Noname manuscript No.
(will be inserted by the editor)

The exact worst-case convergence rate of the alternating direction method of multipliers

Moslem Zamani · Hadi Abbaszadehpeivasti · Etienne de Klerk

Received: date / Accepted: date

Abstract Recently, semidefinite programming performance estimation has been employed as a strong tool for the worst-case performance analysis of first order methods. In this paper, we derive new non-ergodic convergence rates for the alternating direction method of multipliers (ADMM) by using performance estimation. We give some examples which show the exactness of the given bounds. We also study the linear and R-linear convergence of ADMM. We establish that ADMM enjoys a global linear convergence rate if and only if the dual objective satisfies the Polyak-Łojasiewicz (PŁ) inequality in the presence of strong convexity. In addition, we give an explicit formula for the linear convergence rate factor. Moreover, we study the R-linear convergence of ADMM under two new scenarios which are weaker than the existing ones in the literature.

Keywords Alternating direction method of multipliers (ADMM) · Performance estimation · Convergence rate · PŁ inequality

Mathematics Subject Classification (2020) 90C22 · 90C25 · 65K15

This work was supported by the Dutch Scientific Council (NWO) grant OCENW.GROOT.2019.015, *Optimization for and with Machine Learning (OPTIMAL)*.

M. Zamani
Tilburg University, Department of Econometrics and Operations Research, Tilburg, The Netherlands
E-mail: m.zamani_1@tilburguniversity.edu

H. Abbaszadehpeivasti
Tilburg University, Department of Econometrics and Operations Research, Tilburg, The Netherlands
E-mail: h.abbaszadehpeivasti@tilburguniversity.edu

E. de Klerk
Tilburg University, Department of Econometrics and Operations Research, Tilburg, The Netherlands
E-mail: e.deklerk@tilburguniversity.edu

1 Introduction

We consider the optimization problem

$$\min_{(x,z)\in\mathbb{R}^n\times\mathbb{R}^m} f(x)+g(z), \quad \text{s.t. } Ax+Bz=b, \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}\cup\{\infty\}$ and $g:\mathbb{R}^m\to\mathbb{R}\cup\{\infty\}$ are closed proper convex functions, $0\neq A\in\mathbb{R}^{r\times n}$, $0\neq B\in\mathbb{R}^{r\times m}$ and $b\in\mathbb{R}^r$. In addition, without loss of generality, we assume that the matrix $\begin{pmatrix}A & B\end{pmatrix}$ has full row rank. Moreover, we assume that $(x^\star,z^\star)$ is an optimal solution of problem (1) and $\lambda^\star$ is its corresponding Lagrange multipliers.

Problem (1) appears naturally (or after variable splitting) in many applications in statistics, machine learning and image processing to name but a few [5, 19, 24, 34]. The most common method for solving problem (1) is the alternating direction method of multipliers (ADMM). ADMM is dual based approach that exploits separable structure and it may be described as follows.

Algorithm 1 ADMM

Set N and $t>0$ (step length), pick $\lambda^0,\hat{z}$ and $z^0=\hat{z}$.
For $k=1,2,\ldots,N$ perform the following step:
1. $x^k\in\operatorname{argmin} f(x)+\langle\lambda^{k-1},Ax\rangle+\frac{t}{2}\|Ax+Bz^{k-1}-b\|^2$
2. $z^k\in\operatorname{argmin} g(z)+\langle\lambda^{k-1},Bz\rangle+\frac{t}{2}\|Ax^k+Bz-b\|^2$
3. $\lambda^k=\lambda^{k-1}+t(Ax^k+Bz^k-b)$.

ADMM was first proposed in [10,12] for solving nonlinear variational problems. We refer the interested reader to [13] for a historical review of ADMM. The popularity of ADMM is due to its capability to be implemented parallelly and hence can handle large-scale problems [5,18,28,37]. For example, it is used for solving inverse problems governed by partial differential equation forward models [27], and distributed energy resource coordinations [25], to mention but a few.

The convergence of ADMM has been investigated extensively in the literature and there exist many convergence results. However, different performance measures have been used for the computation of convergence rate; see [9, 14, 15, 20, 23, 24, 29, 36]. In this paper, we consider the dual objective value as a performance measure.

Throughout the paper, we assume that each subproblem in steps 1 and 2 of Algorithm 1 attains its minimum. The dual objective of problem (1) is defined as

$$D(\lambda)=\min_{(x,z)\in\mathbb{R}^n\times\mathbb{R}^m} f(x)+g(z)+\langle\lambda,Ax+Bz-b\rangle.$$

We assume that strong duality holds for problem (1), that is

$$\max_{\lambda\in\mathbb{R}^r} D(\lambda) = \min_{Ax+Bz=b} f(x) + g(z).$$

Note that we have strong duality when both functions f and g are real-valued. For extended convex functions, strong duality holds under some mild conditions; see e.g. [3, Chapter 15].

As mentioned earlier, we compute convergence rate in terms of dual objective value. Regarding this criterion, the following convergence rate is known in the literature. This theorem holds for strongly convex functions f and g; recall that f is called strongly convex with modulus $\mu \geq 0$ if the function $f - \frac{\mu}{2}\|\cdot\|^2$ is convex.

Theorem 1 *[15, Theorem 4] Let f and g be strongly convex with moduli $\mu_1 > 0$ and $\mu_2 > 0$, respectively. If $t \leq \sqrt[3]{\frac{\mu_1\mu_2^2}{\lambda_{\max}(A^TA)\lambda_{\max}^2(B^TB)}}$, then*

$$D(\lambda^\star) - D(\lambda^N) \leq \frac{\|\lambda^1 - \lambda^\star\|^2}{2t(N-1)}. \tag{2}$$

In this study we establish that Algorithm 1 has the convergence rate of $O(\frac{1}{N})$ in terms of dual objective value without assuming the strong convexity of g. Under this setting, we also prove that Algorithm 1 has the convergence rate of $O(\frac{1}{N})$ in terms of primal and dual residuals. Moreover, we show that the given bounds are exact.

Outline of our paper

Our paper is structured as follows. We present the semidefinite programming (SDP) performance estimation method in Section 2. We develop performance estimation to handle dual based methods including ADMM, and derive some new non-asymptotic convergence rates by using performance estimation for ADMM in terms of dual function, primal and dual residuals. Furthermore, we show that the given bounds are tight by providing some examples. In Section 3 we proceed with the study of the linear convergence of ADMM. We establish that ADMM enjoys a linear convergence if and only if the dual function satisfies the PŁ inequality when the objective function is strongly convex. Furthermore, we investigate the relation between the PŁ inequality and common conditions used by scholars to prove linear convergence. Section 4 is devoted to the R-linear convergence. We prove that ADMM is R-linear convergent under two new scenarios which are weaker than the existing ones in the literature.

Terminology and notation

In this subsection we review some definitions and concepts from convex analysis. The interested reader is referred to the classical text by Rockafellar [33] for more information. The n-dimensional Euclidean space is denoted by $\mathbb{R}^n$.

We use $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ to denote the Euclidean inner product and norm, respectively. The column vector e_i represents the i-th standard unit vector and I stands for the identity matrix. For a matrix A, A_{ij} denotes its (i,j)-th entry, and A^T represents the transpose of A. The notation $A \succeq 0$ means the matrix A is symmetric positive semidefinite. We use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote the largest and the smallest eigenvalue of symmetric matrix A, respectively.

Suppose that $f:\mathbb{R}^n \to (-\infty,\infty]$ is an extended convex function. The function f is called closed if its epi-graph is closed, that is $\{(x,r): f(x)\le r\}$ is a closed subset of $\mathbb{R}^{n+1}$. The function f is said to be proper if there exists $x\in\mathbb{R}^n$ with $f(x)<\infty$. The subgradients of f at x is denoted and defined as

$$\partial f(x) = \{\xi : f(y) \ge f(x) + \langle \xi, y-x\rangle, \forall y\in\mathbb{R}^n\}.$$

We denote the set of closed proper μ-strongly convex functions on $\mathbb{R}^n$ by $\mathcal{F}_\mu(\mathbb{R}^n)$. We denote the distance function to the set X by $d_X(x) := \inf_{y\in X}\|y-x\|$.

In the following sections we derive some new convergence rates for ADMM by using performance estimation. The main idea of performance estimation is based on interpolablity. Let $\mathcal{I}$ be an index set and let $\{(x^i;g^i;f^i)\}_{i\in\mathcal{I}} \subseteq \mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}$. A set $\{(x^i;\xi^i;f^i)\}_{i\in\mathcal{I}}$ is called $\mathcal{F}_\mu$-interpolable if there exists $f\in\mathcal{F}_\mu(\mathbb{R}^n)$ with

$$f(x^i) = f^i,\ \xi^i \in \partial f(x^i)\quad i\in\mathcal{I}.$$

The next theorem gives necessary and sufficient conditions for $\mathcal{F}_\mu$-interpolablity.

Theorem 2 *[38, Theorem 4.] Let $\mu\in[0,\infty)$ and let $\mathcal{I}$ be an index set. The set $\{(x^i;\xi^i;f^i)\}_{i\in\mathcal{I}} \subseteq \mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}$ is $\mathcal{F}_\mu$-interpolable if and only if for any $i,j\in\mathcal{I}$, we have*

$$\tfrac{\mu}{2}\left\|x^i - x^j\right\|^2 \le f^i - f^j - \left\langle \xi^j, x^i - x^j\right\rangle. \tag{3}$$

Moreover, $\mathcal{F}_0$-interpolable and L-smooth if and only if for any $i,j\in\mathcal{I}$, we have

$$\tfrac{1}{2L}\left\|g^i - g^j\right\|^2 \le f^i - f^j - \left\langle g^j, x^i - x^j\right\rangle. \tag{4}$$

Note that any convex function is 0-strongly convex. We call a differentiable function f L-smooth if for any $x_1,x_2\in\mathbb{R}^n$,

$$\|\nabla f(x_1) - \nabla f(x_2)\| \le L\|x_1 - x_2\|\quad \forall x_1,x_2\in\mathbb{R}^n.$$

Let $f\in\mathcal{F}_\mu(\mathbb{R}^n)$. The conjugate function $f^*:\mathbb{R}^n\to(-\infty,\infty]$ is defined as $f^*(y) = \sup_{x\in\mathbb{R}^n}\langle y,x\rangle - f(x)$. We have the following identity

$$\xi\in\partial f(x)\quad\Leftrightarrow\quad x\in\partial f^*(\xi). \tag{5}$$

Let $f\in\mathcal{F}_\mu(\mathbb{R}^n)$ with $\mu>0$. The function f is μ-strongly convex if and only if f^* is $\frac{1}{\mu}$-smooth. Moreover, $(f^*)^* = f$.

By using conjugate functions, the dual of problem (1) may be written as

$$\begin{aligned} D(\lambda) &= \min_{(x,z)\in\mathbb{R}^n\times\mathbb{R}^m} f(x) + g(z) + \langle \lambda, Ax + Bz - b\rangle \\ &= -\langle \lambda, b\rangle - f^*(-A^T\lambda) - g^*(-B^T\lambda). \end{aligned} \tag{6}$$

The optimality conditions for the subproblems of Algorithm 1 may be written as

$$\begin{aligned} &0 \in \partial f(x^k) + A^T\lambda^{k-1} + tA^T\left(Ax^k + Bz^{k-1} - b\right), \\ &0 \in \partial g(z^k) + B^T\lambda^{k-1} + tB^T\left(Ax^k + Bz^k - b\right). \end{aligned} \tag{7}$$

As $\lambda^k = \lambda^{k-1} + t(Ax^k + Bz^k - b)$, we get

$$0 \in \partial f(x^k) + A^T\lambda^k + tA^TB\left(z^{k-1} - z^k\right), \quad 0 \in \partial g(z^k) + B^T\lambda^k. \tag{8}$$

Hence, (x^k, z^k) is feasible for dual objective at λ^k if and only if $A^TB\left(z^{k-1} - z^k\right) = 0$. We call $A^TB\left(z^{k-1} - z^k\right)$ dual residual.

2 Worst-case convergence rate

In this section, we derive new convergence rates for ADMM by using performance estimation. The performance estimation method introduced by Drori and Teboulle [8] is an SDP-based method for the analysis of first order methods. Since then, many scholars employed this strong tool to derive the worst case convergence rate of different iterative methods; see [1, 22, 35, 38] and the references therein. Moreover, Gu and Yang [16] employed performance estimation to study the extension of the dual step length for ADMM. Note that while there are some similarities between our work and [16] in using performance estimation, the formulations and results are different.

The worst-case convergence rate of Algorithm 1 with respect to criterion (6) may be cast as the following abstract optimization problem,

$$\begin{aligned} \max\ & D(\lambda^\star) - D(\lambda^N) \\ \text{s. t. } & \{x^k, z^k, \lambda^k\}_1^N \text{ is generated by Algorithm 1 w.r.t. } f, g, A, B, b, \lambda^0, \hat{z}, t \\ & (x^\star, z^\star) \text{ is an optimal solution with Lagrangian multipliers } \lambda^\star \\ & \|\lambda^0 - \lambda^\star\|^2 + t^2\left\|B\left(\hat{z} - z^\star\right)\right\|^2 = \Delta \\ & f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n), g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m) \\ & \lambda_{\max}(A^TA) = \nu_1, \lambda_{\max}(B^TB) = \nu_2 \\ & \lambda_0 \in \mathbb{R}^r, \hat{z} \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}, b \in \mathbb{R}^r, \end{aligned} \tag{9}$$

where $f, g, A, B, b, \hat{z}, \lambda^0, x^\star, z^\star, \lambda^\star$ are decision variables and $N, t, \mu_1, \mu_2, \nu_1, \nu_2, \Delta$ are the given parameters.

