \in]0, \frac{1+\sqrt{5}}{2}[$ to $\gamma \in]0, 2[$ remains unsolved and thus Glowinski's question is still open since it was proposed in [11].

The rationale of raising this question can also be explained as follows. Note that if the problem (1) is regarded as a whole and the augmented Lagrangian method (ALM) in [21,22] is directly applied to (1), the iterative scheme becomes:

$$(x^{k+1}, y^{k+1}) = \arg\min_{x\in\mathcal{X},\, y\in\mathcal{Y}} \mathcal{L}_\beta(x, y, z^k), \tag{4a}$$
$$z^{k+1} = z^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{4b}$$

Based on the work [23], the ALM scheme (4) is an application of the proximal point algorithm (PPA) in [24] to the dual of the problem (1), and thus the result in [18] can be applied to modify the scheme (4) as

$$(x^{k+1}, y^{k+1}) = \arg\min_{x\in\mathcal{X},\ y\in\mathcal{Y}} \mathcal{L}_\beta(x, y, z^k), \tag{5a}$$

$$z^{k+1} = z^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), \tag{5b}$$

where $\gamma \in]0, 2[$. Its convergence can be found in, e.g., [19]. Therefore, the ADMM scheme (2) can be regarded as a splitting version of the ALM (4) which splits the (x^{k+1}, y^{k+1})-subproblem (4a) as the surrogates (2a) and (2b) by the Gauss–Seidel manner. Then, with $\gamma \in]0, 2[$ in (5b) for the ALM scheme, it is natural to ask if this property can be maintained for the ADMM scheme (3). Since the ADMM (2) is just an inexact version of the ALM (4), it is not straightforward to claim the validity of this extension and this may explain why Glowinski's question is still open.

In this paper, we restrict our attention to the following quadratic programming problem:

$$\min\left\{\frac{1}{2}x^\top Px + f^\top x + \frac{1}{2}y^\top Qy + g^\top y\ :\ Ax + By = b,\ x \in \mathbb{R}^{n_1},\ y \in \mathbb{R}^{n_2}\right\}, \tag{6}$$

where $P \in \mathbb{R}^{n_1\times n_1}$ and $Q \in \mathbb{R}^{n_2\times n_2}$ are symmetric positive-semidefinite matrices, $A \in \mathbb{R}^{m\times n_1}$, $B \in \mathbb{R}^{m\times n_2}$, $b \in \mathbb{R}^m$, $f \in \mathbb{R}^{n_1}$ and $g \in \mathbb{R}^{n_2}$. We refer to, e.g., [25–31], for various applications that can be modeled as (6) in different fields. The solution set of (6) is assumed to be non-empty throughout our discussion.

3 Preliminaries

3.1 Notations

Given a real number a, $|a|$ represents the absolute value of a. The superscript "$\top$" denotes the transpose, and the superscript "H" denotes the conjugate transpose. A unit vector means its 2-norm is 1, i.e., $x^\top x = 1$. We use $a + b\mathtt{i}$ to denote a complex number, in which "$\mathtt{i}$" represents the imaginary unit. For a complex number a, $|a|$ and $\mathrm{Re}(a)$ denote its modulus and the real part, respectively. Given a vector space $\mathcal{V}$, its dimension is denoted by $\dim(\mathcal{V})$. For a vector $x \in \mathbb{R}^n$, $\|x\|_2$ represents $\sqrt{\sum_{i=1}^n |x_i|^2}$. Given a square matrix $M \in \mathbb{R}^{n\times n}$, $\det(M)$ denotes its determinant. Given a matrix $M \in \mathbb{R}^{m\times n}$ that is not necessarily square, $\mathrm{Rank}(M)$ represents its rank. For a matrix $M \in \mathbb{R}^{n\times n}$, $\mathrm{eig}(M)$ represents all the eigenvalues of M (considering the multiplicity), and $\overline{\mathrm{eig}}(M)$ represents all the nonzero eigenvalues of M (considering the multiplicity). We use the notation $\sigma(M)$ to denote its spectrum, i.e., the set of distinct eigenvalues. For an symmetric matrix M, let $\|M\|_2$ denote its 2-norm. For a nonsymmetric matrix M, $\|M\| := \sqrt{\|M^\top M\|}$ and $\rho(M)$ refers to its spectral radius, i.e., the maximal modulus of its eigenvalues. For a matrix $M \in \mathbb{R}^{n\times n}$ that is not necessarily symmetric, λ_M

denotes any one of its eigenvalues, $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ represent the maximal and minimal eigenvalues of M, respectively. The matrix I_n represents the identity matrix in $\mathbb{R}^{n\times n}$, and we omit the subscript only when its dimension is equal to m. For a matrix $M \in \mathbb{R}^{n\times n}$ that is not necessarily symmetric, the notation $M \succeq 0$ means M is positive semidefinite and $M \succ 0$ means M is positive definite. We use the notation diag$(\cdot)$ to denote a diagonal matrix. The notation $N(\cdot)$ represents the null space. The function δ_{ij} is defined as

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

3.2 Assumptions

Throughout this paper, we make the following assumptions.

Assumption 1 In (6), the symmetric positive-semidefinite matrices P and Q, and the matrices A and B satisfy the conditions:

$$P + A^\top A \succ \mathbf{0} \quad \text{and} \quad Q + B^\top B \succ \mathbf{0}.$$

Assumption 2 The KKT point set of (1) is non-empty.

Remark 3.1 Under Assumptions 1–2, the x- and y-subproblems of the ADMM scheme (3) for solving (6) are well-posed (which means that we have existence and uniqueness). Also, as shown by Corollary 1 in [32], Assumptions 1–2 are necessary for the well-definedness of the ADMM scheme (3) for solving (6).

3.3 Some Lemmas

In the following, we prove a number of lemmas that will be used in later analysis. Some of them are elementary.

Lemma 3.1 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying the conditions*

$$\mathbf{0} \preceq F \preceq I \quad \textit{and} \quad \mathbf{0} \preceq G \preceq I.$$

Then, we have $-I \preceq FG \preceq I$.

Proof Using Cauchy–Schwartz inequality, we have $FG + GF \preceq F^2 + G^2 \preceq 2I$. The second follows from the conditions $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. Analogously, we get $FG + GF \succeq -F^2 - G^2 \succeq -2I$. Thus, the assertion follows directly. □

Lemma 3.2 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying the conditions*

$$\mathbf{0} \preceq F \preceq I \quad \textit{and} \quad \mathbf{0} \preceq G \preceq I.$$

Then, we have

(i) *For any* $x \in \mathbb{R}^m$ *with* $x^T x = 1$*, we have* $|x^\top FGx| \le 1$.
(ii) *For any* $\eta \in \mathbb{C}^m$ *with* $\eta^H \eta = 1$*, if* $\eta^H FG\eta$ *is a real number, then we also have* $|\eta^H FG\eta| \le 1$.

Proof For the first assertion, it follows from Cauchy–Schwarz inequality that

$$|x^\top FGx| = |(Fx)^\top (Gx)| \le \frac{x^\top F^2 x + x^\top G^2 x}{2} \le \frac{x^\top Fx + x^\top Gx}{2} \le 1.$$

For the second assertion, we assume that $\eta := \alpha_1 + \alpha_2 \mathtt{i}$ with $\alpha_1 \in \mathbb{R}^m$, $\alpha_2 \in \mathbb{R}^m$ and

$$\alpha_1^\top \alpha_1 + \alpha_2^\top \alpha_2 = 1.$$

If $\eta^H FG\eta$ is a real number, then we have

$$\begin{aligned}|\eta^H FG\eta| &= |(\alpha_1 + \alpha_2 \mathtt{i})^H FG(\alpha_1 + \alpha_2 \mathtt{i})| = |\alpha_1^\top FG\alpha_1 + \alpha_2^\top FG\alpha_2| \\ &\le \frac{\alpha_1^\top F^2 \alpha_1 + \alpha_1^\top G^2 \alpha_1}{2} + \frac{\alpha_2^\top F^2 \alpha_2 + \alpha_2^\top G^2 \alpha_2}{2} \le \alpha_1^\top \alpha_1 + \alpha_2^\top \alpha_2 = 1.\end{aligned}$$

The proof is complete. □

Lemma 3.3 *Let* U *and* V *be two symmetric matrices in* $\mathbb{R}^{m\times m}$ *satisfying the conditions*

$$-\frac{I}{2} \preceq U \preceq \frac{I}{2} \quad \textit{and} \quad -\frac{I}{2} \preceq V \preceq \frac{I}{2}.$$

Then, for any $x \in \mathbb{R}^m$ *such that* $x^\top x = 1$*, we have*

$$\left| x^\top \frac{(UV + VU)}{2} x \right| \le \frac{1}{4} \quad \textit{and} \quad -\frac{I}{4} \preceq \frac{(UV + VU)}{2} \preceq \frac{I}{4}.$$

Proof First, using Cauchy–Schwarz inequality, we get

$$|x^\top (UV)x| \le \frac{1}{2}\left(x^\top U^2 x + x^\top V^2 x\right), \tag{7}$$

and

$$|x^\top (VU)x| \le \frac{1}{2}\left(x^\top U^2 x + x^\top V^2 x\right). \tag{8}$$

Then, recalling $-\frac{I}{2} \preceq V \preceq \frac{I}{2}$ and $-\frac{I}{2} \preceq U \preceq \frac{I}{2}$, we obtain that $\mathbf{0} \preceq U^2 \preceq \frac{I}{4}$ and $\mathbf{0} \preceq V^2 \preceq \frac{I}{4}$. Substituting these two inequalities into (7) and (8), the first assertion follows immediately. The second assertion follows directly from the first assertion. □

The following lemma is essential in the convergence analysis for the ADMM (3) for the problem (6).

Lemma 3.4 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. Then, we have $\mathbf{0} \preceq I - F - G + 2FG \preceq I$.*

Proof It is equivalent to show that

$$\mathbf{0} \preceq I - F - G + FG + GF \preceq I.$$

Since $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$, we obtain $\mathbf{0} \preceq I - F \preceq I$ and $\mathbf{0} \preceq I - G \preceq I$. With simple calculations, we have

$$I - F - G + FG + GF = (\frac{I}{2} - F)(\frac{I}{2} - G) + (\frac{I}{2} - G)(\frac{I}{2} - F) + \frac{I}{2}. \tag{9}$$

Because of $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$, we obtain

$$-\frac{I}{2} \preceq \frac{I}{2} - F \preceq \frac{I}{2}, \quad -\frac{I}{2} \preceq \frac{I}{2} - G \preceq \frac{I}{2}. \tag{10}$$

Thus, using Lemma 3.3, we get

$$-\frac{I}{2} \preceq (\frac{I}{2} - F)(\frac{I}{2} - G) + (\frac{I}{2} - G)(\frac{I}{2} - F) \preceq \frac{I}{2}. \tag{11}$$

Combining (9) and (11), we have $\mathbf{0} \preceq I - F - G + FG + GF \preceq I$. Thus, the proof is complete. □

The following lemma plays a key role in the convergence analysis of (6).

