FULL LENGTH PAPER

Series A

Accelerated proximal point method for maximally monotone operators

Donghwan Kim[1]

Abstract

This paper proposes an accelerated proximal point method for maximally monotone operators. The proof is computer-assisted via the performance estimation problem approach. The proximal point method includes various well-known convex optimization methods, such as the proximal method of multipliers and the alternating direction method of multipliers, and thus the proposed acceleration has wide applications. Numerical experiments are presented to demonstrate the accelerating behaviors.

Keywords Proximal point method · Acceleration · Maximally monotone operators · Worst-case performance analysis

Mathematics Subject Classification 90C25 · 90C30 · 90C60 · 68Q25 · 49M25 · 90C22

1 Introduction

A fundamental tool for finding a root of a monotone operator is the proximal point method [46,59]. The monotone operator theory is particularly of interest, since it is closely related to convex functions and convex minimization [7,14,60]. For example, the proximal point method is useful when solving ill-conditioned problems or dual problems. In particular, the augmented Lagrangian method (i.e., the method of multipliers) [36,56] and the alternating direction method of multipliers (ADMM) [26,27] are instances of the proximal point method applied to dual problems [23,24,58].

To improve the efficiency of the proximal point method, accelerating its worst-case rate has been of interest both in theory and in applications (see e.g.,

✉ Donghwan Kim
   donghwankim@kaist.ac.kr

[1] Department of Mathematical Sciences, KAIST, Daejeon, Republic of Korea

[1,3,6,15,29,34,43]). In specific, inspired by Nesterov's fast gradient method [51,52], Güler [34] accelerated the worst-case rate of the proximal point method for convex minimization with respect to the cost function. This yields the fast rate $O(1/i^2)$ where i denotes the number of iterations, compared to the $O(1/i)$ rate of the proximal point method. However, this acceleration has not been theoretically generalized to the monotone inclusion problem, and only somewhat empirical accelerations, e.g., via the relaxation and the inertia (i.e., an implicit version of the heavy ball method [55], or equivalently, Nesterov's and Güler's accelerating technique [34,51,52]) in [1,3,6,15,29], have been studied. Therefore, this paper studies accelerating the worst-case rate of the proximal point method with respect to the fixed-point residual for maximally monotone operators. This provides the fast $O(1/i^2)$ rate, which improves upon the rate $O(1/i)$ of the proximal point method [9,32]. The proof is computer-assisted via the performance estimation problem (PEP) approach [21] and its extensions [20,22,31,32,37–41,61,65–67].

Under the additional strong monotonicity condition, the proximal point method has a linear rate in terms of the fixed-point residual [59], while the proposed acceleration is not guaranteed to have such a linear rate. Therefore, this paper further employs a restarting technique (e.g., [50, Section 11.4], [53, Section 5.1]) under the strong monotonicity condition. This has a linear rate, and is faster than the proximal point method for some practical cases.

The proposed acceleration of the proximal point method has wide applications. This provides an acceleration to the proximal method of multipliers [58], the Douglas–Rachford splitting method [19,44], and ADMM [26,27]. The proposed result also applies to a preconditioned proximal point method such as the primal–dual hybrid gradient (PDHG) method [11,13,25,35], (i.e., a preconditioned ADMM), yielding an accelerated PDHG method. This paper then shows that the proposed acceleration applies to a forward method for cocoercive operators. Existing works on accelerating the forward method can be found, for example, in [2,45].

Section 2 reviews maximally monotone operators, the proximal point method and its known accelerations. Section 3 studies the PEP with respect to the fixed-point residual for monotone inclusion problems. Section 4 proposes a new accelerated proximal point method using the PEP. Section 5 considers a restarting technique to yield a linear rate, under the additional strongly monotone assumption. Section 6 applies the proposed acceleration to well-known instances of the proximal point method, such as the proximal method of multipliers, the PDHG method, the Douglas–Rachford splitting method, and ADMM. Section 6 also provides numerical experiments. Section 7 presents that the proposed approach also accelerates the forward method for cocoercive operators, and Sect. 8 concludes.

2 Problem and method

2.1 Monotone inclusion problem

Let $\mathcal{H}$ be a real Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$, and associated norm $|| \cdot ||$. A set-valued operator $\boldsymbol{M} : \mathcal{H} \to 2^{\mathcal{H}}$ is *monotone* if

$$\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{u} - \boldsymbol{v} \rangle \geq 0 \text{ for all } (\boldsymbol{x}, \boldsymbol{u}), (\boldsymbol{y}, \boldsymbol{v}) \in \text{gra } \boldsymbol{M}, \tag{1}$$