By using Theorem 2 and the optimality conditions (7), problem (9) may be reformulated as the finite dimensional optimization problem,

$$\begin{aligned}
\max\ & D(\lambda^\star) - D(\lambda^N) \\
\text{s. t.}\ & \{(x^k;\xi^k;f^k)\}_1^N \cup \{(x^\star;\xi^\star;f^\star)\} \text{ satisfy interpolation constraints (3)} \\
& \{(z^k;\eta^k;g^k)\}_0^N \cup \{(z^\star;\eta^\star;g^\star)\} \text{ satisfy interpolation constraints (3)} \\
& (x^\star, z^\star) \text{ is an optimal solution with Lagrangian multipliers } \lambda^\star \\
& \|\lambda^0 - \lambda^\star\|^2 + t^2 \|B(\hat{z} - z^\star)\|^2 = \Delta \\
& \xi^1 = -A^T\lambda^0 - tA^TAx^1 - tA^TB\hat{z}, \\
& \xi^k = -A^T\lambda^{k-1} - tA^TAx^k - tA^TBz^{k-1}, \quad k \in \{2, ..., N\} \\
& \eta^k = -B^T\lambda^{k-1} - tB^TAx^k - tB^TBz^k, \quad k \in \{1, ..., N\} \\
& \lambda^k = \lambda^{k-1} + t(Ax^k + Bz^k - b), \quad k \in \{1, ..., N\} \\
& \lambda_{\max}(A^TA) = \nu_1, \lambda_{\max}(B^TB) = \nu_2 \\
& \lambda_0 \in \mathbb{R}^r, \hat{z} \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}, b \in \mathbb{R}^r.
\end{aligned} \tag{10}$$

To handle problem (10), we introduce some new variables. As ADMM is invariant under translation of (x, z), we may assume without loss of generality that $b = 0$ and $(x^\star, z^\star) = (0, 0)$. In addition, due to the full row rank of the matrix $\begin{pmatrix} A & B \end{pmatrix}$, we may assume that $\lambda^0 = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} x^0 \\ z^0 \end{pmatrix}$ and $\lambda^\star = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{z} \end{pmatrix}$ for some $\bar{x}, x^0, \bar{z}, z^0$. So,

$$\xi^\star = -A^TA\bar{x} - A^TB\bar{z} \in \partial f(0), \quad \eta^\star = -B^TA\bar{x} - B^TB\bar{z} \in \partial g(0),$$

and $D(\lambda^\star) = f^\star + g^\star$.

By using equality constraints of problem (10) and the newly introduced variables, one can get

$$\begin{aligned}
& \lambda^k = \sum_{i=0}^{k} t_i(Ax^i + Bz^i), \quad k \in \{0, ..., N\}, \\
& -A^TAx^0 - A^TBz^0 - tA^TAx^1 - tA^TB\hat{z} \in \partial f(x^1), \\
& -\sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) - t_kA^TAx^k - t_kA^TBz^{k-1} \in \partial f(x^k), \ k \in \{2, ..., N\}, \\
& -\sum_{i=0}^{k} t_i(B^TAx^i + B^TBz^i) \in \partial g(z^k), \quad k \in \{1, ..., N\},
\end{aligned} \tag{11}$$

where $t_0 = 1$ and $t_i = t$, $i \in \{1, ..., N\}$. Due to the optimality conditions, see (7), we have $D(\lambda^N) = f(x^{N+1}) + g(z^N) + \langle \lambda^N, Ax^{N+1} + Bz^N \rangle$ for some x^{N+1} with $-A^T\lambda^N \in \partial f(x^{N+1})$. Hence, problem (10) may be written as

$$
\begin{aligned}
\max\ & f^\star + g^\star - f^{N+1} - g^N - \langle \sum_{i=0}^{N} t_i(Ax^i + Bz^i), Ax^{N+1} + Bz^N \rangle \\
\text{s.t.}\ & \tfrac{\mu_1}{2}\|x^k - x^1\|^2 \le f^k - f^1 - \\
& \left\langle -A^TAx^0 - A^TBz^0 - tA^TAx^1 - tA^TB\hat{z}, x^k - x^1 \right\rangle,\quad k \in \{2,\dots,N+1\} \\
& \tfrac{\mu_1}{2}\|x^k - x^1\|^2 \le \left\langle \sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) + t_kA^TAx^k + t_kA^TBz^{k-1}, x^1 - x^k \right\rangle \\
& + f^1 - f^k,\ k \in \{2,\dots,N\} \\
& \tfrac{\mu_1}{2}\|x^{N+1} - x^1\|^2 \le \left\langle \sum_{i=0}^{N} t_i(A^TAx^i + A^TBz^i), x^1 - x^{N+1} \right\rangle + f^1 - f^{N+1} \\
& \tfrac{\mu_1}{2}\|x^k - x^j\|^2 \le \left\langle \sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) + t_kA^TAx^k + t_kA^TBz^{k-1}, x^j - x^k \right\rangle \\
& + f^j - f^k,\quad k \in \{2,\dots,N\},\ j \in \{2,\dots,N+1\} \\
& \tfrac{\mu_1}{2}\|x^{N+1} - x^j\|^2 \le \left\langle \sum_{i=0}^{N} t_i(A^TAx^i + A^TBz^i), x^j - x^{N+1} \right\rangle + \\
& f^j - f^{N+1},\quad j \in \{2,\dots,N\} \\
& \tfrac{\mu_2}{2}\|z^k - z^j\|^2 \le g^j - g^k + \left\langle \sum_{i=0}^{k} t_i(B^TAx^i + B^TBz^i), z^j - z^k \right\rangle,\ j,k \in \{1,\dots,N\} \\
& \tfrac{\mu_1}{2}\|x^k\|^2 \le f^k - f^\star - \left\langle -A^TA\bar{x} - A^TB\bar{z}, x^k \right\rangle,\quad k \in \{1,\dots,N+1\} \\
& \tfrac{\mu_1}{2}\|x^1\|^2 \le f^\star - f^1 - \left\langle -A^TAx^0 - A^TBz^0 - tA^TAx^1 - tA^TB\hat{z}, -x^1 \right\rangle, \\
& \tfrac{\mu_1}{2}\|x^k\|^2 \le f^\star - f^k - \\
& \left\langle \sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) + t_kA^TAx^k + t_kA^TBz^{k-1}, x^k \right\rangle,\ k \in \{2,\dots,N\} \\
& \tfrac{\mu_1}{2}\|x^{N+1}\|^2 \le f^\star - f^{N+1} - \left\langle \sum_{i=0}^{N} t_i(A^TAx^i + A^TBz^i), x^{N+1} \right\rangle, \\
& \tfrac{\mu_2}{2}\|z^k\|^2 \le g^k - g^\star - \left\langle -B^TA\bar{x} - B^TB\bar{z}, z^k \right\rangle,\quad k \in \{1,\dots,N\} \\
& \tfrac{\mu_2}{2}\|z^k\|^2 \le g^\star - g^k - \left\langle -\sum_{i=0}^{k} t_i(B^TAx^i + B^TBz^i), -z^k \right\rangle,\quad k \in \{1,\dots,N\} \\
& \|Ax^0 + Bz^0 - (A\bar{x} + B\bar{z})\|^2 + t^2\|B\hat{z}\|^2 = \Delta \qquad (12) \\
& \lambda_{\max}(A^TA) = \nu_1, \lambda_{\max}(B^TB) = \nu_2 \\
& x_0 \in \mathbb{R}^n, \hat{z}, z_0 \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}.
\end{aligned}
$$

In problem (12), $A, B, \{x^k, f^k\}_1^{N+1}, \{z^k, g^k\}_1^N, x^0, z^0, \bar{x}, f^\star, \bar{z}, g^\star, \hat{z}$ are decision variables. By using the Gram matrix method, problem (12) may be relaxed as a semidefinite program as follows. Let

$$U = \left(x^0 \; x^1 \; \dots \; x^{N+1} \; \bar{x}\right),$$
$$V = \left(z^0 \; z^1 \; \dots \; z^N \; \bar{z} \; \hat{z}\right).$$

By introducing matrix variables

$$X = U^T U, \qquad Z = V^T V,$$
$$Y = \left(AU \; BV\right)^T \left(AU \; BV\right),$$

problem (12) may be relaxed as the following SDP,

$$\begin{aligned} \max \; & f^\star + g^\star - f^{N+1} - g^N - \operatorname{tr}(L_o Y) \\ \text{s.t.} \; & \operatorname{tr}(L^f_{i,j} Y) + \operatorname{tr}(O^f_{i,j} X) \leq f^i - f^j, \quad i,j \in \{1, ..., N+1, \star\} \\ & \operatorname{tr}(L^g_{i,j} Y) + \operatorname{tr}(O^g_{i,j} Z) \leq g^i - g^j, \quad i,j \in \{1, ..., N, \star\} \\ & \operatorname{tr}(L_0 Y) + Z_{N+3,N+3} = \Delta \\ & X \succeq 0, Y \succeq 0, Z \succeq 0, \\ & \nu_1 X \succeq Y_{11}, \\ & \nu_2 Z \succeq Y_{22}, \end{aligned} \tag{13}$$

where the constant matrices $L^f_{i,j}, L^g_{i,j}, O^f_{i,j}, O^g_{i,j}, L_o, L_0$ are determined according to the constraints of problem (12) and Y_{11} and Y_{22} are blocks of matrix Y such that

$$Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^T & Y_{22} \end{pmatrix}.$$

Note that the constraint $\nu_1 X \succeq Y_{11}$ resulted from $\lambda_{\max}(A^T A) = \nu_1$ since

$$\nu_1 I \succeq A^T A \;\Rightarrow\; \nu_1 U^T U \succeq U^T A^T A U \;\Rightarrow\; \nu_1 X \succeq Y_{11}.$$

Theorem 3 gives a new convergence rate for ADMM. Before we get to the theorem we need to present a lemma.

Lemma 1 *Let $N \geq 2$ and $t, \mu \in \mathbb{R}$. Let $E(t,\mu)$ be $N \times N$ symmetric matrix given by*

$$\begin{pmatrix} \alpha_1 & \beta_1 & 0 & \dots & 0 & 0 & \dots & 0 & -t \\ \beta_1 & \alpha_2 & \beta_2 & \dots & 0 & 0 & \dots & 0 & t \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \alpha_k & \beta_k & \dots & 0 & t \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & \alpha_{N-1} & \beta_{N-1} \\ -t & t & t & \dots & t & t & \dots & \beta_{N-1} & \alpha_N \end{pmatrix}$$

where

$$\alpha_k = \begin{cases} 6\mu - 3t, & k = 1 \\ 2k(2k+1)\mu - (4k+3)t, & 2 \le k \le N-2, N \ge 4 \\ 2N(N-1)\mu - (2N+1)t, & k = N-1, N \ge 3 \\ 2N\mu - (N+1)t - \frac{(\mu-t)^2}{2\mu}, & k = N, \end{cases}$$

$$\beta_k = \begin{cases} 2(k+1)t - (2k^2+3k)\mu, & 1 \le k \le N-2 \\ 3t - 2(N-1)\mu, & k = N-1, \end{cases}$$

and k denotes row number. If $\mu > 0$ is given, then

$$[0, \mu] \subseteq \{t : E(t, \mu) \succeq 0\}.$$

Proof. As $\{t : E(t,\mu) \succeq 0\}$ is convex set, it suffices to prove the positive semidefiniteness of $E(0,\mu)$ and $E(\mu,\mu)$. Since $E(0,\mu)$ is diagonally dominant, it is positive semidefinite. Now, we establish that the matrix $K = E(1,1) - 2e_1e_1^T$ is positive definite. One can show that the claim holds for $N = 2, 3$. So we investigate $N \ge 4$. To this end, we show that all leading principal minors of K are positive. To compute the leading principal minors, we perform the following elementary row operations on K:

i) Add the first row to the second row;
ii) Add the first row to the last row;
iii) Add the second row to the third row;
iv) For $i = 3 : N-2$
   - Add $i-th$ row to $(i+1)-th$ row;
   - Add $\frac{2-i}{2i^2+i-2}$ times of $i-th$ row to the last row;
v) Add $\frac{N-1}{3N-5}$ times of $(N-1)-th$ row to the last row;

By performing these operations, we get an upper triangular matrix J with diagonal

$$J_{kk} = \begin{cases} 2k^2 + k - 2, & 1 \le k \le N-2 \\ 3N-5, & k = N-1 \\ \frac{2N^2-9N+9}{3N-5} - \sum_{i=3}^{N-2} \frac{(i-2)^2}{2i^2+i-2}, & k = N. \end{cases}$$

It is seen all first $N-1$ diagonal elements of J are positive. We show that J_{NN} is also positive. Indeed, for $i \ge 3$ we have

$$\frac{(i-2)^2}{2i^2+i-2} \le \frac{i^2+4}{2i^2} \le \frac{1}{2} + \frac{2}{i(i-1)}.$$

So,

$$\frac{2N^2-9N+9}{3N-5} - \sum_{i=3}^{N-2} \frac{(i-2)^2}{2i^2+i-2} \ge \frac{(N-2)(N^2-5N+10)}{2N(3N-5)} \ge 0,$$

which implies $J_{NN} > 0$. Since we add a factor of $i-th$ row to $j-th$ row with $i < j$, all leading principal minors of matrices K and J are the same. Hence K is positive definite. As $E(\mu,\mu) = \mu(K + 2e_1e_1^T)$, one can infer the positive definiteness of $E(\mu,\mu)$ and the proof is complete. □

