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

Min Tao[1] and **Xiaoming Yuan** [2]

Abstract. The alternating direction method of multipliers (ADMM) was proposed by Glowinski and Marrocco in 1975; 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 step size for updating the dual variable in ADMM from 1 to $(0, \frac{1+\sqrt{5}}{2})$; and this strategy immediately accelerates the convergence of ADMM for most of its applications. Meanwhile, Glowinski raised the question of whether or not the range can be further enlarged to $(0, 2)$; 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 are quadratic. Glowinski's open question is thus partially answered. We further establish the global linear convergence of the ADMM with the step size range $(0, 2)$ for the quadratic programming case under a condition that turns out to be tight.

Keywords. Alternating direction method of multipliers, Glowinski's open question, quadratic programming, step size, linear convergence
Mathematics Subject Classification: 90C25; 90C30; 65K05; 90C20.

1 Introduction

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

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

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 (1.1) be

$$\mathcal{L}_\beta(x, y, \lambda) = \theta_1(x) + \theta_2(y) - z^\top(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{1.2}$$

with $z \in \mathbb{R}^m$ the Lagrange multiplier and $\beta > 0$ the penalty parameter.

The alternating direction method of multipliers (ADMM) was proposed by Glowinski and Marrocco in [21]; its iterative scheme for (1.1) reads as

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \mathcal{L}_\beta(x, y^k, z^k), & \text{(1.3a)}\\ y^{k+1} = \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y, z^k), & \text{(1.3b)}\\ z^{k+1} = z^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.3c)} \end{cases}$$

Recently, the ADMM has found many applications in a variety of areas because of its simplicity in implementation and usually good numerical performance; and it has received increasing attention in the literature. We refer to, [5, 13, 20, 22], for some review papers of the ADMM.

[1] Department of Mathematics, Nanjing University, JiangSu, China. This author was supported by the NSFC Grant: 11301280. Email: taom@nju.edu.cn.
[2] Department of Mathematics, Hong Kong Baptist University, Hong Kong. This author was supported by the General Research Fund from Hong Kong Research Grants Council: 12313516. Email: xmyuan@hkbu.edu.hk.

In [16][3], Fortin and Glowinski proposed a variant ADMM scheme as follows:

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \mathcal{L}_\beta(x, y^k, z^k), & \text{(1.4a)} \\ y^{k+1} = \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y, z^k), & \text{(1.4b)} \\ z^{k+1} = z^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), & \text{(1.4c)} \end{cases}$$

with $\gamma \in (0, \frac{1+\sqrt{5}}{2})$. The parameter γ is mainly for enlarging the step size for updating the dual variable z in (1.4c) and thus leading to possible faster convergence. It is worthwhile to mention that the parameter γ in (1.4c) is different from the involved parameter in the so-called generalized ADMM that was discussed in [11, 12] based on the idea in [23] (see also [37]). The convergence of (1.4) with $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ has been well studied in various contexts, see, e.g., [17, 18, 19, 26, 28, 39, 36]. Numerically, it has been widely verified that a large value of γ close to $\frac{1+\sqrt{5}}{2}$ can accelerate the convergence ADMM immediately, see, e.g., [19, 27, 35, 40]. Though the convergence of the ADMM scheme (1.4) with $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is known, numerically it has been observed as well that some values exceed the upper bound $\frac{1+\sqrt{5}}{2}$ may still perform convergence. Indeed, the condition $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ is not necessary but just sufficient to ensure the convergence of the ADMM scheme (1.4); and it is natural to ask if there is any theory to ensure the convergence of (1.4) if γ is larger than $\frac{1+\sqrt{5}}{2}$ in (1.4c). Glowinski raised the question in [19] (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 if it can be extended to the more general cases considered here.". The function "G" in [19] corresponds to the function θ_2 in model (1.1); "ALG2" refers to the ADMM scheme (1.4); "[1]" refers to [18] and $\rho := \gamma\beta$ in our setting. With the well studied results for $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ in (1.4), 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 [19].

The rationale of raising this question can also be explained as follows. Note that if the model (1.1) is regarded as a whole and the augmented Lagrangian method (ALM) in [29, 33] is directly applied to (1.1), the iterative scheme becomes:

$$\begin{cases} (x^{k+1}, y^{k+1}) = \arg\min_{x\in\mathcal{X},\, y\in\mathcal{Y}} \mathcal{L}_\beta(x, y, z^k), & \text{(1.5a)} \\ z^{k+1} = z^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(1.5b)} \end{cases}$$

Based on the work [34], the ALM scheme (1.5) is an application of the proximal point algorithm (PPA) in [31] to the dual of the model (1.1) and thus the result in [23] can be applied to modify the scheme (1.5) as

$$\begin{cases} (x^{k+1}, y^{k+1}) = \arg\min_{x\in\mathcal{X},\, y\in\mathcal{Y}} \mathcal{L}_\beta(x, y, z^k), & \text{(1.6a)} \\ z^{k+1} = z^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), & \text{(1.6b)} \end{cases}$$

where $\gamma \in (0, 2)$. Its convergence can be found in, e.g., [37]. Therefore, the ADMM scheme (1.3) can be regarded as a splitting version of the ALM (1.5) which splits the (x^{k+1}, y^{k+1})-subproblem (1.5a) as the surrogates (1.3a) and (1.3b) by the Gauss-Seidel manner. Then, with $\gamma \in (0, 2)$ in (1.6b) for the ALM scheme, it is natural to ask if this property can be maintained for the ADMM scheme (1.4). Since the ADMM (1.3) is just an inexact version of the ALM (1.5), it is not straightforward to claim the validity of this extension and this may explain why Glowinski's question is still open. We also comment that solving the subproblems (1.4a) and (1.4b) dominates the implementation of the ADMM scheme (1.4); and once they are solved, it is meaningful to discuss the theory of enlarging the step size for updating the dual variable as (1.4c), because it may offer an immediate possibility of further accelerating the convergence of (1.4) without additional computation.

In this paper, our main purpose is to show the convergence of the ADMM scheme (1.4) with $\gamma \in (0, 2)$ when both the functions θ_1 and θ_2 in (1.1) are quadratic; hence partially answer Glowinski's open question.

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

The model for further discussion is

$$\min\{\frac{1}{2}x^\top Px + f^\top x + \frac{1}{2}y^\top Qy + g^\top y \mid Ax + By = b,\ x \in \mathbb{R}^{n_1},\ y \in \mathbb{R}^{n_2}\}, \tag{1.7}$$

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 $g \in \mathbb{R}^{n_2}$. The solution set of (1.7) is assumed to be nonempty throughout our discussion. We refer to, e.g., [1, 2, 3, 6, 8, 14, 30, 38], for various applications that can be modeled as the quadratic programming model (1.7) in economics, finance, electromagnetism, electrical circuits and networks, image processing, contact problems, control problems, and intensity modulated radiotherapy problems.

The remaining part of this paper is organized as follows. In Section 2, 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 Section 3, including the specification of the matrix recursion of the ADMM scheme (1.4) for the quadratic programming model (1.7), the KKT condition of (1.7), and an example showing the divergence of (1.4) with $\gamma = 2$ for (1.7). In Section 4, the convergence of the ADMM scheme (1.4) with $\gamma \in (0, 2)$ is proved for (1.7). This is the main result of the paper. The global linear convergence of (1.4) with $\gamma \in (0, 2)$ is also proved for (1.7) in Section 5, under a new condition different from some existing work. Both the convergence and global linear convergence are verified numerically by some examples in Sections 4 and 5 for the case of $\gamma \in (0, 2)$. Finally, some conclusions are drawn in Section 6.

2 Preliminaries

In this section, we summarize some notations and recall some definitions to be used, present some assumptions, and prove some elementary lemmas for further discussion.

2.1 Notation and definitions

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 unite 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, $|a|$ denotes its modulus. Given a vector space $\mathcal{V}$, its dimension is denoted by $\dim(V)$. 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, $\text{Rank}(M)$ represents its rank. For a matrix $M \in \mathbb{R}^{n \times n}$, $\text{eig}(M)$ represents all the eigenvalues of M (considering the multiplicity), and we use the notation $\sigma(M)$ to represent 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 it is abbreviated as I when $n = m$. 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. The notation $N(\cdot)$ represents the null space of N. The function δ_{ij} is defined as

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

2.2 Assumptions

Throughout this paper, we make the following assumptions.

Assumption 1. In (1.7), the symmetric positive semidefinite matrices P and Q and the matrices A and B satisfy the conditions:

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

Assumption 2. The KKT point set of (1.1) is non-empty.