where gra $\boldsymbol{M} := \{(\boldsymbol{x}, \boldsymbol{u}) \in \mathcal{H} \times \mathcal{H} : \boldsymbol{u} \in \boldsymbol{Mx}\}$ denotes the graph of $\boldsymbol{M}$. A monotone operator $\boldsymbol{M}$ is *maximally monotone* if there exists no monotone operator $\boldsymbol{A} : \mathcal{H} \to 2^{\mathcal{H}}$ such that gra $\boldsymbol{A}$ properly contains gra $\boldsymbol{M}$. Let $\mathcal{M}(\mathcal{H})$ be the class of maximally monotone operators on $\mathcal{H}$. In addition, a set-valued operator $\boldsymbol{M} : \mathcal{H} \to 2^{\mathcal{H}}$ is μ-*strongly monotone* for $\mu \in \mathbb{R}_{++}$, if

$$\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{u} - \boldsymbol{v} \rangle \geq \mu ||\boldsymbol{x} - \boldsymbol{y}||^2 \text{ for all } (\boldsymbol{x}, \boldsymbol{u}), (\boldsymbol{y}, \boldsymbol{v}) \in \text{gra } \boldsymbol{M}. \tag{2}$$

Let $\mathcal{M}_\mu(\mathcal{H})$ be the class of maximally and μ-strongly monotone operators on $\mathcal{H}$. Also, define $\mathcal{B}(\mathcal{H}, \mathcal{G}) = \{\boldsymbol{L} : \mathcal{H} \to \mathcal{G} \mid \boldsymbol{L} \text{ is linear and bounded}\}$ for a real Hilbert space $\mathcal{G}$ equipped with inner product $\langle \cdot, \cdot \rangle$, and let $\boldsymbol{L}^* \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ be the adjoint of $\boldsymbol{L} \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ that satisfies $\langle \boldsymbol{Lx}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{L}^* \boldsymbol{y} \rangle$ for all $\boldsymbol{x} \in \mathcal{H}$ and $\boldsymbol{y} \in \mathcal{G}$.

This paper considers the monotone inclusion problem:

$$\text{Find } \boldsymbol{x} \in \mathcal{H} \text{ subject to } \boldsymbol{0} \in \boldsymbol{Mx}, \tag{3}$$

where $\boldsymbol{M} \in \mathcal{M}(\mathcal{H})$ (or $\boldsymbol{M} \in \mathcal{M}_\mu(\mathcal{H})$). This includes convex problems and convex–concave problems; a subdifferential ∂f of a closed proper convex function $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ is maximally monotone [48]. Let $\mathcal{F}(\mathcal{H})$ be the class of closed proper convex functions on $\mathcal{H}$.

We assume that the optimal set $X_*(\boldsymbol{M}) := \{\boldsymbol{x} \in \mathcal{H} : \boldsymbol{0} \in \boldsymbol{Mx}\}$ is nonempty. We also assume that the distance between an initial point $\boldsymbol{x}_0$ and some optimal point $\boldsymbol{x}_* \in X_*(\boldsymbol{M})$ is bounded as

$$||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \leq R \text{ for a constant } R > 0. \tag{4}$$

2.2 Proximal point method and its worst-case rates

Proximal point method was first introduced to convex optimization by Martinet [46], which is based on the proximal mapping by Moreau [49]. The method was later extended to monotone inclusion problem by Rockafellar [59]. The proximal point method for maximally monotone operators includes the augmented Lagrangian [36, 56], the proximal method of multipliers [58], the Douglas–Rachford splitting method [19,44], and the alternating direction method of multipliers (ADMM) [26,27], so studying its worst-case convergence behavior and acceleration is important, which is of main interest in this paper.

The proximal mapping [49] (or the resolvent operator) of an operator $\boldsymbol{M}$ is defined as

$$\boldsymbol{J_M} := (\boldsymbol{I} + \boldsymbol{M})^{-1}, \tag{5}$$

where $\boldsymbol{I} : \mathcal{H} \to \mathcal{H}$ is an identity operator, i.e., $\boldsymbol{I}(\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{H}$. The resolvent operator $\boldsymbol{J_M}$ is single-valued and firmly nonexpansive for $\boldsymbol{M} \in \mathcal{M}(\mathcal{H})$

[47]. The proximal point method [46,59] generates a sequence $\{x_i\}$ by iteratively applying the resolvent operator with a positive real number λ as below.

Proximal Point Method

    Input: $M \in \mathcal{M}(\mathcal{H}), \ x_0 \in \mathcal{H}, \ \lambda \in \mathbb{R}_{++}$.

    For $i = 0, 1, \ldots$

        $x_{i+1} = J_{\lambda M}(x_i)$.