Lemma 3.5 *[33] Let A and B be $m \times m$ Hermitian matrices with eigenvalues*

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m \text{ and } \beta_1 \geq \beta_2 \geq \cdots \geq \beta_m,$$

respectively. Then, we have

$$\min_{\pi} \left\{ \prod_{i=1}^{m} (\alpha_i + \beta_{\pi(i)}) \right\} \leq \det(A + B) \leq \max_{\pi} \left\{ \prod_{i=1}^{m} (\alpha_i + \beta_{\pi(i)}) \right\}.$$

(The minimization and maximization above are taken over all permutations of the indices $1, 2, \ldots, m$). In particular, if $\alpha_m + \beta_m \geq 0$ (which is true if both A and B are positive semidefinite), then we have

$$\prod_{i=1}^{m} (\alpha_i + \beta_i) \leq \det(A + B) \leq \prod_{i=1}^{m} (\alpha_i + \beta_{m+1-i}).$$

Lemma 3.6 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$. Then, we have*

$$\mathrm{Rank}(F + G - 2FG) = \mathrm{Rank}(F + G - 2GF).$$

Proof The proof is elementary; thus we omit here. □

Lemma 3.7 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$ and x is an associated eigenvector, then we have that 1 is also an eigenvalue of the matrix $(I - F - G + 2GF)$ and x is an associated eigenvector.*

Proof Assume that the vector $\hat{x}$ is a unit eigenvector of the matrix $(I - F - G + 2FG)$ associated with 1. It means that $(I - F - G + 2FG)\hat{x} = \hat{x}$. Thus, we get

$$\hat{x}^\top (I - F - G + 2FG)\hat{x} = 1 \text{ with } \hat{x}^\top \hat{x} = 1. \tag{12}$$

On the other hand, for any $x \in \mathbb{R}^m$ with $x^\top x = 1$, we have

$$\begin{aligned} x^\top (I - F - G + 2FG)x &= x^\top \left\{ 2\left[\left(\frac{I}{2} - F\right)\left(\frac{I}{2} - G\right) + \frac{I}{4} \right] \right\} x \\ &\le \left(x^\top \left(\frac{I}{2} - F\right)^2 x + x^\top \left(\frac{I}{2} - G\right)^2 x \right) + \frac{1}{2} \le 1, \end{aligned} \tag{13}$$

where the first and the second inequalities respectively follow from Cauchy–Schwarz inequality and the facts that $\mathbf{0} \preceq (\frac{I}{2} - F)^2 \preceq \frac{I}{4}$ and $\mathbf{0} \preceq (\frac{I}{2} - G)^2 \preceq \frac{I}{4}$ (due to (10)). Moreover, (12) implies that (13) holds with equality. Thus, checking the conditions ensuring the (13) with equality. We have one of the following two arguments:

(i) $(\frac{1}{2}I - F)x = \frac{1}{2}x$ and $(\frac{1}{2}I - G)x = \frac{1}{2}x$;
(ii) $(\frac{1}{2}I - F)x = -\frac{1}{2}x$ and $(\frac{1}{2}I - G)x = -\frac{1}{2}x$.

That is to say that one of the following assertions holds:

(i) $Fx = Gx = \mathbf{0}$;
(ii) $Fx = Gx = x$.

Consequently, we have $(I - F - G + 2GF)x = x$. It implies that the value 1 is an eigenvalue of the matrix $(I - F - G + 2GF)$ and x is its eigenvector. The proof is complete. □

Lemma 3.8 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$ and x is an associated eigenvector, then one of the following assertions hold:*

(i) *Fx=Gx=x;*
(ii) *Fx=Gx=$\mathbf{0}$.*

Conversely, if a vector x satisfies (i) or (ii), then it is an eigenvector of $(I - F - G + 2FG)$ associated with the eigenvalue 1.

Proof The proof is included in the proof for Lemma 3.7. □

The following lemma provides a new way to show that an eigenvalue of a nonsymmetric matrix has the same geometric and algebraic multiplicities; see Theorem 1 in [34].

Lemma 3.9 [34] *Let $A \in \mathbb{C}^{n\times n}$ and let λ be an eigenvalue of A. Then the following two statements are equivalent.*

(i) *There exist bi-orthonormal bases $\{x_1, \ldots, x_J\}$ of $N(A-\lambda I)$ and $\{y_1, \ldots, y_J\}$ of $N(A^H - \bar{\lambda} I)$ in the sense that $y_j^H x_k = \delta_{jk}$, $\forall j, k = 1, \ldots, J$, where J is the geometric multiplicity of λ.*

(ii) *The geometric multiplicity and algebraic multiplicity of λ are equivalent.*

In the following lemma, we show that, if 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, then it has a complete set of eigenvectors.

Lemma 3.10 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, then this eigenvalue has a complete set of eigenvectors. That is, the algebraic and geometric multiplicities are the same, and we denote them by*

$$\ell := m - \mathrm{Rank}(F + G - 2FG). \tag{14}$$

Proof Since 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, its geometric multiplicity of 1 is the ℓ defined in (14). Invoking Lemma 3.7, we know that 1 is also an eigenvalue of the matrix $(I - F - G + 2GF)$. Moreover, any eigenvector x for the matrix $(I - F - G + 2FG)$ associated with 1 is also an eigenvector of the matrix $(I - F - G + 2GF)$ associated with 1. Then, it follows from Lemma 3.6 that the geometric multiplicity of 1 for the matrix $(I - F - G + 2GF)$ is also ℓ defined in (14). Suppose $\{x_1, \ldots, x_l\}$ is a set of orthonormal eigenvectors associated with 1 for the matrix $(I - F - G + 2FG)$ in sense of $x_j^\top x_k = \delta_{jk}, \forall\, j,\ k = 1, \ldots, l$. Then, it is also a set of orthonormal eigenvectors associated with 1 for the matrix $(I - F - G + 2GF)$. According to Lemma 3.9, the algebraic multiplicity of the matrix $(I - F - G + 2FG)$ is ℓ. Indeed, the set of orthonormal vectors $\{x_1, \ldots, x_l\}$ is exactly a complete set of eigenvectors associated with the eigenvalue 1 for the matrix $(I - F - G + 2FG)$. The proof is complete. □

The following lemma can be obtained immediately from Lemma 3.8.

Lemma 3.11 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, then we have*

$$\dim(\{x \;:\; x \in N(F - I) \cap N(G - I)\}) + \dim(\{x \;:\; x \in N(F) \cap N(G)\}) = \ell, \tag{15}$$

where ℓ is the algebraic and geometric multiplicities of $(I - F - G + 2FG)$ defined in (14).

Proof First, invoking Lemma 3.8, we have

$$\begin{aligned} &\{x : x \in N([I - F - G + 2FG] - I)\} \\ &\quad = \{x : x \in N(F - I) \cap N(G - I)\} \bigcup \{x : x \in N(F) \cap N(G)\}. \end{aligned} \tag{16}$$

Note that $\{x\ :\ x \in N(F-I) \cap N(G-I)\} \bigcap \{x\ :\ x \in N(F) \cap N(G)\} = \{\mathbf{0}\}$. From Lemma 3.10, we know that $\dim\{x\ :\ x \in N([I - F - G + 2FG] - I)\} = \ell$ with ℓ defined in (14). Thus, the assertion (15) follows directly. □

Lemma 3.12 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, then there exists an orthogonal matrix $Q \in \mathbb{R}^{m\times m}$ such that*

$$Q^\top F Q = \begin{pmatrix} D & \mathbf{0}_{\ell\times(m-\ell)} \\ \mathbf{0} & \hat{F} \end{pmatrix} \quad \textit{and} \quad Q^\top G Q = \begin{pmatrix} D & \mathbf{0}_{\ell\times(m-\ell)} \\ \mathbf{0} & \hat{G} \end{pmatrix}, \tag{17}$$

where $D \in \mathbb{R}^{\ell\times\ell}$ is a diagonal matrix with the diagonal entries either 0 or 1, and the matrices $\hat{F},\ \hat{G} \in \mathbb{R}^{(m-\ell)\times(m-\ell)}$ satisfy

$$\mathbf{0} \preceq \hat{F} \preceq I_{m-\ell},\quad \mathbf{0} \preceq \hat{G} \preceq I_{m-\ell} \ \textit{ and } \ \hat{F} + \hat{G} - 2\hat{G}\hat{F} \succ \mathbf{0}. \tag{18}$$

Proof Since 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, it follows from Lemma 3.10 that there exists a set of orthonormal eigenvectors associated with 1 for the matrix $(I - F - G + 2FG)$, denoted by $\{x_1, \ldots, x_l\}$. Recall (16). We thus have

$$\{x_1, \ldots, x_l\} = \{x\ :\ x \in N(F-I) \cap N(G-I)\} \bigcup \{x\ :\ x \in N(F) \cap N(G)\}.$$

Let us construct an orthogonal matrix Q with the first ℓ columns as $(x_1; \ldots; x_l)$. Then, we partition Q as $(Q_1; Q_2)$ with $Q_1 = (x_1; x_2; \ldots; x_\ell)$ and $Q_2 \in \mathbb{R}^{m\times(m-\ell)}$. Invoking Lemma 3.8, we have $FQ_1 = Q_1 D$ and $GQ_1 = Q_1 D$, with

$$D = \mathrm{diag}(\lambda_1, \cdots, \lambda_\ell) \ \text{ and } \ \lambda_i = 0 \text{ or } 1,\ i = 1, \ldots, \ell.$$

Therefore, we get

$$Q^\top F Q = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} F \left(Q_1;\ Q_2 \right) = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} \left(Q_1 D;\ F Q_2 \right) = \begin{pmatrix} D & \mathbf{0} \\ \mathbf{0} & Q_2^\top F Q_2 \end{pmatrix}.$$

Analogously, we have the second equation of (17). Thus, the proof is complete.

For the second assertion (18), let us define $\hat{F} = Q_2^\top F Q_2$ and $\hat{G} = Q_2^\top G Q_2$. Then, the first two inequalities in (18) hold. On the other hand, similar to the second inequality in Lemma 3.4, we have $\hat{F} + \hat{G} - 2\hat{G}\hat{F} \succeq \mathbf{0}$. Now, we prove the third inequality in (18) by contradiction. Suppose that 0 is an eigenvalue of the matrix $(\hat{F} + \hat{G} - 2\hat{G}\hat{F})$. Then, 1 is an eigenvalue of the matrix $(I_{m-\ell} + \hat{F} + \hat{G} - 2\hat{G}\hat{F})$. Assume that $\hat{x}$ is an eigenvector of the matrix $(I_{m-\ell} + \hat{F} + \hat{G} - 2\hat{G}\hat{F})$ associated with the eigenvalue 1. Then, similar to the proof of Lemma 3.7, we know that either $\hat{F}\hat{x} = \hat{G}\hat{x} = \hat{x}$ or $\hat{F}\hat{x} = \hat{G}\hat{x} = \mathbf{0}$ holds. This implies that $\hat{F}$ and $\hat{G}$ have the common eigenvector $\hat{x}$ associated with the eigenvalue 1 or 0, which contradicts with the assertion (15). Therefore, we have $\hat{F} + \hat{G} - 2\hat{G}\hat{F} \succ \mathbf{0}$. The proof is complete. □

Remark 3.2 Lemma 3.12 implies that the matrices $\hat{F}$ and $\hat{G}$ have no common eigenvectors for either the eigenvalue 0 or 1.

4 Specification of the ADMM Scheme for Quadratic Programming

In this section, we first specify the application of the ADMM scheme (3) to the quadratic programming problem (6) as a matrix recursion form and then discuss some related issues. This matrix recursion form is the basis of our further analysis.

