SIAM J. OPTIM.
Vol. 30, No. 3, pp. 1905–1921

TIGHT SUBLINEAR CONVERGENCE RATE OF THE PROXIMAL POINT ALGORITHM FOR MAXIMAL MONOTONE INCLUSION PROBLEMS*

GUOYONG GU† AND JUNFENG YANG†

Abstract. The tight sublinear convergence rate of the proximal point algorithm for maximal monotone inclusion problems is established based on the squared fixed point residual. By using the performance estimation framework, the tight sublinear convergence rate problem is written as an infinite dimensional nonconvex optimization problem, which is then equivalently reformulated as a finite dimensional semidefinite programming (SDP) problem. By solving the SDP, the exact sublinear rate is computed numerically. Theoretically, by constructing a feasible solution to the dual SDP, an upper bound is obtained for the tight sublinear rate. On the other hand, an example in two dimensional space is constructed to provide a lower bound. The lower bound matches exactly the upper bound obtained from the dual SDP, which also coincides with the numerical rate computed. Hence, we have established the worst case sublinear convergence rate, which is tight in terms of both the order and the constants involved.

Key words. proximal point algorithm, maximal monotone inclusion, tight sublinear convergence rate, performance estimation framework, semidefinite programming

AMS subject classifications. 65K05, 65K10, 65J22, 90C25

DOI. 10.1137/19M1299049

1. Introduction. The term "proximal point" was originally coined by Moreau [28, 27] and was first introduced to solve optimization problems by Martinet [24, 25]. Later, the proximal point algorithm (PPA) was refined by Rockafellar in [33] for solving maximal monotone operator inclusion problems. Since then, PPA has been playing fundamental roles in the understanding, design, and analysis of the optimization algorithms. It was shown in [32] that the classical method of multipliers of Hestenes [15] and Powell [30] is a dual application of the PPA. Similarly, it was shown in [10] that the Douglas–Rachford operator splitting method [7, 23] is also an application of the PPA to a special splitting operator. As for the rate of convergence, it was shown that PPA converges at least linearly under some regularity conditions, e.g., Lipschitz continuity of the inverse operator at the origin [33] or metric subregularity [22]. For minimizing a proper lower semicontinuous convex function, a nonasymptotic $\mathcal{O}(1/N)$ sublinear convergence rate measured by function value residuals has been established in [11], where N denotes the iteration number. See also [12, 2] for some accelerated proximal-point-like methods designed for solving convex optimization problems by using a Nesterov-type acceleration technique [29].

For maximal monotone operator inclusion problems, an $\mathcal{O}(1/N)$ sublinear convergence rate result measured by the squared fixed point residual was derived in [3, Proposition 8] for the original PPA without regularity conditions. Exactly the same nonasymptotic convergence rate was recently established in [14, Theorem 3.1] for the Douglas–Rachford operator splitting method, a generalization of PPA to treat-

*Received by the editors November 12, 2019; accepted for publication (in revised form) April 20, 2020; published electronically July 16, 2020.
https://doi.org/10.1137/19M1299049
Funding: The first author was supported by NSFC grant 11671195. The second author was supported by NSFC grants 11771208 and 11922111.
†Department of Mathematics, Nanjing University, Nanjing, 210093, China (ggu@nju.edu.cn, jfyang@nju.edu.cn).

ing the sum of two maximal monotone operators. However, the tightness of the results in [3, 14] remains unclear. In this paper, we investigate the tight sublinear convergence rate of PPA for maximal monotone inclusion problems in the absence of regularity conditions. The approach is to formulate the tight convergence rate of PPA by using the performance estimation framework originally proposed by Drori and Teboulle [8, 9] to study optimization algorithms and recently refined by Kim and Fessler [17, 18, 19, 20, 21] and Taylor and colleagues [38, 37, 6, 39]. Very recently, the performance estimation idea has also been extended in [36] to analyze stochastic first order optimization algorithms.

Recently, the performance estimation framework was first extended by Ryu et al. [34] to study operator splitting methods for monotone inclusion problems. Following [34], we express the performance estimation problem of PPA as an infinite dimensional nonconvex optimization problem, which is then equivalently reformulated as a finite dimensional convex semidefinite programming (SDP) problem. By constructing a dual feasible (in fact, optimal) SDP solution, we are able to obtain an upper bound on the tight sublinear convergence rate. A two dimensional example is constructed to show that the upper bound obtained from the dual SDP is nonimprovable, neither in the order nor in any of the constants involved. Specifically, we show that the tight convergence rate is $\frac{1}{(1+\frac{1}{N})^N(N+1)}$ when the underlying Euclidean space has dimension greater than or equal to two. Compared to [3, Proposition 8] and [14, Theorem 3.1], the improvement here is approximately a constant factor of $\exp(-1)$, which eliminates the gap between the known bound and the tight worst case bound. In the one dimensional case, the tight sublinear rate is shown to be $1/(N+1)^2$, one order of magnitude faster than the existing result. Recently, a similar analysis for convex composite optimization problems has been given in [37], where the quality of an approximate solution is measured by the function value residual. In this paper, our focus is the more general maximal monotone inclusion problems, and the quality measure is the squared fixed point residual; see, e.g., [3, 14, 4, 5]. A short while after the release of our work on arXiv, an accelerated PPA for monotone inclusion problems, which achieves $\mathcal{O}(1/N^2)$ convergence measured by squared fixed point residual also, was constructed in [16] relying on the performance estimation and interpolation idea for operators [34].

1.1. Notation. In this paper, we restrict our discussion to the n-dimensional Euclidean space $\mathbb{R}^n$, with inner product denoted by $\langle \cdot, \cdot \rangle$ and the induced norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, although all of the analysis can be easily extended to any finite dimensional real Euclidean spaces. The superscript "$\top$" denotes the matrix or vector transpose operation. For a matrix M and integers p, q, s, and t such that $p < q$ and $s < t$, we let $M(p:q, s:t)$ be the submatrix of M located between the pth and the qth rows and the sth and the tth columns. Similarly, $M(:, s:t)$ (resp., $M(p:q, :)$) represents the submatrix of M containing the sth to the tth columns (resp., the pth to the qth rows). The (i,j)th element of M is denoted by $M_{i,j}$. For two matrices A and B of the same size, we let $\langle A, B \rangle = \sum_{i,j} A_{i,j} B_{i,j}$ be the trace inner product. The set of real symmetric matrices of order n is denoted by $\mathbb{S}^n$. A matrix $A \in \mathbb{S}^n$ that is positive semidefinite is also denoted by $A \succeq 0$.

The set of maximal monotone operators on $\mathbb{R}^n$ is denoted by $\mathcal{M}_n$. The graph of $T \in \mathcal{M}_n$ is given by $\operatorname{graph}(T) := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid y \in T(x)\}$, and the inverse operator T^{-1} of T is defined via its graph $\operatorname{graph}(T^{-1}) := \{(y, x) \in \mathbb{R}^n \times \mathbb{R}^n \mid (x, y) \in \operatorname{graph}(T)\}$. The set of zeros of T is denoted by $T^{-1}(0) := \{w^* \in \mathbb{R}^n \mid 0 \in T(w^*)\}$. Both the identity matrix of appropriate dimension and the identity operator will be denoted by I.

1.2. Organization. The rest of this paper is organized as follows. In section 2, we present our main result, while its proof is postponed to section 4. The discovery of our main result relies on the performance estimation problem (PEP) and its reformulation as SDP via operator interpolation [34], which will be introduced in section 3. Finally, a summary is given in section 5.

2. Main result.

Let $T \in \mathcal{M}_n$ be a maximal monotone operator, and let $\lambda > 0$ be any fixed constant. The resolvent operator of T is defined by $J_{\lambda T} = (I + \lambda T)^{-1}$. A fundamental problem in convex analysis and related fields is to find $w^* \in \mathbb{R}^n$ such that $0 \in T(w^*)$, i.e., a zero of T. It is elementary to show that $w^* \in T^{-1}(0)$ if and only if $w^* = J_{\lambda T}(w^*)$, i.e., w^* is a fixed point of $J_{\lambda T}$. The PPA is an iterative scheme constructed based on this fixed point equation. Specifically, given $w^0 \in \mathbb{R}^n$, PPA iterates as

$$w^{k+1} = J_{\lambda T}(w^k), \quad k = 0, 1, 2, \dots. \tag{2.1}$$

Minty [26] has shown that $J_{\lambda T}$ is single-valued and everywhere defined. Therefore, PPA scheme (2.1) is well defined and generates a unique sequence $\{w^k\}_{k=1}^\infty$ from any given $w^0 \in \mathbb{R}^n$.