Theorem 3 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_0(\mathbb{R}^m)$ with $\mu_1 > 0$. If $t \leq c_1 = \frac{\mu_1}{\lambda_{\max}(A^T A)}$ and $N \geq 2$, then*

$$D(\lambda^\star) - D(\lambda^N) \leq \frac{\|\lambda^0 - \lambda^\star\|^2 + t^2 \left\|B\left(\hat{z} - z^\star\right)\right\|^2}{4Nt}. \tag{14}$$

Proof. The argument is based on weak duality. Indeed, by introducing suitable Lagrangian multipliers, we establish that the given convergence rate is an upper bound for problem (12). By doing some algebra, one can show that

$$\begin{aligned}
& f^\star + g^\star - f^{N+1} - g^N - \langle \sum_{i=0}^{N} t_i(Ax^i + Bz^i), Ax^{N+1} + Bz^N\rangle - \tfrac{1}{4Nt}\Delta + \tfrac{1}{4Nt}\Big(\Delta - \\
& \|Ax^0 + Bz^0 - (A\bar{x} + B\bar{z})\|^2 - t^2\|B\hat{z}\|^2\Big) + \tfrac{1}{2N}\Big(f^{N+1} - f^1 - \Big\langle -A^TAx^0 - A^TBz^0 - \\
& tA^TAx^1 - tA^TB\hat{z}, x^{N+1} - x^1\Big\rangle - \tfrac{\mu_1}{2}\|x^{N+1} - x^1\|^2\Big) + \sum_{k=2}^{N-1}\tfrac{k^2-1}{2N}\Big(f^{k+1} - f^k + \\
& \Big\langle \sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) + t_kA^TAx^k + t_kA^TBz^{k-1}, x^{k+1} - x^k\Big\rangle - \tfrac{\mu_1}{2}\|x^{k+1} - x^k\|^2\Big) + \\
& \tfrac{N-1}{N}\Big(f^{N+1} - f^N + \Big\langle \sum_{i=0}^{N-1} t_i(A^TAx^i + A^TBz^i) + tA^TAx^N + tA^TBz^{N-1}, x^{N+1} - x^N\Big\rangle \\
& - \tfrac{\mu_1}{2}\|x^N - x^{N+1}\|^2\Big) + \sum_{k=2}^{N-1}\tfrac{k(k-1)}{2N}\Big(f^k - f^{k+1} + \Big\langle \sum_{i=0}^{k} t_i(A^TAx^i + A^TBz^i) + \\
& t_{k+1}A^TAx^{k+1} + t_{k+1}A^TBz^k, x^k - x^{k+1}\Big\rangle - \tfrac{\mu_1}{2}\|x^k - x^{k+1}\|^2\Big) + \sum_{k=1}^{N-1}\tfrac{k^2+k}{2N}\Big(g^{k+1} - g^k + \\
& \Big\langle \sum_{i=0}^{k} t_i(B^TAx^i + B^TBz^i), z^{k+1} - z^k\Big\rangle\Big) + \sum_{k=1}^{N-1}\tfrac{k^2}{2N}\Big(g^k - g^{k+1} + \Big\langle \sum_{i=0}^{k+1} t_i(B^TAx^i + \\
& B^TBz^i), z^k - z^{k+1}\Big\rangle\Big) + \tfrac{1}{2N}\sum_{k=1, k\neq N}^{N+1}\Big(f^k - f^\star - \Big\langle -A^TA\bar{x} - A^TB\bar{z}, x^k\Big\rangle - \tfrac{\mu_1}{2}\|x^k\|^2\Big) + \\
& \tfrac{1}{2}\Big(f^N - f^\star - \Big\langle -A^TA\bar{x} - A^TB\bar{z}, x^N\Big\rangle - \tfrac{\mu_1}{2}\|x^N\|^2\Big) + \\
& \tfrac{1}{2N}\sum_{k=1}^{N-1}\Big(g^k - g^\star + \Big\langle B^TA\bar{x} + B^TB\bar{z}, z^k\Big\rangle\Big) + \tfrac{N+1}{2N}\Big(g^N - g^\star + \Big\langle B^TA\bar{x} + B^TB\bar{z}, z^N\Big\rangle\Big) \\
& \leq -\tfrac{c_1}{2N}\Big\|Ax^1 - \tfrac{c_1 - t}{2c_1}Ax^{N+1}\Big\|^2 - \tfrac{t}{4N}\Big\|B\hat{z} + Ax^1 - Ax^{N+1}\Big\|^2 - \\
& \tfrac{1}{4Nt}\Big\|Ax^0 + Bz^0 - A\bar{x} - B\bar{z} + t\sum_{k=1}^{N-1}\Big(Ax^k + Bz^k\Big) + NtAx^N + tAx^{N+1} + (N+1)tAz^N\Big\|^2 \\
& - \tfrac{t}{4N}\sum_{k=1}^{N-2}\Big\|kBz^k - kBz^{k+1} + (k+1)Ax^{k+1} - (k+2)Ax^{k+2} + Ax^{N+1}\Big\|^2 - \\
& \tfrac{t}{4N}\Big\|(N-1)Bz^{N-1} - (N-1)Bz^N + Ax^N - Ax^{N+1}\Big\|^2 - \\
& \tfrac{1}{4N}\operatorname{tr}\Big(E(t, c_1)\left(Ax^2 \ \ldots \ Ax^{N+1}\right)^T\left(Ax^2 \ \ldots \ Ax^{N+1}\right)\Big) \leq 0.
\end{aligned}$$

Note that the first inequality follows from $\|Ax\|^2 \leq \lambda_{\max}(A^TA)\|x\|^2$. The second inequality follows from Lemma 1 as the inner product of positive semidefinite matrices is non-negative. Hence, for any feasible point of problem (12), we have

$$f^\star + g^\star - f^{N+1} - g^N - \langle \sum_{i=0}^{N} t_i(Ax^i + Bz^i), Ax^{N+1} + Bz^N \rangle \leq \tfrac{1}{4Nt}\Delta,$$

which completes the proof. □

In comparison with Theorem 1, we could get a new convergence rate when only f is strongly convex, i.e. g does not need to be strongly convex. Also, the constant does not depend on λ^1. One important question concerning bound (14) is its tightness, that is, if there is an optimization problem which attains the given convergence rate. It turns out that the bound (14) is exact. The following example demonstrates this point.

Example 1 Suppose that $\mu_1 > 0$, $N \geq 2$ and $t \in (0, \mu_1]$. Let $f, g : \mathbb{R} \to \mathbb{R}$ be given as follows,

$$f(x) = \tfrac{1}{2}|x| + \tfrac{\mu_1}{2}x^2, \quad g(z) = \tfrac{1}{2}\max\{\tfrac{N-1}{N}(z - \tfrac{1}{2Nt}) - \tfrac{1}{2Nt}, -z\}.$$

Consider the optimization problem

$$\begin{aligned} \min_{(x,z)\in\mathbb{R}\times\mathbb{R}} \; & f(x) + g(z), \\ \text{s.t.} \; & x + z = 0. \end{aligned}$$

Note that $(x^\star, z^\star) = (0, 0)$ with Lagrangian multiplier $\lambda^\star = \frac{1}{2}$ is an optimal solution and the optimal value is zero. One can check that Algorithm 1 with initial point $\lambda^0 = \frac{-1}{2}$ and $\hat{z} = 0$ generates the following points,

$$\begin{aligned} x^k &= 0 && k \in \{1, \ldots, N\} \\ z^k &= \tfrac{1}{2Nt} && k \in \{1, \ldots, N\} \\ \lambda^k &= \tfrac{-1}{2} + \tfrac{k}{2N} && k \in \{1, \ldots, N\}. \end{aligned}$$

At λ^N, we have $D(\lambda^N) = \frac{-1}{4Nt} = \frac{\|\lambda^0 - \lambda^\star\|^2 + t^2\|B(\hat{z} - z^\star)\|^2}{4Nt}$, which shows the tightness of bound (14).

One important factor concerning dual-based methods that determines an efficiency of an algorithm is primal and dual feasibility (residual) convergence rates. In what follows, we study this subject under the setting of Theorem 3. To measure primal residual convergence rate, similar to problem (12), we formulate the following problem

$$
\begin{aligned}
\max \quad & \|Ax^N + Bz^N - b\|^2 \\
\text{s.t.} \quad & \tfrac{\mu_1}{2}\|x^k - x^1\|^2 \le f^k - f^1 - \\
& \left\langle -A^TAx^0 - A^TBz^0 - tA^TAx^1 - tA^TB\hat{z}, x^k - x^1 \right\rangle, \quad k \in \{2,\dots,N\} \\
& \tfrac{\mu_1}{2}\|x^k - x^1\|^2 \le f^1 - f^k - \\
& \left\langle -\sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) - t_kA^TAx^k - t_kA^TBz^{k-1}, x^1 - x^k \right\rangle, \quad k \in \{2,\dots,N\} \\
& \tfrac{\mu_1}{2}\|x^k - x^j\|^2 \le f^j - f^k - \\
& \left\langle -\sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) - t_kA^TAx^k - t_kA^TBz^{k-1}, x^j - x^k \right\rangle, \ k,j \in \{2,\dots,N\} \\
& \tfrac{\mu_2}{2}\|z^k - z^j\|^2 \le g^j - g^k - \left\langle -\sum_{i=0}^{k} t_i(B^TAx^i + B^TBz^i), z^j - z^k \right\rangle, \ k,j \in \{1,\dots,N\} \\
& \tfrac{\mu_1}{2}\|x^k\|^2 \le f^k - f^\star - \left\langle -A^TA\bar{x} - A^TB\bar{z}, x^k \right\rangle, \ k \in \{1,\dots,N\} \\
& \tfrac{\mu_1}{2}\|x^1\|^2 \le f^\star - f^1 - \left\langle -A^TAx^0 - A^TBz^0 - tA^TAx^1 - tA^TB\hat{z}, -x^1 \right\rangle, \\
& \tfrac{\mu_1}{2}\|x^k\|^2 \le f^\star - f^k - \\
& \left\langle -\sum_{i=0}^{k-1} t_i(A^TAx^i + A^TBz^i) - t_kA^TAx^k - t_kA^TBz^{k-1}, -x^k \right\rangle, \quad k \in \{2,\dots,N\} \\
& \tfrac{\mu_2}{2}\|z^k\|^2 \le g^k - g^\star - \left\langle -B^TA\bar{x} - B^TB\bar{z}, z^k \right\rangle, \quad k \in \{1,\dots,N\} \\
& \tfrac{\mu_2}{2}\|z^k\|^2 \le g^\star - g^k - \left\langle -\sum_{i=0}^{k} t_i(B^TAx^i + B^TBz^i), -z^k \right\rangle, \quad k \in \{1,\dots,N\} \\
& \|Ax^0 + Bz^0 - (A\bar{x} + B\bar{z})\|^2 + t^2\|B\hat{z}\|^2 = \Delta \qquad (15) \\
& \lambda_{\max}(A^TA) = \nu_1, \lambda_{\max}(B^TB) = \nu_2 \\
& x_0 \in \mathbb{R}^n, \hat{z}, z_0 \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}.
\end{aligned}
$$

The next theorem gives a convergence rate in terms of primal residual under the setting of Theorem 3.

Theorem 4 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_0(\mathbb{R}^m)$ with $\mu_1 > 0$. If $t \le c_1 = \frac{\mu_1}{\lambda_{\max}(A^TA)}$ and $N \ge 3$, then*

$$\|Ax^N + Bz^N - b\| \le \frac{\sqrt{\|\lambda^0 - \lambda^\star\|^2 + t^2\,\|B\,(\hat{z} - z^\star)\|^2}}{tN}. \qquad (16)$$

Proof. The argument is similar to that used in Theorem 3. Indeed, we show that the convergence rate is an upper bound for problem (15). By some calculus, one can show that