Obviously, applying the ADMM scheme (3) to the quadratic programming problem (6) results in the iterative scheme:

$$\begin{aligned} (P+\beta A^\top A)x^{k+1} &= A^\top z^k - \beta A^\top B y^k + \beta A^\top b - f, \\ (Q+\beta B^\top B)y^{k+1} &= B^\top z^k - \beta B^\top A x^{k+1} + \beta B^\top b - g, \\ z^{k+1} &= z^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b). \end{aligned} \tag{19}$$

4.1 Matrix Recursion Form

Notice that the variable x in the ADMM (3) plays an intermediate role in the sense that x^k is not involved in the iteration to generate the next iterate. That is, a new iterate $(x^{k+1}, y^{k+1}, z^{k+1})$ can be generated by (y^k, z^k). Therefore, we first eliminate the variable x from the matrix recursion form (19) and obtain a more compact matrix recursion in a lower-dimension space. For this purpose, introducing an auxiliary variable $\mu^k := z^k/\beta$, we can recast the scheme (19) as

$$(P/\beta + A^\top A)x^{k+1} = A^\top \mu^k - A^\top B y^k + A^\top b - f/\beta, \tag{20a}$$

$$(Q/\beta + B^\top B)y^{k+1} = B^\top \mu^k - B^\top A x^{k+1} + B^\top b - g/\beta, \tag{20b}$$

$$\mu^{k+1} = \mu^k - \gamma(Ax^{k+1} + By^{k+1} - b). \tag{20c}$$

Note that (20a) can be written as

$$x^{k+1} = (P/\beta + A^\top A)^{-1}\left[A^\top \mu^k - A^\top B y^k + A^\top b - f/\beta\right]. \tag{21}$$

Then, substituting (21) into (20b) and (20c), we eliminate x^{k+1} from (20) and obtain

$$\begin{aligned} \hat{Q}y^{k+1} &= B^\top A\hat{P}^{-1}A^\top B y^k + (B^\top - B^\top A\hat{P}^{-1}A^\top)\mu^k + \alpha_1, \\ \gamma B y^{k+1} + \mu^{k+1} &= (I - \gamma A\hat{P}^{-1}A^\top)\mu^k + \gamma A\hat{P}^{-1}A^\top B y^k + \alpha_2, \end{aligned} \tag{22}$$

with

$$\hat{P} = P/\beta + A^\top A, \quad \hat{Q} = Q/\beta + B^\top B, \tag{23}$$

$$\alpha_1 = -B^\top A\hat{P}^{-1}A^\top b + B^\top A\hat{P}^{-1} f/\beta + B^\top b - g/\beta, \tag{24}$$

$$\alpha_2 = \gamma b - \gamma A\hat{P}^{-1}A^\top b + \gamma A\hat{P}^{-1} f/\beta. \tag{25}$$

Then, with simple calculations, the iterative scheme (22) can be written compactly as follows:

$$v^{k+1} = T(\gamma)v^k + q, \tag{26}$$

with

$$T(\gamma) = \begin{pmatrix} \hat{Q}^{-1}B^\top A\hat{P}^{-1}A^\top B & \hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \\ \gamma(I - B\hat{Q}^{-1}B^\top)A\hat{P}^{-1}A^\top B & I - \gamma A\hat{P}^{-1}A^\top - \gamma B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \end{pmatrix} \tag{27}$$

and

$$v^k = \begin{pmatrix} y^k \\ \mu^k \end{pmatrix}, \quad q = \begin{pmatrix} q_1 := \hat{Q}^{-1}\alpha_1 \\ q_2 := \alpha_2 - \gamma B\hat{Q}^{-1}\alpha_1 \end{pmatrix}. \tag{28}$$

Thus, the application of the ADMM scheme (3) to the quadratic programming problem (6) can be written as the matrix recursion form (26)–(28).

To establish the convergence of the ADMM (19) with $\gamma \in]0, 2[$, we only need to conduct a spectral analysis for the iterative matrix $T(\gamma)$ defined in (27). First, note that the matrix $T(\gamma)$ can be factorized as

$$T(\gamma) = \begin{pmatrix} \hat{Q}^{-1}B^\top A\hat{P}^{-1}A^\top & \hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \\ \gamma(I - B\hat{Q}^{-1}B^\top)A\hat{P}^{-1}A^\top & I - \gamma A\hat{P}^{-1}A^\top - \gamma B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \end{pmatrix} \cdot \begin{pmatrix} B & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}. \tag{29}$$

Thus, switching the order of the products by moving the first component to the last, we have a new matrix defined as

$$\tilde{T}(\gamma) = \begin{pmatrix} B\hat{Q}^{-1}B^\top A\hat{P}^{-1}A^\top & B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \\ \gamma(I - B\hat{Q}^{-1}B^\top)A\hat{P}^{-1}A^\top & I - \gamma A\hat{P}^{-1}A^\top - \gamma B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \end{pmatrix}. \tag{30}$$

For any two matrices X and Y (not necessarily square) with an appropriate dimension, we have

$$\overline{\text{eig}}(XY) = \overline{\text{eig}}(YX).$$

Hence, we obtain

$$\overline{\text{eig}}(T(\gamma)) = \overline{\text{eig}}(\tilde{T}(\gamma)). \tag{31}$$

Therefore, we only need to conduct the spectral analysis in terms of the matrix $\tilde{T}(\gamma)$.

4.2 KKT Condition

In this subsection, we show several equivalent forms to characterize the KKT condition of the quadratic programming problem (6) which will be useful for later analysis. These are necessary preparation for the convergence analysis in the next section. Recall that Assumptions 1–2 hold in our analysis.

Let (x^*, y^*, z^*) be a KKT point of the problem (6). That is, (x^*, y^*, z^*) satisfies the following equations:

$$Px^* = A^\top z^* - f,\ \ Qy^* = B^\top z^* - g,\ \ Ax^* + By^* - b = \mathbf{0}.$$

Furthermore, we denote

$$\mu^* := z^*/\beta. \tag{32}$$

Then, the pair (x^*, y^*, μ^*) satisfies the following equations:

$$Px^* = \beta A^\top \mu^* - f, \tag{33a}$$
$$Qy^* = \beta B^\top \mu^* - g, \tag{33b}$$
$$Ax^* + By^* - b = \mathbf{0}. \tag{33c}$$

In the following, we present another equivalent form of the KKT condition of the problem (6) represented by x^* and (y^*, μ^*) separately. This form helps us better reveal the relationship between a fixed point of the iterative matrix given in (26) and the KKT point of (6). We first prove a lemma that turns out to be essential for the convergence analysis to be presented.

Lemma 4.1 *Suppose that Assumptions 1–2 hold. Let $\gamma \neq 0$. Then, the pair (x^*, y^*, z^*) is a KKT point of (6) if and only if it satisfies the following equations:*

$$x^* = \hat{P}^{-1}\left[A^\top \mu^* - A^\top B y^* + A^\top b - f/\beta\right] \ \ \textit{and} \ \ \left[I - T(\gamma)\right] v^* = q, \tag{34}$$

where $(v^*)^\top := ((y^*)^\top, (\mu^*)^\top)$ and the vectors μ^*, q, the matrix $T(\gamma)$ are defined in (32), (28) and (27), respectively.

Proof Recall that the pair (x^*, y^*, z^*) is a KKT point of (6) if and only if it satisfies (33). First, we show that Eq. (34) holds when (x^*, y^*, μ^*) satisfies (33). For this purpose, we multiply both sides of (33c) by $\beta A^\top$ from the left and add the resulting equation to (33a). This manipulation yields the equation:

$$(P + \beta A^\top A)x^* = \beta(A^\top \mu^* - A^\top B y^* + A^\top b - f/\beta).$$

Dividing both sides of the above equation by β and using the definition of $\hat{P}$ in (23) and then multiplying it by $\hat{P}^{-1}$ from the left, we obtain

$$x^* = \hat{P}^{-1}\left[A^\top \mu^* - A^\top B y^* + A^\top b - f/\beta\right]. \tag{35}$$

Next, we multiply (33c) by $\beta B^\top$ from the left and adding the resulting equation to (33b), which is further divided in both sides by β. These operations enable us to have

$$(Q/\beta + B^\top B)y^* = B^\top \mu^* - B^\top Ax^* + B^\top b - g/\beta. \tag{36}$$

Then, substituting (35) into the above equality and recalling the definitions of $\hat{Q}$ in (23) and α_1 in (24), we get

$$\hat{Q}y^* = B^\top A\hat{P}^{-1}A^\top By^* + (B^\top - B^\top A\hat{P}^{-1}A^\top)\mu^* + \alpha_1. \tag{37}$$

On the other hand, it follows from (35) and the definition of α_2 in (25) that (33c) can be reformulated as

$$\begin{aligned}\mu^* = \mu^* - \gamma(Ax^* + By^* - b) = \gamma A\hat{P}^{-1}A^\top By^* - \gamma By^* \\ +(I - \gamma A\hat{P}^{-1}A^\top)\mu^* + \alpha_2.\end{aligned} \tag{38}$$

Combining (37) with the equation above, we have

$$\begin{pmatrix} \hat{Q} & \mathbf{0} \\ \gamma B & I \end{pmatrix} v^* = \begin{pmatrix} B^\top A\hat{P}^{-1}A^\top B & B^\top - B^\top A\hat{P}^{-1}A^\top \\ \gamma A\hat{P}^{-1}A^\top B & I - \gamma A\hat{P}^{-1}A^\top \end{pmatrix} v^* + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}. \tag{39}$$

Moreover, with simple calculations and recalling the definitions of $T(\gamma)$ in (27) and q in (28), we get

$$v^* = T(\gamma)v^* + q. \tag{40}$$

Then, combining (40) and (35), the assertion (34) is proved.

Next, we verify the assertion of the other direction. That is, if (34) holds with one $\gamma \neq 0$, then (33) is true. Since (40) holds, we know that (39) is true. Furthermore, we get (37) and the second equality in (38) because of (39). Substituting (35) into the second equality of (38), we prove the first equality in (38). Also, substituting (35) into (37), we get (36). Recall the definition $\hat{P}$ in (23). Then, we have

$$\begin{aligned}(P/\beta + A^\top A)x^* &= A^\top \mu^* - A^\top By^* + A^\top b - f/\beta, \\ (Q/\beta + B^\top B)y^* &= B^\top \mu^* - B^\top Ax^* + B^\top b - g/\beta, \\ \mu^* &= \mu^* - \gamma(Ax^* + By^* - b).\end{aligned}$$

Because of $\gamma \neq 0$, it is equivalent to

$$\begin{aligned}(P/\beta + A^\top A)x^* &= A^\top \mu^* - A^\top By^* + A^\top b - f/\beta, \\ (Q/\beta + B^\top B)y^* &= B^\top \mu^* - B^\top Ax^* + B^\top b - g/\beta, \\ Ax^* + By^* - b &= 0.\end{aligned}$$

Then, substituting the last equality into the first and second equations of the above system, we obtain (33). Thus, the conclusion of this lemma follows directly and the proof is complete. □

Remark 4.1 For the "only if" part, the equations in (34) hold for any γ. For the "if" part, if there exists one $\gamma \neq 0$ such that (34) holds, then the pair (x^*, y^*, μ^*) satisfies (33). That is, the pair (x^*, y^*, z^*) is a KKT point of (6).

4.3 Divergence for $\gamma = 2$

In [19], we have shown that the ALM (5) is not necessarily convergent if $\gamma = 2$. Hence, it is intuitive to assert that the convergence of the ADMM scheme (3), as an inexact version of the ALM (5), is not ensured with $\gamma = 2$ for the generic case (1), either. Before we prove the convergence for the scheme (19) with $\gamma \in]0, 2[$, we construct an example to show that the convergence of (19) with $\gamma = 2$ is not guaranteed. Hence, we just need to focus on $\gamma \in]0, 2[$ for the discussion. That is, the range $\gamma \in]0, 2[$ is optimal and further enlargement of this range is not practical when the convergence of the ADMM scheme (3) is discussed.

More specifically, let us take

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} 0.4 & 0.3 \\ 0.5 & 2.2 \end{pmatrix}, \quad B = \begin{pmatrix} 1.2 & -0.2 \\ 1.6 & 0.1 \end{pmatrix},$$

and $f = g = b = \mathbf{0}_{2\times 1}$. This is a special case of the quadratic programming (6), and it has a unique solution $x = y = \mathbf{0}_{2\times 1}$.

With $\gamma = 2$ and $\beta = 1$, the iterative matrix in (27) is specified as

$$T(2) = \begin{pmatrix} 0.7897 & 0.0267 & 0.2142 & -0.0292 \\ 0.1610 & 0.0111 & -0.1639 & 0.0224 \\ -1.1706 & -0.0810 & 0.1923 & -0.1626 \\ 0.8780 & 0.0608 & -0.8942 & -0.8781 \end{pmatrix}.$$

By elementary calculations, we have $\rho(T(2)) = 1$ and one of its eigenvalues is -1. Suppose η is the eigenvector corresponding to the eigenvalue -1. Let the sequence $\{v^k\}$ be generated by (26) with the starting point $v^0 := \eta$. Then, it is easy to verify that the sequence $\{v^k\}$ is 2-periodic satisfying

$$v^k = \begin{cases} \eta, & \text{if } k \text{ is even,} \\ -\eta, & \text{if } k \text{ is odd.} \end{cases}$$

Hence, the convergence of the scheme (19) with $\gamma = 2$ is not guaranteed.

5 Convergence Analysis

In this section, we establish the convergence of the scheme (19) with $\gamma \in]0, 2[$. The analysis still relies on the spectral analysis for the corresponding iterative matrix $T(\gamma)$

defined in (27). But we would emphasize that our analysis is different from current approaches in the literature which are based on the strict contraction property of certain distance function to the solution set measured by matrix norms with positive-definite or positive-semidefinite matrices (e.g., [10,13,14,35]) or the non-expansiveness property of certain maximal monotone operator (e.g., [9,16,36]). Indeed, it is easy to show that the so-called strict contraction property in these mentioned work does not hold for the case where $\gamma \geq \frac{1+\sqrt{5}}{2}$ and hence it is difficult to apply directly these existing techniques to establish the convergence of the ADMM scheme (3) with $\gamma \geq \frac{1+\sqrt{5}}{2}$. This may be explained as a difficulty of answering Glowinski's open question under consideration.