In [9, Proposition 8], the worst-case rate of the proximal point method with respect to the fixed-point residual

$$||x - J_{\lambda M}(x)||^2 \tag{6}$$

was found to satisfy

$$||x_i - x_{i-1}||^2 \leq \frac{R^2}{i} \tag{7}$$

for $i \geq 1$. Very recently in [32], this was improved to

$$||x_i - x_{i-1}||^2 \leq \left(1 - \frac{1}{i}\right)^{i-1} \frac{R^2}{i}, \tag{8}$$

which is exact when $\dim \mathcal{H} \geq 2$. Such exact worst-case with $\dim \mathcal{H} = 2$ given in [32] will be visited at the end of Sect. 4. The bound (8) is asymptotically e-times lower than (7), where e is Euler's number. When we additionally assume the μ-strong monotonicity, the proximal point method has a linear rate [7, Example 23.40] [59]

$$||x_{i+1} - x_i||^2 \leq \left(\frac{1}{1 + \lambda\mu}\right)^2 ||x_i - x_{i-1}||^2 \tag{9}$$

for $i \geq 1$, which is exact considering the case $Mx = \mu x$ with $\dim \mathcal{H} = 1$.

For a convex minimization of $f \in \mathcal{F}(\mathcal{H})$, [66, Conjecture 4.2] conjectures that the proximal point method satisfies

$$||x_i - x_{i-1}||^2 \leq \frac{R^2}{i^2} \tag{10}$$

for $i \geq 1$, which is faster than (8) for maximally monotone operators. In addition, the $O(1/i)$ worst-case rate of the proximal point method with respect to the cost function was studied in [33, Theorem 2.1], and this was improved by a constant 2 in [66, Theorem 4.1]

$$f(\boldsymbol{x}_i) - f(\boldsymbol{x}_*) \leq \frac{R^2}{4\lambda i} \tag{11}$$

for $i \geq 1$ and some $\boldsymbol{x}_* \in X_*(\partial f)$ with $\|\boldsymbol{x}_0 - \boldsymbol{x}_*\| \leq R$.

Remark 2.1 The results for the proximal point method can be applied to a *preconditioned* proximal point method. Let $\boldsymbol{L} \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ be invertible. Then, $\boldsymbol{L}^*\boldsymbol{ML}$ is maximally monotone for $\boldsymbol{M} \in \mathcal{M}(\mathcal{H})$ [7, Proposition 23.25], and the corresponding proximal point method is

$$\tilde{\boldsymbol{x}}_{i+1} = \boldsymbol{J}_{\lambda \boldsymbol{L}^*\boldsymbol{ML}}(\tilde{\boldsymbol{x}}_i) = (\boldsymbol{I} + \lambda \boldsymbol{L}^*\boldsymbol{ML})^{-1}\tilde{\boldsymbol{x}}_i. \tag{12}$$

Introducing $\boldsymbol{x}_i = \boldsymbol{L}\tilde{\boldsymbol{x}}_i$ and $\boldsymbol{P} = (\boldsymbol{LL}^*)^{-1}$ yields the following equivalent *preconditioned* proximal point method

$$\boldsymbol{x}_{i+1} = (\boldsymbol{P} + \lambda \boldsymbol{M})^{-1}\boldsymbol{Px}_i. \tag{13}$$

So, for example, the inequality (7) leads to the *preconditioned* fixed-point residual bound for the *preconditioned* proximal point method

$$\langle \boldsymbol{P}(\boldsymbol{x}_i - \boldsymbol{x}_{i-1}), \, \boldsymbol{x}_i - \boldsymbol{x}_{i-1} \rangle \leq \frac{R^2}{i} \tag{14}$$

for $i \geq 1$, and for some $\boldsymbol{x}_* \in X_*(\boldsymbol{M})$ with $\langle \boldsymbol{P}(\boldsymbol{x}_0 - \boldsymbol{x}_*), \, \boldsymbol{x}_0 - \boldsymbol{x}_* \rangle \leq R^2$. This is particularly useful when considering the PDHG method [11,13,25], [35, Lemma 2.2], which is an instance of a *preconditioned* proximal point method. We will revisit this in Sect. 6.2.

2.3 Existing accelerations for proximal point method

This section reviews existing accelerations of proximal point method for convex minimization with respect to the cost function. To the best of our knowledge, there is no other type of proximal point methods that guarantees accelerated worst-case rates.

For convex minimization, Güler [34] developed the following two accelerated versions, inspired by Nesterov's fast gradient method [51,52]. The following is the first accelerated version of the proximal point method in [34] which is an instance of FISTA [8]. The original version in [34] includes some variation with an iteration-dependent λ_i, rather than a fixed constant λ (see also [5] for choosing λ_i appropriate for further acceleration). This paper focuses on a fixed constant λ, and we leave its extension to a varying constant λ_i as future work.

Güler's First Accelerated Proximal Point Method in [34, Sec. 2]

Input: $f \in \mathcal{F}(\mathcal{H})$, $\boldsymbol{x}_0 = \boldsymbol{y}_0 \in \mathcal{H}$, $\lambda \in \mathbb{R}_{++}$, $t_0 = 1$.