After $N \geq 1$ iterations, w^N is generated, whose quality is measured by

$$e(w, \lambda) := w - J_{\lambda T}(w).$$

The quantity $\|e(w^N, \lambda)\|$ is referred to as fixed point residual and has been frequently used in the literature; see, e.g., [4, 5]. Assume that $T^{-1}(0)$ is nonempty, and let $w^* \in T^{-1}(0)$. It was first shown in [3, Proposition 8] for PPA and later extended in [14, Theorem 3.1] to the Douglas–Rachford operator splitting method that

$$\frac{\|e(w^N, \lambda)\|^2}{\|w^0 - w^*\|^2} \leq \frac{1}{N+1} \quad \forall\, N \geq 1,\ n \geq 1. \tag{2.2}$$

The main result of this work, to tighten the bound (2.2), is summarized in the following theorem.

THEOREM 2.1. *Let $T \in \mathcal{M}_n$ be a maximal monotone operator on $\mathbb{R}^n$ such that $T^{-1}(0)$ is nonempty, and let $\lambda > 0$ be a constant. Then, for any integer $N \geq 1$ and $w^* \in T^{-1}(0)$, the sequence $\{w^k\}_{k=0}^\infty$ generated by the PPA (2.1) with an arbitrary starting point $w^0 \notin T^{-1}(0)$ satisfies*

$$\frac{\|e(w^N, \lambda)\|^2}{\|w^0 - w^*\|^2} \leq \begin{cases} \frac{1}{(N+1)^2} & \text{if } n = 1, \\ \frac{1}{(1+\frac{1}{N})^N (N+1)} & \text{if } n \geq 2. \end{cases} \tag{2.3}$$

Furthermore, these bounds are tight.

Theorem 2.1 is discovered based on the idea of performance estimation [8] in the setting of maximal monotone operators [34], which is introduced in section 3. The proof is postponed to section 4.

3. Performance estimation and SDP.

Consider the following class of operators:

$$\mathcal{P}_n = \{T \in \mathcal{M}_n \mid T^{-1}(0) \neq \emptyset\}.$$

Let $T \in \mathcal{P}_n$. Initialized at any given $w^0 \in \mathbb{R}^n$, the first N iterations of PPA generate a unique sequence of points $\{w^1, \dots, w^N\}$. We denote this procedure by

$$\{w^1, \dots, w^N\} = \mathrm{PPA}(T, w^0, \lambda, N).$$

The quantity on the left-hand side of (2.2) is dependent on $(T, w^0, w^*, \lambda, N)$. For convenience, we define

$$\varepsilon(T, w^0, w^*, \lambda, N) := \frac{\|e(w^N, \lambda)\|^2}{\|w^0 - w^*\|^2} = \frac{\|w^N - w^{N+1}\|^2}{\|w^0 - w^*\|^2},$$

where $w^{N+1} = J_{\lambda T}(w^N)$. In the performance estimation framework, the exact or tight worst case convergence rate, which we denote by $\zeta_n(N)$, is formulated as the following problem:

(3.1)
$$\zeta_n(N) := \sup_{T, w^0, w^*, \lambda} \left\{ \varepsilon(T, w^0, w^*, \lambda, N) \;\middle|\; \begin{array}{l} T \in \mathcal{P}_n,\ w^0 \in \mathbb{R}^n,\ 0 \in T(w^*),\ \lambda > 0, \\ \{w^1, \dots, w^{N+1}\} = \mathrm{PPA}(T, w^0, \lambda, N+1) \end{array} \right\}.$$

We note that the supremum value (3.1) is dependent on n, the dimension of the underlying Euclidean space, and this fact is indicated by the subscript n in $\zeta_n(N)$.

3.1. Simplification. In the following, we argue that one can always set $\lambda = 1$, $w^* = 0$, and $w^0 \in \mathbb{R}^n$ such that $\|w^0\| = 1$ in (3.1) without affecting the value of $\zeta_n(N)$.

Fact 1. For any $\lambda > 0$ and $T \in \mathcal{M}_n$, define $T_\lambda = T \circ (\lambda I)$. Then, $T \in \mathcal{P}_n$ if and only if $T_\lambda \in \mathcal{P}_n$, and $u = J_{\lambda T} z$ if and only if $u/\lambda = J_{T_\lambda}(z/\lambda)$.

Based on Fact 1, it is easy to show that the sequence $\{w^1, \dots, w^N, w^{N+1}\}$ generated by $\mathrm{PPA}(T, w^0, \lambda, N+1)$ also satisfies

$$\{w^1/\lambda, \dots, w^N/\lambda, w^{N+1}/\lambda\} = \mathrm{PPA}(T_\lambda, w^0/\lambda, 1, N+1).$$

Furthermore, it is apparent that $0 \in T(w^*)$ if and only if $0 \in T_\lambda(w^*/\lambda)$. As a result, we have

$$\varepsilon(T, w^0, w^*, \lambda, N) = \varepsilon(T_\lambda, w^0/\lambda, w^*/\lambda, 1, N).$$

Since $T \in \mathcal{P}_n$ if and only if $T_\lambda \in \mathcal{P}_n$ and the focus is the worst case bound $\zeta_n(N)$, it is thus without loss of generality to assume $\lambda = 1$.

Fact 2. For any $\gamma > 0$ and $T \in \mathcal{M}_n$, define $T^\gamma = (\frac{1}{\gamma} I) \circ T \circ (\gamma I)$. Then, $T \in \mathcal{P}_n$ if and only if $T^\gamma \in \mathcal{P}_n$, and $u = J_T z$ if and only if $u/\gamma = J_{T^\gamma}(z/\gamma)$.

Based on Fact 2, it is easy to show that the sequence $\{w^1, \dots, w^N, w^{N+1}\}$ generated by $\mathrm{PPA}(T, w^0, 1, N+1)$ also satisfies

$$\{w^1/\gamma, \dots, w^N/\gamma, w^{N+1}/\gamma\} = \mathrm{PPA}(T^\gamma, w^0/\gamma, 1, N+1).$$

Furthermore, $0 \in T(w^*)$ if and only if $0 \in T^\gamma(w^*/\gamma)$, and $T \in \mathcal{P}_n$ if and only if $T^\gamma \in \mathcal{P}_n$. These together imply that, for any $\gamma > 0$, (T, w^0, w^*) (together with $\lambda = 1$) is a solution of (3.1) if and only if $(T^\gamma, w^0/\gamma, w^*/\gamma)$ is a solution. Therefore, it is also without loss of generality to assume $\|w^0 - w^*\| = 1$.

Fact 3. For any $T \in \mathcal{M}_n$ and $u, v \in \mathbb{R}^n$, define $T' = T(\cdot + u) + v$. Then, $T \in \mathcal{P}_n$ if and only if $T' \in \mathcal{P}_n$. Furthermore, $J_{T'} = J_T(\cdot + u - v) - u$.

Based on Fact 3, it is easy to show that $w^* = J_T(w^*)$ if and only if $0 = J_{T'}(0)$, where $T' = T(\cdot + w^*)$. Therefore, it is without loss of generality to assume $w^* = 0$.

In the following, we always assume $\lambda = 1$, $w^* = 0$, and $\|w^0 - w^*\| = \|w^0\| = 1$. As a result, the PEP given in (3.1) can be simplified to

$$\zeta_n(N) = \sup_{T, w^0} \left\{ \|w^N - w^{N+1}\|^2 \;\middle|\; \begin{array}{l} T \in \mathcal{P}_n,\ \|w^0\| = 1,\ 0 \in T(0), \\ \{w^1, \dots, w^{N+1}\} = \mathrm{PPA}(T, w^0, 1, N+1) \end{array} \right\}. \tag{3.2}$$

The PEP given in (3.2) is an infinite dimensional nonconvex optimization problem due to the presence of the constraint $T \in \mathcal{P}_n$. This seemingly makes it intractable. However, this is not the case. In fact, by following the idea of [34] we can reformulate (3.2) as a finite dimensional convex SDP by using the maximal monotone operator interpolation theorem.