5.1 Theoretical Analysis

Even for the specific quadratic programming problem (6), the resulting iterative matrix $T(\gamma)$ defined in (27) is complicated at least in the following senses. (1) It is nonsymmetric; hence very few analytical tools are available for the spectral radius analysis. (2) It may have complex eigenvalues and eigenvectors. (3) The penalty parameter β is coupled in the iterative matrix. All these problems prohibit us to apply typical spectral analysis techniques to this challenging case. Hence, the spectral analysis is more complicated than the typical case of $\gamma = 1$. This is also why in Sect. 4 we suggest first eliminating the variable x from the matrix recursion form and obtaining a non-homogeneous matrix recurrence in a lower-dimension space. Then, some operations such as a matrix transform with the same eigenvalues (accounting for multiplicity) should be conducted to achieve a block-structure matrix so that a spectral analysis can be applied.

In what follows, we shall verify that the eigenvalue $\lambda_{T(\gamma)}$ of the iterative matrix $T(\gamma)$ defined in (27) is satisfied with $|\lambda_{T(\gamma)}| < 1$ or $\lambda_{T(\gamma)} = 1$; and if 1 is an eigenvalue of the iterative matrix, then it has a complete set of eigenvectors. We first define a matrix and study its eigenstructure before investigating the spectral radius analysis for the iterative matrix $T(\gamma)$ in (27).

Theorem 5.1 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. For $\gamma \in]0, 2[$, let us define*

$$M(\gamma) = \begin{pmatrix} GF & G - GF \\ \gamma F - \gamma GF & I - \gamma F - \gamma G + \gamma GF \end{pmatrix}. \tag{41}$$

Then, for any eigenvalue of $M(\gamma)$, denoted by λ, we have $|\lambda| < 1$ or $\lambda = 1$.

Proof Note that the matrix $M(\gamma)$ can be factorized as:

$$M(\gamma) = \begin{pmatrix} G & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} F & I - F \\ \gamma(I - G)F & I - \gamma F - \gamma G + \gamma GF \end{pmatrix}.$$

Switching the order of the products by moving the first component to the last in $M(\gamma)$, we obtain the matrix, denoted by $M'(\gamma)$, as follows:

$$\begin{aligned}M'(\gamma) &= \begin{pmatrix} F & I-F \\ \gamma(I-G)F & I-\gamma F-\gamma G+\gamma GF \end{pmatrix}\begin{pmatrix} G & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \\ &= \begin{pmatrix} FG & I-F \\ \gamma(I-G)FG & I-\gamma F-\gamma G+\gamma GF \end{pmatrix}.\end{aligned}$$

Clearly, $\mathrm{eig}(M(\gamma)) = \mathrm{eig}(M'(\gamma))$. Let λ be any eigenvalue of the matrix $M(\gamma)$. Then, it is also an eigenvalue of the matrix $M'(\gamma)$. Let $(u^\top, w^\top)^\top$ be an eigenvector of $M'(\gamma)$ associated with λ. Then, we have

$$\begin{pmatrix} FG & I-F \\ \gamma(I-G)FG & I-\gamma F-\gamma G+\gamma GF \end{pmatrix}\begin{pmatrix} u \\ w \end{pmatrix} = \lambda \begin{pmatrix} u \\ w \end{pmatrix}.$$

This is equivalent to

$$FGu + (I-F)w = \lambda u, \tag{42a}$$

$$\gamma(I-G)FGu + \gamma(I-G)(I-F)w + (1-\gamma)w = \lambda w. \tag{42b}$$

Multiplying both sides in (42a) by $\gamma(I-G)$ from the left and then subtracting it by (42b), we get

$$[\lambda - 1 + \gamma]w = \gamma\lambda(I-G)u.$$

If $\lambda - 1 + \gamma = 0$, then $|\lambda| = |\gamma - 1| < 1$ because of $\gamma \in]0, 2[$. The assertion is proved.

In the following, we assume that $\lambda - 1 + \gamma \neq 0$. Dividing both sides of the above equation by $(\lambda - 1 + \gamma)$ and then substituting it into (42a), we obtain

$$FGu + (I-F)\frac{\lambda\gamma}{\lambda+\gamma-1}(I-G)u = \lambda u. \tag{43}$$

Without loss of generality, we assume that $\|u\|_2 = 1$. Note that λ might be complex number. Thus, the associated eigenvector $(u^\top, w^\top)^\top$ might be a complex vector. Let us define two constants as follows:

$$\xi_1 := u^H FGu \quad \text{and} \quad \xi_2 = u^H(I-F)(I-G)u.$$

Then, multiplying both sides of (43) by $(\lambda + \gamma - 1)u^H$ from the left yields

$$\lambda^2 + (\gamma - 1 - \xi_1 - \gamma\xi_2)\lambda + (1-\gamma)\xi_1 = 0. \tag{44}$$

For convenience, we define

$$f(\lambda) = \lambda^2 + (\gamma - 1 - \xi_1 - \gamma\xi_2)\lambda + (1-\gamma)\xi_1. \tag{45}$$

Note that $f(\lambda) = 0$ is a quadratic equation with one variable λ. Thus, we define

$$\Delta := (\gamma - 1 - \xi_1 - \gamma\xi_2)^2 - 4(1-\gamma)\xi_1.$$

The remaining part of the proof should be divided into two cases.

Case 1. λ is a real eigenvalue. The corresponding eigenvector $(u^\top, w^\top)^\top$ is also real. Thus, both ξ_1 and ξ_2 are real numbers. This means that the equation $f(\lambda) = 0$ has real coefficients. First, invoking Lemma 3.4, we have

$$0 \le \xi_1 + \xi_2 \le 1.$$

Thus, $f(1) = \gamma - \gamma(\xi_1 + \xi_2) \ge 0$. The following discussion is divided into three cases.

I) $\gamma \in]0, 1[$.
Note that $f(-1) = (2-\gamma)(1+\xi_1) + \gamma\xi_2 > 0$. If $\xi_1 < 0$, then $f(0) < 0$. It implies that one of the roots of $f(\lambda) = 0$ belongs to $]0, 1]$, and the other belongs to $]-1, 0[$. If $\xi_1 \ge 0$, then $f(0) \ge 0$. It implies that the two roots of $f(\lambda) = 0$ have the same sign. Moreover, we have $|(1-\gamma)\xi_1| < 1$ and $|\gamma - 1 - \xi_1 - \gamma\xi_2| \le 2$. We conclude that Eq. (44) has two real roots, and both of them belong to either $[0, 1]$ or $]-1, 0]$.

II) $\gamma = 1$.
The equation $f(\lambda) = 0$ has two roots: $\lambda_1 = 0$ and $\lambda_2 = \xi_1 + \xi_2$. Invoking Lemma 3.4, we have $0 \le \lambda_2 \le 1$.

III) $\gamma \in]1, 2[$.
If $\xi_1 \le 0$, then we have $\xi_2 \ge 0$ because of $\xi_1 + \xi_2 \ge 0$. Thus, we have $f(-1) > 0$, $f(0) \ge 0$ and $f(1) \ge 0$. Also, we have $|(1-\gamma)\xi_1| < 1$ and $|\gamma - 1 - \xi_1 - \gamma\xi_2| < 2$. We conclude that Eq. (44) has two real roots, and both of them belong to either $[0, 1]$ or $]-1, 0]$.
If $\xi_1 > 0$, then $f(1) \ge 0$ and $f(0) < 0$. Note $f(-1) = 2-\gamma+2\xi_1-\gamma\xi_1+\gamma\xi_2$. If $f(-1) > 0$, then we know that one of the roots of $f(\lambda) = 0$ belongs to $]0, 1]$, and the other belongs to $]-1, 0[$. If $f(-1) \le 0$, it implies that Eq. (44) has a root $\lambda_2 \le -1$. In the following, we show that λ_2 is an extraneous root by contradiction. Without loss of generality, we assume that $F \succ \mathbf{0}$ and $G \succ \mathbf{0}$. If λ_2 is not an extraneous root, then it is an eigenvalue of (41). It follows from (43) that λ_2 is a root of the following equation:

$$\det\left(FG + (I-F)\frac{\lambda\gamma}{\lambda+\gamma-1}(I-G) - \lambda I\right) = 0. \tag{46}$$

We denote

$$\kappa := \frac{\lambda\gamma}{\lambda+\gamma-1}. \tag{47}$$

Since $\lambda \leq -1$ and $\gamma \in]1, 2[$, $\kappa > 0$ and the matrix $((\kappa - \lambda)I - \kappa G)$ is nonsingular. Thus, it follows from (46) that

$$\det(F) \cdot \det\left([(1+\kappa)G - \kappa I]\,[(\kappa - \lambda)I - \kappa G]^{-1} + F^{-1}\right) \cdot \det\left((\kappa - \lambda)I - \kappa G\right) = 0. \tag{48}$$

Note that $\left([(1+\kappa)G - \kappa I]\,[(\kappa - \lambda)I - \kappa G]^{-1}\right)$ is a real symmetric matrix. Denote the eigenvalues of F and G by $1 \geq f_1 \geq f_2 \geq \cdots \geq f_m > 0$ and $1 \geq g_1 \geq g_2 \geq \cdots \geq g_m > 0$, respectively. Then, using Lemma 3.5, we get

$$\begin{aligned} q(\lambda) &:= \det(F) \cdot \det\left([(1+\kappa)G - \kappa I]\,[(\kappa - \lambda)I - \kappa G]^{-1} + F^{-1}\right) \\ &\quad \cdot \det\left((\kappa - \lambda)I - \kappa G\right) \\ &\geq \left(\prod_{i=1}^{m} f_i\right)\left\{\min_{\pi} \prod_{i=1}^{m}\left(\frac{(1+\kappa)g_i - \kappa}{(\kappa - \lambda) - \kappa g_i} + \frac{1}{f_{\pi(i)}}\right)\right\}\left(\prod_{i=1}^{m}[(\kappa - \lambda) - \kappa g_i]\right) \\ &\geq \left(\prod_{i=1}^{m} f_i\right)\left\{\prod_{i=1}^{m}\left(\frac{(1+\kappa)g_i - \kappa}{(\kappa - \lambda) - \kappa g_i} + \frac{1}{f_{m+1-i}}\right)\right\}\left(\prod_{i=1}^{m}[(\kappa - \lambda) - \kappa g_i]\right) \\ &\geq \prod_{i=1}^{m}\left\{-\frac{1}{\lambda + \gamma - 1}\left[\lambda^2 + \lambda\left(-1 + \gamma g_i - (1+\gamma) g_i f_{m+1-i}\right.\right.\right. \\ &\quad \left.\left.\left. + \gamma f_{m+1-i}\right) + (1-\gamma) f_{m+1-i} g_i\right]\right\}. \end{aligned} \tag{49}$$

The second inequality is due to the fact that the function $h(x) = \frac{(1+\kappa)x - \kappa}{(\kappa - \lambda) - \kappa x}$ is an increasing function with respect to x and Lemma 3.5, and the last follows from (47). If we denote

$$\begin{aligned} q_i(\lambda) := \lambda^2 + \lambda\left(-1 + \gamma g_i - (1+\gamma) g_i f_{m+1-i} + \gamma f_{m+1-i}\right) \\ + (1-\gamma) f_{m+1-i} g_i, \end{aligned}$$

then we have

$$q(\lambda) \geq \prod_{i=1}^{m}\left[-\frac{1}{\lambda + \gamma - 1} q_i(\lambda)\right]. \tag{50}$$

Indeed, for any $i = 1, \ldots, m$, we have

$$\begin{aligned} q_i(1) = \gamma(g_i + f_{m+1-i} - 2 f_{m+1-i} g_i) \geq 0,\ q_i(0) \\ = (1-\gamma) g_i f_{m+1-i} < 0, \end{aligned} \tag{51}$$

and

$$\begin{aligned} q_i(-1) = 2 + 2 f_{m+1-i} g_i - \gamma f_{m+1-i} - \gamma g_i > 2 \\ + 2 f_{m+1-i} g_i - 2 f_{m+1-i} - 2 g_i \geq 0. \end{aligned} \tag{52}$$

Combining (51) and (52), we have

$$q_i(\lambda) > 0 \text{ when } \lambda \leq -1, \quad \forall i = 1, \ldots, m. \tag{53}$$

Also note that

$$-\frac{1}{\lambda + \gamma - 1} > 0 \text{ when } \lambda \leq -1 \text{ and } \gamma \in]1, 2[. \tag{54}$$

Combining (50), (53) and (54), we get $q(\lambda) > 0$, when $\lambda \leq -1$ and $\gamma \in]1, 2[$.