Remark 2.1. Under Assumptions 1-2, the x- and y-subproblems of the ADMM scheme (1.4) for solving (1.7) are well-defined and each has a unique solution. Also, as shown by Corollary 1 in [7], Assumptions 1-2 are necessary for the well-definedness of the ADMM scheme (1.4) for solving (1.7).

2.3 Some lemmas

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

Lemma 2.2. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy the conditions $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. 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 2.3. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy the conditions $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. Then, we have*

1.) *For any $x \in \mathbb{R}^m$ with $x^T x = 1$, we have $|x^\top FGx| \le 1$.*

2.) *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

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

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

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

The proof is complete. □

Lemma 2.4. *Let U and V be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy the conditions $-\frac{I}{2} \preceq U \preceq \frac{I}{2}$ and $-\frac{I}{2} \preceq V \preceq \frac{I}{2}$. Then, for any $x \in \mathbb{R}^m$ such that $x^\top x = 1$, we have*

$$\left| x^\top \frac{(UV + VU)}{2} x \right| \le \frac{1}{4}$$

and

$$-\frac{I}{4} \preceq \frac{(UV + VU)}{2} \preceq \frac{I}{4}.$$

Proof. First, using Cauchy-Schwarz inequality, we get

$$|x^\top (UV)x| \leq \frac{1}{2}\left(x^\top U^2 x + x^\top V^2 x\right), \tag{2.1}$$

and

$$|x^\top (VU)x| \leq \frac{1}{2}\left(x^\top U^2 x + x^\top V^2 x\right). \tag{2.2}$$

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 (2.1) and (2.2), 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 (1.4) for the model (1.7).

Lemma 2.5. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ 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

$$I - F - G + FG + GF = (\frac{I}{2} - F)(\frac{I}{2} - G) + (\frac{I}{2} - G)(\frac{I}{2} - F) + \frac{I}{2}. \tag{2.3}$$

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

$$-\frac{I}{2} \preceq \frac{I}{2} - F \preceq \frac{I}{2}, \quad -\frac{I}{2} \preceq \frac{I}{2} - G \preceq \frac{I}{2}. \tag{2.4}$$

Thus, using Lemma 2.4, we get

$$-\frac{I}{2} \preceq (\frac{I}{2} - F)(\frac{I}{2} - G) + (\frac{I}{2} - G)(\frac{I}{2} - F) \preceq \frac{I}{2}. \tag{2.5}$$

Combining (2.3) and (2.5), 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 (1.7).

Lemma 2.6. *[15] Let A and B be $m \times m$ Hermitian matrices with eigenvalues $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_m$ and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_m$, respectively. Then, we have*

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

(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

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

Lemma 2.7. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$. Then, we have*

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

Proof. The assertion follows immediately from the following two equations:

$$F + G - 2FG = I - 2\left((\frac{I}{2} - F)(\frac{I}{2} - G) + \frac{I}{4}\right) \text{ and } F + G - 2GF = I - 2\left((\frac{I}{2} - G)(\frac{I}{2} - F) + \frac{I}{4}\right).$$

The proof is complete. □

Lemma 2.8. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$ and x is an eigenvector associated with 1, then we have that 1 is also an eigenvalue of the matrix $(I - F - G + 2GF)$ and x is its eigenvector associated with 1. Conversely, if 1 is an eigenvalue of the matrix $(I - F - G + 2GF)$ and x is an associated eigenvector with 1, then 1 is also an eigenvalue of the matrix $(I - F - G + 2FG)$ and x is an eigenvector associated with 1.*

Proof. Assume that the vector $\hat{x}$ is a unite 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

$$\hat{x}^\top (I - F - G + 2FG)\hat{x} = 1 \text{ with } \hat{x}^\top \hat{x} = 1. \tag{2.6}$$

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

$$\begin{aligned} x^\top (I - F - G + 2FG)x &= x^\top \left\{2\left[(\frac{I}{2} - F)(\frac{I}{2} - G) + \frac{I}{4}\right]\right\} x \\ &\le \left(x^\top (\frac{I}{2} - F)^2 x + x^\top (\frac{I}{2} - G)^2 x\right) + \frac{1}{2} \le 1, \end{aligned} \tag{2.7}$$

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 (2.4)). Moreover, (2.6) implies that (2.7) holds with equality. Thus, checking the conditions ensuring the (2.7) with equality, we have the following three arguments:

a.) $x^\top (\frac{I}{2} - F)^2 x = 1/4$;

b.) $x^\top (\frac{I}{2} - G)^2 x = 1/4$;

c.) $(\frac{I}{2} - F)x = (\frac{I}{2} - G)x$.

Consequently, it means that one of the following arguments holds:

a.) The matrices $(\frac{1}{2}I - F)$ and $(\frac{1}{2}I - G)$ have the common eigenvalue $1/2$, and x is a common eigenvector associated with $1/2$;

b.) The matrices $(\frac{1}{2}I - F)$ and $(\frac{1}{2}I - G)$ have the common eigenvalue $-1/2$, and x is a common eigenvector associated with $-1/2$.

Or, equivalently, one of the following assertions is true:

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

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

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

a). $Fx = Gx = \mathbf{0}$;

b). $Fx = Gx = x$.

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 converse direction can be proved in a similar way, thus omitted. The proof is complete. □

Lemma 2.9. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$ and x is an eigenvector associated with 1, then one of the following assertions hold:*

1.) *Fx=Gx=x;*

2.) *Fx=Gx=$\mathbf{0}$.*

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

Proof. The proof can be found in the proof of Lemma 2.8. □

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 [9].

Lemma 2.10. *[9] Let $A \in \mathbb{C}^{n\times n}$ and let λ be an eigenvalue of A. Then the following two statements are equivalent.*

1.) *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.*

2.) *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 2.11. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ 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*

$$\ell := m - Rank(F + G - 2FG). \tag{2.8}$$

Proof. Since 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, its geometric multiplicity of 1 is the ℓ defined in (2.8). Invoking Lemma 2.8, 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 2.7 that the geometric multiplicity of 1 for the matrix $(I - F - G + 2GF)$ is also ℓ defined in (2.8). 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 2.10, 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 2.9.

Lemma 2.12. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$, then we have*

$$dim(\{x \mid x \in N(F - I) \cap N(G - I)\}) + dim(\{x \mid x \in N(F) \cap N(G)\}) = \ell, \tag{2.9}$$

where ℓ is the algebraic and geometric multiplicities of $I - F - G + 2FG$ defined in (2.8).

Proof. First, invoking Lemma 2.9, we have

$$\{x \mid x \in N\left([I-F-G+2FG]-I\right)\} = \{x \mid x \in N(F-I)\cap N(G-I)\}\bigcup\{x \mid x \in N(F)\cap N(G)\}. \tag{2.10}$$

Note that

$$\{x \mid x \in N(F-I)\cap N(G-I)\}\bigcap\{x \mid x \in N(F)\cap N(G)\} = \{\mathbf{0}\}.$$

From Lemma 2.11, we know that $\dim\{x \mid x \in N\left([I-F-G+2FG]-I\right)\} = \ell$ with ℓ defined in (2.8). Thus, the assertion (2.9) follows directly. □

Lemma 2.13. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ 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*

$$Q^\top FQ = \begin{pmatrix} D & R_{\ell\times(m-\ell)} \\ \mathbf{0} & \hat{F} \end{pmatrix} \quad \textit{and} \quad Q^\top GQ = \begin{pmatrix} D & S_{\ell\times(m-\ell)} \\ \mathbf{0} & \hat{G} \end{pmatrix}, \tag{2.11}$$

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

$$\mathbf{0} \preceq \hat{F} \preceq I_{m-\ell}, \ \ \mathbf{0} \preceq \hat{G} \preceq I_{m-\ell} \ \ \textit{and} \ \ \hat{F}+\hat{G}-2\hat{G}\hat{F} \succ \mathbf{0}. \tag{2.12}$$

Proof. Since 1 is an eigenvalue of the matrix $(I-F-G+2FG)$, it follows from Lemma 2.11 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 (2.10). We thus have

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

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

$$Q_1 = (x_1;x_2;\ldots;x_\ell) \text{ and } Q_2 \in \mathbb{R}^{m\times(m-\ell)}.$$

Invoking Lemma 2.9, we have

$$FQ_1 = Q_1D \text{ and } GQ_1 = Q_1D,$$

with

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_\ell \end{pmatrix} \quad \text{and} \quad \lambda_i = 0 \text{ or } 1, \ i = 1,\ldots,\ell.$$

Therefore, we get

$$Q^\top FQ = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} F \begin{pmatrix} Q_1; & Q_2 \end{pmatrix} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix}\begin{pmatrix} Q_1D; & FQ_2 \end{pmatrix} = \begin{pmatrix} D & Q_1^\top FQ_2 \\ \mathbf{0} & Q_2^\top FQ_2 \end{pmatrix}.$$

Analogously, we have

$$Q^\top GQ = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} G \begin{pmatrix} Q_1; & Q_2 \end{pmatrix} = \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix}\begin{pmatrix} Q_1D; & GQ_2 \end{pmatrix} = \begin{pmatrix} D & Q_1^\top GQ_2 \\ \mathbf{0} & Q_2^\top GQ_2 \end{pmatrix}.$$

Thus, the assertion (2.11) is proved by setting $R := Q_1^\top FQ_2$ and $S := Q_1^\top GQ_2$.