For $i = 0, 1, \ldots$

$$\boldsymbol{x}_{i+1} = \boldsymbol{J}_{\lambda \partial f}(\boldsymbol{y}_i),$$

$$t_{i+1} = \frac{1 + \sqrt{1 + 4t_i^2}}{2},$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_{i+1} + \frac{t_i - 1}{t_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i).$$

The sequence generated by the Güler's first accelerated proximal point method satisfies [34, Theorem 2.3], [8, Theorem 4.4]

$$f(\boldsymbol{x}_i) - f(\boldsymbol{x}_*) \leq \frac{R^2}{2\lambda t_{i-1}^2} \leq \frac{2R^2}{\lambda(i + 1)^2} \tag{15}$$

for $i \geq 1$ and for some $\boldsymbol{x}_* \in X_*(\partial f)$ with $||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \leq R$. The following is another accelerated proximal point method by Güler [34], which the formulation is similar to those of the optimized gradient methods [37,39,40].

Güler's Second Accelerated Proximal Point Method in [34, Appendix]

Input: $f \in \mathcal{F}(\mathcal{H})$, $\boldsymbol{x}_0 = \boldsymbol{y}_0 \in \mathcal{H}$, $\lambda \in \mathbb{R}_{++}$, $t_0 = 1$.

For $i = 0, 1, \ldots$

$$\boldsymbol{x}_{i+1} = \boldsymbol{J}_{\lambda \partial f}(\boldsymbol{y}_i),$$

$$t_{i+1} = \frac{1 + \sqrt{1 + 4t_i^2}}{2},$$

$$\boldsymbol{y}_{i+1} = \boldsymbol{x}_{i+1} + \frac{t_i - 1}{t_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) + \frac{t_i}{t_{i+1}}(\boldsymbol{x}_{i+1} - \boldsymbol{y}_i).$$

The sequence generated by Güler's second accelerated proximal point method satisfies [34, Theorem 6.1] for $i \geq 1$

$$f(\boldsymbol{x}_i) - f(\boldsymbol{x}_*) \leq \frac{R^2}{4\lambda t_{i-1}^2} \leq \frac{R^2}{\lambda(i + 1)^2}, \tag{16}$$

which is twice smaller than (15).

2.4 Main contribution

To accelerate the worst-case rate of the proximal point method for maximally monotone operators, the relaxation and the inertia (i.e., an implicit version of the heavy ball method [55], or equivalently, Nesterov's and Güler's accelerating technique

[34,51,52]) have been studied in [1,3,6,15,29]. However, none of them guarantee accelerated rates. Therefore, the main contribution of this paper is to develop a method that has a fast $O(1/i^2)$ rate with respect to the fixed-point residual, improving upon the $O(1/i)$ rate of the proximal point method in (7) and (8).

This paper considers the following general proximal point method with step coefficients $\{h_{i+1,k+1}\}_{k=0}^{i}$ for reusing previous and current updates $\{x_{k+1} - y_k\}_{k=0}^{i}$. This includes the proximal point method, the accelerated methods via the relaxation and the inertia [1,3,6,15,29], and the proposed accelerated method.

General Proximal Point Method

Input: $M \in \mathcal{M}(\mathcal{H})$, $y_0 \in \mathcal{H}$, $\lambda \in \mathbb{R}_{++}$.

For $i = 0, 1, \ldots$

$$x_{i+1} = J_{\lambda M}(y_i),$$

$$y_{i+1} = y_i + \sum_{k=0}^{i} h_{i+1,k+1}(x_{k+1} - y_k).$$

This paper next uses the PEP approach to find the choice of $\{h_{i+1,k+1}\}_{k=0}^{i}$ that guarantees an accelerated rate. While the formulation of the general proximal point method is inefficient in general, the proposed accelerated method with the specific choice of $\{h_{i+1,k+1}\}_{k=0}^{i}$ found by PEP has an efficient equivalent form. This form is similar to the other accelerated methods with the relaxation and/or the inertia [1,3,6, 15,29].

3 Performance estimation problem for maximally monotone operators

This section uses the performance estimation problem (PEP) approach [21,66,67] to analyze the general proximal point method for maximally monotone operators, in terms of the fixed-point residual (6). This was recently studied in [32] for the proximal point method, providing the exact rate (8). The same authors [31] also used the PEP to study the exact worst-case rate for the *ergodic* sequence of the (relaxed) proximal point method for the variational inequalities. Similarly, [66] used PEP to analyze the worst-case rate of the proximal point method for convex minimization in terms of the fixed-point residual and the cost function, yielding (10) and (11), respectively.