Recalling the definition of $q(\lambda)$ in (49), the above result is contradicted with (46). Therefore, we verify that λ_2 is an extraneous root when $F \succ \mathbf{0}$ and $G \succ \mathbf{0}$. Finally, if $F \succeq \mathbf{0}$ and $G \succeq \mathbf{0}$, we can take two positive definite matrix sequences $\{F_n\}$ and $\{G_n\}$ which converge to F and G in the Frobenius norm, respectively. Then, using the fact that the eigenvalue of a matrix is continuous with respect to the matrix's entries (see, e.g., [37]), we can also show that the eigenvalues of the matrix $M(\gamma)$ are not less than or equal to -1.

Case 2. λ is a complex eigenvalue of the matrix $M(\gamma)$. Indeed, we only need to focus on $\gamma > 1$. For this case, the matrix $M(\gamma)$ contains only real numbers and thus its complex eigenvalues must occur in conjugate pairs (see, e.g., [37]). If the equation $f(\lambda) = 0$ still has real coefficients, then it has a pair of conjugate complex roots, i.e., λ and $\bar{\lambda}$. Then, $\lambda \cdot \bar{\lambda} = (1 - \gamma)\xi_1 < 1$ which is due to Lemma 3.2. Thus, we have $|\lambda| < 1$. If the equation $f(\lambda) = 0$ has complex coefficients, it implies that λ must be complex. For this case, the equation (48) still holds. Consequently,

$$\det\left([(1+\kappa)G - \kappa I][(\kappa - \lambda)I - \kappa G]^{-1} + F^{-1}\right) = 0.$$

Since G is symmetric, there exists an orthogonal matrix S such that $G = S^{-1}\Lambda S$ where $\Lambda := \text{diag}(g_1, \ldots, g_m)$. Then, we get

$$\det\left([(1+\kappa)\Lambda - \kappa I][(\kappa - \lambda)I - \kappa\Lambda]^{-1} + \tilde{F}^{-1}\right) = 0, \tag{55}$$

with $\tilde{F} = SFS^{-1}$. We denote

$$\Lambda_1 + \Lambda_2 i := [(1+\kappa)\Lambda - \kappa I][(\kappa - \lambda)I - \kappa\Lambda]^{-1},$$

where both of Λ_1 and Λ_2 are real diagonal matrices. It follows from (55), we get

$$\left(\Lambda_1 + \tilde{F}^{-1}\right) \not\succ \mathbf{0}.$$

Thus, we have $\lambda_m(\Lambda_1 + \tilde{F}^{-1}) \leq 0$, where $\lambda_m(\cdot)$ denotes the smallest eigenvalue. Invoking Theorem III.2.1 of [38], we have $0 \geq \lambda_m(\Lambda_1 + \tilde{F}^{-1}) \geq$

$\lambda_m(\Lambda_1) + \lambda_m(\tilde{F}^{-1})$. In the following, we represent the complex eigenvalue λ of the matrix $M(\gamma)$ as $a + bi$. With simple calculations, we get $\Lambda_1 := \text{diag}(\Delta(g_1), \ldots, \Delta(g_m))$, where

$$\Delta(g) := \frac{\Delta_1(g)}{\Delta_2(g)},$$

with

$$\begin{aligned}\Delta_1(g) &:= ((a\gamma + a + \gamma - 1)g - a\gamma)(-a^2 + a + b^2 - ag\gamma) \\ &\quad + ((b\gamma + b)g - b\gamma)(-2ab + b - b\gamma g), \\ \Delta_2(g) &:= (b^2 - a^2 - ag\gamma + a)^2 + (-2ab + b - b\gamma g)^2.\end{aligned}$$

Note that $\Delta_2(g) > 0$ because λ is complex. Consequently, we obtain that

$$\min_{g\in[0,1]} \left(\Delta(g) + \lambda_m(\tilde{F}^{-1})\right) \leq 0.$$

Recalling $\lambda_m(\tilde{F}^{-1}) \geq 1$, we obtain that

$$\min_{g\in[0,1]} (\Delta(g) + 1) \leq 0.$$

which is equivalent to the following minimization problem:

$$\min_{g\in[0,1]} (\Delta_1(g) + \Delta_2(g)) \leq 0. \tag{56}$$

By some elementary calculations, we know that

$$\Delta_1(g) + \Delta_2(g) := \delta_2 g^2 + \delta_1 g + \delta_0 \text{ with } \delta_2 = -|\lambda|^2\gamma - a\gamma(\gamma - 1).$$

If $\delta_2 \geq 0$, we can easily show that $|\lambda| < 1$. If $\delta_2 < 0$, we see that the minimizer of (56) can be attained only at the end points of $[0, 1]$. If it is attained at $g = 0$, it implies that

$$\text{Re}\left(\frac{-\kappa}{\kappa - \lambda}\right) + 1 \leq 0.$$

Then, we can get $\gamma(1 - a) \geq (a - 1)^2 + b^2$. Obviously, $a < 1$. Furthermore, we get that $\gamma - 1 + (2 - \gamma)a \geq a^2 + b^2$. Then, it follows that $-1 < a < 1$. Thus, we obtain that $|\lambda| < 1$. On the other hand, if the minimizer of (56) is attained at $g = 1$, it means that

$$\text{Re}\left(\frac{1}{-\lambda}\right) + 1 \leq 0.$$

Thus, we have $\frac{a}{a^2+b^2} \geq 1$. Furthermore, we obtain that $0 < a < 1$. Thus, combining these two inequalities, we get $|\lambda| < 1$.

The proof is complete. □

Remark 5.1 From the proof of Theorem 5.1, we know that if 1 is an eigenvalue of $M(\gamma)$, then $f(1) = 0$ where $f(\lambda)$ is defined in (45). It implies that $\xi_1 + \xi_2 = 1$. That is, 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$.

Theorem 5.1 enables us to study the spectral property of the matrix $M(\gamma)$. Moreover, it will be shown in the following theorem that if 1 is an eigenvalue of $M(\gamma)$, then it has a complete set of eigenvectors. The proof is partially inspired by Lemma 6 in [32].

Theorem 5.2 *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ satisfying $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $M(\gamma)$ defined in (41), then the algebraic multiplicity of 1 for $M(\gamma)$ equals its geometric multiplicity.*

Proof It follows from the definition of $M(\gamma)$ in (41) that

$$\begin{aligned}
&\det(\lambda I - M(\gamma)) \\
&\quad = (-1)^m \det \begin{pmatrix} GF - G - \frac{1}{\lambda\gamma+\lambda-1}(\lambda I - G)\left[(\lambda-1)I + \gamma F\right] & \lambda I - G \\ \mathbf{0} & (\lambda\gamma + \lambda - 1)I \end{pmatrix} \\
&\quad = \det\left[\lambda^2 I - \lambda I + \gamma\,(F + G - 2GF)\,\lambda + (\gamma - 1)GF(\lambda - 1)\right]. \qquad (57)
\end{aligned}$$

Since 1 is an eigenvalue of the matrix $M(\gamma)$, it is also an eigenvalue of the matrix $(I - F - G + 2FG)$. Then, invoking Lemma 3.12, we know that there exists an orthogonal matrix $Q \in \mathbb{R}^{m\times m}$ such that (17) holds. Consequently, we have

$$Q^\top(F + G - 2GF)Q = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{F} + \hat{G} - 2\hat{G}\hat{F} \end{pmatrix} \quad \text{and} \quad Q^\top(GF)Q = \begin{pmatrix} D^2 & \mathbf{0} \\ \mathbf{0} & \hat{G}\hat{F} \end{pmatrix},$$

in which the first equality is due to $2D - 2D^2 = \mathbf{0}$. Then, it yields

$$\begin{aligned}
&Q^\top\left(\lambda^2 I - \lambda I + \gamma\,(F + G - 2GF)\,\lambda + (\gamma - 1)GF(\lambda - 1)\right)Q \\
&\quad = \begin{pmatrix} \lambda(\lambda - 1)I_\ell + (\gamma - 1)D^2(\lambda - 1) & \mathbf{0} \\ \mathbf{0} & (\lambda^2 - \lambda)I_{m-\ell} + \gamma(\hat{F} + \hat{G} - 2\hat{G}\hat{F})\lambda + (\gamma - 1)\hat{G}\hat{F}(\lambda - 1) \end{pmatrix}.
\end{aligned}$$

Taking the determinant on both sides in the above equation, we get

$$\begin{aligned}
&\det\left[\lambda(\lambda - 1)I + \gamma\,(F + G - 2GF)\,\lambda + (\gamma - 1)GF(\lambda - 1)\right] \\
&\quad = (\lambda - 1)^\ell \left[\Pi_{i=1}^{\ell}\left(\lambda + (\gamma - 1)D_{ii}^2\right)\right] q(\lambda)
\end{aligned}$$

with $q(\lambda) := \det\left[(\lambda^2 - \lambda)I_{m-\ell} + \gamma\left(\hat{F} + \hat{G} - 2\hat{G}\hat{F}\right)\lambda + (\gamma - 1)\hat{G}\hat{F}(\lambda - 1)\right]$. Invoking Lemma 3.12, we have

$$\mathbf{0} \prec \hat{F} + \hat{G} - 2\hat{G}\hat{F} \preceq I_{m-\ell}. \qquad (58)$$

We can actually conclude that $(\lambda - 1) \nmid q(\lambda)$. Let us prove it by contradiction. If $(\lambda - 1) | q(\lambda)$, then it implies that $q(1) = 0$, i.e., $\det\left[\gamma\left(\hat{F} + \hat{G} - 2\hat{G}\hat{F}\right)\right] = 0$. This contradicts with (58). Moreover, note that

$$\lambda + (\gamma - 1)D_{ii}^2 = \begin{cases} \lambda, & \text{if } D_{ii} = 0, \\ \lambda + \gamma - 1, & \text{if } D_{ii} = 1. \end{cases}$$

Therefore, $(\lambda - 1) \nmid \left\{[\Pi_{i=1}^{\ell}(\lambda + (\gamma - 1)D_{ii}^2)]q(\lambda)\right\}$. It implies that the algebraic multiplicity of 1 for $M(\gamma)$ is ℓ defined in (14). On the other hand, the geometric multiplicity of 1 for $M(\gamma)$ is identical with the following quality:

$$\begin{aligned} &2m - \text{Rank}\begin{pmatrix} I - GF & -G + GF \\ -\gamma F + \gamma GF & \gamma(F + G - GF) \end{pmatrix} \\ &= 2m - \text{Rank}\begin{pmatrix} I - GF & -G + GF \\ -F + GF & F + G - GF \end{pmatrix} \\ &= 2m - \text{Rank}\begin{pmatrix} I - GF & -G + GF \\ I - F & F \end{pmatrix} = 2m - \text{Rank}\begin{pmatrix} I - G & -G + GF \\ I & F \end{pmatrix} \\ &= 2m - \text{Rank}\begin{pmatrix} -G + GF & I - G \\ F & I \end{pmatrix} = 2m - \text{Rank}\begin{pmatrix} -G - F + 2GF & I - G \\ \mathbf{0} & I \end{pmatrix} \\ &= m - \text{Rank}(-G - F + 2GF). \end{aligned}$$

Invoking Lemma 3.6, we conclude that the geometric multiplicity of eigenvalue 1 for $M(\gamma)$ is also ℓ. The proof is complete. □

Remark 5.2 Note that, if 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, then 0 is an eigenvalue of the matrix $(F + G - 2GF)$ because of Lemma 3.7. From the proof of Theorem 5.2 (see (57)), we know that 1 is an eigenvalue of $M(\gamma)$ if 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$. Therefore, because of Remark 5.1, we know that 1 is an eigenvalue of $M(\gamma)$ if and only if 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$.