3.2. Operator interpolation. Let K be an index set and $\mathcal{Q}$ be a set of operators on $\mathbb{R}^n$. A set $\{(x_j, y_j)\}_{j\in K} \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is said to be $\mathcal{Q}$-interpolable if there exists an operator $T \in \mathcal{Q}$ such that $\{(x_j, y_j)\}_{j\in K} \subseteq \mathrm{graph}(T)$. Recall that $\mathcal{M}_n$ denotes the set of maximally monotone operators on $\mathbb{R}^n$. Based on [34, Fact 1] and the monotone extension theorem [1, Theorem 20.21], the set $\{(x_j, y_j)\}_{j\in K}$ is $\mathcal{M}_n$-interpolable if and only if

$$\langle x_i - x_j,\ y_i - y_j\rangle \geq 0 \quad \forall i, j \in K.$$

Let $S := \{(w^{k+1}, w^k - w^{k+1})\}_{k=0}^N \cup \{(0, 0)\}$. According to (2.1), the constraints

$$T \in \mathcal{P}_n,\ \ 0 \in T(0),\ \ \text{and}\ \ \{w^1, \ldots, w^{N+1}\} = \mathrm{PPA}(T, w^0, 1, N+1) \tag{3.3}$$

in (3.2) can be restated as $S \subseteq \mathrm{graph}(T)$ for some $T \in \mathcal{P}_n$, i.e., the set S is $\mathcal{P}_n$-interpolable. Since $\mathcal{P}_n \subseteq \mathcal{M}_n$, the set S is also $\mathcal{M}_n$-interpolable. As a result, the following set of inequalities are satisfied:

$$\langle w^i - w^j, (w^{i-1} - w^i) - (w^{j-1} - w^j)\rangle \geq 0 \quad \forall\, 1 \leq i < j \leq N+1, \tag{3.4a}$$

$$\langle w^i, w^{i-1} - w^i\rangle \geq 0 \quad \forall\, 1 \leq i \leq N+1. \tag{3.4b}$$

On the other hand, if the inequalities in (3.4) are satisfied, then the set S is also $\mathcal{M}_n$-interpolable [34, Fact 1], [1, Theorem 20.21], i.e., $S \subseteq \mathrm{graph}(T)$ for some $T \in \mathcal{M}_n$. Since $(0, 0) \in S$, it follows that $T \in \mathcal{P}_n$. In summary, the three constraints given in (3.3) can be replaced by the set of inequalities given in (3.4). As a result, (3.2) is equivalent to

$$\zeta_n(N) = \sup_{\{w^0,\ldots,w^{N+1}\}} \left\{ \|w^N - w^{N+1}\|^2 \;\middle|\; \begin{array}{l} \langle w^0, w^0\rangle = 1, \\ \{w^0, \ldots, w^{N+1}\}\ \text{ satisfies }\ (3.4) \end{array} \right\}. \tag{3.5}$$

The equivalence of (3.2) and (3.5) lies in the fact that an optimal solution to either problem corresponds to a feasible solution to the other. Clearly, (3.5) is finite dimensional but still nonconvex. Fortunately, both the objective and the constraints of (3.5) are linear functions of the Grammian matrix $X := (\langle w^i, w^j\rangle)_{i,j=0,1,\ldots,N+1}$. As long as the dimension n is greater than or equal to $N + 2$ (the dimension of the Grammian matrix X), (3.5) is equivalent to a linear SDP, in which case the exact worst case bound $\zeta_n(N)$ defined in (3.2) can be computed numerically (e.g., using standard SDP solvers); see, e.g., [38, Theorem 5], [37, Proposition 2.6], [34, Lemma 1] for similar discussions.

3.3. Grammian representation. In the following, we let $P = (w^0, w^1, \ldots, w^N, w^{N+1}) \in \mathbb{R}^{n\times(N+2)}$ and $X = P^\top P \succeq 0$. Then, $w^{i-1} = Pe_i$ for $i = 1, 2, \ldots, N+2$, where e_i denotes the ith unit vector of length $N + 2$.

For $1 \leq i < j \leq N + 1$, we let $\xi_{ij} := (e_i - e_{i+1}) - (e_j - e_{j+1})$. Then, the set of conditions in (3.4a) can be equivalently stated as

$$\begin{cases} \langle w^i - w^j,\ (w^{i-1} - w^i) - (w^{j-1} - w^j)\rangle = \langle P(e_{i+1} - e_{j+1}),\ P\xi_{ij}\rangle = \langle A_{i,j},\ X\rangle/2 \geq 0, \\ A_{i,j} := (e_{i+1} - e_{j+1})\xi_{ij}^\top + \xi_{ij}(e_{i+1} - e_{j+1})^\top, \quad 1 \leq i < j \leq N+1. \end{cases} \tag{3.6}$$

Note that here $A_{i,j}$ is a symmetric matrix dependent on i and j, instead of the (i,j)th entry of matrix A. Although it is a slight abuse of notation, no confusion occurs. On the other hand, the set of conditions in (3.4b) can be restated as

$$\begin{cases} \langle w^i, w^{i-1} - w^i \rangle = \langle Pe_{i+1}, P(e_i - e_{i+1}) \rangle = \langle B_i, X \rangle / 2 \geq 0, \\ B_i := e_{i+1}(e_i - e_{i+1})^\top + (e_i - e_{i+1}) e_{i+1}^\top, \quad 1 \leq i \leq N+1. \end{cases} \tag{3.7}$$

Furthermore, we let

$$E_{11} := e_1 e_1^\top \quad \text{and} \quad C := (e_{N+1} - e_{N+2})(e_{N+1} - e_{N+2})^\top. \tag{3.8}$$

Then, $\langle w^0, w^0 \rangle = 1$ and $\|w^N - w^{N+1}\|^2$ can be rewritten, respectively, as

$$\langle E_{11}, X \rangle = 1 \quad \text{and} \quad \|w^N - w^{N+1}\|^2 = \langle C, X \rangle. \tag{3.9}$$

3.4. The SDP reformulation. With the discussions in section 3.2 and the matrices defined in section 3.3, the PEP given in (3.5) can be equivalently reformulated as the following SDP with rank constraint:

$$\zeta_n(N) = \max_{X \in \mathbb{S}^{N+2}} \left\{ \langle C, X \rangle \;\middle|\; \begin{array}{ll} \langle A_{i,j}, X \rangle \geq 0, & 1 \leq i < j \leq N+1, \\ \langle B_i, X \rangle \geq 0, & 1 \leq i \leq N+1, \\ \langle E_{11}, X \rangle = 1, & X \succeq 0, \ \mathrm{rank}(X) \leq n \end{array} \right\}, \tag{3.10}$$

where the matrices $\{A_{i,j} : 1 \leq i < j \leq N+1\}$, $\{B_i : 1 \leq i \leq N+1\}$, E_{11}, and C are as defined in section 3.3. For large problems with $n \geq N+2$, the rank constraint $\mathrm{rank}(X) \leq n$ is redundant since $X \in \mathbb{S}^{N+2}$ and can be safely dropped, resulting in a standard linear SDP

$$\max_{X \in \mathbb{S}^{N+2}} \left\{ \langle C, X \rangle \;\middle|\; \begin{array}{ll} \langle A_{i,j}, X \rangle \geq 0, & 1 \leq i < j \leq N+1, \\ \langle B_i, X \rangle \geq 0, & 1 \leq i \leq N+1, \\ \langle E_{11}, X \rangle = 1, & X \succeq 0 \end{array} \right\}. \tag{3.11}$$

For small problems with $n < N+2$, the SDP problem (3.11) might be a relaxation of (3.10) since the rank constraint may be violated, in which case the optimum value of (3.11) is only an upper bound of $\zeta_n(N)$. In any case, the SDP problem (3.11) is a reformulation but not a relaxation whenever $n \geq \mathrm{rank}(X^*)$, where X^* is any optimal solution of (3.11).

SeDuMi [35] is used to solve the above SDP for $N = 1, 2, \ldots, 100$. The tight worst case iteration bound just computed is compared in Figure 1 to the existing bound [3, Proposition], [14, Theorem 3.1], i.e., the one given in (2.2).

4. Proof of Theorem 2.1. In this section, we prove Theorem 2.1 for $n \geq 2$ by establishing an upper bound on $\zeta_n(N)$ and providing a two dimensional example which matches the upper bound, implying its tightness. The one dimensional case is discussed separately in subsection 4.3.

4.1. An upper bound on $\zeta_n(N)$. Let $N \geq 1$ be any integer, and let $\{A_{i,j} : 1 \leq i < j \leq N+1\}$, $\{B_i : 1 \leq i \leq N+1\}$, and E_{11} and C be defined, respectively, in (3.6), (3.7), and (3.8). For convenience, we define

$$M := \eta E_{11} - C - \sum_{1 \leq i < j \leq N+1} \lambda_{i,j} A_{i,j} - \sum\nolimits_{i=1}^{N+1} \mu_i B_i, \tag{4.1}$$

where $\{\lambda_{i,j} \in \mathbb{R} : 1 \leq i < j \leq N+1\}$, $\{\mu_i \in \mathbb{R} : 1 \leq i \leq N+1\}$, and $\eta \in \mathbb{R}$ are parameters. We summarize the upper bound on $\zeta_n(N)$ in the following proposition.