For the second assertion (2.12), let us define $\hat{F} = Q_2^\top FQ_2$ and $\hat{G} = Q_2^\top GQ_2$. Then, the first two inequalities in (2.12) hold. On the other hand, similar to the second inequality in Lemma 2.5, we have

$$\hat{F}+\hat{G}-2\hat{G}\hat{F} \succeq \mathbf{0}.$$

Now, we prove the third inequality in (2.12) 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 2.8, 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 (2.9). Therefore, we have $\hat{F}+\hat{G}-2\hat{G}\hat{F}\succ\mathbf{0}$. The proof is complete. □

Remark 2.14. Lemma 2.13 implies that the matrices $\hat{F}$ and $\hat{G}$ have no common eigenvectors for either the eigenvalue 0 or 1.

3 Specification of the ADMM scheme (1.4) for (1.7)

In this section, we first specify the application of the ADMM scheme (1.4) to the quadratic programming model (1.7) 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 (1.4) to the quadratic programming model (1.7) results in the iterative scheme:

$$\begin{cases} (P+\beta A^\top A)x^{k+1}=A^\top z^k-\beta A^\top By^k+\beta A^\top b-f, \\ (Q+\beta B^\top B)y^{k+1}=B^\top z^k-\beta B^\top Ax^{k+1}+\beta B^\top b-g, \\ z^{k+1}=z^k-\gamma\beta(Ax^{k+1}+By^{k+1}-b). \end{cases} \tag{3.1}$$

3.1 Matrix recursion form of (3.1)

Notice that the variable x in the ADMM (1.4) 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 (3.1) 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 (3.1) as

$$\begin{cases} (P/\beta+A^\top A)x^{k+1}=A^\top\mu^k-A^\top By^k+A^\top b-f/\beta, & (3.2a)\\ (Q/\beta+B^\top B)y^{k+1}=B^\top\mu^k-B^\top Ax^{k+1}+B^\top b-g/\beta, & (3.2b)\\ \mu^{k+1}=\mu^k-\gamma(Ax^{k+1}+By^{k+1}-b). & (3.2c)\end{cases}$$

Note that (3.2a) can be written as

$$x^{k+1}=(P/\beta+A^\top A)^{-1}\left[A^\top\mu^k-A^\top By^k+A^\top b-f/\beta\right]. \tag{3.3}$$

Then, substituting (3.3) into (3.2b) and (3.2c), we eliminate x^{k+1} from (3.2) and obtain

$$\begin{cases} \hat{Q}y^{k+1}=B^\top A\hat{P}^{-1}A^\top By^k+(B^\top-B^\top A\hat{P}^{-1}A^\top)\mu^k+\alpha_1, \\ \gamma By^{k+1}+\mu^{k+1}=(I-\gamma A\hat{P}^{-1}A^\top)\mu^k+\gamma A\hat{P}^{-1}A^\top By^k+\alpha_2, \end{cases} \tag{3.4}$$

with

$$\hat{P}=P/\beta+A^\top A,\quad \hat{Q}=Q/\beta+B^\top B, \tag{3.5}$$

$$\alpha_1=-B^\top A\hat{P}^{-1}A^\top b+B^\top A\hat{P}^{-1}f/\beta+B^\top b-g/\beta, \tag{3.6}$$

$$\alpha_2=\gamma b-\gamma A\hat{P}^{-1}A^\top b+\gamma A\hat{P}^{-1}f/\beta. \tag{3.7}$$

It follows from (3.4) that

$$\begin{pmatrix} \hat{Q} & \mathbf{0} \\ \gamma B & I \end{pmatrix}\begin{pmatrix} y^{k+1} \\ \mu^{k+1} \end{pmatrix}=\begin{pmatrix} B^\top A\hat{P}^{-1}A^\top B & B^\top-B^\top A\hat{P}^{-1}A^\top \\ \gamma A\hat{P}^{-1}A^\top B & I-\gamma A\hat{P}^{-1}A^\top \end{pmatrix}\begin{pmatrix} y^k \\ \mu^k \end{pmatrix}+\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

Moreover, with simple calculations, the iterative scheme (3.4) can be written compactly as follows:

$$v^{k+1} = T(\gamma)v^k + q, \tag{3.8}$$

with

$$T(\gamma) = \begin{pmatrix} \hat{Q}^{-1}B^\top A\hat{P}^{-1}A^\top B & \hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \\ \gamma(I - B\hat{Q}^{-1}B^\top)A\hat{P}^{-1}A^\top B & I - \gamma A\hat{P}^{-1}A^\top - \gamma B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \end{pmatrix} \tag{3.9}$$

and

$$v^k = \begin{pmatrix} y^k \\ \mu^k \end{pmatrix}, \quad q = \begin{pmatrix} q_1 := \hat{Q}^{-1}\alpha_1 \\ q_2 := \alpha_2 - \gamma B\hat{Q}^{-1}\alpha_1 \end{pmatrix}. \tag{3.10}$$

Thus, the application of the ADMM scheme (1.4) to the quadratic programming model (1.7) can be written as the matrix recursion form (3.8)-(3.10).

To establish the convergence of the ADMM (3.1) with $\gamma \in (0, 2)$, we only need to conduct a spectral analysis for the iterative matrix $T(\gamma)$ defined in (3.9). First, note that the matrix $T(\gamma)$ can be factorized as

$$T(\gamma) \;=\; \begin{pmatrix} \hat{Q}^{-1}B^\top A\hat{P}^{-1}A^\top & \hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \\ \gamma(I - B\hat{Q}^{-1}B^\top)A\hat{P}^{-1}A^\top & I - \gamma A\hat{P}^{-1}A^\top - \gamma B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \end{pmatrix} \cdot \begin{pmatrix} B & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}. \tag{3.11}$$

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

$$\begin{aligned} &\tilde{T}(\gamma) \\ &= \begin{pmatrix} B & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix} \cdot \begin{pmatrix} \hat{Q}^{-1}B^\top A\hat{P}^{-1}A^\top & \hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \\ \gamma(I - B\hat{Q}^{-1}B^\top)A\hat{P}^{-1}A^\top & I - \gamma A\hat{P}^{-1}A^\top - \gamma B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \end{pmatrix} \\ &= \begin{pmatrix} B\hat{Q}^{-1}B^\top A\hat{P}^{-1}A^\top & B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \\ \gamma(I - B\hat{Q}^{-1}B^\top)A\hat{P}^{-1}A^\top & I - \gamma A\hat{P}^{-1}A^\top - \gamma B\hat{Q}^{-1}B^\top(I - A\hat{P}^{-1}A^\top) \end{pmatrix}. \end{aligned} \tag{3.12}$$

For any two square matrices X and Y with an appropriate dimension, we have $\mathrm{eig}(XY) = \mathrm{eig}(YX)$. Hence, we obtain

$$\mathrm{eig}(T(\gamma)) = \mathrm{eig}(\tilde{T}(\gamma)). \tag{3.13}$$

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

3.2 KKT Condition of (1.7)

In this subsection, we show several equivalent forms to characterize the KKT condition of the quadratic programming model (1.7) which will be useful for later analysis. These are necessary preparations 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 model (1.7). That is, (x^*, y^*, z^*) satisfies the following equations:

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

Furthermore, we denote

$$\mu^* := z^*/\beta. \tag{3.14}$$

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

$$\begin{cases} Px^* = \beta A^\top \mu^* - f, & (3.15a) \\ Qy^* = \beta B^\top \mu^* - g, & (3.15b) \\ Ax^* + By^* - b = \mathbf{0}. & (3.15c) \end{cases}$$

In the following, we present another equivalent form of the KKT condition of the model (1.7) represented by x^* and (y^*, μ^*) separately. This form helps us better reveal the relationship between a fix point of the iterative matrix given in (3.8) and the KKT point of (1.7). We first prove a lemma that turns out to be essential for the convergence analysis to be presented.