Now, we proceed to the spectral analysis for the iterative matrix $T(\gamma)$ defined in (29). This is the essential pillar for proving the convergence of the scheme (19) with $\gamma \in]0, 2[$. A lemma is proved first.

Lemma 5.1 *Assumptions 1–2 hold; the matrices $\hat{Q}$ and $\hat{P}$ are defined in (23); $\gamma \in]0, 2[$; the matrix $T(\gamma)$ is defined in (27). Then, we have $|\lambda_{T(\gamma)}| < 1$ or $\lambda_{T(\gamma)} = 1$. Furthermore, if 1 is an eigenvalue of $T(\gamma)$, then it is complete.*

Proof Setting $G = B\hat{Q}^{-1}B^\top$ and $F = A\hat{P}^{-1}A^\top$ in $M(\gamma)$ (see (41)), and invoking the definition of $\tilde{T}(\gamma)$ in (30), we have

$$M(\gamma) = \tilde{T}(\gamma).$$

Thus, it holds $\text{eig}(M(\gamma)) = \text{eig}(\tilde{T}(\gamma))$. Indeed, according to (31) , we have

$$\overline{\text{eig}}(T(\gamma)) = \overline{\text{eig}}(M(\gamma)). \tag{59}$$

It follows from Theorem 5.1 that $|\lambda_{T(\gamma)}| < 1$ or $\lambda_{T(\gamma)} = 1$.

If 1 is an eigenvalue of the matrix $T(\gamma)$, using Theorem 5.2 and (59), we know that it is also an eigenvalue of $M(\gamma)$ and its algebraic and geometric multiplicities are the same for $T(\gamma)$. The proof is complete. □

Now, we are at the stage to prove the convergence of the scheme (19) with $\gamma \in]0, 2[$.

Theorem 5.3 *Assumptions 1–2 hold. Let $\{(x^k, y^k, z^k)\}$ be the sequence generated by the scheme (19), i.e., the application of the ADMM scheme (3) with $\gamma \in]0, 2[$ to the quadratic programming problem (6). Then, the sequence $\{(x^k, y^k, z^k)\}$ converges to a KKT point of (6).*

Proof The proof is similar as Theorem 3 in [32]; thus we omit it here. □

5.2 Numerical Verification of the Convergence

In this section, we construct a simple particular case of problem (6) and verify numerically the convergence of (19) with $\gamma \in]0, 2[$. In particular, as well observed in the literature, e.g., [9–15], it is advantageous to employ larger value of γ closer to 2 to accelerate the convergence in the scheme (20). The codes were written by MATLAB 7.8 (R2009a) and were run on a X1 Carbon notebook with the Intel Core i7-4600U CPU at 2.1 GHz and 8 GB of memory.

Let us set $f = g = b = \mathbf{0}$ in (6); the resulting problem has a unique solution $x^* = y^* = \mathbf{0}$. The matrix P and Q in (6) are generated by

$$\begin{aligned} &P1 = randn(n_1, n_1);\;\; P = P1' * P1;\;\; a = eigs(P, 1, 'sm');\\ &\;P = P - (a - (1e-4)) * eye(n_1) \end{aligned}$$

and

$$\begin{aligned} &Q1 = randn(n_2, n_2);\quad Q = Q1' * Q1;\;\; b = eigs(Q, 1, 'sm');\\ &\;Q = Q - (b - (1e-4)) * eye(n_1), \end{aligned}$$

respectively. Furthermore, the matrices $A \in \mathbb{R}^{m\times n_1}$ and $B \in \mathbb{R}^{m\times n_2}$ in (6) are generated independently, and their elements are *i.i.d.* uniformly distributed in the interval $[0, 1]$. Note that both P and Q are symmetric positive-semidefinite matrices; and both P and Q are seriously ill conditioned. To implement the scheme (19), let us fix $\beta = 1$, $y^0 = randn(n_2, 1)$, $z^0 = randn(m, 1)$, and the stopping criterion is (see [2])

$$\texttt{err} := \max\{\|B(y^k - y^{k+1})\|_2, \|z^k - z^{k+1}\|_2\} \le 10^{-6}. \tag{60}$$

We test different scenarios of this example where $m = n_1 = n_2 = 50, 100, 200, 500$, respectively. The parameter γ varies from 0.2 to 1.8 with an equal distance of 0.2. Moreover, the step size proposed by Glowinski $\gamma = 1.618 \approx \frac{\sqrt{5}+1}{2}$ is compared as a benchmark and several values larger than 1.618 are tested, i.e., $\gamma = 1.65, 1.7, 1.75$. In

Table 1, we report the error on x^k (measured by $\|\bar{x}-x^*\|_2$), the error on y^k (measured by $\|\bar{y}-y^*\|_2$), the number of iteration ("Itr.") and the CPU time in seconds ("Time(s)"). Here, $\bar{x}$ and $\bar{y}$ represent the last iterate satisfying the criterion (60). The condition numbers of P and Q ("Cond(P)" and "Cond(Q)" respectively) are also included in Table 1. Results in this table verify the convergence of (19) with $\gamma \in]0, 2[$ and the acceleration with γ close to 2. In particular, it is shown that the values of $\gamma > 1$ significantly speed up the convergence compared to $\gamma = 1$, i.e., the original ADMM, and that some values larger than 1.618 also result in faster convergence considerately. Hence, it is verified that it is worth considering larger values for γ in Glowinski's ADMM scheme (3).

6 Global Linear Convergence

In addition to the main purpose of establishing the convergence of the scheme (19) with $\gamma \in]0, 2[$ and answering Glowinski's open question partially, we will show in this section the global linear convergence of scheme (19) with $\gamma \in]0, 2[$ under a condition. This is a supplementary result to the main convergence result in Section 5.

6.1 Review of Existing Results

The linear convergence of the ADMM (3) with the special case of $\gamma = 1$ has been discussed in the quadratic programming context in [39,40] under different conditions. Let us briefly review them. In [40], the local linear convergence of a generalized version of the ADMM proposed in [16], which reduces to the original ADMM (3) when the parameter is taken as 1, is established for the quadratic programming problem (6) under some local error bound conditions. In [39], the following convex quadratic programming problem is considered:

$$\begin{array}{ll} \min_x & \frac{1}{2}x^\top Qx + c^\top x + g(y) \\ s.t. & Ax = b, \;\; x = y, \end{array} \quad \text{with} \;\; g(y) = \begin{cases} 0, & \text{if } y \geq 0, \\ +\infty, & \text{if } y \ngeq 0. \end{cases} \tag{61}$$

Then, the following ADMM scheme is suggested in [39]:

$$\begin{aligned} x^{k+1} &= \arg\min_{Ax=b} \mathcal{L}'_\beta(x, y^k, z^k), \\ y^{k+1} &= \arg\min \mathcal{L}'_\beta(x^{k+1}, y, z^k), \\ z^{k+1} &= z^k - \beta(x^{k+1} - y^{k+1}), \end{aligned} \tag{62}$$

where

$$\mathcal{L}'_\beta(x, y, z) = \frac{1}{2}x^\top Qx + c^\top x + g(y) - z^\top(x - y) + \frac{\beta}{2}\|x - y\|^2.$$

Note that the equation $Ax = b$ is considered as a constraint in the x-subproblem of (61). Then, based on the typical spectral analysis for a homogeneous linear equation

Table 1 Convergence of the ADMM

m	n_1	n_2	γ	$\lVert\bar{x}-x^*\rVert_2$	$\lVert\bar{y}-y^*\rVert_2$	Itr.	Time(s)	Condition numbers
50	50	50	0.2	5.005e−6	3.835e−6	1361	0.03	Cond(P):1.6358e+6
			0.4	2.274e−6	1.648e−6	728	0.06	Cond(Q):2.0175e+6
			0.6	1.644e−6	1.398e−6	623	0.03	
			0.8	1.216e−6	1.067e−6	473	0.03	
			1	9.557e−7	8.504e−7	396	0.04	
			1.2	7.870e−7	6.766e−7	343	0.02	
			1.4	6.710e−7	5.875e−7	324	0.00	
			1.6	5.921e−7	5.222e−7	277	0.03	
			1.618	5.757e−7	5.006e−7	263	0.03	
			1.65	5.681e−7	4.863e−7	218	0.00	
			1.7	5.659e−7	4.757e−7	221	0.03	
			1.75	5.205e−7	4.114e−7	198	0.02	
			1.8	4.828e−7	4.689e−7	192	0.00	
100	100	100	0.2	3.258e−6	3.457e−6	1946	0.12	Cond(P):3.540e+6
			0.4	1.530e−6	1.755e−6	921	0.06	Cond(Q):3.7540e+6
			0.6	1.084e−6	1.138e−6	688	0.09	
			0.8	7.921e−7	8.697e−7	529	0.08	
			1	6.383e−7	6.742e−7	440	0.06	
			1.2	4.640e−7	5.495e−7	317	0.06	
			1.4	4.421e−7	4.845e−7	366	0.06	
			1.6	3.861e−7	4.268e−7	327	0.03	
			1.618	3.817e−7	4.225e−7	270	0.06	
			1.65	3.795e−7	3.887e−7	227	0.03	
			1.7	3.444e−7	3.834e−7	223	0.03	
			1.75	2.937e−7	3.323e−7	210	0.03	
			1.8	3.427e−7	3.724e−7	234	0.06	
200	200	200	0.2	2.278e−6	2.299e−6	1838	0.28	Cond(P):7.5130e+6
			0.4	1.316e−6	1.241e−6	1027	0.19	Cond(Q):7.9452e+6
			0.6	8.600e−7	7.272e−7	877	0.16	
			0.8	6.435e−7	5.618e−7	644	0.16	
			1	5.094e−7	4.431e−7	604	0.12	
			1.2	4.248e−7	3.887e−7	492	0.12	
			1.4	3.821e−7	3.608e−7	453	0.12	
			1.6	3.674e−7	2.993e−7	380	0.18	
			1.618	3.894e−7	3.221e−7	378	0.12	
			1.65	5.606e−7	4.825e−7	356	0.12	
			1.7	2.360e−7	2.208e−7	288	0.18	
			1.75	3.091e−7	2.706e−7	298	0.12	
			1.8	3.858e−7	3.767e−7	276	0.12	

Table 1 continued

m	n1	n2	γ	$\|\bar{x}-x^*\|_2$	$\|\bar{y}-y^*\|_2$	Itr.	Time(s)	Condition numbers
500	500	500	0.2	1.529e−6	1.597e−6	2396	2.54	Cond(P):2.0165e+7
			0.4	7.664e−7	8.019e−7	1300	1.74	Cond(Q):1.9426e+7
			0.6	5.136e−7	5.423e−7	890	1.59	
			0.8	3.880e−7	4.253e−7	653	1.65	
			1	2.943e−7	2.976e−7	699	1.72	
			1.2	3.960e−7	4.025e−7	691	1.43	
			1.4	2.476e−7	2.606e−7	619	1.47	
			1.6	4.857e−7	5.089e−7	668	1.48	
			1.618	4.977e−7	5.215e−7	662	1.57	
			1.65	5.111e−7	5.356e−7	640	1.90	
			1.7	5.317e−7	5.573e−7	674	1.66	
			1.75	5.565e−7	5.883e−7	657	1.32	
			1.8	5.711e−7	5.987e−7	575	1.40	

characterizing the corresponding matrix recursion form, the local linear convergence is established for the scheme (62). It is worthwhile to mention that the iterative matrix considered in the homogeneous linear equation varies iteratively and as analyzed in [39], four regimes occur. Assuming the convergence (e.g., by results in [2]), the uniqueness of solution, and the strict complementarity condition (See Theorem 6.4 in [39]), it is proved in [39] that the iterative matrices finally become fixed with a spectral radius less than 1, and hence, the local linear convergence is derived therein for (61).