$$
\begin{aligned}
&\left\|Ax^N+Bz^N\right\|^2-\tfrac{1}{(Nt)^2}\Delta+\tfrac{1}{(Nt)^2}\Big(\Delta-\|Ax^0+Bz^0-(A\bar{x}+B\bar{z})\|^2-t^2\|B\hat{z}\|^2\Big)+\\
&\tfrac{2}{tN^2}\Big(f^N-f^1-\Big\langle-A^TAx^0-A^TBz^0-tA^TAx^1-tA^TB\hat{z},x^N-x^1\Big\rangle-\tfrac{\mu_1}{2}\|x^N-x^1\|^2\Big)+\\
&\sum_{k=2}^{N-1}\tfrac{2(k-1)(k+1)}{tN^2}\Big(f^{k+1}-f^k-\Big\langle-\sum_{i=0}^{k-1}t_i(A^TAx^i+A^TBz^i)-t_kA^TAx^k-t_kA^TBz^{k-1},\\
&x^{k+1}-x^k\Big\rangle-\tfrac{\mu_1}{2}\|x^{k+1}-x^k\|^2\Big)+\sum_{k=2}^{N-1}\tfrac{2k(k-1)}{tN^2}\Big(f^k-f^{k+1}+\Big\langle\sum_{i=0}^{k}t_i(A^TAx^i+A^TBz^i)+\\
&t_{k+1}A^TAx^{k+1}+t_{k+1}A^TBz^k,x^k-x^{k+1}\Big\rangle-\tfrac{\mu_1}{2}\|x^k-x^{k+1}\|^2\Big)+\sum_{k=1}^{N-1}\tfrac{2k(k+1)}{tN^2}\Big(g^{k+1}-g^k-\\
&\Big\langle-\sum_{i=0}^{k}t_i(B^TAx^i+B^TBz^i),z^{k+1}-z^k\Big\rangle\Big)+\sum_{k=1}^{N-1}\tfrac{2k^2}{tN^2}\Big(g^k-g^{k+1}-\Big\langle-\sum_{i=0}^{k+1}t_i(B^TAx^i+\\
&B^TBz^i),z^k-z^{k+1}\Big\rangle\Big)+\tfrac{2}{tN^2}\sum_{k=1}^{N}\Big(f^k-f^\star-\Big\langle-A^TA\bar{x}-A^TB\bar{z},x^k\Big\rangle-\tfrac{\mu_1}{2}\|x^k\|^2\Big)+\\
&\frac{2}{tN}\Big(f^\star-f^N-\Big\langle-\sum_{i=0}^{N-1}t_i(A^TAx^i+A^TBz^i)-t_NA^TAx^N-t_NA^TBz^{N-1},-x^N\Big\rangle-\\
&\tfrac{\mu_1}{2}\|x^N\|^2\Big)+\tfrac{2}{tN^2}\sum_{k=1}^{N}\Big(g^k-g^\star-\Big\langle-B^TA\bar{x}-B^TB\bar{z},z^k\Big\rangle\Big)+\\
&\frac{2}{tN}\Big(g^\star-g^N-\Big\langle-\sum_{i=0}^{N}t_i(B^TAx^i+B^TBz^i),-z^N\Big\rangle\Big)\\
&\le\tfrac{-1}{(Nt)^2}\left\|Ax^0+Bz^0-A\bar{x}-B\bar{z}+t\sum_{k=1}^{N}\Big(Ax^k+Bz^k\Big)\right\|^2-\tfrac{1}{N^2}\left\|B\hat{z}+Ax^1-Ax^N\right\|^2-\\
&\sum_{k=1}^{N-2}\Big(\tfrac{k}{N}\Big)^2\left\|Bz^k-Bz^{k+1}+\tfrac{k+1}{k}Ax^{k+1}-\tfrac{k+2}{k}Ax^{k+2}+\tfrac{1}{k}Ax^N\right\|^2-\\
&\tfrac{2c_1}{tN^2}\left\|Ax^1-\tfrac{c_1-t}{2c_1}Ax^N\right\|^2-\Big(\tfrac{N-1}{N}\Big)^2\left\|Bz^{N-1}-Bz^N+\tfrac{N}{(N-1)^2}Ax^N\right\|^2-\\
&\tfrac{1}{tN^2}\operatorname{tr}\Big(D(t,c_1)\left(Ax^2\ \dots\ Ax^N\right)^T\left(Ax^2\ \dots\ Ax^N\right)\Big)\le 0,
\end{aligned}
$$

where the matrix $D(t,c_1)$ is as follows,

$$
D(t,c_1)=\begin{pmatrix}
\alpha_1 & \beta_1 & 0 & \dots & 0 & 0 & \dots & 0 & -t\\
\beta_1 & \alpha_2 & \beta_2 & \dots & 0 & 0 & \dots & 0 & t\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & \dots & \alpha_k & \beta_k & \dots & 0 & t\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
0 & 0 & 0 & \dots & 0 & 0 & \dots & \alpha_{N-2} & \beta_{N-2}\\
-t & t & t & \dots & t & t & \dots & \beta_{N-2} & \alpha_{N-1}
\end{pmatrix},
$$

and

$$\alpha_k = \begin{cases} 6c_1 - 3t, & k = 1 \\ 2k(2k+1)c_1 - (4k+3)t, & 2 \le k \le N-2, \\ \frac{1}{2}c_1(4(N-2)N+7) - \frac{t^2}{2c_1} + \frac{(N((10-3N)N-13)+5)t}{(N-1)^2}, & k = N-1, \end{cases}$$

$$\beta_k = \begin{cases} 2(k+1)t - k(2k+3)c_1, & 1 \le k \le N-3, \\ (2N-1)t - (N-2)(2N-1)c_1, & k = N-2,\ N \ge 4 \\ 3t - 5c_1, & k = N-2,\ N = 3. \end{cases}$$

We have the first inequality, because $\|Ax\|^2 \le \lambda_{\max}(A^TA)\|x\|^2$. By analysis similar to that in the proof of Lemma 1, one can show $D(t, c_1) \succeq 0$. Hence, the second inequality follows from positive semidefinitness of $D(t, c_1)$ and this completes the proof. □

The following example shows the exactness of bound (16).

Example 2 Let $\mu_1 > 0$, $N \ge 2$ and $t \in (0, \mu_1]$. Consider functions $f, g : \mathbb{R} \to \mathbb{R}$ given by the formulae follows,

$$f(x) = \tfrac{1}{2}|x| + \tfrac{\mu_1}{2}x^2,$$
$$g(z) = \max\{\left(\tfrac{1}{2} - \tfrac{1}{N}\right)\left(z - \tfrac{1}{Nt}\right), \tfrac{1}{2}\left(\tfrac{1}{Nt} - z\right)\}.$$

We formulate the following optimization problem,

$$\min_{(x,z)\in\mathbb{R}\times\mathbb{R}} f(x) + g(z),$$
$$\text{s.t. } Ax + Bz = 0,$$

where $A = B = I$. One can verify that $(x^\star, z^\star) = (0, 0)$ with Lagrangian multiplier $\lambda^\star = \frac{1}{2}$ is an optimal solution and the optimal value is zero. Algorithm 1 with initial point $\lambda^0 = \frac{-1}{2}$ and $\hat{z} = 0$ generates the following points,

$$x^k = 0 \qquad k \in \{1, \ldots, N\}$$
$$z^k = \tfrac{1}{Nt} \qquad k \in \{1, \ldots, N\}$$
$$\lambda^k = \tfrac{2k-N}{2N} \qquad k \in \{1, \ldots, N\}.$$

At iteration N, we have $\|Ax^N + Bz^N\| = \frac{1}{tN} = \frac{\sqrt{\|\lambda^0-\lambda^\star\|^2 + t^2\|B(\hat{z}-z^\star)\|^2}}{tN}$, which shows the tightness of bound (16).

In line with our discussion in the section, we cast the following problem to measure the convergence rate of the sequence $\{B\left(z^{k-1}-z^k\right)\}$,

$$
\begin{aligned}
\max\ & \left\|B\left(z^N-z^{N-1}\right)\right\|^2\\
\text{s.t. } & \tfrac{\mu_1}{2}\|x^k-x^1\|^2 \le f^k-f^1-\\
& \left\langle -A^TAx^0-A^TBz^0-tA^TAx^1-tA^TB\hat{z}, x^k-x^1\right\rangle,\quad k\in\{2,\dots,N\}\\
& \tfrac{\mu_1}{2}\|x^k-x^1\|^2 \le f^1-f^k-\\
& \left\langle -\sum_{i=0}^{k-1} t_i(A^TAx^i+A^TBz^i)-t_kA^TAx^k-t_kA^TBz^{k-1}, x^1-x^k\right\rangle,\quad k\in\{2,\dots,N\}\\
& \tfrac{\mu_1}{2}\|x^k-x^j\|^2 \le f^j-f^k-\\
& \left\langle -\sum_{i=0}^{k-1} t_i(A^TAx^i+A^TBz^i)-t_kA^TAx^k-t_kA^TBz^{k-1}, x^j-x^k\right\rangle,\ k,j\in\{2,\dots,N\}\\
& \tfrac{\mu_2}{2}\|z^k-z^j\|^2 \le g^j-g^k-\left\langle -\sum_{i=0}^{k} t_i(B^TAx^i+B^TBz^i), z^j-z^k\right\rangle,\ k,j\in\{1,\dots,N\}\\
& \tfrac{\mu_1}{2}\|x^k\|^2 \le f^k-f^\star-\left\langle -A^TA\bar{x}-A^TB\bar{z}, x^k\right\rangle,\quad k\in\{1,\dots,N\}\\
& \tfrac{\mu_1}{2}\|x^1\|^2 \le f^\star-f^1-\left\langle -A^TAx^0-A^TBz^0-tA^TAx^1-tA^TB\hat{z}, -x^1\right\rangle,\\
& \tfrac{\mu_1}{2}\|x^k\|^2 \le f^\star-f^k-\\
& \left\langle -\sum_{i=0}^{k-1} t_i(A^TAx^i+A^TBz^i)-t_kA^TAx^k-t_kA^TBz^{k-1}, -x^k\right\rangle,\quad k\in\{2,\dots,N\}\\
& \tfrac{\mu_2}{2}\|z^k\|^2 \le g^k-g^\star-\left\langle -B^TA\bar{x}-B^TB\bar{z}, z^k\right\rangle,\quad k\in\{1,\dots,N\}\\
& \tfrac{\mu_2}{2}\|z^k\|^2 \le g^\star-g^k-\left\langle -\sum_{i=0}^{k} t_i(B^TAx^i+B^TBz^i), -z^k\right\rangle,\quad k\in\{1,\dots,N\}\\
& \|Ax^0+Bz^0-(A\bar{x}+B\bar{z})\|^2+t^2\|B\hat{z}\|^2=\Delta \qquad (17)\\
& \lambda_{\max}(A^TA)=\nu_1, \lambda_{\max}(B^TB)=\nu_2\\
& x_0\in\mathbb{R}^n, \hat{z}, z_0\in\mathbb{R}^m, A\in\mathbb{R}^{r\times n}, B\in\mathbb{R}^{r\times m}.
\end{aligned}
$$

The next theorem provides a convergence rate for the sequence $\{B\left(z^{k-1}-z^k\right)\}$.

Theorem 5 *Let assumptions Theorem 3 hold and $N\ge 4$. Then*

$$
\left\|B\left(z^N-z^{N-1}\right)\right\| \le \frac{\sqrt{\|\lambda^0-\lambda^\star\|^2+t^2\left\|B\left(\hat{z}-z^\star\right)\right\|^2}}{(N-1)t}. \qquad (18)
$$

Proof. We consider problem (17) and follow the proof of the previous theorems. By some algebra, one can verify that

$$\left\|B\left(z^N-z^{N-1}\right)\right\|^2-\tfrac{1}{((N-1)t)^2}\Delta+\tfrac{1}{((N-1)t)^2}\left(\Delta-\|Ax^0+Bz^0-(A\bar{x}+B\bar{z})\|^2-t^2\|B\hat{z}\|^2\right)$$
$$+\tfrac{2}{t(N-1)^2}\left(f^N-f^1+\left\langle A^TAx^0+A^TBz^0+tA^TAx^1+tA^TB\hat{z},x^N-x^1\right\rangle-\tfrac{\mu_1}{2}\|x^N-x^1\|^2\right)$$
$$+\sum_{k=2}^{N-2}\tfrac{2(k-1)(k+1)}{t(N-1)^2}\left(f^{k+1}-f^k-\left\langle-\sum_{i=0}^{k-1}t_i(A^TAx^i+A^TBz^i)-t_kA^TAx^k-t_kA^TBz^{k-1},\right.\right.$$
$$\left.\left.x^{k+1}-x^k\right\rangle-\tfrac{\mu_1}{2}\|x^{k+1}-x^k\|^2\right)+\tfrac{2(N^2-3N+1)}{t(N-1)^2}\left(f^N-f^{N-1}+\left\langle\sum_{i=0}^{N-2}t_i(A^TAx^i+A^TBz^i)\right.\right.$$
$$\left.\left.+t_{N-1}A^TAx^{N-1}+t_{N-1}A^TBz^{N-2},x^N-x^{N-1}\right\rangle-\tfrac{\mu_1}{2}\|x^N-x^{N-1}\|^2\right)+\sum_{k=2}^{N-1}\tfrac{2k(k-1)}{t(N-1)^2}$$
$$\left(f^k-f^{k+1}-\left\langle-\sum_{i=0}^{k}t_i(A^TAx^i+A^TBz^i)-t_{k+1}A^TAx^{k+1}-t_{k+1}A^TBz^k,x^k-x^{k+1}\right\rangle-\right.$$
$$\left.\tfrac{\mu_1}{2}\|x^k-x^{k+1}\|^2\right)+\sum_{k=1}^{N-1}\tfrac{2k(k+1)}{t(N-1)^2}\left(g^{k+1}-g^k+\left\langle\sum_{i=0}^{k}t_i(B^TAx^i+B^TBz^i),z^{k+1}-z^k\right\rangle\right)+$$
$$\sum_{k=1}^{N-1}\tfrac{2k^2}{t(N-1)^2}\left(g^k-g^{k+1}-\left\langle-\sum_{i=0}^{k+1}t_i(B^TAx^i+B^TBz^i),z^k-z^{k+1}\right\rangle\right)+\tfrac{2}{t(N-1)^2}$$
$$\sum_{k=1}^{N-1}\left(f^k-f^\star-\left\langle-A^TA\bar{x}-A^TB\bar{z},x^k\right\rangle-\tfrac{\mu_1}{2}\|x^k\|^2\right)+\frac{2}{t(N-1)}\left(f^\star-f^{N-1}-\right.$$
$$\left.\left\langle\sum_{i=0}^{N-2}t_i(A^TAx^i+A^TBz^i)+t_{N-1}A^TAx^{N-1}+t_{N-1}A^TBz^{N-2},x^{N-1}\right\rangle-\tfrac{\mu_1}{2}\|x^{N-1}\|^2\right)+$$
$$\frac{2}{t(N-1)}\left(g^\star-g^N-\left\langle-\sum_{i=0}^{N}t_i(B^TAx^i+B^TBz^i),-z^N\right\rangle\right)+$$
$$\tfrac{2}{t(N-1)^2}\sum_{k=1}^{N-1}\left(g^k-g^\star-\left\langle-B^TA\bar{x}-B^TB\bar{z},z^k\right\rangle\right)$$
$$\leq\tfrac{-1}{((N-1)t)^2}\left\|Ax^0+Bz^0-A\bar{x}-B\bar{z}+t\sum_{k=1}^{N-1}\left(Ax^k+Bz^k\right)\right\|^2-\tfrac{1}{(N-1)^2}\left\|B\hat{z}+Ax^1-Ax^N\right\|^2-$$
$$\tfrac{N+1}{N-1}\left\|\tfrac{N}{N+1}Ax^N+Bz^N\right\|^2-\sum_{k=1}^{N-3}\left(\tfrac{k}{N-1}\right)^2\left\|Bz^k-Bz^{k+1}+\tfrac{k+1}{k}Ax^{k+1}-\tfrac{k+2}{k}Ax^{k+2}+\tfrac{1}{k}Ax^N\right\|^2$$
$$-\left(\tfrac{N-2}{N-1}\right)^2\left\|Bz^{N-2}-Bz^{N-1}+\tfrac{N-1}{N-2}Ax^{N-1}-\left(1-\tfrac{1}{(N-2)^2}\right)Ax^N\right\|^2-$$
$$\frac{2c_1}{t(N-1)^2}\left\|Ax^1-\tfrac{c_1-t}{2c_1}Ax^N\right\|^2-\tfrac{1}{t(N-1)^2}\operatorname{tr}\left(F(t,c_1)\left(Ax^2\ \ldots\ Ax^N\right)^T\left(Ax^2\ \ldots\ Ax^N\right)\right)\leq 0,$$