Lemma 3.1. *Suppose that Assumptions 1-2 hold. Let $\gamma \neq 0$. Then, the pair (x^*, y^*, z^*) is a KKT point of (1.7) if and only if it satisfies the following equations:*

$$x^* = \hat{P}^{-1}\left[A^\top \mu^* - A^\top B y^* + A^\top b - f/\beta\right] \text{ and } [I - T(\gamma)] \begin{pmatrix} y^* \\ \mu^* \end{pmatrix} = q, \tag{3.16}$$

where the vectors μ^*, q and the matrix $T(\gamma)$ are defined in (3.14), (3.10) and (3.9), respectively.

Proof. Recall that the pair (x^*, y^*, z^*) is a KKT point of (1.7) if and only if it satisfies (3.15). First, we show that the equation (3.16) holds when (x^*, y^*, μ^*) satisfies (3.15). For this purpose, we multiply both sides of (3.15c) by $\beta A^\top$ from the left and add the resulting equation to (3.15a). This manipulation yields the equation:

$$(P + \beta A^\top A)x^* = \beta(A^\top \mu^* - A^\top B y^* + A^\top b - f/\beta). \tag{3.17}$$

Dividing both sides of the above equation by β and using the definition of $\hat{P}$ in (3.5), and then multiplying it by $\hat{P}^{-1}$ from the left, we obtain

$$x^* = \hat{P}^{-1}\left[A^\top \mu^* - A^\top B y^* + A^\top b - f/\beta\right]. \tag{3.18}$$

Next, we multiply (3.15c) by $\beta B^\top$ from the left and adding the resulting equation to (3.15b), which is further divided in both sides by β. These operations enable us to have

$$(Q/\beta + B^\top B)y^* = B^\top \mu^* - B^\top A x^* + B^\top b - g/\beta. \tag{3.19}$$

Then, substituting (3.18) into the above equality and recalling the definitions of $\hat{Q}$ in (3.5) and α_1 in (3.6), we get

$$\hat{Q}y^* = B^\top A \hat{P}^{-1} A^\top B y^* + (B^\top - B^\top A \hat{P}^{-1} A^\top)\mu^* + \alpha_1. \tag{3.20}$$

On the other hand, it follows from (3.18) and the definition of α_2 in (3.7) that (3.15c) can be reformulated as

$$\mu^* = \mu^* - \gamma(Ax^* + By^* - b) = \gamma A \hat{P}^{-1} A^\top B y^* - \gamma B y^* + (I - \gamma A \hat{P}^{-1} A^\top)\mu^* + \alpha_2. \tag{3.21}$$

Combining (3.20) with the equation above, we have

$$\begin{pmatrix} \hat{Q} & \mathbf{0} \\ \gamma B & I \end{pmatrix} \begin{pmatrix} y^* \\ \mu^* \end{pmatrix} = \begin{pmatrix} B^\top A \hat{P}^{-1} A^\top B & B^\top - B^\top A \hat{P}^{-1} A^\top \\ \gamma A \hat{P}^{-1} A^\top B & I - \gamma A \hat{P}^{-1} A^\top \end{pmatrix} \begin{pmatrix} y^* \\ \mu^* \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}. \tag{3.22}$$

Then, multiplying the matrix

$$\begin{pmatrix} \hat{Q} & \mathbf{0} \\ \gamma B & I \end{pmatrix}^{-1} = \begin{pmatrix} \hat{Q}^{-1} & \mathbf{0} \\ -\gamma B \hat{Q}^{-1} & I \end{pmatrix},$$

to both sides of the equation (3.22) from the left, and recalling the definitions of $T(\gamma)$ in (3.9) and q in (3.10), we get

$$\begin{pmatrix} y^* \\ \mu^* \end{pmatrix} = T(\gamma) \begin{pmatrix} y^* \\ \mu^* \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}. \tag{3.23}$$

Thus, combining (3.18) and (3.23), the assertion (3.16) is proved.

Next, we verify the assertion of the other direction. That is, if (3.16) holds with one $\gamma \neq 0$, then (3.15) is true. Since (3.23) holds, we know that (3.22) is true. Furthermore, we get (3.20) and the second equality in (3.21) because of (3.22). Substituting (3.18) into the second equality of (3.21), we prove the first equality in (3.21). Also, substituting (3.18) into (3.20), we get (3.19). Recall the definition $\hat{P}$ in (3.5). Then, we have

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

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

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

Then, substituting the last equality into the first and second equations of the above system, we obtain (3.15). Thus, the conclusion of this lemma follows directly and the proof is complete. □

Remark 3.2. For the "only if" direction, the equations in (3.16) hold for any $\gamma \neq 0$. For the "if" direction, if there exists one $\gamma \neq 0$ such that (3.16) holds, then the pair (x^*, y^*, μ^*) satisfies (3.15). That is, the pair (x^*, y^*, z^*) is a KKT point of (1.7).

3.3 Divergence of $\gamma = 2$

In [37], we have shown that the ALM (1.6) is not necessarily convergent if $\gamma = 2$. Hence, it is intuitive to assert that the convergence of the ADMM scheme (1.4), as an inexact version of the ALM (1.6), is not ensured with $\gamma = 2$ for the generic case (1.1), either. Before we prove the convergence for the scheme (3.1) with $\gamma \in (0, 2)$, we construct an example to show that the convergence of (3.1) 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 (1.4) is discussed.

More specifically, let us take

$$F = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{3.24}$$

and

$$A = \begin{pmatrix} 0.4 & 0.3 \\ 0.5 & 2.2 \end{pmatrix}, \quad B = \begin{pmatrix} 1.2 & -0.2 \\ 1.6 & 0.1 \end{pmatrix}, \quad f = g = b = \mathbf{0}_{2\times 1}. \tag{3.25}$$

This is a special case of the quadratic programming (1.7) and it has a unique solution $x = y = \mathbf{0}_{2\times 1}$.

With $\gamma = 2$ and $\beta = 1$, the iterative matrix in (3.9) 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 (3.8) 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 (3.1) with $\gamma = 2$ is not guaranteed.

4 Convergence analysis of (3.1)

In this section, we establish the convergence of the scheme (3.1) with $\gamma \in (0, 2)$. The analysis still replies on the spectral analysis for the corresponding iterative matrix $T(\gamma)$ defined in (3.9). 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., [18, 25, 28, 39]) or the non-expansiveness property of certain maximal monotone operator (e.g., [10, 11, 17]). 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 directly apply these existing techniques to establish the convergence of the ADMM scheme (1.4) with $\gamma \geq \frac{1+\sqrt{5}}{2}$. This may be explained as a difficulty of answering Glowinski's open question under consideration.

4.1 Theoretical analysis

Even for the specific quadratic programming model (1.7), the resulting iterative matrix $T(\gamma)$ defined in (3.9) is complicated at least in the following sense. (1) It is non-symmetric; 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 Section 3 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 eigenvalue (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 (3.9) 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 (3.9).

Theorem 4.1. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. For $\gamma \in (0, 2)$, let us define*

$$M(\gamma) = \begin{pmatrix} GF & G - GF \\ \gamma F - \gamma GF & I - \gamma F - \gamma G + \gamma GF \end{pmatrix}. \tag{4.1}$$

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:

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

It is clear that

$$\text{eig}(M(\gamma)) = \text{eig}(M'(\gamma)).$$

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

$$\begin{cases} FGu + (I - F)w = \lambda u, & (4.2a) \\ \gamma(I - G)FGu + \gamma(I - G)(I - F)w + (1 - \gamma)w = \lambda w. & (4.2b) \end{cases}$$

Multiplying both sides in (4.2a) by $\gamma(I - G)$ from the left and then subtracting it by (4.2b), we get

$$[\lambda - 1 + \gamma]w = \gamma\lambda(I - G)u. \tag{4.3}$$

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 (4.2a), we obtain

$$FGu + (I - F)\frac{\lambda\gamma}{\lambda + \gamma - 1}(I - G)u = \lambda u. \tag{4.4}$$

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:

$$\xi_1 := u^H FGu \ \text{ and } \ \xi_2 = u^H(I - F)(I - G)u. \tag{4.5}$$

Then, multiplying both sides of (4.4) by $(\lambda + \gamma - 1)u^H$ from the left yields

$$\lambda^2 + (\gamma - 1 - \xi_1 - \gamma\xi_2)\lambda + (1 - \gamma)\xi_1 = 0. \tag{4.6}$$

For convenience, we define

$$f(\lambda) = \lambda^2 + (\gamma - 1 - \xi_1 - \gamma\xi_2)\lambda + (1 - \gamma)\xi_1. \tag{4.7}$$