6.2 Global Linear Convergence Under a Tight Condition

In this section, we establish the global linear convergence of the scheme (20) with $\gamma \in]0,2[$ under a new assumption different from those in [39,40]. Recall the iterative matrix $T(\gamma)$ defined in (27). If $\rho(T(\gamma)) < 1$, then the linear convergence of the sequence $\{v^k\}$ generated by (26)–(27) follows immediately. Hence, the new condition to be presented is to ensure the property $\rho(T(\gamma)) < 1$, and we shall show that this condition is tight.

Theorem 6.1 *Assumptions 1–2 hold. Assume that*

$$N(B\hat{Q}^{-1}B^\top - I)\bigcap N(A\hat{P}^{-1}A^\top - I) = \{\mathbf{0}\} \quad \text{and}$$
$$N(B\hat{Q}^{-1}B^\top)\bigcap N(A\hat{P}^{-1}A^\top) = \{\mathbf{0}\}, \tag{63}$$

with $\hat{Q}$ and $\hat{P}$ defined in (23). Then, the sequence $\{(x^k, y^k, z^k)\}$ generated by the scheme (19) with $\gamma \in]0,2[$ converges linearly to a KKT point of (6).

Proof Setting $G = B\hat{Q}^{-1}B^\top$ and $F = A\hat{P}^{-1}A^\top$ in $M(\gamma)$ (see (41)), and combining the proof in Lemma 5.1, we know that $\sigma(T(\gamma)) = \sigma(M(\gamma))$. Invoking Lemma 3.8 and Remark 5.2, we have

$$1 \notin \sigma(M(\gamma)) \Leftrightarrow 1 \notin \sigma(I + 2GF - F - G) \Leftrightarrow (63).$$

Consequently, $\rho(M(\gamma)) < 1$ when (63) is satisfied. Then, there is a matrix norm $\|\cdot\|_G$ such that $\rho(M(\gamma)) \leq \|M(\gamma)\|_G < \rho(M(\gamma)) + \epsilon \leq 1$. Thus, the sequence $\{(y^k, \mu^k)\}$ converges linearly to a point $\{(y^*, \mu^*)\}$. It implies that the sequence $\{(y^k, z^k)\}$ converges linearly to the point $\{(y^*, z^*)\}$ with $z^* = \beta\mu^*$ as well. Define x^* like (35), and recall (21) and (35). We obtain that

$$\|x^{k+1} - x^*\| = \|\hat{P}^{-1}\left[A^\top(\mu^k - \mu^*) - A^\top B(y^k - y^*)\right]\| \leq \|\hat{P}^{-1}\| \left(\|A\|\|\mu^k - \mu^*\| + \|A^\top B\|\|y^k - y^*\|\right).$$

Thus, the sequence $\{x^k\}$ converges linearly and the linear convergence of the sequence $\{(x^k, y^k, z^k)\}$ follows as well. The proof is complete. □

Remark 6.1 The condition (63) for ensuring the linear convergence of (19) is indeed tight. To see it, notice that (63) implies that for either the eigenvalue 1 or 0, the matrices F and G do not have any common eigenvector. If the condition (63) does not hold, this means that there exists at least one common eigenvector associated with either 1 or 0 for the matrices F and G. Hence, 1 is an eigenvalue of the iterative matrix $T(\gamma)$ and this invalidates the linear convergence of the sequence of $\{v^k\}$ defined in (26).

Remark 6.2 Note that $0 \leq \lambda_{B\hat{Q}^{-1}B^\top} \leq 1$ and $0 \leq \lambda_{A\hat{P}^{-1}A^\top} \leq 1$. Thus, it is easy to verify that the conditions $0 < \lambda_{B\hat{Q}^{-1}B^\top} < 1$ and $0 < \lambda_{A\hat{P}^{-1}A^\top} < 1$ suffice to ensure the condition (63) and hence the linear convergence of the sequence $\{(x^k, y^k, z^k)\}$ generated by the scheme (19).

Remark 6.3 The linear convergence rate result in Theorem 6.1 differs from those in [39,40] in the following aspects. (1) The linear convergence rate in Theorem 6.1 is global, while those in [39,40] are local. (2) The condition (63) is different from those in [39,40]; and it is tight. (3) Here we consider the scheme (19) with $\gamma \in]0, 2[$ and the targeted problem is (6), while in [39] only the special case $\gamma = 1$ is considered, and its targeted problem is (61); and in [40] the generalized ADMM in [16] is considered. Moreover, the conditions in (63) depend only on the matrices P, Q, A and B in the problem (6) per se; they do not involve any local information near the solution point such as the local error bound in [40] or the identification of the regimes of the corresponding iterative matrix in [39]. Indeed, the conditions (63) are equivalent to the nonsingularity of the coefficient matrix in the KKT system of the problem (6). Some necessary and sufficient conditions can be found in (See Theorem 3.1 in [41]).

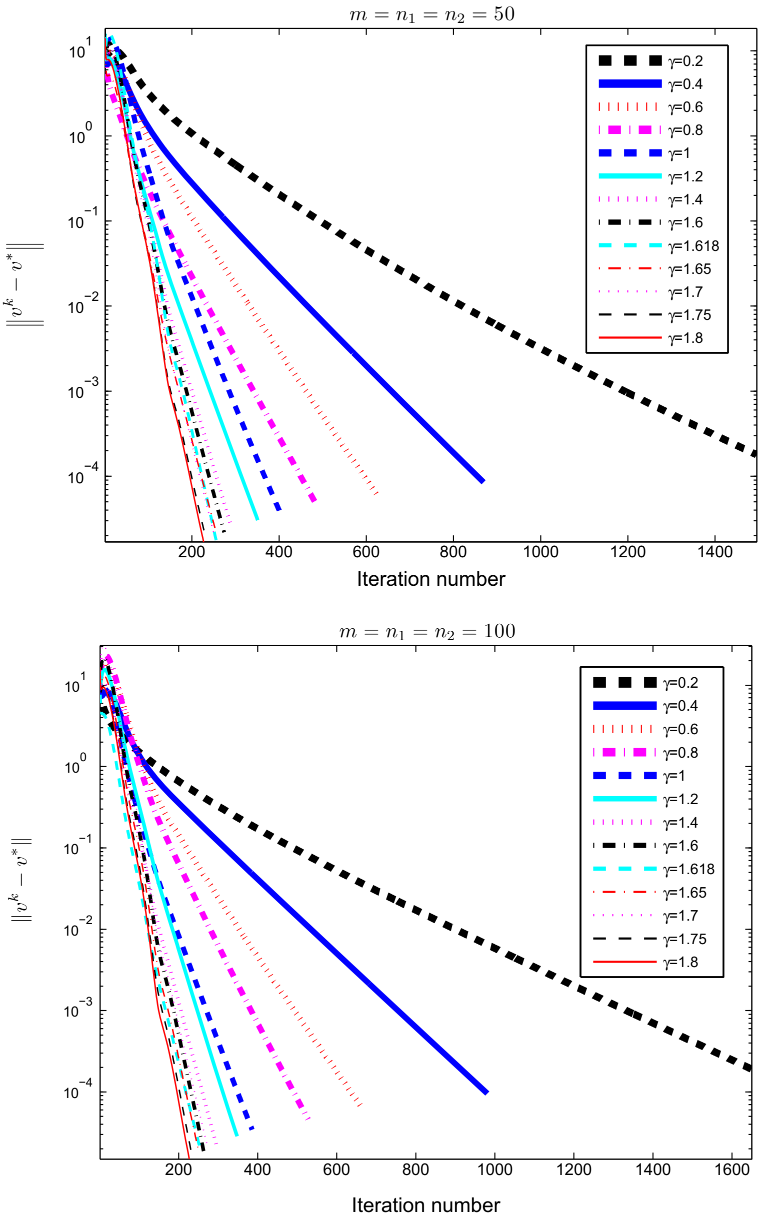

Fig. 1 Linear convergence of the ADMM for different γ

6.3 Numerical Verification of the Global Linear Convergence

In this subsection we verify numerically the global linear convergence of scheme (19) with $\gamma \in]0, 2[$ under the condition (63) by an example.

Table 2 Linear convergence of the ADMM

m	n_1	n_2	γ	$\\|\bar{x}-x^*\\|_2$	$\\|\bar{y}-y^*\\|_2$	Itr.	Time (s)	Condition numbers
50	50	50	$\lambda_{\min}(A\hat{P}^{-1}A^\top)$ =4.9815e−5, $\lambda_{\max}(A\hat{P}^{-1}A^\top) = 0.9995$					
			$\lambda_{\min}(B\hat{Q}^{-1}B^\top)$ =4.7014e−5, $\lambda_{\max}(B\hat{Q}^{-1}B^\top)$ =0.9996					
			0.2	3.376e−6	2.819e−6	1496	0.06	Cond(P):2.2940e+4
			0.4	1.447e−6	1.563e−6	869	0.06	Cond(Q):1.4170e+5
			0.6	1.023e−6	1.066e−6	627	0.03	
			0.8	9.064e−7	6.940e−7	486	0.03	
			1	7.419e−7	6.946e−7	404	0.04	
			1.2	6.272e−7	5.130e−7	351	0.04	
			1.4	5.129e−7	4.357e−7	291	0.04	
			1.6	4.467e−7	3.844e−7	274	0.04	
			1.618	3.949e−7	4.265e−7	256	0.04	
			1.65	4.417e−7	3.683e−7	256	0.04	
			1.7	4.301e−7	3.649e−7	251	0.04	
			1.75	4.230e−7	3.775e−7	232	0.03	
			1.8	4.866e−7	3.979e−7	227	0.03	
100	100	100	$\lambda_{\min}(A\hat{P}^{-1}A^\top)$ =1.4930e−7, $\lambda_{\max}(A\hat{P}^{-1}A^\top)$ =0.9998					
			$\lambda_{\min}(B\hat{Q}^{-1}B^\top)$ =3.0493e−5, $\lambda_{\max}(B\hat{Q}^{-1}B^\top)$ =0.9993					
			0.2	3.765e−6	3.044e−6	1650	0.06	Cond(P):1.7514e+5
			0.4	1.951e−6	1.557e−6	981	0.07	Cond(Q):4.6216e+4
			0.6	1.272e−6	9.991e−7	666	0.09	
			0.8	9.583e−7	7.604e−7	527	0.03	
			1.0	6.413e−7	7.229e−7	386	0.03	
			1.2	5.757e−7	5.326e−7	348	0.03	
			1.4	4.195e−7	4.772e−7	295	0.03	
			1.6	4.064e−7	4.079e−7	264	0.03	
			1.618	4.131e−7	3.273e−7	253	0.03	
			1.65	4.426e−7	4.502e−7	251	0.03	
			1.7	4.202e−7	3.516e−7	274	0.00	
			1.75	4.459e−7	4.978e−7	234	0.03	
			1.8	4.441e−7	4.864e−7	227	0.03	

The details of constructing the example is nearly the same as those in Section 5.2 except

$$\mathtt{P1} = \mathtt{randn}(\mathtt{n}_1, \mathtt{n}_1); \quad \mathtt{P} = \mathtt{P1}' * \mathtt{P1}; \quad \text{and}$$
$$\mathtt{Q1} = \mathtt{randn}(\mathtt{n}_2, \mathtt{n}_2); \quad \mathtt{Q} = \mathtt{Q1}' * \mathtt{Q1};.$$