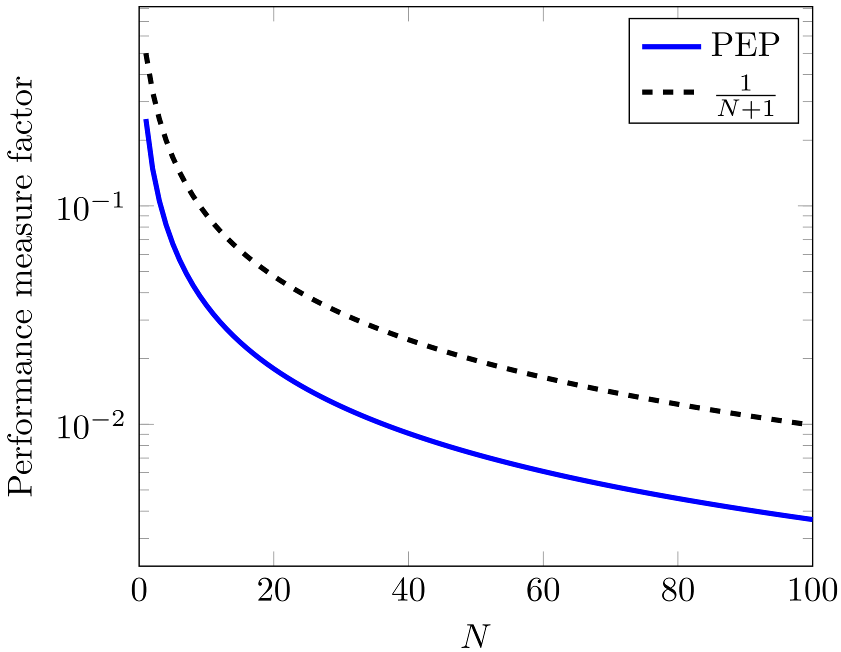

FIG. 1. *Comparison between the tight worst case bound computed by PEP (the lower curve) and the existing results in* [3, 14] *(the dashed curve).*

PROPOSITION 4.1 (upper bound). *Let $N \geq 1$ be any integer. Then, it holds that*

$$\zeta_n(N) \leq \eta = \frac{1}{(1+\frac{1}{N})^N (N+1)} \quad \forall\, N \geq 1,\ n \geq 1. \tag{4.2}$$

Proof. First, we define

$$\lambda_{i,j} := \begin{cases} \frac{N^{N-i}}{(N+1)^{N-i}} \frac{i}{N+1}, & j-1 = i = 1,2,\ldots,N, \\ 0 & \text{otherwise}, \end{cases} \tag{4.3a}$$

$$\mu_i := \begin{cases} \frac{N^{N-i}}{(N+1)^{N-i}} \frac{N-i}{(N+1)^2}, & i = 1,2,\ldots,N, \\ \frac{1}{N+1}, & i = N+1, \end{cases} \tag{4.3b}$$

$$\eta := \frac{N^N}{(N+1)^N} \frac{1}{N+1}. \tag{4.3c}$$

Since the $\lambda_{i,j}$'s and μ_i's are all nonnegative, it follows from (3.4) and $\|w^0\| = 1$ that

$$\begin{aligned} \|w^N - w^{N+1}\|^2 \leq \|w^N - w^{N+1}\|^2 + 2\sum_{i=1}^{N+1} \mu_i \langle w^i, w^{i-1} - w^i\rangle - \eta\big(\|w^0\|^2 - 1\big) \\ + 2\sum_{1\leq i<j\leq N+1} \lambda_{i,j} \langle w^i - w^j,\ (w^{i-1} - w^i) - (w^{j-1} - w^j)\rangle. \end{aligned} \tag{4.4}$$

Here, the scalars 2 aim to cancel out the factors 1/2 in (3.6) and (3.7) so that the right-hand side of (4.4) matches the definition of M in (4.1). In fact, it follows from

(3.6)–(3.9) and (4.1) that the right-hand side of (4.4) can be rewritten as $\eta - \langle M, X\rangle$, where $X = (\langle w^i, w^j\rangle)_{i,j=0,1,\ldots,N+1}$, and thus (4.4) becomes

$$\|w^N - w^{N+1}\|^2 \le \eta - \langle M, X\rangle.$$

It can be shown that

$$\langle M, X\rangle = \sum_{i=1}^{N} \frac{N^{N-i}}{(N+1)^{N-i}} \frac{i}{N} \left\| \frac{N}{N+1} w^{i-1} - \frac{2N}{N+1} w^i + w^{i+1} \right\|^2. \tag{4.5}$$

Consequently, we have $\|w^N - w^{N+1}\|^2 \le \eta$. ∎

We mention that the sum-of-squares reformulation of $\langle M, X\rangle$ in (4.5) can be motivated by Cholesky decomposition of M. A similar approach for establishing tight convergence rates has been adopted in, e.g., [6, 34, 36, 39]. In fact, the coefficients (4.3) are an optimal solution to the dual SDP, and by checking the dual feasibility, Proposition 4.1 follows from weak duality. This alternative proof for the upper bound is included in the appendix.

The upper bound (4.2) holds for all $n \ge 1$, and it is tight when $n \ge 2$ since an example matching the upper bound exists in $\mathbb{R}^2$.

4.2. A two dimensional example. According to our experiments, the solutions of (3.11) with different values of N seem always to have rank 2. This suggests that a worst case example may exist in $\mathbb{R}^2$. Indeed, we managed to construct such an example, which provides a lower bound on $\zeta_n(N)$ in the case $n \ge 2$.

Let $\theta \in (0, \pi/2)$, and define a rotation matrix in $\mathbb{R}^2$ as

$$\Theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{4.6}$$

It is easy to check that $\Theta^\top = \Theta^{-1}$ and for any integer k,

$$\Theta^k + \Theta^{-k} = \begin{pmatrix} \cos(k\theta) & -\sin(k\theta) \\ \sin(k\theta) & \cos(k\theta) \end{pmatrix} + \begin{pmatrix} \cos(k\theta) & \sin(k\theta) \\ -\sin(k\theta) & \cos(k\theta) \end{pmatrix} = 2\cos(k\theta) I. \tag{4.7}$$

To simplify the notation, we denote $\beta = \cos\theta$. The example we construct is summarized below. Its geometry for $N = 5$ and $N = 100$ is shown in Figure 2.

Example 4.2. *Given an integer $N \ge 1$, define $w^0 := (1, 0)^\top$ and $w^* := u^* := (0, 0)^\top$. For $k = 1, 2, \ldots, N+1$, define*

$$w^k := \beta\Theta w^{k-1} = \beta^k \Theta^k w^0 \quad \textit{and} \quad u^k := w^{k-1} - w^k.$$

Proposition 4.3. *Let $\theta \in (0, \pi/2)$ and Θ be as given in* (4.6)*. Then $\{(w^k, u^k) : k = 1, \ldots, N+1\}$ and (w^*, u^*) defined in Example* 4.2 *satisfy*

$$\langle w^j - w^i, u^j - u^i\rangle = 0 \quad \forall\, 1 \le i < j \le N+1, \tag{4.8a}$$

$$\langle w^i - w^*, u^i - u^*\rangle = 0 \quad \forall\, 1 \le i \le N+1. \tag{4.8b}$$

Proof. First, for any $1 \le i \le N+1$, we have

$$\begin{aligned}\langle w^i - w^*, u^i - u^*\rangle &= \langle w^i, w^{i-1} - w^i\rangle \\ &= \langle \beta^i \Theta^i w^0, \beta^{i-1}\Theta^{i-1}(I - \beta\Theta) w^0\rangle\end{aligned}$$

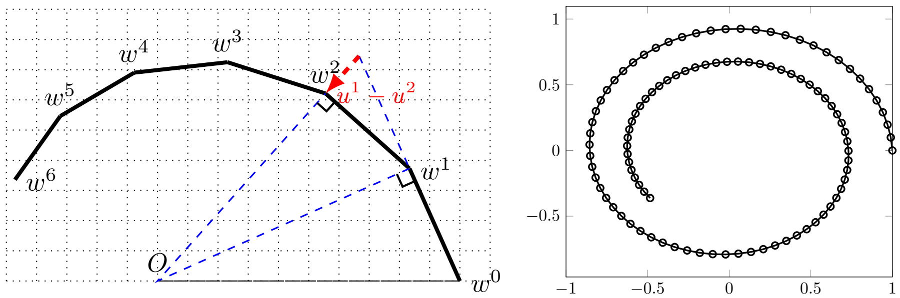

FIG. 2. *Geometry of the constructed example. Left:* $N = 5$. *Right:* $N = 100$.