where the matrix $F(t,c_1)$ is defined as follows,

$$F(t,c_1)=\begin{pmatrix}\alpha_1 & \beta_1 & 0 & \dots & 0 & 0 & \dots & 0 & -t\\ \beta_1 & \alpha_2 & \beta_2 & \dots & 0 & 0 & \dots & 0 & t\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \alpha_k & \beta_k & \dots & 0 & t\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & \alpha_{N-2} & \beta_{N-2}\\ -t & t & t & \dots & t & t & \dots & \beta_{N-2} & \alpha_{N-1}\end{pmatrix},$$

and

$$\alpha_k=\begin{cases}6c_1-3t, & k=1\\ 2k(2k+1)c_1-(4k+3)t, & 2\le k\le N-2,\\ \frac{1}{2}\left(c_1(4(N-3)N+7)-\frac{t^2}{c_1}-\frac{2(N(N((N-6)N+7)+13)-19)t}{(N-2)^2(N+1)}\right), & k=N-1,\end{cases}$$

$$\beta_k=\begin{cases}2(k+1)t-k(2k+3)c_1, & 1\le k\le N-3,\\ (N+\frac{1}{2-N}-1)t-(2(N-3)N+3)c_1, & k=N-2.\end{cases}$$

The rest of the proof runs as before. □

Note that one could infer the convergence rate of dual residual from Theorem 5 as $\left\|A^TB\left(z^{k-1}-z^k\right)\right\|\le\|A\|\left\|B\left(z^{k-1}-z^k\right)\right\|$. The following example shows the tightness of this bound.

Example 3 Assume that $\mu_1>0$, $N\ge 4$ and $t\in(0,\mu_1]$ are given, and $f,g:\mathbb{R}\to\mathbb{R}$ are defined by,

$$f(x)=\tfrac{1}{2}\max\left\{-\tfrac{N+1}{N-1}x,x\right\}+\tfrac{\mu_1}{2}x^2,$$
$$g(z)=\tfrac{1}{2}\max\left\{\tfrac{1}{t(N-1)}-z,\tfrac{N-3}{N-1}\left(z-\tfrac{1}{t(N-1)}\right)\right\}.$$

Consider the optimization problem

$$\min_{(x,z)\in\mathbb{R}\times\mathbb{R}} f(x)+g(z),$$
$$\text{s.t. } Ax+Bz=0.$$

where $A=B=1$. The point $(x^\star,z^\star)=(0,0)$ with Lagrangian multiplier $\lambda^\star=\frac{1}{2}$ is an optimal solution with the optimal value zero. After performing N iterations of Algorithm 1 with setting $\lambda^0=\frac{-1}{2}$ and $\hat{z}=0$, we have

$$x^k=0,\qquad k\in\{1,\dots,N\},$$
$$z^k=\begin{cases}\frac{1}{t(N-1)}, & k\in\{1,\dots,N-1\},\\ 0, & k=N,\end{cases}$$
$$\lambda^k=\begin{cases}\frac{2k+1-N}{2(N-1)}, & k\in\{1,\dots,N-1\},\\ \frac{1}{2}, & k=N.\end{cases}$$

It can be seen that $\left\| A^T B \left(z^N - z^{N-1} \right) \right\| = \frac{1}{(N-1)t} = \frac{\sqrt{\|\lambda^0 - \lambda^\star\|^2 + t^2 \|B(\hat{z} - z^\star)\|^2}}{(N-1)t}$, which shows that the bound is tight.

Theorem 3 and 4 address the case that f is strongly convex and g is convex. Based on numerical results by solving problem (12) and 15 we conjecture, under the assumptions of Theorem 3, if g_2 is μ_2-strongly convex, Algorithm 1 enjoys the following convergence rates

$$D(\lambda^\star) - D(\lambda^N) \le \frac{\|\lambda^0 - \lambda^\star\|^2 + t^2 \|B\left(\hat{z} - z^\star\right)\|^2}{4Nt + \frac{2}{\frac{\lambda_{\max}(A^T A)}{\mu_1} + \frac{\lambda_{\max}(B^T B)}{\mu_2}}},$$

$$\|Ax^N + Bz^N - b\| \le \frac{\sqrt{\|\lambda^0 - \lambda^\star\|^2 + t^2 \|B\left(\hat{z} - z^\star\right)\|^2}}{Nt + \frac{1}{\frac{\lambda_{\max}(A^T A)}{\mu_1} + \frac{\lambda_{\max}(B^T B)}{\mu_2}}}.$$

We have verified these conjectures numerically for many specific values of the parameters.

3 Linear convergence of ADMM

In this section we study the linear convergence of ADMM. The linear convergence of ADMM has been addressed by some authors and some conditions for linear convergence have been proposed, see [7, 17, 18, 21, 26, 31, 39]. Two common types of assumptions employed for proving the linear convergence of ADMM are error bound property and L-smoothness. To the best knowledge of authors, most scholars investigated the linear convergence of the sequence $\{(x^k, z^k, \lambda^k)\}$ to a saddle point and there is no result in terms of dual objective value for ADMM. In line with the previous section, we study the linear convergence in terms of dual objective value and we derive some formulas for linear convergence rate by using performance estimation.

As mentioned earlier, error bound property is used by scholars for establishing the linear convergence; see e.g. [17, 21, 26, 32, 39]. Let

$$D^a(\lambda) := \min f(x) + g(z) + \langle \lambda, Ax + Bz - b \rangle + \tfrac{a}{2}\|Ax + Bz - b\|^2, \qquad (19)$$

stands for augmented dual objective for the given $a > 0$ and $\Lambda^\star$ denotes the optimal solution set of the dual problem. Note that function D^a is an a-smooth function on its domain without assuming strong convexity; see [21, Lemma 2.2].

Definition 1 The function D^a satisfies the error bound if we have

$$d_{\Lambda^\star}(\lambda) \le \tau \|\nabla D^a(\lambda)\|, \quad \lambda \in \mathbb{R}^r, \qquad (20)$$

for some $\tau > 0$.

Hong et al. [21] established the linear convergence by employing error bound property (20).

Recently, some scholars established the linear convergence of gradient methods for L-smooth convex functions by replacing strong convexity with some mild conditions, see [2, 4, 30] and references therein. Inspired by these results, we prove the linear convergence of ADMM by using the so-called PŁ inequality. Concerning differentiability of dual objective, by (6), we have

$$b - A\partial f^*(-A^T\lambda) - B\partial g^*(-B^T\lambda) \subseteq \partial\left(-D(\lambda)\right). \tag{21}$$

Note that inclusion (21) holds as an equality under some mild conditions, see e.g. [3, Chapter 3].

Definition 2 The function D is said to satisfy the PŁ inequality if there exists an $L_p > 0$ such that for any $\lambda \in \mathbb{R}^r$ we have

$$D(\lambda^\star) - D(\lambda) \leq \tfrac{1}{2L_p}\|\xi\|^2, \quad \xi \in b - A\partial f^*(-A^T\lambda) - B\partial g^*(-B^T\lambda). \tag{22}$$

Note that if f and g are strongly convex, then D is an L-smooth convex function with $L \leq \frac{\lambda_{\max}(A^TA)}{\mu_1} + \frac{\lambda_{\max}(B^TB)}{\mu_2}$. Under this setting, we have $L_p \leq \frac{\lambda_{\max}(A^TA)}{\mu_1} + \frac{\lambda_{\max}(B^TB)}{\mu_2}$. In the next proposition, we show that definitions (20) and (22) are equivalent.

Proposition 1 *Let L_a denote the Lipschitz constant of ∇D^a, where D^a is given in* (19)*.*

i) If D^a satisfies the error bound (20)*, then D satisfies the PŁ inequality with $L_p = \frac{1}{L_a\tau^2}$.*
ii) If D satisfies the PŁ inequality and $\frac{1}{L_p} > a$, then D^a satisfies the error bound (20) *with $L_a = \frac{L_p}{1-aL_p}$.*

Proof. First we prove *i*). Suppose $\lambda \in \mathbb{R}^r$ and $\xi \in b - A\partial f^*(-A^T\lambda) - B\partial g^*(-B^T\lambda)$. Due to identity (5), we have $\xi = b - A\bar{x} - B\bar{z}$ for some $(\bar{x}, \bar{z}) \in \operatorname{argmin} f(x) + g(z) + \langle \lambda, Ax + Bz - b\rangle$. Due to the smoothness of D^a, we get

$$D^a(\lambda^\star) - D^a(\nu) \leq \tfrac{L_a\tau^2}{2}\|\nabla D^a(\nu)\|^2, \quad \nu \in \mathbb{R}^r. \tag{23}$$

Suppose that $\bar{\nu} = \lambda - a(A\bar{x} + B\bar{z} - b)$. Note that $D^a(\lambda^\star) = D(\lambda^\star)$. By the optimality conditions of problem (19), we have

$$(\bar{x}, \bar{z}) \in \operatorname{argmin} f(x) + g(z) + \langle \bar{\nu}, Ax + Bz - b\rangle + \tfrac{a}{2}\|Ax + Bz - b\|^2.$$

By [21, Lemma 2.1], we have $\nabla D^a(\bar{\nu}) = A\bar{x} + B\bar{z} - b$. This equality with (23) imply

$$D(\lambda^\star) - D(\lambda) \leq D^a(\lambda^\star) - D^a(\bar{\nu}) \leq \tfrac{L_a\tau^2}{2}\|A\bar{x} + B\bar{z} - b\|^2,$$

and the proof of i) is complete.
Now we establish ii). Let λ be in domain of ∇D^a. Analogous to the proof of part i), one can get

$$D^a(\lambda^\star) - D^a(\lambda) \leq \left(\frac{1}{2L_p} - \frac{a}{2}\right) \|\nabla D^a(\lambda)\|^2.$$

This inequality says that D^a satisfies the PŁ inequality. On the other hand, the PŁ inequality implies the error bound with the same constant, see [4], and the proof is complete. □

In what follows, we employ performance estimation to derive a linear convergence rate for ADMM in terms of dual objective when the PŁ inequality holds. To this end, we compare the value of dual problem in two consecutive iterations, that is, $\frac{D(\lambda^\star)-D(\lambda^2)}{D(\lambda^\star)-D(\lambda^1)}$. The following optimization problem gives the worst-case convergence rate,

$$\begin{aligned}
\max\ & \tfrac{D(\lambda^\star)-D(\lambda^2)}{D(\lambda^\star)-D(\lambda^1)} \\
\text{s. t. } & \{x^2, z^2, \lambda^2\} \text{ is generated by Algorithm 1 w.r.t. } f, g, A, B, b, \lambda^1, z^1 \\
& (x^\star, z^\star) \text{ is an optimal solution and its Lagrangian multipliers is } \lambda^\star \\
& D \text{ satisfies the PŁ inequality} \\
& f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n), g \in \mathcal{F}_{\mu_2}(\mathbb{R}^n) \\
& \lambda_{\max}(A^TA) = \nu_1, \lambda_{\max}(B^TB) = \nu_2 \\
& \lambda^1 \in \mathbb{R}^r, z^1 \in \mathbb{R}^m, A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}, b \in \mathbb{R}^r.
\end{aligned} \tag{24}$$

Analogous to our discussion in Section 2, we may assume without loss of generality $b = 0$, $\lambda^1 = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} x^0 \\ z^0 \end{pmatrix}$ and $\lambda^\star = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} \bar{x} \\ \bar{z} \end{pmatrix}$ for some $\bar{x}, x^0, \bar{z}, z^0$. In addition, we assume that $\hat{x}^1 \in \operatorname{argmin} f(x) + \langle \lambda^1, Ax \rangle$ and $\hat{x}^2 \in \operatorname{argmin} f(x) + \langle \lambda^2, Ax \rangle$. Hence,

$$D(\lambda^1) = f(\hat{x}^1) + g(z^1) + \langle \lambda^1, A\hat{x}^1 + Bz^1 \rangle, \quad D(\lambda^2) = f(\hat{x}^2) + g(z^2) + \langle \lambda^2, A\hat{x}^2 + Bz^2 \rangle,$$

and

$$\begin{aligned}
&-A^T\lambda^1 \in \partial f(\hat{x}^1), && -B^T\lambda^1 \in \partial g(z^1), \\
&-A^T\lambda^2 \in \partial f(\hat{x}^2), && -B^T\lambda^2 \in \partial g(z^2).
\end{aligned} \tag{25}$$