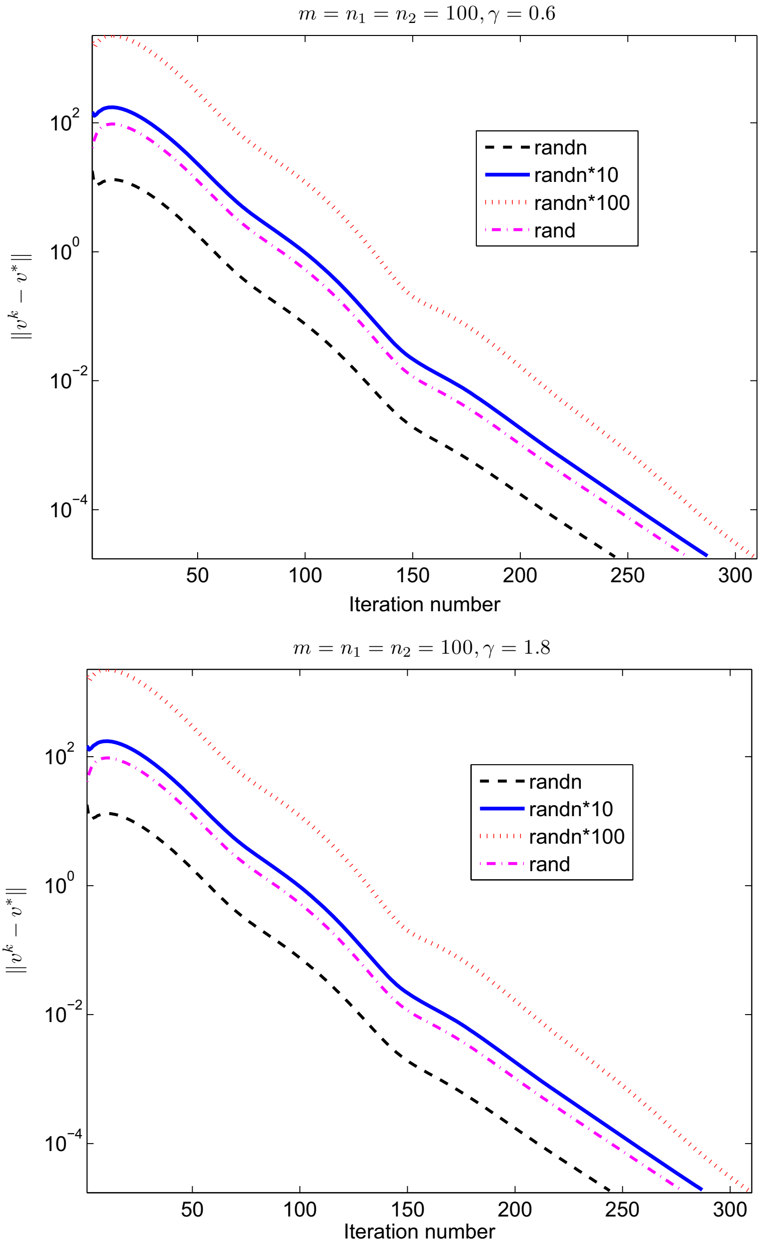

Fig. 2 Linear convergence of the ADMM for different initial points

Table 3 Global convergence of the ADMM with different initial points

γ	Initial.	$\|\bar{x} - x^*\|_2$	$\|\bar{y} - y^*\|_2$	Itr.	Time (s)
0.6	randn	1.142e−6	9.213e−7	657	0.05
	randn*10	1.272e−6	1.041e−6	812	0.06
	randn*100	1.270e−6	1.021e−6	942	0.06
	rand	1.226e−6	1.077e−6	716	0.03
1.8	randn	4.272e−7	3.766e−7	244	0.03
	randn*10	3.803e−7	3.040e−7	287	0.03
	randn*100	2.983e−7	3.445e−7	310	0.03
	rand	3.920e−7	3.277e−7	277	0.03

We set $\beta = 1$, and thus we just need to check the following conditions to ensure (63):

$$0 < \lambda_{\min}(A\hat{P}^{-1}A^\top) \text{ and } \lambda_{\max}(A\hat{P}^{-1}A^\top) < 1;\ 0 < \lambda_{\min}(B\hat{Q}^{-1}B^\top) \text{ and } \lambda_{\max}(B\hat{Q}^{-1}B^\top) < 1,$$

with $\hat{P}$ and $\hat{Q}$ defined in (23). If the generated matrices P, Q, A and B are satisfied with above conditions, it implies that neither 0 nor 1 is the common eigenvalues of $A\hat{P}^{-1}A^\top$ and $B\hat{Q}^{-1}B^\top$. Therefore, the condition (63) in Theorem 6.1 is ensured. The implementation details of the scheme (19) are the same as those in Sect. 5.2,

We first test the performance with the initial point $y^0 = randn(n_2, 1)$ and $z^0 = randn(m, 1)$ for the scenarios where $m = n_1 = n_2 = 50, 100$. A number of γ's values varying from 0.2 to 1.8 with an equal distance of 0.2 are tested. Again, the value of 1.618 suggested by Glowinski is tested as a benchmark and several values larger than 1.618, i.e, $\gamma = 1.65, 1.7, 1.75$, are compared to show the possible acceleration with larger values of γ. In Fig. 1, we plot the evolution of the errors to the exact solution point, i.e., $\|v^k - v^*\|_2$ with v^k defined in (28)), with respect to iteration numbers. The linear convergence of the scheme (19) is displayed in this figure for different choices of $\gamma \in]0, 2[$. The errors on x^k (measured by $\|\bar{x} - x^*\|_2$), y^k (measured by $\|\bar{y} - y^*\|_2$), the number of iteration ("Itr.")and the CPU time in seconds ("Time(s)") are reported in Table 2. Results in Table 2 demonstrate the global convergence of the scheme (19) under the condition (63).

To further see the global feature of the linear convergence in Theorem 6.1, with different initial points we focus on the case where $m = n_1 = n_2 = 100$, and the values of γ are 0.6 and 1.8, respectively. We report the numerical performance with several different initial points. We generate the initial points (y^0, z^0) by different ways as listed in Table 3. The evolution of the errors to the exact solution point, i.e., $\|v^k - v^*\|_2$ with v^k defined in (26), with respect to iteration numbers is plotted for these different initial points in Fig. 2. The curves in Fig. 2 clearly show the linear convergence of the scheme (19) under the condition (63) with different initial points. The numerical results are also summarized in Table 3.

7 Conclusions

In this paper, we prove the convergence of the alternating direction method of multipliers (ADMM) with a factor $\gamma \in]0, 2[$ for updating its dual variable when the objective function is the sum of two quadratic functions. Glowinski's open question in 1984 is thus partially answered. Because of the quadratic programming context under discussion, the spectral analysis plays a crucial role in the analysis. But our analysis is featured by a nonsymmetric matrix involving the factor $\gamma \in]0, 2[$, and hence, more complicated analysis than the typical case of $\gamma = 1$ in the original ADMM is needed. The setting under our discussion seems to be by now the most general one regarding the answer to Glowinski's open question. Answering this question completely for the generic case where the objective function is the sum of two general convex functions seems to need more advanced analytic tools, rather than just the spectral analysis in numerical linear algebra. We hope the new analysis presented in the paper will favor this ultimate goal. A by-product of our analysis is the global linear convergence rate of the ADMM with $\gamma \in]0, 2[$ for the quadratic programming case under a tight condition. This result differs from existing results in the literature.

Acknowledgements Min Tao was supported by the NSFC Grant: 11301280 and the Fundamental Research Funds for the Central Universities: 14380019. Xiaoming Yuan was supported by the General Research Fund from Hong Kong Research Grants Council: 12313516.

References

1. Glowinski, R., Marrocco, A.: Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. Revue Fr. Autom. Inf. Rech. Opér. Anal. Numér **2**, 41–76 (1975)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–122 (2010)
3. Glowinski, R.: On alternating direction methods of multipliers: a historical perspective. In: Fitzgibbon, W., Kuznetsov, Y.A., Neittaanmaki, P., Pironneau, O. (eds.) Modeling, Simulation and Optimization for Science and Technology, Computational Methods in Applied Sciences, vol. 34, pp. 59–82. Springer, Dordrecht (2014)
4. Glowinski, R., Osher, S.J., Yin, W.T. (eds.): Splitting Methods for Communications and Imaging, Science and Engineering. Springer, Switzerland (2016)
5. Sun, J., Zhang, S.: A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs. Eur. J. Oper. Res. **207**, 1210–1220 (2010)
6. Wen, Z.W., Goldfarb, D., Yin, W.T.: Alternating direction augmented Lagrangian methods for semidefinite programming. Math. Program. Comput. **2**, 203–230 (2010)
7. Eckstein, J., Yao, W.: Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results. Pac. J. Optim. **11**, 619–644 (2015)
8. Fortin, M., Glowinski, R.: On decomposition-coordination methods using an augmented Lagrangian. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Solution of Boundary Problems, pp. 97–146. North-Holland, Amsterdam (1983)
9. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems, pp. 299–331. North Holland, Amsterdam (1983)
10. 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)
11. Glowinski, R.: Numerical Methods for Nonlinear Variational Problems. Springer, New York (1984)

12. He, B.S., Ma, F., Yuan, X.M.: Convergence analysis of the symmetric version of ADMM. SIAM J. Imaging Sci. **9**, 1467–1501 (2016)
13. He, B.S., Yang, H.: Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities. Oper. Res. Lett. **23**, 151–161 (1998)
14. Xu, M.H.: Proximal alternating directions method for structured variational inequalities. J. Optim. Theory Appl. **134**, 107–117 (2007)
15. Tao, M., Yuan, X.M.: On the $O(1/t)$ convergence rate of alternating direction method with logarithmic-quadratic proximal regularization. SIAM J. Optim. **22**, 1431–1448 (2012)
16. 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)
17. Eckstein, J., Fukushima, M.: Reformulations and applications of the alternating direction method of multipliers. In: Hager, W.W., Hearn, D.W., Pardalos, P.M. (eds.) Large Scale Optimization: State of the Art, pp. 115–134. Kluwer Academic Publishers, Dordrecht (1994)
18. Gol'shtein, E.G., Tret'yakov, N.V.: Modified Lagrangian in convex programming and their generalizations. Math. Program. Stud. **10**, 86–97 (1979)
19. Tao, M., Yuan, X.M.: The generalized proximal point algorithm with step size 2 is not necessarily convergent. Comput. Optim. Appl. **70**, 827–839 (2018)
20. He, B.S., Xu, M.H., Yuan, X.M.: Solving large-scale least squares covariance matrix problems by alternating direction methods. SIAM J. Matrix Anal. Appl. **32**, 136–152 (2011)
21. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
22. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)
23. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 877–898 (1976)
24. Martinet, B.: Regularisation, d'inéquations variationelles par approximations succesives. Rev. Francaise d'Inform. Recherche Oper. **4**, 154–159 (1970)
25. Bergen, A.R.: Power Systems Analysis. Prentice Hall, Englewood Cliffs (1986)
26. Björck, A.: Numerical Methods for Least Squares Problems. SIAM, Philadelphia (1996)
27. Bruckstein, A.M., Donoho, D.L., Elad, M.: From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev. **51**, 34–81 (2009)
28. Chua, L.O., Desoer, C.A., Kuh, E.S.: Linear and Nonlinear Circuits. McGraw-Hill, New York (1987)
29. Ehrgott, M., Winz, I.: Interactive decision support in radiation therapy treatment planning. OR Spectr. **30**, 311–329 (2008)
30. Markowitz, H.M.: Porfolio Selection: Efficient Diversification of Investments. Wiley, New York (1959)
31. Tikhonov, A., Arsenin, V.: Solution of Ill-Posed problems. Winston, Washington (1977)
32. Chen, C.H., Li, M., Liu, X., Ye, Y.Y.: Extended ADMM and BCD for nonseparable convex minimization models with quadratic coupling terms: convergence analysis and insights. Math. Program. (2017). https://doi.org/10.1007/s10107-017-1205-9
33. Fiedler, M.: Bounds for the determinant of the sum of Hermitian matrices. Proc. Am. Math. Soc. **30**, 27–31 (1971)
34. Ding, J., Rhee, N.H.: On the equality of algebraic and geometric multiplicities of matrix eigenvalues. Appl. Math. Lett. **24**, 2211–2215 (2011)
35. He, B.S., Liao, L.Z., Han, D.R., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. Math. Program. **92**, 103–118 (2002)
36. Eckstein, J.: Some saddle-function splitting methods for convex programming. Optim. Methods Softw. **4**, 75–83 (1994)
37. Meyer, C.D.: Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia (2006)
38. Bhatia, R.: Matrix Analysis. Springer, New York (1997)
39. Boley, D.: Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs. SIAM J. Optim. **23**, 2183–2207 (2013)
40. Han, D.R., Yuan, X.M.: Local linear convergence of the alternating direction method of multipliers for quadratic programs. SIAM J. Numer. Anal. **51**, 3446–3457 (2013)
41. Bai, Z.Z., Tao, M.: Rigorous convergence analysis of alternating variable minimization with multiplier methods for quadratic programming problems with equality constraints. BIT **56**, 399–422 (2016)