$$
\begin{aligned}
&= \beta^{2i-1}\langle \Theta w^0, (I - \beta\Theta) w^0\rangle \\
&= \beta^{2i-1}\left(\langle \Theta w^0, w^0\rangle - \beta\langle \Theta w^0, \Theta w^0\rangle\right) \\
&= \beta^{2i-1}\left(\langle \Theta w^0, w^0\rangle - \beta\right) \\
&= 0,
\end{aligned}
$$

where in the third "=" we used the fact that $\Theta^\top = \Theta^{-1}$, while the last "=" follows from $\langle \Theta w^0, w^0\rangle = \|\Theta w^0\|\|w^0\|\cos\theta = \beta$. Then, for $1 \le i < j \le N+1$, we have

$$
\begin{aligned}
&\langle w^j - w^i, u^j - u^i\rangle \\
&= \langle \beta^{j-i}\Theta^{j-i} w^i - w^i, \beta^{j-i}\Theta^{j-i} u^i - u^i\rangle \\
&= \langle \beta^{j-i}\Theta^{j-i} w^i, \beta^{j-i}\Theta^{j-i} u^i\rangle - \langle \beta^{j-i}\Theta^{j-i} w^i, u^i\rangle - \langle w^i, \beta^{j-i}\Theta^{j-i} u^i\rangle + \langle w^i, u^i\rangle \\
&= \beta^{2(j-i)}\langle w^i, u^i\rangle - \beta^{j-i}\langle (\Theta^{j-i} + \Theta^{i-j}) w^i, u^i\rangle + \langle w^i, u^i\rangle \\
&= (\beta^{2(j-i)} + 1)\langle w^i, u^i\rangle - \beta^{j-i}\langle (\Theta^{j-i} + \Theta^{i-j}) w^i, u^i\rangle \\
&= (\beta^{2(j-i)} + 1)\langle w^i, u^i\rangle - 2\beta^{j-i}\cos((j-i)\theta)\langle w^i, u^i\rangle \\
&= 0,
\end{aligned}
$$

where the second-to-last "=" follows from (4.7) and the last "=" is because $\langle w^i, u^i\rangle = 0$ for $1 \le i \le N+1$. ∎

Now, we are ready to establish a lower bound on $\zeta_n(N)$, which is summarized below.

PROPOSITION 4.4 (lower bound). *Let $N \ge 1$ be any integer, $n \ge 2$, and let $\zeta_n(N)$ be as defined in, e.g.,* (3.1), (3.2), *or* (3.5). *Then, it holds that*

$$
\zeta_n(N) \ge \frac{1}{(1+\frac{1}{N})^N (N+1)}. \tag{4.9}
$$

Proof. It follows from (4.8), [34, Fact 1], and [1, Theorem 20.21] that there exists a maximally monotone operator $T \in \mathcal{M}_n$ such that the set $\{(w^k, u^k) : k = 1, 2, \ldots, N+1\} \cup \{(w^*, u^*)\}$ defined in Example 4.2 satisfies

$$
\{(w^k, u^k) : k = 1, 2, \ldots, N+1\} \cup \{(w^*, u^*)\} \subseteq \operatorname{graph}(T),
$$

i.e., $\{(w^k, u^k) : k = 1, 2, \ldots, N+1\} \cup \{(w^*, u^*)\}$ is $\mathcal{M}_n$-interpolable. Since $u^k = w^{k-1} - w^k \in T(w^k)$ for $k = 1, 2, \ldots, N+1$ and T is maximally monotone, this implies that $\{w^k : k = 1, 2, \ldots, N+1\}$ obeys the PPA scheme $w^{k+1} = J_T(w^k)$ for $k = 0, \ldots, N$.

In hindsight, we found that T admits an explicit formula given by

$$T(w) = \tan(\theta) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} w, \quad w \in \mathbb{R}^2. \tag{4.10}$$

Clearly, T is continuous and satisfies $\langle Tw, w\rangle = 0$ for any $w \in \mathbb{R}^2$, which is sufficient to guarantee that T is maximally monotone. Furthermore, it is easy to show that $J_T = \beta\Theta$. Then, for any $k \geq 0$, it is elementary to verify that

$$\begin{aligned} \|w^k - w^{k+1}\|^2 &= \|\beta^k \Theta^k w^0 - \beta^{k+1}\Theta^{k+1} w^0\|^2 \\ &= \beta^{2k}\langle w^0 - \beta\Theta w^0, w^0 - \beta\Theta w^0\rangle \\ &= \beta^{2k}(1 - 2\beta\langle w^0, \Theta w^0\rangle + \beta^2\langle \Theta w^0, \Theta w^0\rangle) \\ &= \beta^{2k}(1-\beta^2). \end{aligned} \tag{4.11}$$

Therefore, it holds that $\|w^N - w^{N+1}\|^2 = \beta^{2N}(1-\beta^2)$. By maximizing over β, we obtain that $\beta = \cos\theta = \sqrt{\frac{N}{N+1}}$ and thus

$$\|w^N - w^{N+1}\|^2 = \frac{1}{(1+\frac{1}{N})^N (N+1)}.$$

Thus, we have shown the lower bound (4.9) by constructing an example. Since the constructed example exists in $\mathbb{R}^2$, (4.9) is valid only for $n \geq 2$. ∎

By combining Propositions 4.1 and 4.4, we obtain

$$\zeta_n(N) = \frac{1}{(1+\frac{1}{N})^N (N+1)} \quad \forall\, N \geq 1,\ n \geq 2,$$

which, by further considering Facts 1–3 in section 3.1, completes the proof of Theorem 2.1 in the case $n \geq 2$.

4.3. Proof of Theorem 2.1 for $n = 1$.

Proof. Let w^∞ be the Euclidean projection of w^0 onto $T^{-1}(0)$. Since T is maximal monotone, it follows from [1, Props. 20.22 and 20.31] that $T^{-1}(0)$ is closed and convex. Furthermore, by assumption $T^{-1}(0)$ is nonempty, and thus w^∞ is well defined.

We first show that the sequence $\{w^k\}_{k=0}^\infty$ generated by PPA is monotonically convergent to w^∞. In fact, the convergence of $\{w^k\}_{k=0}^\infty$ to an element of $T^{-1}(0)$ is well known, and we only need to clarify the monotonicity of $\{w^k\}_{k=0}^\infty$ and $\lim_{k\to\infty} w^k = w^\infty$. It follows from the firm nonexpansiveness of $J_{\lambda T}$ (see, e.g., [1]) that, for any $w^* \in T^{-1}(0)$, we have

$$|w^{k+1} - w^*|^2 \leq |w^k - w^*|^2 - |w^{k+1} - w^k|^2 \quad \forall\, k \geq 0.$$

This implies that $\{|w^k - w^*|\}_{k=0}^\infty$ is monotonically nonincreasing and $|w^{k+1} - w^k| \leq |w^k - w^*|$. By letting $w^* = w^\infty$ in case $T^{-1}(0)$ is not a singleton, we obtain $|w^{k+1} - w^\infty| \leq |w^k - w^\infty|$ and $|w^{k+1} - w^k| \leq |w^k - w^\infty|$, the right-hand side of which represents the distance of w^k to $T^{-1}(0)$. Then, the monotonicity of $\{w^k\}_{k=0}^\infty$ and $\lim_{k\to\infty} w^k = w^\infty$ follow simultaneously from $|w^{k+1} - w^\infty| \leq |w^k - w^\infty|$ and

$$w^k - |w^k - w^\infty| \leq w^{k+1} \leq w^k + |w^k - w^\infty| \quad \forall\, k \geq 0.$$

Since $\{w^k\}_{k=0}^{\infty}$ is monotone, the $(w^k - w^{k+1})$'s have the same sign from which we obtain

$$\sum\nolimits_{k=0}^{\infty} |w^k - w^{k+1}| = \left|\sum\nolimits_{k=0}^{\infty} (w^k - w^{k+1})\right| = |w^0 - w^\infty|. \tag{4.12}$$

Furthermore, $\{|w^k - w^{k+1}|\}_{k=0}^{\infty}$ is monotonically nonincreasing (see [14, Lemma 2.4]). It then follows from (4.12) that

$$|e(w^N, \lambda)| = |w^N - w^{N+1}| \le \frac{|w^0 - w^\infty|}{N+1} \quad \forall\, N \ge 1. \tag{4.13}$$

Then, by the definition of w^∞, (2.3) holds for any $w^* \in T^{-1}(0)$ (for $n = 1$).