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 complex eigenvalue of the matrix $M(\gamma)$. For this case, the matrix $M(\gamma)$ contains only real numbers and thus its complex eigenvalues must occur in conjugate pairs (see, e.g., [32]). Thus, $\bar{\lambda}$ is also an eigenvalue of the matrix $M(\gamma)$. Moreover, it can be easily shown that $\left(\bar{\lambda}, \begin{pmatrix} u \\ w \end{pmatrix}\right)$ is also an eigenpair of $M(\gamma)$. Consequently, both λ and $\bar{\lambda}$ are roots of the equation $f(\lambda) = 0$. Thus, $\Delta < 0$. For this case, although the vectors u and w are complex vectors, we can show that $(1-\gamma)\xi_1$ is a real number due to $(1 - \gamma)\xi_1 = \lambda \cdot \bar{\lambda}$. It further implies that ξ_1 is a real number, i.e.,

$$\bar{\xi}_1 = (u^H FGu)^H = u^H FGu = \xi_1.$$

Recall that F and G are two real symmetric matrices. Then, using the above equality, we get

$$\bar{\xi}_2 = [u^H(I - F - G - FG)u]^H = u^H(I - F - G - FG)u = \xi_2,$$

which implies that ξ_2 is also a real number. For this case, the equation $f(\lambda) = 0$ still has real coefficients. Finally, since $(1 - \gamma)\xi_1 \leq 1$, i.e., $\lambda \cdot \bar{\lambda} = (1 - \gamma)\xi_1 < 1$ which is due to Lemma 2.3, and λ is a complex eigenvalue, we have $|\lambda| < 1$.

Case 2. 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 2.5, we have

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

Thus, $f(1) = \gamma - \gamma(\xi_1 + \xi_2) \geq 0$. 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 $\xi_1 \geq 0$, then $f(0) \geq 0$. It implies that the two roots of $f(\lambda) = 0$ have the same sign. Moreover, we have $|(1-\gamma)\xi_1| < 1$ and $|\gamma - 1 - \xi_1 - \gamma\xi_2| \leq 2$. We conclude that the equation (4.6) 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 2.5, we have $0 \leq \lambda_2 \leq 1$.

III) $\gamma \in (1,2)$.
If $\xi_1 \leq 0$, then we have $\xi_2 \geq 0$ because of $\xi_1 + \xi_2 \geq 0$. Thus, we have $f(-1) > 0$, $f(0) \geq 0$ and $f(1) \geq 0$. Also, we have $|(1-\gamma)\xi_1| < 1$ and $|\gamma - 1 - \xi_1 - \gamma\xi_2| < 2$. We conclude that the equation (4.6) has two real roots, and both of them belong to either $[0,1]$ or $(-1,0]$.

If $\xi_1 > 0$, then $f(1) \geq 0$ 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 $f(-1) \leq 0$, it implies that the equation (4.6) has a root $\lambda_2 \leq -1$. 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 (4.1). It follows from (4.4) that λ_2 is a root of the following equation:

$$\det\left(FG + (I-F)\frac{\lambda\gamma}{\lambda+\gamma-1}(I-G) - \lambda I\right) = 0. \tag{4.8}$$

We denote

$$\kappa := \frac{\lambda\gamma}{\lambda+\gamma-1}. \tag{4.9}$$

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 (4.8) that

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

Note that $\left([(1+\kappa)G - \kappa I]\left[(\kappa-\lambda)I - \kappa G\right]^{-1}\right)$ 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 2.6, we get

$$\begin{aligned} q(\lambda) &:= \det(F)\cdot\det\left([(1+\kappa)G - \kappa I]\left[(\kappa-\lambda)I - \kappa G\right]^{-1} + F^{-1}\right)\cdot\det\left((\kappa-\lambda)I - \kappa G\right) \\ &\geq \left(\prod_{i=1}^m f_i\right)\left\{\min_{\pi}\prod_{i=1}^m\left(\frac{(1+\kappa)g_i - \kappa}{(\kappa-\lambda) - \kappa g_i} + \frac{1}{f_{\pi(i)}}\right)\right\}\left(\prod_{i=1}^m\left[(\kappa-\lambda) - \kappa g_i\right]\right) \\ &\geq \left(\prod_{i=1}^m f_i\right)\left\{\prod_{i=1}^m\left(\frac{(1+\kappa)g_i - \kappa}{(\kappa-\lambda) - \kappa g_i} + \frac{1}{f_{m+1-i}}\right)\right\}\left(\prod_{i=1}^m\left[(\kappa-\lambda) - \kappa g_i\right]\right) \\ &\geq \prod_{i=1}^m\left\{-\frac{1}{\lambda+\gamma-1}\left[\lambda^2 + \lambda\left(-1 + \gamma g_i - (1+\gamma)g_i f_{m+1-i} + \gamma f_{m+1-i}\right) + (1-\gamma)f_{m+1-i}g_i\right]\right\}. \end{aligned} \tag{4.10}$$

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 2.6, and the last follows from (4.9). If we denote

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

then we have

$$q(\lambda) \geq \prod_{i=1}^m\left[-\frac{1}{\lambda+\gamma-1}q_i(\lambda)\right]. \tag{4.11}$$

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

$$q_i(1) = \gamma(g_i + f_{m+1-i} - 2f_{m+1-i}g_i) \geq 0,\ q_i(0) = (1-\gamma)g_i f_{m+1-i} < 0, \tag{4.12}$$

and

$$q_i(-1) = 2 + 2f_{m+1-i}g_i - \gamma f_{m+1-i} - \gamma g_i > 2 + 2f_{m+1-i}g_i - 2f_{m+1-i} - 2g_i \geq 0. \tag{4.13}$$

Combining (4.12) and (4.13), we have

$$q_i(\lambda) > 0 \text{ when } \lambda \leq -1,\ \ \forall i = 1, \ldots, m. \tag{4.14}$$

Also note that

$$-\frac{1}{\lambda + \gamma - 1} > 0 \text{ when } \lambda \leq -1 \text{ and } \gamma \in (1,2). \tag{4.15}$$

Combining (4.11), (4.14) and (4.15), we get

$$q(\lambda) > 0,\ \text{when } \lambda \leq -1,\ \gamma \in (1,2).$$

Recalling the definition of $q(\lambda)$ in (4.10), the above result is contradicted with (4.8). 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., [32]), we can also show that the eigenvalues of the matrix $M(\gamma)$ are not less than or equal to -1.

The proof is complete. □

Remark 4.2. From the proof of Theorem 4.1, we know that if 1 is an eigenvalue of $M(\gamma)$, then $f(1) = 0$ where $f(\lambda)$ is defined in (4.7). It implies that $\xi_1 + \xi_2 = 1$. That is, 1 is an eigenvalue of the matrix $(I - F - G + 2FG)$.

Theorem 4.1 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 [7].

Theorem 4.3. *Let F and G be two symmetric matrices in $\mathbb{R}^{m\times m}$ and they satisfy $\mathbf{0} \preceq F \preceq I$ and $\mathbf{0} \preceq G \preceq I$. If 1 is an eigenvalue of the matrix $M(\gamma)$ defined in (4.1), then the algebraic multiplicity of 1 for $M(\gamma)$ equals its geometric multiplicity.*

Proof. It follows from the definition of $M(\gamma)$ in (4.1) that

$$\begin{aligned}
&\det(\lambda I - M(\gamma)) \\
&= \det\begin{pmatrix} \lambda I - GF & -G + GF \\ -\gamma F + \gamma GF & (\lambda-1)I + \gamma F + \gamma G - \gamma GF \end{pmatrix} \\
&= \det\begin{pmatrix} \lambda I - GF & -G + GF \\ \lambda\gamma I - \gamma F & (\lambda-1)I + \gamma F \end{pmatrix} \\
&= \det\begin{pmatrix} \lambda I - G & -G + GF \\ (\lambda\gamma + \lambda - 1)I & (\lambda-1)I + \gamma F \end{pmatrix} \\
&= (-1)^m \det\begin{pmatrix} -G + GF & \lambda I - G \\ (\lambda-1)I + \gamma F & (\lambda\gamma + \lambda - 1)I \end{pmatrix} \\
&= (-1)^m \det\begin{pmatrix} GF - G - \frac{1}{\lambda\gamma+\lambda-1}(\lambda I - G)\left[(\lambda-1)I + \gamma F\right] & \lambda I - G \\ \mathbf{0} & (\lambda\gamma + \lambda - 1)I \end{pmatrix} \\
&= \det\left[\lambda^2 I - \lambda I + (\gamma F + \gamma G - (1+\gamma)GF)\lambda + (1-\gamma)GF\right] \\
&= \det\left[\lambda^2 I - \lambda I + \gamma\left(F + G - 2GF\right)\lambda + (\gamma - 1)GF(\lambda - 1)\right].
\end{aligned} \tag{4.16}$$