Building upon [21,31,32,66,67], the worst-case rate of the general proximal point method after N iterations for decreasing the fixed-point residual (6) under the initial distance condition (4) can be computed by

$$\max_{\substack{M \in \mathcal{M}(\mathcal{H})}} \max_{\substack{x_1,\ldots,x_N \in \mathcal{H}, \\ y_0,\ldots,y_{N-1} \in \mathcal{H}, \\ x_* \in X_*(M)}} \frac{1}{R^2} ||x_N - y_{N-1}||^2$$

$$\text{subject to } x_{i+1} = J_{\lambda M}(y_i), \quad i = 0, \ldots, N-1,$$

$$y_{i+1} = y_i + \sum_{k=0}^{i} h_{i+1,k+1}(x_{k+1} - y_k), \quad i = 0, \dots, N-2,$$

$$\|y_0 - x_*\|^2 \leq R^2. \tag{17}$$

This is an infinite-dimensional problem due to the constraint $M \in \mathcal{M}(\mathcal{H})$, which is impractical to solve. PEP in [21] further introduced a series of steps that reformulate such impractical problem into a tractable problem, which we apply to (17) step by step below.

The first step is to reformulate the problem (17) into a finite-dimensional problem. [61, Fact 1] implies that one can replace $M \in \mathcal{M}(\mathcal{H})$ in (17) by a set of inequality constraints (1) for $M \in \mathcal{M}(\mathcal{H})$ on the finite number of pairs of points $\{x_1, \dots, x_N, x_*\}$ without strictly relaxing the problem (17). In specific, such constraints are

$$\langle x_i - x_j, \, q_i - q_j \rangle \geq 0, \tag{18}$$

for all $i, j \in \{1, \dots, N, *\}$, with additional variables $q_i \in Mx_i$ for $i = 1, \dots, N$ and $q_* = 0 \in Mx_*$. Then the resulting equivalent problem of (17) is

$$\max_{\substack{x_1,\dots,x_N,x_*\in\mathcal{H}, \\ y_0,\dots,y_{N-1}\in\mathcal{H}, \\ q_1,\dots,q_N\in\mathcal{H}}} \frac{1}{R^2}\|x_N - y_{N-1}\|^2$$

$$\text{subject to} \quad \langle x_i - x_j, \, q_i - q_j \rangle \geq 0, \quad i < j = 1, \dots, N,$$

$$\langle x_i - x_*, \, q_i \rangle \geq 0, \quad i = 1, \dots, N,$$

$$x_{i+1} = y_i - \lambda q_{i+1}, \quad i = 0, \dots, N-1,$$

$$y_{i+1} = y_i - \lambda \sum_{k=0}^{i} h_{i+1,k+1} q_{k+1}, \quad i = 0, \dots, N-2,$$

$$\|y_0 - x_*\|^2 \leq R^2. \tag{19}$$

Further removing x_i and using the change of variables

$$g_i := \frac{\lambda}{R} q_i, \quad i = 1, \dots, N, \tag{20}$$

simplify the problem (19) as

$$\max_{\substack{y_0,\dots,y_{N-1},x_*\in\mathcal{H}, \\ g_1,\dots,g_N\in\mathcal{H}}} \|g_N\|^2$$

$$\text{subject to} \quad \frac{1}{R} \langle y_{i-1} - Rg_i - y_{j-1} + Rg_j, \, g_i - g_j \rangle \geq 0, \quad i < j = 1, \dots, N,$$

$$\frac{1}{R} \langle y_{i-1} - Rg_i - x_*, \, g_i \rangle \geq 0, \quad i = 1, \dots, N,$$

$$y_{i+1} = y_i - R \sum_{k=0}^{i} h_{i+1,k+1} g_{k+1}, \quad i = 0, \ldots, N-2,$$

$$\|y_0 - x_*\|^2 \leq R^2. \tag{21}$$

As in [21,31,32,66,67], we next introduce the Gram matrix

$$Z = \begin{bmatrix} \|g_1\|^2 & \langle g_1, g_2 \rangle \cdots & \langle g_1, g_N \rangle & \frac{1}{R} \langle g_1, y_0 - x_* \rangle \\ \langle g_1, g_2 \rangle & \|g_2\|^2 \cdots & \langle g_2, g_N \rangle & \frac{1}{R} \langle g_2, y_0 - x_* \rangle \\ \vdots & \vdots \ddots & \vdots & \vdots \\ \langle g_1, g_N \rangle & \cdots & \|g_N\|^2 & \frac{1}{R} \langle g_N, y_0 - x_* \rangle \\ \frac{1}{R} \langle g_1, y_0 - x_* \rangle & \cdots & \frac{1}{R} \langle g_N, y_0 - x_* \rangle & \frac{1}{R^2} \|y_0 - x_*\|^2 \end{bmatrix} \tag{22}$$

to relax the problem as

$$\max_{Z \in \mathbb{S}_+^{N+1}} \text{tr}\{u_N u_N^\top Z\}$$

$$\text{subject to } \text{tr}\{A_{i,j}(h)Z\} \leq 0, \quad i < j = 1, \ldots, N,$$