The tightness can be illustrated by an example. Let $w^0 = 1$, and let T be the subdifferential operator of $f(w) = \frac{1}{N+1}|w|$, $w \in \mathbb{R}$. Then it is easy to verify that $w^k = J_T(w^{k-1}) = 1 - k/(N+1)$, $k = 1, \dots, N+1$, and thus equality holds in (4.13) with $w^\infty = 0$ and $\lambda = 1$. For $\lambda \ne 1$, it suffices to let T be the subdifferential operator of $g(w) = \frac{1}{\lambda(N+1)}|w|$, $w \in \mathbb{R}$. Therefore, the upper bound (4.13) is tight for any $\lambda > 0$, and so is (2.3) for $n = 1$. ∎

The tight sublinear rate $\frac{1}{(N+1)^2}$ for $n = 1$ is attained by the subdifferential operator of a convex function and coincides with [37, Conjecture 4.2], where PPA for convex optimization is considered. This is not by accident. In fact, it follows from [31, Theorems 24.3 and 24.9] that, when $n = 1$, monotone mappings and cyclically monotone mappings are the same. As a result, [31, Theorem 24.8] implies that, when $n = 1$, maximal monotone operators and subdifferential operators of closed convex functions coincide. This is not true when $n \ge 2$.

5. Concluding remarks. PPA is a fundamental algorithmic framework in convex optimization and monotone operator theory. Aided by the performance estimation framework [8, 34, 38, 37, 6, 39] and SDP, we have established the following best dimension-independent bound:

$$\|w^N - w^{N+1}\|^2 \le \frac{\|w^0 - w^*\|^2}{(1 + \frac{1}{N})^N (N+1)} \quad \forall\, N \ge 1,\ n \ge 1,\ w^* \in T^{-1}(0). \tag{5.1}$$

An example in $\mathbb{R}^2$ is constructed, which matches the upper bound, implying the tightness of (5.1) in the case $n \ge 2$. The improvement over the known bound (2.2) is approximately a constant factor of $\exp(-1)$. For $n = 1$, $|w^N - w^{N+1}|^2$ converges one order of magnitude faster, namely, at the rate of $1/(N+1)^2$. In both cases, the sublinear rates are nonimprovable without imposing regularity conditions.

Appendix A. An alternative proof of Proposition 4.1. This appendix is devoted to an alternative proof of Proposition 4.1. Let M be defined by (4.1). The dual problem of (3.11) is given by

$$\min_{\{\lambda_{i,j}\},\{\mu_i\},\eta} \left\{ \eta \;\middle|\; \begin{array}{lll} M \succeq 0, & \lambda_{i,j} \ge 0, & 1 \le i < j \le N+1, \\ \mu_i \ge 0, & i = 1, 2, \dots, N+1, & \eta \in \mathbb{R} \end{array} \right\}. \tag{A.1}$$

Note that M is dependent on the Lagrange multipliers $\lambda_{i,j}$, μ_i, and η. The approach is to show that the coefficients defined in (4.3) are feasible for the dual problem of (3.11), and thus, by following weak duality, the upper bound (4.2) is established. Further considering the two dimensional matching example given in section 4.2, it is then clear that the coefficients defined in (4.3) are in fact an optimal solution of the dual SDP problem (A.1).

Clearly, all the quantities in (4.3a), (4.3b) are nonnegative. It remains to show that the choices of the Lagrange multipliers specified in (4.3) are such that the matrix M is positive semidefinite.

From the definitions of $\{A_{i,j} : 1 \le i < j \le N+1\}$, $\{B_i : 1 \le i \le N+1\}$, and C in (3.6), (3.7), and (3.8), respectively, we have

$$A_{i,i+1}(i : i+2, i : i+2) = \begin{pmatrix} 0 & 1 & -1 \\ 1 & -4 & 3 \\ -1 & 3 & -2 \end{pmatrix}, \quad i = 1, 2, \ldots, N, \tag{A.2a}$$

$$B_i(i : i+1, i : i+1) = \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix}, \quad i = 1, 2, \ldots, N+1, \tag{A.2b}$$

$$C(N+1 : N+2, N+1 : N+2) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{A.2c}$$

It is then not difficult to verify from (4.1), (4.3), and (A.2) that M is a pentadiagonal matrix whose nonzero entries are given by

(A.3)

$$\begin{cases} M_{11} = \eta = \frac{N^N}{(N+1)^N} \frac{1}{N+1}, \\ M_{ii} = \begin{cases} 4\lambda_1 + 2\mu_1, & i = 2, \\ 2\lambda_{i-2} + 4\lambda_{i-1} + 2\mu_{i-1}, & i = 3, \ldots, N, \end{cases} \\ \qquad = \frac{N^{N+1-i}}{(N+1)^{N+1-i}} \frac{(i-1)(6N+2)}{(N+1)^2}, \quad i = 2, 3, \ldots, N, \\ M_{N+1,N+1} = 2\lambda_{N-1} + 4\lambda_N + 2\mu_N - 1 = \frac{5N^2-1}{(N+1)^2}, \\ M_{N+2,N+2} = 2\lambda_N + 2\mu_{N+1} - 1 = 1, \\ M_{1,2} = M_{2,1} = -\lambda_1 - \mu_1 = -\frac{N^N}{(N+1)^N} \frac{2}{N+1}, \\ M_{i,i+1} = M_{i+1,i} = -3\lambda_{i-1} - \lambda_i - \mu_i = -\frac{N^{N+1-i}}{(N+1)^{N+1-i}} \frac{4i-2}{N+1}, \quad i = 2, 3, \ldots, N, \\ M_{N+1,N+2} = M_{N+2,N+1} = -3\lambda_N - \mu_{N+1} + 1 = -\frac{2N}{N+1}, \\ M_{i,i+2} = M_{i+2,i} = \lambda_i = \frac{N^{N-i}}{(N+1)^{N-i}} \frac{i}{N+1}, \quad i = 1, 2, \ldots, N. \end{cases}$$

If $N = 1$, then

$$M = \begin{pmatrix} 1/4 & -1/2 & 1/2 \\ -1/2 & 1 & -1 \\ 1/2 & -1 & 1 \end{pmatrix},$$

which is positive semidefinite. In the following, we assume that $N > 1$. Our proof of $M \succeq 0$ heavily relies on the following Schur complement lemma.

Lemma A.1 (Schur complement [13]). *Let $k > 0$ be an integer, $H \in \mathbb{S}^k$ be a real symmetric matrix, $b \in \mathbb{R}^k$ be a column vector, and $c > 0$ be a scalar. Then*

$$\begin{pmatrix} H & b \\ b^\top & c \end{pmatrix} \succeq 0 \quad \textit{if and only if} \quad H - \frac{bb^\top}{c} \succeq 0.$$

The Schur complement lemma has several variants and generalizations, and Lemma A.1 is only one special case that we will use repeatedly. For this purpose, we fix our notation first. Applying the Schur complement reduction as specified in Lemma A.1

to $M \in \mathbb{S}^{N+2}$ recursively for j times, we obtain a matrix of order $N+2-j$, which we denote by $M^{[N+2-j]} \in \mathbb{S}^{N+2-j}$, $j = 0, 1, \ldots, N-1$. Note that here the superscript k in $M^{[k]}$ corresponds to the order of $M^{[k]}$. Since M is a pentadiagonal matrix, applying Schur complement reduction as specified in Lemma A.1 to $M^{[k]}$ once is equivalent to deleting the kth row and the kth column of the matrix while simultaneously updating $M^{[k]}(k-2:k-1, k-2:k-1)$ appropriately. Specifically, $M^{[k]}$ and $M^{[k-1]}$ appear as

(A.4)

$$M^{[k]} = \begin{pmatrix} * & * & * & & & \\ * & \ddots & \ddots & \ddots & & \\ * & \ddots & * & * & * & \\ & \ddots & * & \times & \times & \alpha \\ & & * & \times & \times & \beta \\ & & & \alpha & \beta & \gamma \end{pmatrix}_{k\times k}, \quad M^{[k-1]} = \begin{pmatrix} * & * & * & & \\ * & \ddots & \ddots & \ddots & \\ * & \ddots & * & * & * \\ & \ddots & * & \triangle & \triangle \\ & & * & \triangle & \triangle \end{pmatrix}_{(k-1)\times(k-1)},$$

where $\alpha = M^{[k]}_{k-2,k}$, $\beta = M^{[k]}_{k-1,k}$, $\gamma = M^{[k]}_{k,k}$, and

$$M^{[k-1]}(k-2:k-1, k-2:k-1) = \begin{pmatrix} M^{[k]}_{k-2,k-2} - \frac{\alpha^2}{\gamma} & M^{[k]}_{k-2,k-1} - \frac{\alpha\beta}{\gamma} \\ M^{[k]}_{k-1,k-2} - \frac{\alpha\beta}{\gamma} & M^{[k]}_{k-1,k-1} - \frac{\beta^2}{\gamma} \end{pmatrix}.$$

Note that all the $*$'s in both $M^{[k]}$ and $M^{[k-1]}$ are never touched and remain the same as those corresponding elements in the original matrix M.