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 2.13, we know that there exists an orthogonal matrix $Q\in\mathbb{R}^{m\times m}$ such that (2.11) holds. Consequently, we have

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

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

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

where the matrix $\Upsilon:=\gamma\lambda(R+S-2(DR+S\hat{F}))+(\gamma-1)(\lambda-1)(DR+S\hat{F})$. Taking the determinant on both sides in the above equation, we get

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

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 2.13, we have

$$\mathbf{0}\prec\hat{F}+\hat{G}-2\hat{G}\hat{F}\preceq I_{m-\ell}. \tag{4.17}$$

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 (4.17). 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 (2.8). On the other hand, the geometric multiplicity of 1 for $M(\gamma)$ is identical with the following quality:

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

$$= m - \text{Rank}(-G - F + 2GF).$$

Invoking Lemma 2.7, we conclude that the geometric multiplicity of eigenvalue 1 for $M(\gamma)$ is also ℓ. The proof is complete. □

Remark 4.4. 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 2.8. From the proof of Theorem 4.3 (see (4.16)), 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 4.2, 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 (3.11). This is the essential pillar for proving the convergence of the scheme (3.1) with $\gamma \in (0, 2)$. A lemma is proved first.

Lemma 4.5. *Assumptions 1-2 hold; the matrices $\hat{Q}$ and $\hat{P}$ are defined in (3.5); $\gamma \in (0, 2)$; the matrix $T(\gamma)$ is defined in (3.9). Then, we have $|\lambda_{T(\gamma)}| < 1$ or $\lambda_{T(\gamma)} = 1$. Furthermore, if 1 is an eigenvalue of $T(\gamma)$, then the algebraic and geometric multiplicities of 1 for $T(\gamma)$ are the same.*

Proof. Setting $G = B\hat{Q}^{-1}B^\top$ and $F = A\hat{P}^{-1}A^\top$ in $M(\gamma)$ (see (4.1)), and invoking the definition of $\tilde{T}(\gamma)$ in (3.12), we have

$$M(\gamma) = \tilde{T}(\gamma). \tag{4.18}$$

Thus, it holds $\text{eig}(M(\gamma)) = \text{eig}(\tilde{T}(\gamma))$. Indeed, according to (3.13) , we have $\text{eig}(T(\gamma)) = \text{eig}(M(\gamma))$. It follows from Theorem 4.1 that $|\lambda_{T(\gamma)}| < 1$ or $\lambda_{T(\gamma)} = 1$.

If 1 is an eigenvalue of the matrix $T(\gamma)$, using Theorem 4.3 and (4.18), 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 (3.1) with $\gamma \in (0, 2)$. The proof of the following theorem is inspired by Theorem 3 in [7].

Theorem 4.6. *Assumptions 1-2 hold. Let $\{(x^k, y^k, z^k)\}$ be the sequence generated by the scheme (3.1), i..e, the application of the ADMM scheme (1.4) with $\gamma \in (0, 2)$ to the quadratic programming model (1.7). Then, the sequence $\{(x^k, y^k, z^k)\}$ converges to a KKT point of (1.7).*

Proof. Let (x^*, y^*, z^*) be a KKT point of (1.7). As shown in Lemma 3.1, we have

$$\begin{pmatrix} y^* \\ \mu^* \end{pmatrix} = T(\gamma) \begin{pmatrix} y^* \\ \mu^* \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \tag{4.19}$$

where $T(\gamma)$, q_1 and q_2 are defined in (3.9), (3.10), respectively. It follows from (3.8) and (4.19) that

$$(v^{k+1} - v^*) = T(\gamma)(v^k - v^*).$$

Recall the definition of μ^k and the matrix recursion form (3.8)-(3.10). It follows from Lemma 4.5 that $|\lambda_{T(\gamma)}| < 1$ or $\lambda_{T(\gamma)} = 1$. Next, we consider two cases to complete the proof.

The first case is where $|\lambda_{T(\gamma)}| < 1$. For this case, it holds that $v^k \to v^*$ as $k \to \infty$ and the assertion of the theorem is obvious. The second case is where $\lambda_{T(\gamma)} = 1$. For this case, by lemma 4.5, we know that the eigenvalue 1 for $T(\gamma)$ has a complete set of eigenvectors. As a result, there exists a nonsingular matrix J such that $T(\gamma)$ admits the following Jordan decomposition:

$$T(\gamma) = J^{-1} \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \rho_1 & * & \\ & & & & \ddots & * \\ & & & & & \rho_t \end{pmatrix} J,$$

where $|\rho_i| < 1$ for all $i = 1, \ldots, t$. Moreover, it can be shown that

$$(T(\gamma))^k = J^{-1} \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} J,$$

when $k \to \infty$. Therefore, the sequence $\{v^k - v^*\}$ converges to an eigenvector of the matrix $T(\gamma)$ associated with the eigenvalue 1, denoted by $\bar{v} = (\bar{y}; \bar{\mu})$. Then, we have

$$(I - T(\gamma)) \begin{pmatrix} \bar{y} \\ \bar{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

Adding the above equation to (4.19), we obtain

$$(I - T(\gamma)) \begin{pmatrix} \bar{y} + y^* \\ \bar{\mu} + \mu^* \end{pmatrix} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}.$$

Let us denote $(\bar{v} + v^*)$ by $\hat{v}$. Then, we have that $\{v^k\}$ converges to $\hat{v}$ because $(v^k - v^*) \to \bar{v}$ and $\hat{v}$ also satisfies

$$(I - T(\gamma))\hat{v} = q. \tag{4.20}$$

Let $k \to \infty$ in (3.3). Since $v^k \to \hat{v}$, there exists a vector $\hat{x} \in \mathbb{R}^{n_1}$ such that $x^k \to \hat{x}$ and $\hat{x}$ is satisfied with

$$\hat{P}\hat{x} = A^\top \hat{\mu} - A^\top B\hat{y} + A^\top b - f/\beta. \tag{4.21}$$

Finally, combining (4.20) and (4.21), and invoking Lemma 3.1, we conclude that $(\hat{x}, \hat{y}, \hat{z})$ is a KKT point of the model (1.7). The proof is complete. □

4.2 Numerical verification of the convergence

In this section, we construct a simple example of (1.7) and numerically verify the convergence of (3.1) with $\gamma \in (0, 2)$. In particular, as well observed in the literature, e.g., [17, 18, 19, 26, 28, 39, 36], it is advantageous to employ larger value of γ closer to 2 to accelerate the convergence in the scheme (3.2). 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 (1.7), and the resulting model has a unique solution $x^* = y^* = \mathbf{0}$. The matrix P and Q in (1.7) are generated by

$$P1 = randn(n_1, n_1);\ P = P1' * P1;\ a = eigs(P, 1, 'sm');\ P = P - (a - (1e-4)) * eye(n_1)$$

and

$$Q1 = randn(n_2, n_2);\quad Q = Q1' * Q1;\ b = eigs(Q, 1, 'sm');\ Q = Q - (b - (1e-4)) * eye(n_1)$$

respectively. Furthermore, the matrices $A \in \mathbb{R}^{m \times n_1}$ and $B \in \mathbb{R}^{m \times n_2}$ in (1.7) 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 (3.1), let us fix $\beta = 1$, $y^0 = randn(n_2, 1)$, $z^0 = randn(m, 1)$, and the stopping criterion is (see [5])

$$\texttt{err} := \max\{\|B(y^k - y^{k+1})\|_2, \|z^k - z^{k+1}\|_2\} \le 10^{-6}. \tag{4.22}$$

We test different scenarios of this example where $m = n_1 = n_2 = 50$, 100, 200, 500, respectively. The values of γ from 0.2 to 1.8 with an equal distance of 0.2 are tested. 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 of x^k (measured by $\|\bar{x} - x^*\|_2$), the error of 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 (4.22). The condition numbers of P and Q ("Cond(P)" and "Cond(Q)", respectively) are also included in Table 1. Data in this table verify the convergence of (3.1) with $\gamma \in (0, 2)$ and the acceleration with γ close to 2. In particular, it is shown that the cases of $\gamma > 1$ can easily accelerate the case of $\gamma = 1$, i.e., the original ADMM; and that some values larger than 1.618 also result in faster convergence considerately. Hence, it is verified to consider larger values for γ in Glowinski's ADMM scheme (1.4).