$$\text{tr}\{B_i(h)Z\} \leq 0, \quad i = 1, \ldots, N,$$

$$\text{tr}\{CZ\} \leq 1, \tag{23}$$

where $\{u_i\}_{i=1}^{N+1}$ is the canonical basis of $\mathbb{R}^{N+1}$ and

$$\begin{cases} A_{i,j}(h) := (u_i - u_j) \odot (u_i - u_j) - (u_i - u_j) \odot \sum_{l=i-1}^{j-2} \sum_{k=0}^{l} h_{l+1,k+1} u_{k+1}, & i < j = 1, \ldots, N, \\ B_i(h) := u_i u_i^\top - u_i \odot u_{N+1} + u_i \odot \sum_{l=0}^{i-2} \sum_{k=0}^{l} h_{l+1,k+1} u_{k+1}, & i = 1, \ldots, N, \\ C := u_{N+1} u_{N+1}^\top \end{cases}$$

with the outer product operator $u \odot v := \frac{1}{2}(uv^\top + vu^\top)$. If $\dim \mathcal{H} \geq N+1$, the problems (17) and (23) are equivalent, based on the following lemma similar to [61, Lemma 1].

Lemma 3.1 *If $\dim \mathcal{H} \geq N+1$, then*

$$Z \in \mathbb{S}_+^{N+1} \quad \Leftrightarrow \quad \exists \, g_1, g_2, \ldots, g_N, \frac{1}{R}(y_0 - x_*) \in \mathcal{H}$$

such that $Z = $ expression of (22).

For simplicity in later analysis, we discard some constraints as

$$\max_{Z \in \mathbb{S}_+^{N+1}} \text{tr}\{u_N u_N^\top Z\}$$

$$\text{subject to } \text{tr}\{A_{i-1,i}(h)Z\} \leq 0, \quad i = 2, \ldots, N,$$

$$\text{tr}\{B_N(h)Z\} \leq 0,$$

$$\text{tr}\{CZ\} \leq 1, \tag{24}$$

which does not affect the result in the paper, i.e., the optimal values of (23) and (24) are found to be numerically equivalent for the method proposed in this paper. Finally, we construct the associated Lagrangian dual of (24)

$$\mathcal{B}_D(\boldsymbol{h}) := \min_{a_2,\dots,a_N,b_N,c \in \mathbb{R}} c \tag{D}$$

$$\text{subject to } \sum_{i=2}^{N} a_i \boldsymbol{A}_{i-1,i}(\boldsymbol{h}) + b_N \boldsymbol{B}_N(\boldsymbol{h}) + c\boldsymbol{C} - \boldsymbol{u}_N \boldsymbol{u}_N^\top \succeq \boldsymbol{0},$$

$$a_2,\dots,a_N,b_N,c \geq 0,$$

where $a_2,\dots,a_N,b_N,c$ are dual variables associated with the constraints of (24), respectively. Then, for any given $\boldsymbol{h}$ for the general proximal point method, one can compute its (upper bound of) worst-case fixed-point residual by numerically solving (D) using any SDP solver. For some choices of $\boldsymbol{h}$ as for the proximal point method in [32], it might be possible to analytically solve (D); [32] analytically solved (D) for the proximal point method yielding the rate (8). This paper provides another choice of $\boldsymbol{h}$ that provides an analytical solution to (D) with an accelerated rate.

4 Accelerating the proximal point method for maximally monotone operators

Using the dual problem (D), this section develops an accelerated version of the proximal point method via PEP:

$$\min_{\boldsymbol{h}} \mathcal{B}_D(\boldsymbol{h}), \tag{HD}$$

which is studied in [20–22,37–40] for certain classes of problems and methods. The problem is non-convex but convex for the variables $(a_2,\dots,a_N,b_N,c)$ given $\boldsymbol{h}$ and for the variables $(c,\boldsymbol{h})$ given $(a_2,\dots,a_N,b_N)$. Therefore, we used a variant of alternating minimization that alternatively optimizes over $(a_2,\dots,a_N,b_N,c)$ given $\boldsymbol{h}$ and over $(c,\boldsymbol{h})$ given $(a_2,\dots,a_N,b_N)$ to find a minimizer using a SDP solver [16,30]. Inspired by numerical results, the following lemma specifies a feasible point of (HD) analytically. We do not have a guarantee that such point is a (unique) minimizer of (HD).

Lemma 4.1 *The following*

$$h_{i,k} = \begin{cases} -\frac{2k}{i(i+1)}, & i = 1,\dots,N-1, \ k = 1,\dots,i-1, \\ \frac{2i}{i+1}, & i = 1,\dots,N-1, \ k = i, \end{cases} \tag{25}$$

$$a_i = \frac{2(i-1)i}{N^2}, \quad i = 2,\dots,N, \quad b_N = \frac{2}{N}, \quad c = \frac{1}{N^2} \tag{26}$$

is a feasible point of (D) *and* (HD).