For $j = 0$, we have

$$M^{[N+2-j]} = M^{[N+2]} = M = \begin{pmatrix} M(1:N+1, 1:N+1) & M(1:N+1, N+2) \\ M(1:N+1, N+2)^\top & M_{N+2,N+2} \end{pmatrix}.$$

Since $M_{N+2,N+2} = 1 > 0$, according to Lemma A.1, $M^{[N+2]} \succeq 0$ if and only if

$$M^{[N+1]} := M(1:N+1, 1:N+1) - \frac{M(1:N+1, N+2)M(1:N+1, N+2)^\top}{M_{N+2,N+2}} \succeq 0.$$

Direct computation indicates that

(A.5)

$$\begin{aligned} & M^{[N+1]}(:,\ N-1:N+1) = M^{[N+1]}(N-1:N+1,\ :)^\top \\ = & \begin{pmatrix} & O_{N-4,3} & \\ M_{N-3,N-1} & 0 & 0 \\ M_{N-2,N-1} & M_{N-2,N} & 0 \\ M_{N-1,N-1} & M_{N-1,N} & M_{N-1,N+1} \\ M_{N,N-1} & M_{N,N} - \frac{M^2_{N,N+2}}{M_{N+2,N+2}} & M_{N,N+1} - \frac{M_{N,N+2}M_{N+1,N+2}}{M_{N+2,N+2}} \\ M_{N+1,N-1} & M_{N,N+1} - \frac{M_{N,N+2}M_{N+1,N+2}}{M_{N+2,N+2}} & M_{N+1,N+1} - \frac{M^2_{N+1,N+2}}{M_{N+2,N+2}} \end{pmatrix} \end{aligned}$$

$$
= \begin{pmatrix} & O_{N-4,3} & \\ \frac{N^3}{(N+1)^3}\frac{N-3}{N+1} & 0 & 0 \\ -\frac{N^3}{(N+1)^3}\frac{(4N-10)}{N+1} & \frac{N^2}{(N+1)^2}\frac{N-2}{N+1} & 0 \\ \frac{N^2}{(N+1)^2}\frac{(N-2)(6N+2)}{(N+1)^2} & -\frac{N^2}{(N+1)^2}\frac{(4N-6)}{N+1} & \frac{N}{N+1}\frac{N-1}{N+1} \\ -\frac{N^2}{(N+1)^2}\frac{(4N-6)}{N+1} & \frac{N}{N+1}\frac{5N^2-5N-2}{(N+1)^2} & -\frac{N(2N-2)}{(N+1)^2} \\ \frac{N}{N+1}\frac{N-1}{N+1} & -\frac{N(2N-2)}{(N+1)^2} & \frac{N-1}{N+1} \end{pmatrix}.
$$

It follows from $N > 1$ that $M^{[N+1]}_{N+1,N+1} = \frac{N-1}{N+1} > 0$. Thus, $M^{[N+1]} \succeq 0$ if and only if $M^{[N]} \succeq 0$. In the following, we apply Schur complement reduction as specified in Lemma A.1 to $M^{[k]}$ and show inductively for $k = N+1, N, \ldots, 5$ that $M^{[k]}_{k,k} > 0$ and

(A.6)

$$
\begin{aligned}
&M^{[k]}(:,\ k-2:k) = M^{[k]}(k-2:k,\ :)^\top \\
&= \begin{pmatrix} & O_{k-5,3} & \\ \frac{N^{N+4-k}}{(N+1)^{N+4-k}}\frac{k-4}{N+1} & 0 & 0 \\ -\frac{N^{N+4-k}}{(N+1)^{N+4-k}}\frac{(4k-14)}{N+1} & \frac{N^{N+3-k}}{(N+1)^{N+4-k}}\frac{k-3}{N+1} & 0 \\ \frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{(k-3)(6N+2)}{(N+1)^2} & -\frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{(4k-10)}{N+1} & \frac{N^{N+2-k}}{(N+1)^{N+2-k}}\frac{k-2}{N+1} \\ -\frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{(4k-10)}{N+1} & \frac{N^{N+2-k}}{(N+1)^{N+2-k}}\frac{(5k-11)N+(k-3)}{(N+1)^2} & -\frac{N^{N+2-k}}{(N+1)^{N+2-k}}\frac{2k-4}{N+1} \\ \frac{N^{N+2-k}}{(N+1)^{N+2-k}}\frac{k-2}{N+1} & -\frac{N^{N+2-k}}{(N+1)^{N+2-k}}\frac{2k-4}{N+1} & \frac{N^{N+1-k}}{(N+1)^{N+1-k}}\frac{k-2}{N+1} \end{pmatrix}.
\end{aligned}
$$

It is already noted that $M^{[N+1]}_{N+1,N+1} > 0$. By comparing with (A.5), we see that (A.6) holds for $k = N+1$. Assume that (A.6) holds for $5 < k \le N+1$. We will show that (A.6) also holds for $k-1$. Since $k > 5$, we have $M^{[k]}_{k,k} = \frac{N^{N+1-k}}{(N+1)^{N+1-k}}\frac{k-2}{N+1} > 0$. Furthermore, direct calculations show that

$$
\begin{aligned}
&M^{[k-1]}(:,\ k-3:k-1) = M^{[k-1]}(k-3:k-1,:)^\top \\
&= \begin{pmatrix} & O_{k-6,3} & \\ M_{k-5,k-3} & 0 & 0 \\ M_{k-4,k-3} & M_{k-4,k-2} & 0 \\ M_{k-3,k-3} & M_{k-3,k-2} & M_{k-3,k-1} \\ M_{k-2,k-3} & M_{k-2,k-2} - \frac{M^2_{k-2,k}}{M^{[k]}_{k,k}} & M_{k-2,k-1} - \frac{M_{k-2,k}M^{[k]}_{k-1,k}}{M^{[k]}_{k,k}} \\ M_{k-1,k-3} & M_{k-1,k-2} - \frac{M_{k-2,k}M^{[k]}_{k-1,k}}{M^{[k]}_{k,k}} & M^{[k]}_{k-1,k-1} - \frac{(M^{[k]}_{k-1,k})^2}{M^{[k]}_{k,k}} \end{pmatrix} \\
&= \begin{pmatrix} & O_{k-6,3} & \\ \frac{N^{N+5-k}}{(N+1)^{N+5-k}}\frac{k-5}{N+1} & 0 & 0 \\ -\frac{N^{N+5-k}}{(N+1)^{N+5-k}}\frac{(4k-18)}{N+1} & \frac{N^{N+4-k}}{(N+1)^{N+4-k}}\frac{k-4}{N+1} & 0 \\ \frac{N^{N+4-k}}{(N+1)^{N+4-k}}\frac{(k-4)(6N+2)}{(N+1)^2} & -\frac{N^{N+4-k}}{(N+1)^{N+4-k}}\frac{(4k-14)}{N+1} & \frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{k-3}{N+1} \\ -\frac{N^{N+4-k}}{(N+1)^{N+4-k}}\frac{(4k-14)}{N+1} & \frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{(5k-16)N+(k-4)}{(N+1)^2} & -\frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{2k-6}{N+1} \\ \frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{k-3}{N+1} & -\frac{N^{N+3-k}}{(N+1)^{N+3-k}}\frac{2k-6}{N+1} & \frac{N^{N+2-k}}{(N+1)^{N+2-k}}\frac{k-3}{N+1} \end{pmatrix}.
\end{aligned}
$$