5 Global linear convergence of (3.1)

In addition to the main purpose of establishing the convergence of the scheme (3.1) with $\gamma \in (0, 2)$ and answering Glowinski's open question partially, in this section we show the global linear convergence of the scheme (3.1) with $\gamma \in (0, 2)$ under a condition. This is a supplementary result to the main convergence result in Section 4.

5.1 Review of existing results

The linear convergence of the ADMM (1.4) with the special case of $\gamma = 1$ has been discussed in the quadratic programming context in [4, 24] under different conditions. Let us briefly review them. In [24], the local linear convergence of a generalized version of the ADMM proposed in [11], which reduces to the original ADMM (1.4) when the parameter is taken as 1, is established for the quadratic programming model (1.7) under some local error bound conditions. In [4], the following convex quadratic programming model is considered:

$$\begin{cases} \min_x & \frac{1}{2}x^\top Qx + c^\top x + g(y) \\ s.t. & Ax = b, \\ & x = y, \end{cases} \quad \text{with} \;\; g(y) = \begin{cases} 0 & \text{if } y \geq 0; \\ +\infty & \text{if } y \ngeq 0. \end{cases} \tag{5.1}$$

Then, the following ADMM scheme is suggested in [4]:

$$\begin{cases} x^{k+1} = \arg\min_{Ax=b} \mathcal{L}'_\beta(x, y^k, z^k), \\ y^{k+1} = \arg\min \mathcal{L}'_\beta(x^{k+1}, y, z^k), \\ z^{k+1} = z^k - \beta(x^{k+1} - y^{k+1}), \end{cases} \tag{5.2}$$

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 (5.1). Then, based on the typical spectral analysis for a homogeneous linear equation characterizing the corresponding matrix recursion form, the local linear convergence is established for the scheme (5.2). It is worthwhile to mention that the iterative matrix considered in the homogeneous linear equation varies iteratively and as analyzed in [4], four regimes occur. Assuming the convergence (e.g., by results in [5]), the uniqueness of solution, and the strict complementarity condition (See Theorem 6.4 in [4]), it is proved in [4] that the iterative matrices finally become fixed with a spectral radius less than 1 and hence the local linear convergence is derived therein for (5.1).

Table 1: Convergence of (3.1)

m	n_1	n_2	γ	$\Vert\bar{x} - x^*\Vert_2$	$\Vert\bar{y} - y^*\Vert_2$	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	
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	

5.2 Global linear convergence under a tight condition

In this section, we establish the global linear convergence of the scheme (3.2) with $\gamma \in (0,2)$ under a new assumption different from those in [4, 24]. Recall the iterative matrix $T(\gamma)$ defined in (3.9). If $\rho(T(\gamma)) < 1$, then the linear convergence of the sequence $\{v^k\}$ generated by (3.8)-(3.9) 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 5.1. *Assumptions 1-2 hold. Assume that*

$$N(B\hat{Q}^{-1}B^\top - I)\bigcap N(A\hat{P}^{-1}A^\top - I) = \{\mathbf{0}\} \ \text{ and } \ N(B\hat{Q}^{-1}B^\top)\bigcap N(A\hat{P}^{-1}A^\top) = \{\mathbf{0}\}, \tag{5.3}$$

with $\hat{Q}$ and $\hat{P}$ defined in (3.5). Then, the sequence $\{(x^k, y^k, z^k)\}$ generated by the scheme (3.1) with $\gamma \in (0,2)$ converges linearly to a KKT point of (1.7).

Proof. Setting $G = B\hat{Q}^{-1}B^\top$ and $F = A\hat{P}^{-1}A^\top$ in $M(\gamma)$ (see (4.1)), and combining the proof in Lemma 4.5, we know that

$$\sigma(T(\gamma)) = \sigma(M(\gamma)).$$

Invoking Lemma 2.9 and Remark 4.4, we have

$$1 \notin \sigma(M(\gamma)) \Leftrightarrow 1 \notin \sigma(I + 2GF - F - G) \Leftrightarrow (5.3).$$

Consequently, $\rho(M(\gamma)) < 1$ when (5.3) is satisfied. Then, there is a matrix norm $\|\cdot\|_G$ such that $\rho(M(\gamma)) \le \|M(\gamma)\|_G < \rho(M(\gamma)) + \epsilon \le 1$. 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 (3.18), and recall (3.3) and (3.18). We obtain that

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

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 5.2. The condition (5.3) for ensuring the linear convergence of (3.1) is indeed tight. To see it, notice that (5.3) implies that for either the eigenvalue 1 or 0, the matrices F and G do not have any common eigenvector. If the condition (5.3) 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 (3.8).

Remark 5.3. Note that $0 \le \lambda_{B\hat{Q}^{-1}B^\top} \le 1$ and $0 \le \lambda_{A\hat{P}^{-1}A^\top} \le 1$. Thus, it is easy to verify that the conditions

$$\begin{cases} 0 < \lambda_{B\hat{Q}^{-1}B^\top} < 1 \\ 0 < \lambda_{A\hat{P}^{-1}A^\top} < 1 \end{cases}$$

suffice to ensure the condition (5.3) and hence the linear convergence of the sequence $\{(x^k, y^k, z^k)\}$ generated by the scheme (3.1).

Remark 5.4. The linear convergence rate result in Theorem 5.1 differs from those in [4, 24] in the following aspects. (1) The linear convergence rate in Theorem 5.1 is global, while those in [4, 24] are local. (2) The condition (5.3) is different from those in [4, 24]; and it is tight. (3) Here we consider the scheme (3.1) with $\gamma \in (0,2)$ and the targeted model is (1.7); while in [4] only the special case $\gamma = 1$ is considered and its targeted model is (5.1); and in [24] the generalized ADMM in [11] is considered. Moreover, the conditions in (5.3) depend only on the matrices P, Q, A and B in the model (1.7) per se and the penalty β; it does not require any local information near the solution point such as the local error bound in [24] or the identification of the regimes of the corresponding iterative matrix in [4]. Indeed, with Assumption 1, we know that the condition "$N(B\hat{Q}^{-1}B^\top)\bigcap N(A\hat{P}^{-1}A^\top) = \{\mathbf{0}\}$" in (5.3) is equivalent to $N(B^\top)\bigcap N(A^\top) = \{\mathbf{0}\}$ and thus it does not involve the parameter β. For the condition "$N(B\hat{Q}^{-1}B^\top - I)\bigcap N(A\hat{P}^{-1}A^\top - I) = \{\mathbf{0}\}$" in (5.3), it is easy to see that the parameter β is not involved if "$P \succ \mathbf{0}$ and $Q \succ \mathbf{0}$" or "$A^\top A \succ \mathbf{0}$ and $B^\top B \succ \mathbf{0}$" is further assumed in Assumption 1.

Table 2: Linear convergence of (3.1)

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	

5.3 Numerical verification of the global linear convergence

In this subsection we numerically verify the global linear convergence of the scheme (3.1) with $\gamma \in (0,2)$ under the condition (5.3) by an example.

The details of constructing the example is nearly the same as those in Section 4.2 except for

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

We set $\beta = 1$ and thus we just need to check the following conditions to ensure (5.3):

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

with $\hat{P}$ and $\hat{Q}$ defined in (3.5). 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 (5.3) in Theorem 5.1 is ensured. The implementation details of the scheme (3.1) are the same as those in Section 4.2,

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 Figure 1, we plot the evolution of the errors to the exact solution point, i.e., $\|v^k - v^*\|_2$ with v^k defined in (3.8)), with respect to iteration numbers. The linear convergence of the scheme (3.1) is displayed in this figure for different choices of $\gamma \in (0,2)$.

To see the global feature of the linear convergence in Theorem 5.1, 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 errors of x^k (measured by $\|\bar{x} - x^*\|_2$), y^k (measured by $\|\bar{y} - y^*\|_2$), the number of iteration ("Itr.") and the CPU

Table 3: Global convergence of (3.1) 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

time in seconds ("Time(s)") are reported in Table 3. Data in Table 3 demonstrates the global convergence of the scheme (3.1) under the condition (5.3). The evolution of the errors to the exact solution point, i.e., $\|v^k - v^*\|_2$ with v^k defined in (3.8), with respect to iteration numbers are plotted for these different initial points in Figure 2. The curves in Figure 2 clearly show the linear convergence of the scheme (3.1) under the condition (5.3).