Moreover, by (25) and (21), we get

$$-A\hat{x}^1 - Bz^1 \in \partial\left(-D(\lambda^1)\right), \qquad -A\hat{x}^2 - Bz^2 \in \partial\left(-D(\lambda^2)\right).$$

On the other hand, $\lambda^2 = \lambda^1 + tAx^2 + tBz^2$. Therefore, by using Theorem 2, problem (24) may be relaxed as follows,

$$
\begin{aligned}
\max\ & \frac{f^\star + g^\star - \hat{f}^2 - g^2 - \langle Ax^0 + Bz^0 + tAx^2 + tBz^2, A\hat{x}^2 + Bz^2\rangle}{f^\star + g^\star - \hat{f}^1 - g^1 - \langle Ax^0 + Bz^0, A\hat{x}^1 + Bz^1\rangle} \\
\text{s.t.}\ & \Big\{ \left(\hat{x}^1, -A^TAx^0 - A^TBz^0, \hat{f}^1\right), \left(x^2, -A^TAx^0 - A^TBz^0 - tA^TAx^2 - tA^TBz^1, f^2\right), \\
& \left(\hat{x}^2, -A^TAx^0 - A^TBz^0 - tA^TAx^2 - tA^TBz^2, \hat{f}^2\right), \left(0, -A^TA\bar{x} - A^TB\bar{z}, f^\star\right) \Big\} \\
& \text{satisfy interpolation constraints (3)} \\
& \Big\{ \left(z^1, -B^TAx^0 - B^TBz^0, g^1\right), \left(z^2, -B^TAx^0 - B^TBz^0 - tB^TAx^2 - tB^TBz^2, g^2\right), \\
& \left(0, -B^TA\bar{z} - B^TB\bar{z}, g^\star\right) \Big\} \text{ satisfy interpolation constraints (3)} \\
& f^* + g^* - \left\langle Ax^0 + Bz^0, A\hat{x}^1 + Bz^1\right\rangle - \hat{f}^1 - g^1 \le \tfrac{1}{2L_p}\left\|A\hat{x}^1 + Bz^1\right\|^2 \qquad (26) \\
& f^* + g^* - \left\langle Ax^0 + Bz^0 + tAx^2 + tBz^2, A\hat{x}^2 + Bz^2\right\rangle - \hat{f}^2 - g^2 \le \tfrac{1}{2L_p}\left\|A\hat{x}^2 + Bz^2\right\|^2 \\
& \lambda_{\max}(A^TA) = \nu_1, \lambda_{\max}(B^TB) = \nu_2 \\
& A \in \mathbb{R}^{r\times n}, B \in \mathbb{R}^{r\times m}.
\end{aligned}
$$

By deriving an upper bound for the optimal value of problem (26) in the next theorem, we establish the linear convergence of ADMM in the presence of the PŁ inequality.

Theorem 6 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m)$ with $\mu_1, \mu_2 > 0$, and let D satisfies the PŁ inequality with L_p. Suppose that $t \le \sqrt{c_1c_2}$, where $c_1 = \frac{\mu_1}{\lambda_{\max}(A^TA)}$ and $c_2 = \frac{\mu_2}{\lambda_{\max}(B^TB)}$.*

(i) If $c_1 \ge c_2$, then

$$
\frac{D(\lambda^\star) - D(\lambda^2)}{D(\lambda^\star) - D(\lambda^1)} \le \frac{2c_1c_2 - t^2}{2c_1c_2 - t^2 + L_pt\left(4c_1c_2 - c_2t - 2t^2\right)}, \tag{27}
$$

in particular, if $t = \sqrt{c_1c_2}$,

$$
\frac{D(\lambda^\star) - D(\lambda^2)}{D(\lambda^\star) - D(\lambda^1)} \le \frac{1}{1 + L_p\left(2\sqrt{c_1c_2} - c_2\right)}.
$$

(ii) If $c_1 < c_2$, then

$$
\frac{D(\lambda^\star) - D(\lambda^2)}{D(\lambda^\star) - D(\lambda^1)} \le \tag{28}
$$
$$
\frac{4c_2^2 - 2c_2\sqrt{c_1c_2} - t^2}{4c_2^2 - 2c_2\sqrt{c_1c_2} - t^2 + L_pt\left(8c_2^2 + 5c_2t - 2\sqrt{c_1c_2}\left(1 + \frac{t}{c_1}\right)(2c_2 + t)\right)}.
$$

Proof. The proof is analogous to that of Theorem 3. First, we prove (i). Assume that α denotes the right hand side of inequality (27). As $2c_1c_2 - t^2 > 0$ and $4c_1c_2 - c_2t - 2t^2 > 0$, we have $0 < \alpha < 1$. With some algebra, one can show that

$$
\begin{aligned}
&f^\star + g^\star - \hat{f}^2 - g^2 - \langle Ax^0 + Bz^0 + tAx^2 + tBz^2, A\hat{x}^2 + Bz^2 \rangle - \\
&\alpha \left(f^\star + g^\star - \hat{f}^1 - g^1 - \langle Ax^0 + Bz^0, A\hat{x}^1 + Bz^1 \rangle \right) + \\
&\alpha \left(\hat{f}^2 - \hat{f}^1 + \langle A^TAx^0 + A^TBz^0, \hat{x}^2 - \hat{x}^1 \rangle - \tfrac{\mu_1}{2} \left\| \hat{x}^2 - \hat{x}^1 \right\|^2 \right) + \\
&\alpha \left(f^2 - \hat{f}^2 + \langle A^TAx^0 + A^TBz^0 + tA^TAx^2 + tA^TBz^2, x^2 - \hat{x}^2 \rangle - \tfrac{\mu_1}{2} \left\| x^2 - \hat{x}^2 \right\|^2 \right) + \\
&\alpha \left(\hat{f}^2 - f^2 + \langle A^TAx^0 + A^TBz^0 + tA^TAx^2 + tA^TBz^1, \hat{x}^2 - x^2 \rangle - \tfrac{\mu_1}{2} \left\| \hat{x}^2 - x^2 \right\|^2 \right) + \\
&\alpha \left(g^2 - g^1 + \langle B^TAx^0 + B^TBz^0, z^2 - z^1 \rangle - \tfrac{\mu_2}{2} \left\| z^2 - z^1 \right\|^2 \right) + \\
&(1-\alpha)\big(- f^\star - g^\star + \langle Ax^0 + Bz^0 + tAx^2 + tBz^2, A\hat{x}^2 + Bz^2 \rangle + \hat{f}^2 + \\
&\quad + g^2 + \tfrac{1}{2L_p} \left\| A\hat{x}^2 + Bz^2 \right\|^2 \big) \\
&\le \tfrac{-c_1\alpha}{2} \left\| A\hat{x}^1 - A\hat{x}^2 \right\| - \tfrac{c_2\alpha}{2} \left\| Bz^1 - Bz^2 + \tfrac{t}{c_2} Ax^2 - \tfrac{t}{c_2} A\hat{x}^2 \right\| - \\
&\alpha (c_1 - \tfrac{t^2}{2c_2}) \left\| Ax^2 + \tfrac{tc_2}{2c_1c_2 - t^2} Bz^2 - \tfrac{tc_2 - 2c_1c_2 + t^2}{2c_1c_2 - t^2} A\hat{x}^2 \right\|.
\end{aligned}
$$

Note that the inequality follows from $\|Ax\|^2 \le \lambda_{\max}(A^TA)\|x\|^2$, $\|Bz\|^2 \le \lambda_{\max}(B^TB)\|z\|^2$ and rearranging terms. Hence, we get

$$
\begin{aligned}
&f^\star + g^\star - \hat{f}^2 - g^2 - \langle Ax^0 + Bz^0 + tAx^2 + tBz^2, A\hat{x}^2 + Bz^2 \rangle \le \\
&\quad \alpha \left(f^\star + g^\star - \hat{f}^1 - g^1 - \langle Ax^0 + Bz^0, A\hat{x}^1 + Bz^1 \rangle \right)
\end{aligned}
$$

for any feasible point of problem (24) and the proof of the first part is complete. For (ii), we proceed analogously to the proof of (i), but with different Lagrange multipliers. Let β denote the right hand side of inequality (28). It is seen that $0 < \beta < 1$. By doing some calculus, we have

$$
\begin{aligned}
&f^\star + g^\star - \hat{f}^2 - g^2 - \langle Ax^0 + Bz^0 + tAx^2 + tBz^2, A\hat{x}^2 + Bz^2\rangle - \\
&\beta\left(f^\star + g^\star - \hat{f}^1 - g^1 - \langle Ax^0 + Bz^0, A\hat{x}^1 + Bz^1\rangle\right) + \\
&\beta\left(\hat{f}^2 - \hat{f}^1 + \langle A^TAx^0 + A^TBz^0, \hat{x}^2 - \hat{x}^1\rangle - \tfrac{\mu_1}{2}\left\|\hat{x}^2 - \hat{x}^1\right\|^2\right) + \\
&\sqrt{\tfrac{c_2}{c_1}}\beta\left(f^2 - \hat{f}^2 + \langle A^TAx^0 + A^TBz^0 + tA^TAx^2 + tA^TBz^2, x^2 - \hat{x}^2\rangle - \tfrac{\mu_1}{2}\left\|x^2 - \hat{x}^2\right\|^2\right) + \\
&\sqrt{\tfrac{c_2}{c_1}}\beta\left(\hat{f}^2 - f^2 + \langle A^TAx^0 + A^TBz^0 + tA^TAx^2 + tA^TBz^1, \hat{x}^2 - x^2\rangle - \tfrac{\mu_1}{2}\left\|\hat{x}^2 - x^2\right\|^2\right) + \\
&\sqrt{\tfrac{c_2}{c_1}}\beta\left(g^2 - g^1 + \langle B^TAx^0 + B^TBz^0, z^2 - z^1\rangle - \tfrac{\mu_2}{2}\left\|z^2 - z^1\right\|^2\right) + \\
&\left(\sqrt{\tfrac{c_2}{c_1}} - 1\right)\beta\Big(g^1 - g^2 + \langle B^TAx^0 + B^TBz^0 + tB^TAx^2 + tB^TBz^2, z^1 - z^2\rangle - \\
&\tfrac{\mu_2}{2}\left\|z^1 - z^2\right\|^2\Big) + (1-\beta)\big(-f^\star - g^\star + \langle Ax^0 + Bz^0 + tAx^2 + tBz^2, A\hat{x}^2 + Bz^2\rangle + \\
&\hat{f}^2 + g^2 + \tfrac{1}{2L_p}\left\|A\hat{x}^2 + Bz^2\right\|^2\big) \\
&\leq -\tfrac{c_1\beta}{2}\left\|A\hat{x}^1 - A\hat{x}^2\right\|^2 - (\sqrt{c_1c_2}\beta)\left\|Ax^2 - \left(1 - \frac{t}{2\sqrt{c_1c_2}}\right)A\hat{x}^2 + \frac{t}{2\sqrt{c_1c_2}}Bz^1\right\|^2 - \\
&\left(\frac{\beta-1}{2L_p} + \beta t\left(1 - \frac{t}{4\sqrt{c_1c_2}}\right)\right)\left\|A\hat{x}^2 - \left(\frac{\beta L_p\left(-2c_2\sqrt{c_1c_2} + 4c_2^2 - t^2\right)}{-\beta L_p t^2 + 2\sqrt{c_1c_2}(2\beta L_p t + \beta - 1)}\right)^{\frac{1}{2}}Bz^1 + \right. \\
&\left.\left(\frac{2\left(2\beta c_2 L_p\left(t + c_2\right) + \sqrt{c_1c_2}\left(\beta - \beta L_p c_2 - 1\right)\right)}{-\beta L_p t^2 + 2\sqrt{c_1c_2}(2\beta L_p t + \beta - 1)}\right)^{\frac{1}{2}}Bz^2\right\|^2.
\end{aligned}
$$

The rest of the proof is similar to that of the former case. □

We computed the bounds in Theorem 6 by selecting suitable Lagrangian multipliers and solving the semidefinite formulation of problem (26) by hand. The semidefinite formulation is formed analogous to problem (15). Note that the optimal value of problem (26) may be smaller than the bounds introduced in Theorem 6. Indeed, our aim was to provide a concrete mathematical proof for the linear convergence rate. However, the linear convergence rate factor is not necessarily tight. Needless to say that the optimal value of problem (26) also does not necessarily give the tight convergence factor as it is just a relaxation of problem (24).

Recently the authors showed that the PŁ inequality is necessary and sufficient conditions for the linear convergence of the gradient method with constant step lengths for L-smooth function; see [2, Theorem 5]. In what follows, we establish that the PŁ inequality is a necessary condition for the linear convergence of ADMM. Firstly, we present a lemma that is very useful for our proof.