Proof It is obvious that $a_2, \ldots, a_N, b_N, c$ are nonnegative, so we are only left to show the positive semidefinite condition in (D). Since

$$
\sum_{i=2}^{N} a_i A_{i-1,i}(\boldsymbol{h}) + b_N \boldsymbol{B}_N(\boldsymbol{h}) + c\boldsymbol{C} - \boldsymbol{u}_N \boldsymbol{u}_N^\top
$$

$$
= \sum_{i=2}^{N} \frac{2(i-1)i}{N^2} \Big[(\boldsymbol{u}_{i-1} - \boldsymbol{u}_i) \odot (\boldsymbol{u}_{i-1} - \boldsymbol{u}_i)
$$

$$
- (\boldsymbol{u}_{i-1} - \boldsymbol{u}_i) \odot \left(\frac{2(i-1)}{i} \boldsymbol{u}_{i-1} - \sum_{k=0}^{i-3} \frac{2(k+1)}{(i-1)i} \boldsymbol{u}_{k+1} \right) \Big]
$$

$$
+ \frac{2}{N} \Big[\boldsymbol{u}_N \boldsymbol{u}_N^\top - \boldsymbol{u}_N \odot \boldsymbol{u}_{N+1} + \boldsymbol{u}_N \odot \sum_{l=0}^{N-2}
$$

$$
\left(\frac{2(l+1)}{l+2} \boldsymbol{u}_{l+1} - \sum_{k=0}^{l-1} \frac{2(k+1)}{(l+1)(l+2)} \boldsymbol{u}_{k+1} \right) \Big]
$$

$$
+ \frac{1}{N^2} \boldsymbol{u}_{N+1} \boldsymbol{u}_{N+1}^\top - \boldsymbol{u}_N \boldsymbol{u}_N^\top
$$

$$
= \sum_{i=2}^{N-1} \Big[\frac{2(i-1)i}{N^2} + \frac{2i(i+1)}{N^2} \left(1 - \frac{2i}{i+1} \right) \Big] \boldsymbol{u}_i \boldsymbol{u}_i^\top + \Big[\frac{2(N-1)N}{N^2}
$$

$$
+ \frac{2}{N} - 1 \Big] \boldsymbol{u}_N \boldsymbol{u}_N^\top + \frac{1}{N^2} \boldsymbol{u}_{N+1} \boldsymbol{u}_{N+1}^\top
$$

$$
+ \sum_{i=2}^{N-1} \Big[\frac{2(i-1)i}{N^2} \left(-2 + \frac{2(i-1)}{i} \right) + \frac{2i(i+1)}{N^2} \frac{2(i-1)}{i(i+1)} \Big] \boldsymbol{u}_{i-1} \odot \boldsymbol{u}_i
$$

$$
+ \Big[\frac{2(N-1)}{N} \left(-2 + \frac{2(N-1)}{N} \right) + \frac{2}{N} \frac{2(N-1)}{N} \Big] \boldsymbol{u}_{N-1} \odot \boldsymbol{u}_N - \frac{2}{N} \boldsymbol{u}_N \odot \boldsymbol{u}_{N+1}
$$

$$
+ \sum_{i=3}^{N-1} \sum_{k=0}^{i-3} \Big[-\frac{2(i-1)i}{N^2} \frac{2(k+1)}{(i-1)i} + \frac{2i(i+1)}{N^2} \frac{2(k+1)}{i(i+1)} \Big] \boldsymbol{u}_{k+1} \odot \boldsymbol{u}_i
$$

$$
+ \sum_{k=0}^{N-3} \Big[-\frac{2(N-1)N}{N^2} \frac{2(k+1)}{(N-1)N}
$$

$$
+ \frac{2}{N} \left(\frac{2(k+1)}{k+2} - \sum_{l=k+1}^{N-2} \frac{2(k+1)}{(l+1)(l+2)} \right) \Big] \boldsymbol{u}_{k+1} \odot \boldsymbol{u}_N
$$

$$
= \boldsymbol{u}_N \boldsymbol{u}_N^\top + \frac{1}{N^2} \boldsymbol{u}_{N+1} \boldsymbol{u}_{N+1}^\top - \frac{2}{N} \boldsymbol{u}_N \odot \boldsymbol{u}_{N+1}
$$

$$
= \left(\boldsymbol{u}_N - \frac{1}{N} \boldsymbol{u}_{N+1} \right) \left(\boldsymbol{u}_N - \frac{1}{N} \boldsymbol{u}_{N+1} \right)^\top \succeq \boldsymbol{0},
$$

the given point is a feasible point of (HD). □

Before providing the worst-case rate of the general proximal point method with h in (25), we develop its efficient formulation below. This has a low computational cost per iteration, comparable to that of the proximal point method. Note that this may not be the only efficient form for h in (25).