That is, (A.6) also holds for $k-1$. In summary, (A.6) holds for all $k = N+1, N, \ldots, 5$, and $M \succeq 0$ if and only if $M^{[5]} \succeq 0$. Since $M_{55}^{[5]} = \frac{N^{N-4}}{(N+1)^{N-4}} \frac{3}{N+1} > 0$, $M^{[5]} \succeq 0$ if and only if

$$M^{[4]} = \begin{pmatrix} \frac{N^N}{(N+1)^N}\frac{1}{N+1} & -\frac{N^N}{(N+1)^N}\frac{2}{N+1} & \frac{N^{N-1}}{(N+1)^{N-1}}\frac{1}{N+1} & 0 \\ -\frac{N^N}{(N+1)^N}\frac{2}{N+1} & \frac{N^{N-1}}{(N+1)^{N-1}}\frac{(6N+2)}{(N+1)^2} & -\frac{N^{N-1}}{(N+1)^{N-1}}\frac{6}{N+1} & \frac{N^{N-2}}{(N+1)^{N-2}}\frac{2}{N+1} \\ \frac{N^{N-1}}{(N+1)^{N-1}}\frac{1}{N+1} & -\frac{N^{N-1}}{(N+1)^{N-1}}\frac{6}{N+1} & \frac{N^{N-2}}{(N+1)^{N-2}}\frac{9N+1}{(N+1)^2} & -\frac{N^{N-2}}{(N+1)^{N-2}}\frac{4}{N+1} \\ 0 & \frac{N^{N-2}}{(N+1)^{N-2}}\frac{2}{N+1} & -\frac{N^{N-2}}{(N+1)^{N-2}}\frac{4}{N+1} & \frac{N^{N-3}}{(N+1)^{N-3}}\frac{2}{N+1} \end{pmatrix} \succeq 0.$$

Since $M_{4,4}^{[4]} > 0$, $M^{[4]} \succeq 0$ if and only if $M^{[3]} \succeq 0$. One more step computation shows that

$$M^{[3]} = \begin{pmatrix} \frac{N^N}{(N+1)^N}\frac{1}{N+1} & -\frac{N^N}{(N+1)^N}\frac{2}{N+1} & \frac{N^{N-1}}{(N+1)^{N-1}}\frac{1}{N+1} \\ -\frac{N^N}{(N+1)^N}\frac{2}{N+1} & \frac{N^{N-1}}{(N+1)^{N-1}}\frac{4N}{(N+1)^2} & -\frac{N^{N-1}}{(N+1)^{N-1}}\frac{2}{N+1} \\ \frac{N^{N-1}}{(N+1)^{N-1}}\frac{1}{N+1} & -\frac{N^{N-1}}{(N+1)^{N-1}}\frac{2}{N+1} & \frac{N^{N-2}}{(N+1)^{N-2}}\frac{1}{N+1} \end{pmatrix}.$$

Since $M_{3,3}^{[3]} > 0$, $M^{[3]} \succeq 0$ if and only if $M^{[2]} \succeq 0$. One further step computation shows that $M^{[2]} = O_{2,2}$, the 2×2 zero matrix, which is positive semidefinite. In summary, we have shown that, for any integer $N \geq 1$, M is positive semidefinite by recursively using the Schur complement reduction procedure in Lemma A.1.

Acknowledgments. We sincerely appreciate the useful comments of the associate editor and three anonymous referees, which were very helpful in improving the exactness and readability of the paper. We also thank Dr. Felix Lieder of Heinrich Heine Universität Düsseldorf, who, after the acceptance of this paper, brought to our attention his Ph.D. thesis, which studied the tight convergence rate of the Krasnoselski–Mann iteration.

REFERENCES

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., with a foreword by Hédy Attouch, CMS Books Math./Ouvrages Math. SMC, Springer, Cham, 2017, https://doi.org/10.1007/978-3-319-48311-5.

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202, https://doi.org/10.1137/080716542.

[3] H. Brézis and P.-L. Lions, *Produits infinis de résolvantes*, Israel J. Math., 29 (1978), pp. 329–345, https://doi.org/10.1007/BF02761171.

[4] D. Davis and W. Yin, *Convergence rate analysis of several splitting schemes*, in Splitting Methods in Communication, Imaging, Science, and Engineering, Sci. Comput., Springer, Cham, 2016, pp. 115–163.

[5] D. Davis and W. Yin, *Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions*, Math. Oper. Res., 42 (2017), pp. 783–805, https://doi.org/10.1287/moor.2016.0827.

[6] E. de Klerk, F. Glineur, and A. B. Taylor, *On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions*, Optim. Lett., 11 (2017), pp. 1185–1199, https://doi.org/10.1007/s11590-016-1087-4.

[7] J. Douglas, Jr., and H. H. Rachford, Jr., *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[8] Y. Drori and M. Teboulle, *Performance of first-order methods for smooth convex minimization: A novel approach*, Math. Program., 145 (2014), pp. 451–482, https://doi.org/10.1007/s10107-013-0653-0.

[9] Y. Drori and M. Teboulle, *An optimal variant of Kelley's cutting-plane method*, Math. Program., 160 (2016), pp. 321–351, https://doi.org/10.1007/s10107-016-0985-7.
[10] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming, 55 (1992), pp. 293–318, https://doi.org/10.1007/BF01581204.
[11] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., 29 (1991), pp. 403–419, https://doi.org/10.1137/0329022.
[12] O. Güler, *New proximal point algorithms for convex minimization*, SIAM J. Optim., 2 (1992), pp. 649–664, https://doi.org/10.1137/0802032.
[13] E. V. Haynsworth, *On the Schur Complement*, Basel Mathematical Notes, BMN 20, 1968, University of Basel.
[14] B. He and X. Yuan, *On the convergence rate of Douglas-Rachford operator splitting method*, Math. Program., 153 (2015), pp. 715–722, https://doi.org/10.1007/s10107-014-0805-x.
[15] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.
[16] D. Kim, *Accelerated Proximal Point Method for Maximally Monotone Operators*, preprint, https://arxiv.org/abs/1905.05149, 2019.
[17] D. Kim and J. A. Fessler, *Optimized first-order methods for smooth convex minimization*, Math. Program., 159 (2016), pp. 81–107, https://doi.org/10.1007/s10107-015-0949-3.
[18] D. Kim and J. A. Fessler, *On the convergence analysis of the optimized gradient method*, J. Optim. Theory Appl., 172 (2017), pp. 187–205, https://doi.org/10.1007/s10957-016-1018-7.
[19] D. Kim and J. A. Fessler, *Another look at the fast iterative shrinkage/thresholding algorithm (FISTA)*, SIAM J. Optim., 28 (2018), pp. 223–250, https://doi.org/10.1137/16M108940X.
[20] D. Kim and J. A. Fessler, *Generalizing the optimized gradient method for smooth convex minimization*, SIAM J. Optim., 28 (2018), pp. 1920–1950, https://doi.org/10.1137/17M112124X.
[21] D. Kim and J. A. Fessler, *Optimizing the Efficiency of First-order Methods for Decreasing the Gradient of Smooth Convex Functions*, preprint, https://arxiv.org/abs/1803.06600, 2018.
[22] D. Leventhal, *Metric subregularity and the proximal point method*, J. Math. Anal. Appl., 360 (2009), pp. 681–688, https://doi.org/10.1016/j.jmaa.2009.07.012.
[23] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, https://doi.org/10.1137/0716071.
[24] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), pp. 154–158.
[25] B. Martinet, *Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox*, C. R. Acad. Sci. Paris Sér. A-B, 274 (1972), pp. A163–A165.
[26] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J., 29 (1962), pp. 341–346.
[27] J.-J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C. R. Acad. Sci. Paris, 255 (1962), pp. 2897–2899.
[28] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[29] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.
[30] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization (Sympos., Univ. Keele, Keele, 1968), Academic Press, London, 1969, pp. 283–298.
[31] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, 1970.
[32] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.
[33] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898, https://doi.org/10.1137/0314056.
[34] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson, *Operator Splitting Performance Estimation: Tight Contraction Factors and Optimal Parameter Selection*, https://arxiv.org/abs/1812.00146, 2018.
[35] J. F. Sturm, *Using SeDuMi* 1.02*, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods Softw., 11/12 (1999), pp. 625–653.
[36] A. B. Taylor and F. Bach, *Stochastic first-order methods: Non-asymptotic and computer-aided analyses via potential functions*, Proc. Mach. Learn. Res., 99 (2019), pp. 1–59.
[37] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Exact worst-case performance of first-order methods for composite convex optimization*, SIAM J. Optim., 27 (2017), pp. 1283–1313, https://doi.org/10.1137/16M108104X.

[38] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Smooth strongly convex interpolation and exact worst-case performance of first-order methods*, Math. Program., 161 (2017), pp. 307–345, https://doi.org/10.1007/s10107-016-1009-3.

[39] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Exact worst-case convergence rates of the proximal gradient method for composite convex minimization*, J. Optim. Theory Appl., 178 (2018), pp. 455–476, https://doi.org/10.1007/s10957-018-1298-1.