Lemma 2 *Let* $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ *and* $g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m)$. *Consider Algorithm 1. If* $(\hat{x}^1, z^1) \in \operatorname{argmin} f(x) + g(z) + \langle\lambda^1, Ax + Bz - b\rangle$, *then*

$$\langle A\hat{x}^1 + Bz^1 - b, Ax^2 + Bz^2 - b\rangle \leq \left\|A\hat{x}^1 + Bz^1 - b\right\|^2. \tag{29}$$

Proof. Without loss of generality we assume that $\mu_1 = \mu_2 = 0$. By optimality conditions, we have

$$f(\hat{x}^1) - \langle A^T\lambda^1, x^2 - \hat{x}^1\rangle \le f(x^2), \quad g(z^1) - \langle B^T\lambda^1, z^2 - z^1\rangle \le g(z^2),$$
$$f(x^2) - \langle A^T\lambda^1 + tA^T(Ax^2 + Bz^1 - b), \hat{x}^1 - x^2\rangle \le f(\hat{x}^1),$$
$$g(z^2) - \langle B^T\lambda^1 + tB^T(Ax^2 + Bz^2 - b), z^1 - z^2\rangle \le g(z^1).$$

By using these inequities, we get

$$\begin{aligned}
0 \le & \tfrac{1}{t}\left(f(x^2) - f(\hat{x}^1) + \left\langle A^T\lambda^1, x^2 - \hat{x}^1\right\rangle\right) + \tfrac{1}{t}\left(g(z^2) - g(z^1) + \left\langle B^T\lambda^1, z^2 - z^1\right\rangle\right) + \\
& \tfrac{1}{t}\left(f(\hat{x}^1) - f(x^2) + \left\langle A^T\lambda^1 + tA^T(Ax^2 + Bz^1 - b), \hat{x}^1 - x^2\right\rangle\right) + \\
& \tfrac{1}{t}\left(g(z^1) - g(z^2) + \left\langle B^T\lambda^1 + tB^T(Ax^2 + Bz^2 - b), z^1 - z^2\right\rangle\right) \\
= & \left\|A\hat{x}^1 + Bz^1 - b\right\|^2 - \langle A\hat{x}^1 + Bz^1 - b, Ax^2 + Bz^2 - b\rangle - \tfrac{3}{4}\left\|B\left(z^1 - z^2\right)\right\|^2 - \\
& \left\|A\left(\hat{x}^1 - x^2\right) + \tfrac{1}{2}B\left(z^1 - z^2\right)\right\|^2.
\end{aligned}$$

Hence, we have

$$\frac{\langle A\hat{x}^1 + Bz^1 - b, Ax^2 + Bz^2 - b\rangle}{\left\|A\hat{x}^1 + Bz^1 - b\right\|^2} \le 1,$$

which completes the proof. □

The next theorem establishes that the PŁ inequality is a necessary condition for the linear convergence of ADMM.

Theorem 7 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m)$. If Algorithm 1 is linearly convergent with respect to the dual objective value, then D satisfies the PŁ inequality.*

Proof. Consider $\lambda^1 \in \mathbb{R}^r$ and $\xi \in b - A\partial f^*(-A^T\lambda^1) - B\partial g^*(-B^T\lambda^1)$. Hence, $\xi = b - A\hat{x}^1 - Bz^1$ for some $(\hat{x}^1, z^1) \in \operatorname{argmin} f(x) + g(z) + \langle \lambda, Ax + Bz - b\rangle$. If one sets $\hat{z} = z^1$ and $\lambda^0 = \lambda^1 - t(A\hat{x}^1 + Bz^1 - b)$ in Algorithm 1, the algorithm may generate λ^1. As Algorithm 1 is linearly convergent, there exist $\gamma \in [0, 1)$ with

$$D(\lambda^\star) - D(\lambda^2) \le \gamma\left(D(\lambda^\star) - D(\lambda^1)\right).$$

So, we have

$$(1-\gamma)\left(D(\lambda^\star) - D(\lambda^1)\right) \le D(\lambda^2) - D(\lambda^1) \le \left\langle A\hat{x}^1 + Bz^1 - b, \lambda^2 - \lambda^1\right\rangle,$$

where the last inequality follows from the concavity of the function D. Since $\lambda^2 - \lambda^1 = t(Ax^2 + Bz^2 - b)$, Lemma 2 implies that

$$D(\lambda^\star) - D(\lambda^1) \le \tfrac{t}{1-\gamma}\|\xi\|^2,$$

so D satisfies the PŁ inequality. □

By combining Theorem 6 and 7, we get the following corollary.

Corollary 1 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ and $g \in \mathcal{F}_{\mu_2}(\mathbb{R}^m)$ with $\mu_1, \mu_2 > 0$ and let $t \leq \sqrt{\frac{\mu_1\mu_2}{\lambda_{\max}(A^TA)\lambda_{\max}(B^TB)}}$. Algorithm 1 is linearly convergent with respect to the dual objective value if and only if D satisfies the PŁ inequality.*

Another assumption used for establishing linear convergence is L-smoothness; see for example [6,7,11,31]. Deng et al. [7] show that the sequence $\{(x^k, z^k, \lambda^k)\}$ is convergent linearly to a saddle point under the two scenarios given in Table 1.

Table 1: Scenarios leading to linear convergence rates

Scenario	Strong convexity	Lipschitz continuity	Full row rank
1	f, g	∇f	A
2	f, g	$\nabla f, \nabla g$	-
3	f	$\nabla f, \nabla g$	B^T

It is worth mentioning that Scenario 1 or Scenario 2 implies strong convexity of the dual objective function and therefore the PŁ inequality is resulted, see [2]. Hence, Theorem 6 implies the linear convergence in terms of dual value under Scenario 1 or Scenario 2. Deng et al. [7] studied the linear convergence under Scenario 3, but they just proved the linear convergence of the sequence $\{(x^k, Bz^k, \lambda^k)\}$. In the next section, we investigate the R-linear convergence without assuming L-smoothness of f. Indeed, we establish the R-linear convergence when f is strongly convex, g is L-smooth and B has full row rank.

Note that the PŁ inequality does not imply necessarily Scenario 1 or Scenario 2. Indeed, consider the following optimization problem,

$$\begin{aligned} \min\ & f(x) + g(z), \\ \text{s.t.}\ & x + z = 0, \\ & x, z \in \mathbb{R}^n, \end{aligned}$$

where $f(x) = \frac{1}{2}\|x\|^2 + \|x\|_1$ and $g(z) = \frac{1}{2}\|z\|^2 + \|z\|_1$. With some algebra, one may show that $D(\lambda) = \sum_{i=1}^n h(\lambda_i)$ with

$$h(s) = \begin{cases} -(s-1)^2, & s > 1 \\ 0, & |s| \leq 1 \\ -(s+1)^2, & s < -1. \end{cases}$$

Hence, the PŁ inequality holds for $L_p = \frac{1}{2}$ while neither f nor g is L-smooth.

As mentioned earlier the performance estimation problem including the PŁ inequality at finite set of points is a relaxation for computing the worst-case convergence rate. Contrary to Theorem 6, we could not manage to prove the linear convergence of primal and dual residuals under the assumptions of Theorem 6 by employing performance estimation.

4 R-linear convergence of ADMM

In this section, we study the R-linear convergence of ADMM. Recall that ADMM enjoys R-linear convergent in terms of dual objective value if

$$D(\lambda^\star) - D(\lambda^N) \le c\gamma^N,$$

for some $c \ge 0$ and $\gamma \in [0, 1)$. Note that linear convergence implies R-linear convergence.

We investigate the R-linear convergence under the following scenarios:

- (S1): $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ is L-smooth with $\mu_1 > 0$ and A has full row rank;
- (S2): $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ with $\mu_1 > 0$, g is L-smooth and B has full row rank.

Our technique for proving the R-linear convergence is based on establishing the linear convergence of the sequence $\{V^k\}$ given by

$$V^k = \|\lambda^k - \lambda^\star\|^2 + t^2 \left\| B\left(z^k - z^\star\right)\right\|^2. \tag{30}$$

Note that V^k is called Lyapunov function for ADMM; see [5].

First we consider the case that f is μ-strongly convex and L-smooth. The following proposition establishes the linear convergence of $\{V^k\}$.

Proposition 2 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ be L-smooth with $\mu_1 > 0$, $g \in \mathcal{F}_0(\mathbb{R}^m)$ and let A has full row rank. If $t < \sqrt{\frac{\mu_1 L}{\lambda_{\min}(AA^T)\lambda_{\max}(A^TA)}}$, then*

$$V^{k+1} \le \left(1 - \tfrac{2\lambda_{\min}(AA^T)\mu_1 t}{L\mu_1 + 2\lambda_{\min}(AA^T)\mu_1 t + \lambda_{\min}(AA^T)\lambda_{\max}(A^TA)t^2}\right) V^k. \tag{31}$$

Proof. By optimality conditions, we have

$$\begin{aligned}
\nabla f(x^{k+1}) &= -A^T\left(\lambda^k + tAx^{k+1} + tBz^k\right), \qquad & \eta^k &= -B^T\lambda^{k+1},\\
\nabla f(x^\star) &= -A^T\lambda^\star, & \eta^\star &= -B^T\lambda^\star,
\end{aligned}$$

for some $\eta^k \in \partial g(z^{k+1})$ and $\eta^\star \in \partial g(z^\star)$. Analogous to Section 2, we may assume without loss of generality $x^\star = 0$ and $z^\star = 0$. Let $c = \frac{\mu_1}{\lambda_{\max}(A^TA)}$, $d = \frac{L}{\lambda_{\min}(AA^T)}$ and $\alpha = \frac{2t}{c^2d^2 + 2cdt^2 - 4c^2t^2 + t^4}$. By Theorem 2, we get

$$\begin{aligned}
0 \le& \alpha\left(t^2 + cd\right)^2\Bigg(f(x^{k+1}) + \left\langle A^T\lambda^\star, x^{k+1}\right\rangle - \tfrac{1}{2L}\left\|A^T\left(\lambda^k + tAx^{k+1} + tBz^k - \lambda^\star\right)\right\|^2 -\\
&f(x^\star)\Bigg) + \alpha\left(c^2d^2 - t^4\right)\Bigg(f(x^\star) - f(x^{k+1}) - \left\langle A^T\left(\lambda^k + tAx^{k+1} + tBz^k\right), x^{k+1}\right\rangle - \tfrac{1}{2L}\\
&\left\|A^T\left(\lambda^k + tAx^{k+1} + tBz^k - \lambda^\star\right)\right\|^2\Bigg) + 2\alpha t^2\left(cd + t^2\right)\Bigg(f(x^\star) - f(x^{k+1}) - \tfrac{\mu_1}{2}\left\|x^{k+1}\right\|^2\\
&- \left\langle A^T\left(\lambda^k + tAx^{k+1} + tBz^k\right), x^{k+1}\right\rangle\Bigg) + 2t\Big(g(z^{k+1}) - g(z^\star) + \left\langle B^T\lambda^\star, z^{k+1}\right\rangle\Big) +\\
&2t\Big(g(z^\star) - g(z^{k+1}) + \left\langle B^T\lambda^{k+1}, -z^{k+1}\right\rangle\Big)\\
\le& -2\alpha c^2 t\left\|\lambda^k - \lambda^\star + \tfrac{t^2 + 2ct + cd}{2c}Ax^{k+1} + \tfrac{t^2 + cd}{2c}Bz^k\right\|^2 + \left(1 - \tfrac{2ct}{cd + 2ct + t^2}\right)\\
&\left(\left\|\lambda^k - \lambda^\star\right\|^2 + t^2\left\|Bz^k\right\|^2\right) - \left(\left\|\lambda^{k+1} - \lambda^\star\right\|^2 + t^2\left\|Bz^{k+1}\right\|^2\right).
\end{aligned}$$

Note that the last inequality follows from $\|x\|^2 \geq \frac{c}{\mu_1}\|Ax\|^2$, $\|A^T\lambda\|^2 \geq \frac{L}{d}\|\lambda\|^2$ and $\lambda^{k+1} = \lambda^k + tAx^{k+1} + tBz^{k+1}$. The above inequality implies that

$$V^{k+1} \leq \left(1 - \frac{2ct}{cd+2ct+t^2}\right) V^k,$$

and the proof is complete. □

Note that one can improve bound (31) under the assumptions of Proposition 2 by employing the following inequality

$$\frac{1}{2(1-\frac{\mu_1}{L})}\left(\frac{1}{L}\left\|\nabla f(x) - \nabla f(y)\right\|^2 + \mu_1\left\|x-y\right\|^2 - \frac{2\mu_1}{L}\left\langle \nabla f(x) - \nabla f(y), x-y\right\rangle\right) \\ \leq f(y) - f(x) - \left\langle \nabla f(x), y-x\right\rangle.$$

Indeed, we employed the given inequality but we could not manage to obtain a closed form formula for the convergence rate. The next theorem establishes the R-linear convergence of ADMM in terms of dual objective value under the assumptions of Proposition 2.

Theorem 8 *Let $N \geq 2$ and let A has full row rank. Suppose that $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ is L-smooth with $\mu_1 > 0$ and $g \in \mathcal{F}_0(\mathbb{R}^m)$. If*

$$t < \min\{\tfrac{\mu_1}{\lambda_{\max}(A^TA)}, \sqrt{\tfrac{\mu_1 L}{\lambda_{\min}(AA^T)\lambda_{\max}(A^TA)}}\},$$

then

$$D(\lambda^\star) - D(\lambda^N) \leq \frac{V^0}{8t}\left(1 - \frac{2\lambda_{\min}(AA^T)\mu_1 t}{L\mu_1 + 2\lambda_{\min}(AA^T)\mu_1 t + \lambda_{\min}(AA^T)\lambda_{\max}(A^TA)t^2}\right)^{N-2}.$$

Proof. By Theorem 3 and Proposition 2, one can infer the following inequalities,

$$\begin{aligned} D(\lambda^\star) - D(\lambda^N) &\leq \frac{V^{N-2}}{8t} \\ &\leq \frac{V^0}{8t}\left(1 - \frac{2\lambda_{\min}(AA^T)\mu_1 t}{L\mu_1 + 2\lambda_{\min}(AA^T)\mu_1 t + \lambda_{\min}(AA^T)\lambda_{\max}(A^TA)t^2}\right)^{N-2}, \end{aligned}$$

which shows the desired inequality. □

Nishihara et al. [31] showed the R-linear convergence of ADMM in terms of $\{x^k, z^k, \lambda^k\}$ under the following conditions:

i) The function $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ is L-smooth with $\mu_1 > 0$;
ii) The matrix A is invertible and that B has full column rank.