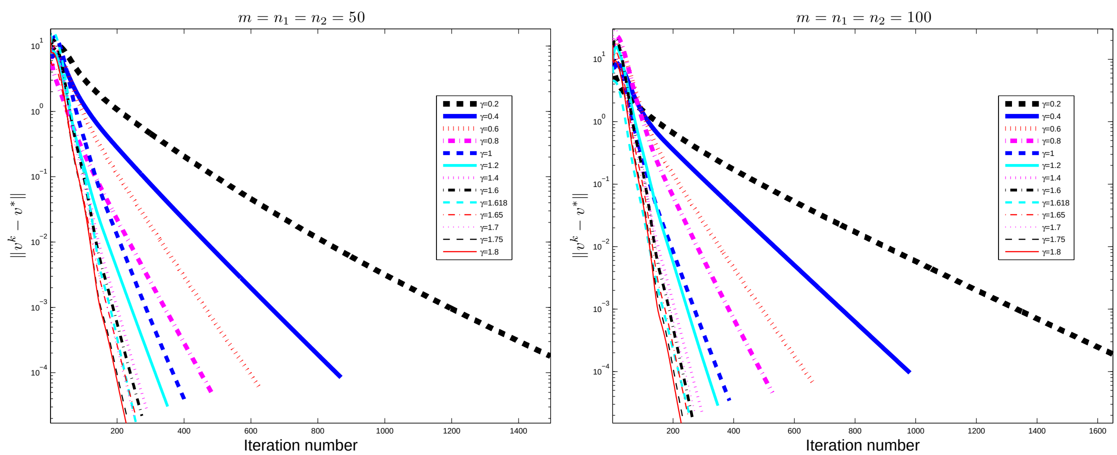

Figure 1: Linear convergence of the ADMM (3.1) for different γ

6 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 non-symmetric 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.

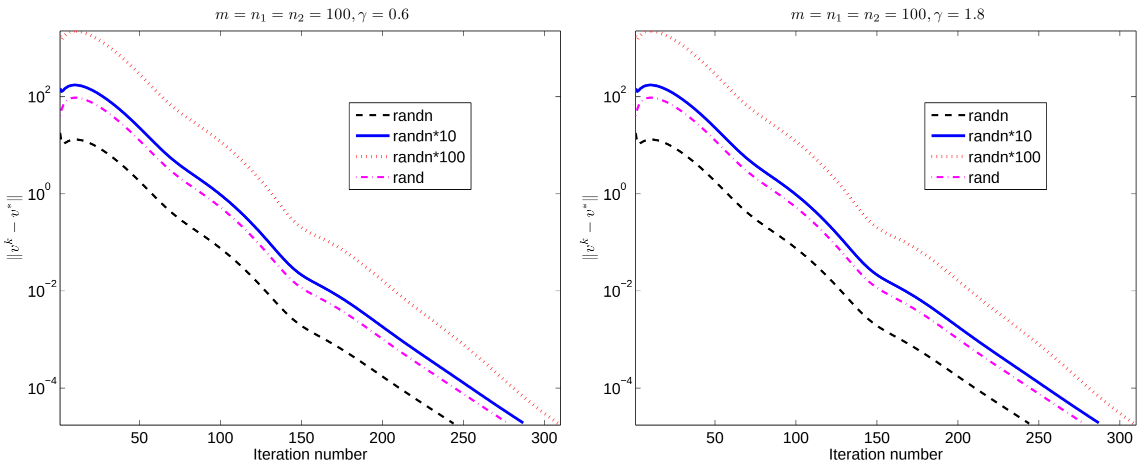

Figure 2: Linear convergence of the ADMM (3.1) for different initial points

References

[1] K. Arrow, L. Hurwicz and H. Uzawa, *Studies in Nonlinear Programming*, Stanford University Press, Stanford, 1958.

[2] A. R. Bergen, *Power Systems Analysis*, Prentice Hall, Englewood Cliffs, 1986.

[3] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, 1996.

[4] D. Boley, *Local Linear convergence of the alternating direction method of multipliers on quadratic or linear programs*, SIAM Journal on Optimization, 23(2013), pp. 2183-2207.

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

[6] A. M. Bruckstein, D. L. Donoho and M. Elad, *From sparse solutions of systems of equations to sparse modeling of signals and images*, SIAM Review, 51(2009), pp. 34-81.

[7] C. H. Chen, M. Li, X. Liu and Y. Y. Ye, *Extended ADMM and BCD for nonseparable convex minimization models with quadratic coupling terms: convergence analysis and insights*, available at http://arxiv.org/pdf/1508.00193v3.pdf.

[8] L. O. Chua, C. A. Desoer and E. S. Kuh, *Linear and Nonlinear Circuits*, McGraw-Hill, New York, 1987.

[9] J. Ding and N. H. Rhee, *On the equality of algebraic and geometric multiplicities of matrix eigenvalues*, Applied Mathematics Letters, 24(2011), pp. 2211-2215.

[10] J. Eckstein, *Some saddle-function splitting methods for convex programming*, Optimization Methods and Software, 4(1994), pp. 75-83.

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

[12] J. Eckstein and M. Fukushima, *Reformulations and applications of the alternating direction method of multipliers*, Large Scale Optimization, W. W. Hager, D. W. Hearn and P. M. Pardalos, eds., Springer US, pp. 115-134, 1994.

[13] J. Eckstein and W. Yao, *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, Pacific Journal of Optimization, 11(2015), pp. 619-644.

[14] M. Ehrgott and I. Winz, *Interactive decision support in radiation therapy treatment planning*, OR Spectrum 30(2008), pp. 311–329.

[15] M. Fiedler, *Bounds for the determinant of the sum of Hermitian matrices*, Proceedings of the American Mathematical Society, 30(1971), pp. 27-31.

[16] M. Fortin and R. Glowinski, Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems, Studies in Applied Mathematics, 15, North-Holland, Amsterdam, 1983.

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

[18] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, Computers & Mathematics with Applications, 2(1976), pp. 17-40.

[19] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[20] R. Glowinski, *On alternating direction methods of multipliers: A historical perspective*, In Modeling, Simulation and Optimization for Science and Technology, W. Fitzgibbon, Y.A. Kuznetsov, P. Neittaanmaki & O. Pironneau, eds., Computational Methods in Applied Sciences, Vol. 34, Springer, Dordrecht, 2014, pp. 59-82.

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

[22] R. Glowinski, S. J. Osher and W. Yin (eds), Splitting Methods for Communications and Imaging, Science and Engineering, Springer, 2016.

[23] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Mathematical programming studies, 10(1979), pp. 86-97.

[24] D. R. Han and X. M. Yuan, *Local linear convergence of the alternating direction method of multipliers for quadratic programs*, SIAM Journal on Numerical Analysis, 51(2013), pp. 3446-3457.

[25] B. S. He, L. Z. Liao, D. Han and H. Yang, *A new inexact alternating directions method for monontone variational inequalities*, Mathematical Programming, 92(2002), pp. 103-118.

[26] B. S. He, F. Ma and X. M. Yuan, *Convergence analysis of the symmetric version of ADMM*, SIAM Journal on Imaging Sciences, 9(2016), pp. 1467-1501.

[27] B. S. He, M. H. Xu and X. M. Yuan, *Solving large-scale least squares covariance matrix problems by alternating direction methods*, SIAM Journal on Matrix Analysis and Applications, 32(2011), pp. 136-152.

[28] B. S. He and H. Yang, *Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities*, Operations Research Letters, 23(1998), pp. 151-161.

[29] M. R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, 4(1969), pp. 303-320.

[30] H. M. Markowitz, *Porfolio Selection: Efficient Diversification of Inverstments*, Wiley, New York, 1959.

[31] B. Martinet, *Regularisation, d'inéquations variationelles par approximations succesives*, Rev. Francaise d'Inform. Recherche Oper., 4(1970), pp. 154-159.

[32] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2006.

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

[34] R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Mathematical Methods of Operations Research, 1(1976), pp. 877-898.

[35] J. Sun and S. Zhang, *A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs*, European Journal of Operational Research, 207(2010), pp. 1210-1220.

[36] M. Tao and X. M. Yuan, *On the O(1/t) convergence rate of alternating direction method with logarithmic-quadratic proximal regularization*, SIAM Journal on Optimization, 22(2012), pp. 1431-1448.

[37] M. Tao and X. M. Yuan, *The generalized proximal point algorithm with step size* 2 *is not necessarily convergent*, manuscript, 2016.

[38] A. Tikhonov and V. Arsenin, *Solution of Ill-Posed problems*, Winston, Washington, DC, 1977.

[39] M. H. Xu, *Proximal alternating directions method for structured variational inequalities*, Journal of Optimization Theory and Applications, 134(2007), pp. 107-117.

[40] Z. W. Wen, D. Goldfarb and W. T. Yin, *Alternating direction augmented Lagrangian methods for semidefinite programming*, Mathematical Programming Computation, 2(2010), pp. 203-230.