In Theorem 8, we obtain the R-linear convergence under weaker assumptions. Indeed, we replace condition *ii*) with the matrix A having full row rank.

In the sequel, we investigate the R-linear convergence under the hypotheses of scenario (S2). The next proposition shows the linear convergence of $\{V^k\}$.

Proposition 3 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ with $\mu_1 > 0$ and let $g \in \mathcal{F}_0(\mathbb{R}^m)$ be an L-smooth function. Suppose that B has full row rank and $k \geq 1$. If $t \leq \frac{1}{2}\min\{\frac{\mu_1}{\lambda_{\max}(A^TA)}, \frac{L}{\lambda_{\min}(BB^T)}\}$, then*

$$V^{k+1} \leq \left(\tfrac{L}{L+t\lambda_{\min}(BB^T)}\right)^2 V^k. \tag{32}$$

Proof. Due to the optimality conditions, we have

$$\begin{aligned}
&\xi^{k+1} = -A^T\left(\lambda^k + tAx^{k+1} + tBz^k\right), && && \xi^\star = -A^T\lambda^\star,\\
&\nabla g(z^k) = -B^T\lambda^k, && \nabla g(z^{k+1}) = -B^T\lambda^{k+1}, && \nabla g(z^{k+1}) = -B^T\lambda^\star,
\end{aligned}$$

for some $\xi^{k+1} \in \partial f(x^{k+1})$ and $\xi^\star \in \partial f(x^\star)$. Without loss of generality, we assume that $x^\star = 0$ and $z^\star = 0$. Suppose that $c = \frac{\mu_1}{\lambda_{\max}(A^TA)}$, $d = \frac{L}{\lambda_{\min}(BB^T)}$ and $\alpha = \frac{2dt}{d+t}$. By Theorem 2, we obtain

$$\begin{aligned}
0 \leq &\left(\frac{\alpha\left(d^2+t^2\right)}{d^2-t^2}\right)\left(f(x^\star) - f(x^{k+1}) - \left\langle A^T\left(\lambda^k + tAx^{k+1} + tBz^k\right), x^{k+1}\right\rangle - \tfrac{\mu_1}{2}\left\|x^{k+1}\right\|^2\right)\\
&+\left(\frac{\alpha\left(d^2+t^2\right)}{d^2-t^2}\right)\left(f(x^{k+1}) - f(x^\star) + \left\langle A^T\lambda^\star, x^{k+1}\right\rangle - \tfrac{\mu_1}{2}\left\|x^{k+1}\right\|^2\right)\\
&+\alpha\left(g(z^{k+1}) - g(z^\star) + \left\langle B^T\lambda^\star, z^{k+1}\right\rangle - \tfrac{1}{2L}\left\|B^T\left(\lambda^\star - \lambda^{k+1}\right)\right\|^2\right)\\
&+\alpha\left(g(z^\star) - g(z^{k+1}) - \left\langle B^T\lambda^{k+1}, z^{k+1}\right\rangle - \tfrac{1}{2L}\left\|B^T\left(\lambda^\star - \lambda^{k+1}\right)\right\|^2\right)\\
&+\alpha\left(g(z^k) - g(z^{k+1}) + \left\langle B^T\lambda^{k+1}, z^k - z^{k+1}\right\rangle - \tfrac{1}{2L}\left\|B^T\left(\lambda^{k+1} - \lambda^k\right)\right\|^2\right)\\
&+\alpha\left(g(z^{k+1}) - g(z^k) + \left\langle B^T\lambda^k, z^{k+1} - z^k\right\rangle - \tfrac{1}{2L}\left\|B^T\left(\lambda^{k+1} - \lambda^k\right)\right\|^2\right)\\
\leq &\tfrac{-\alpha^2}{4}\left\|\left(\frac{2t^2}{d^2-dt}\right)Ax^{k+1} + Bz^k - \left(1+\tfrac{t}{d}\right)Bz^{k+1}\right\|^2 - \frac{2t\left(d^2+t^2\right)\left(cd^2-dt(c+t)-t^3\right)}{\left(d^2-t^2\right)^2}\\
&\left\|Ax^{k+1}\right\|^2 - \tfrac{\alpha^2}{4d^2}\left\|\lambda^k - \lambda^\star + \left(\frac{2d^2-(d-t)^2}{d-t}\right)Ax^{k+1} + (d+t)\,Bz^{k+1}\right\|^2\\
&+\left(\tfrac{d}{d+t}\right)^2\left(\left\|\lambda^k - \lambda^\star\right\|^2 + t^2\left\|Bz^k\right\|^2\right) - \left(\left\|\lambda^{k+1} - \lambda^\star\right\|^2 + t^2\left\|Bz^{k+1}\right\|^2\right).
\end{aligned}$$

The last inequality is obtained by using $\|x\|^2 \geq \frac{c}{\mu_1}\|Ax\|^2$, $\|B^T\lambda\|^2 \geq \frac{L}{d}\|\lambda\|^2$ and $\lambda^{k+1} = \lambda^k + tAx^{k+1} + tBz^{k+1}$. Hence, we have

$$V^{k+1} \leq \left(\tfrac{d}{d+t}\right)^2 V^k,$$

and the proof is complete. □

As the sequence $\{V^k\}$ is not increasing [5], we have $V^1 \leq V^0$. Thus, by using Theorem 3 and Proposition 3, one can infer the following theorem.

Theorem 9 *Let $f \in \mathcal{F}_{\mu_1}(\mathbb{R}^n)$ with $\mu_1 > 0$ and let $g \in \mathcal{F}_0(\mathbb{R}^m)$ be L-smooth. Assume that $N \geq 3$ and B has full row rank. If $t < \min\{\frac{\mu_1}{2\lambda_{\max}(A^TA)}, \frac{L}{2\lambda_{\min}(BB^T)}\}$, then*

$$D(\lambda^\star) - D(\lambda^N) \leq \frac{V^0}{8t}\left(\frac{L}{L+t\lambda_{\min}(BB^T)}\right)^{2N-6}. \quad (33)$$

In the same line, one can infer the R-linear convergence in terms of primal and dual residuals under the assumptions of Theorem 8 and Theorem 9. In the section, we prove the linear convergence of $\{V^k\}$ under two scenarios (S1) and (S2). By (6), it is readily seen that the dual objective function is strongly convex under the hypotheses of both scenarios (S1) and (S2). Therefore, both scenarios imply the PŁ inequality. One may wonder that if the PŁ inequality and the strong convexity of f imply the linear of $\{V^k\}$. By using performance estimation, we could not establish such an implication.

Concluding remarks

In this paper we developed performance estimation framework to handle dual-based methods. Thanks to this framework, we could obtain some tight convergence rates for ADMM. This framework may be exploited for the analysis of other variants of ADMM in the ergodic and non-ergodic sense. Moreover, similarly to [22], one can apply this framework for introducing and analyzing new accelerated ADMM variants.

It is worth pointing out that all given results in this paper hold under a weaker assumption on strong convexity. Indeed, all the arguments work in the same way if one replaces the strong convexity of f (or g) with the convexity of function $f - c_1\|\cdot\|_A^2$ (or $g - c_2\|\cdot\|_B^2$) for some $c_1 \geq 0$ (or $c_2 \geq 0$), where the seminorm $\|\cdot\|_A$ is defined as $\|x\|_A = \|Ax\|$. Moreover, most results hold for positive step length, t, but we managed to get closed form formulas for some interval of positive numbers.

References

1. Abbaszadehpeivasti, H., de Klerk, E., Zamani, M.: On the rate of convergence of the Difference-of-Convex Algorithm (DCA). arXiv preprint arXiv:2109.13566 (2021)
2. Abbaszadehpeivasti, H., de Klerk, E., Zamani, M.: Conditions for linear convergence of the gradient method for non-convex optimization. arXiv preprint arXiv:2204.00647 (2022)
3. Beck, A.: First-order methods in optimization. SIAM (2017)
4. Bolte, J., Nguyen, T.P., Peypouquet, J., Suter, B.W.: From error bounds to the complexity of first-order descent methods for convex functions. Mathematical Programming **165**(2), 471–507 (2017)
5. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3**(1), 1–122 (2011)
6. Davis, D., Yin, W.: Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions. Mathematics of Operations Research **42**(3), 783–805 (2017)

7. Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. Journal of Scientific Computing **66**(3), 889–916 (2016)
8. Drori, Y., Teboulle, M.: Performance of first-order methods for smooth convex minimization: a novel approach. Mathematical Programming **145**(1), 451–482 (2014)
9. Franca, G., Robinson, D., Vidal, R.: ADMM and accelerated ADMM as continuous dynamical systems. In: International Conference on Machine Learning, pp. 1559–1567. PMLR (2018)
10. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers & mathematics with applications **2**(1), 17–40 (1976)
11. Giselsson, P., Boyd, S.: Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. IEEE Transactions on Automatic Control **62**(2), 532–544 (2016)
12. Glowinski, R., Marroco, 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. ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique **9**(R2), 41–76 (1975)
13. Glowinski, R., Osher, S.J., Yin, W.: Splitting methods in communication, imaging, science, and engineering. Springer (2017)
14. Goldfarb, D., Ma, S., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. Mathematical Programming **141**(1), 349–382 (2013)
15. Goldstein, T., O'Donoghue, B., Setzer, S., Baraniuk, R.: Fast alternating direction optimization methods. SIAM Journal on Imaging Sciences **7**(3), 1588–1623 (2014)
16. Gu, G., Yang, J.: On the dual step length of the alternating direction method of multipliers. arXiv preprint arXiv:2006.08309 (2020)
17. Han, D., Sun, D., Zhang, L.: Linear rate convergence of the alternating direction method of multipliers for convex composite programming. Mathematics of Operations Research **43**(2), 622–637 (2018)
18. Han, D.R.: A survey on some recent developments of alternating direction method of multipliers. Journal of the Operations Research Society of China **10**(1), 1–52 (2022)
19. Hastie, T., Tibshirani, R., Wainwright, M.: Statistical learning with sparsity. Monographs on statistics and applied probability **143**, 143 (2015)
20. He, B., Yuan, X.: On the O(1/n) convergence rate of the Douglas-Rachford alternating direction method. SIAM Journal on Numerical Analysis **50**(2), 700–709 (2012)
21. Hong, M., Luo, Z.Q.: On the linear convergence of the alternating direction method of multipliers. Mathematical Programming **162**(1), 165–199 (2017)
22. Kim, D., Fessler, J.A.: Optimized first-order methods for smooth convex minimization. Mathematical programming **159**(1), 81–107 (2016)
23. Li, H., Lin, Z.: Accelerated alternating direction method of multipliers: An optimal O(1/k) nonergodic analysis. Journal of Scientific Computing **79**(2), 671–699 (2019)
24. Lin, Z.: Alternating Direction Method of Multipliers for Machine Learning. Springer Nature
25. Liu, H., Shi, Y., Wang, Z., Ran, L., Lü, Q., Li, H.: A distributed algorithm based on relaxed ADMM for energy resources coordination. International Journal of Electrical Power & Energy Systems **135**, 107482 (2022)
26. Liu, Y., Yuan, X., Zeng, S., Zhang, J.: Partial error bound conditions and the linear convergence rate of the alternating direction method of multipliers. SIAM Journal on Numerical Analysis **56**(4), 2095–2123 (2018)
27. Lozenski, L., Villa, U.: Consensus ADMM for inverse problems governed by multiple PDE models. arXiv preprint arXiv:2104.13899 (2021)
28. Madani, R., Kalbat, A., Lavaei, J.: ADMM for sparse semidefinite programming with applications to optimal power flow problem. In: 2015 54th IEEE Conference on Decision and Control (CDC), pp. 5932–5939. IEEE (2015)
29. Monteiro, R.D., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. SIAM Journal on Optimization **23**(1), 475–507 (2013)
30. Necoara, I., Nesterov, Y., Glineur, F.: Linear convergence of first order methods for non-strongly convex optimization. Mathematical Programming **175**(1), 69–107 (2019)

31. Nishihara, R., Lessard, L., Recht, B., Packard, A., Jordan, M.: A general analysis of the convergence of ADMM. In: International Conference on Machine Learning, pp. 343–352. PMLR (2015)
32. Peña, J., Vera, J.C., Zuluaga, L.F.: Linear convergence of the Douglas-Rachford algorithm via a generic error bound condition. arXiv preprint arXiv:2111.06071 (2021)
33. Rockafellar, R.T.: Convex analysis. In: Convex analysis. Princeton university press (1970)
34. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena **60**(1-4), 259–268 (1992)
35. Ryu, E.K., Taylor, A.B., Bergeling, C., Giselsson, P.: Operator splitting performance estimation: Tight contraction factors and optimal parameter selection. SIAM Journal on Optimization **30**(3), 2251–2271 (2020)
36. Sabach, S., Teboulle, M.: Faster Lagrangian-based methods in convex optimization. SIAM Journal on Optimization **32**(1), 204–227 (2022)
37. Stellato, B., Banjac, G., Goulart, P., Bemporad, A., Boyd, S.: Osqp: An operator splitting solver for quadratic programs. Mathematical Programming Computation **12**(4), 637–672 (2020)
38. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Smooth strongly convex interpolation and exact worst-case performance of first-order methods. Mathematical Programming **161**(1-2), 307–345 (2017)
39. Yuan, X., Zeng, S., Zhang, J.: Discerning the linear convergence of admm for structured convex optimization through the lens of variational analysis. J. Mach. Learn. Res. **21**, 83–1 (2020)