Proposed Accelerated Proximal Point Method for Maximally Monotone Operators

Input: $M \in \mathcal{M}(\mathcal{H})$, $x_0 = y_0 = y_{-1} \in \mathcal{H}$, $\lambda \in \mathbb{R}_{++}$.

For $i = 0, 1, \dots$

$$x_{i+1} = J_{\lambda M}(y_i),$$

$$y_{i+1} = x_{i+1} + \frac{i}{i+2}(x_{i+1} - x_i) - \frac{i}{i+2}(x_i - y_{i-1}).$$

Proposition 4.1 *The sequences $\{x_i\}$ and $\{y_i\}$ generated by the general proximal point method with step coefficients $\{h_{i,k}\}$ in (25) are identical to the corresponding sequence generated by the proposed accelerated proximal point method starting from the same initial point.*

Proof We use induction, and for clarity we use the notation $x'_1, x'_2, \dots$ and $y'_0, y'_1, \dots$ for the general proximal point method with (25). It is obvious that $x_0 = y'_0 = y_0$, $x'_1 = x_1 = y_1$, and we have

$$y'_1 = y'_0 + h_{1,1}(x'_1 - y'_0) = x'_1 = y_1.$$

Similarly, it is obvious that $x'_2 = x_2$, and we have

$$y'_2 = y'_1 + \sum_{k=0}^{1} h_{2,k+1}(x'_{k+1} - y'_k) = y_1 + \frac{4}{3}(x_2 - y_1) - \frac{1}{3}(x_1 - y_0)$$

$$= x_2 + \frac{1}{3}(x_2 - x_1) - \frac{1}{3}(x_1 - y_0) = y_2.$$

It is then also obvious that $x'_3 = x_3$. Assuming $x'_l = x_l$ for $l = 1, \dots, i+1$ and $y'_l = y_l$ for $l = 0, \dots, i$, for some $i \geq 2$, we have

$$y'_{i+1} = y'_i + \sum_{k=0}^{i} h_{i+1,k+1}(x'_{k+1} - y'_k)$$

$$= y_i + \frac{2(i+1)}{i+2}(x_{i+1} - y_i) + \sum_{k=0}^{i-1}\left(-\frac{2(k+1)}{(i+1)(i+2)}\right)(x_{k+1} - y_k)$$

$$= y_i + \left(1 + \frac{i}{i+2}\right)(x_{i+1} - y_i) + \frac{i}{i+2}\sum_{k=0}^{i-1}\left(-\frac{2(k+1)}{i(i+1)}\right)(x_{k+1} - y_k)$$

$$= x_{i+1} + \frac{i}{i+2}(x_{i+1} - y_i) + \frac{i}{i+2}(y_i + y_{i-1} - 2x_i)$$

$$= x_{i+1} + \frac{i}{i+2}(x_{i+1} - x_i) - \frac{i}{i+2}(x_i - y_{i-1}) = y_{i+1},$$

where the fourth equality uses

$$\boldsymbol{y}_i = \boldsymbol{y}_{i-1} + \frac{2i}{i+1}(\boldsymbol{x}_i - \boldsymbol{y}_{i-1}) + \sum_{k=0}^{i-2}\left(-\frac{2(k+1)}{i(i+1)}\right)(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k)$$

$$= \boldsymbol{y}_{i-1} + 2(\boldsymbol{x}_i - \boldsymbol{y}_{i-1}) + \sum_{k=0}^{i-1}\left(-\frac{2(k+1)}{i(i+1)}\right)(\boldsymbol{x}_{k+1} - \boldsymbol{y}_k).$$

□

The proposed accelerated method has the inertia term $\frac{i}{i+2}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i)$, similar to Nesterov's acceleration [51,52] and Güler's methods [34]. However, the proposed method also has a correction term $-\frac{i}{i+2}(\boldsymbol{x}_i - \boldsymbol{y}_{i-1})$, which is essential to guarantee an accelerated rate. Without such correction term, the accelerated method can diverge, for which we provide an example at the end of this section. We leave further understanding the role of the proposed correction term as future work, possibly via a differential equation perspective as in [64] for Nesterov's acceleration. Note that a different correction term for Nesterov's acceleration has been studied via the differential equation analysis for convex minimization [4,63].

The following theorem provides an accelerated rate of the proposed method in terms of the fixed-point residual.[1]

Theorem 4.1 *Let $\boldsymbol{M} \in \mathcal{M}(\mathcal{H})$ and let $\boldsymbol{x}_0, \boldsymbol{y}_0, \boldsymbol{x}_1, \boldsymbol{y}_1, \ldots \in \mathcal{H}$ be generated by the proposed accelerated proximal point method. Assume that $||\boldsymbol{x}_0 - \boldsymbol{x}_*|| \leq R$ for a constant $R > 0$ and for some $\boldsymbol{x}_* \in X_*(\boldsymbol{M})$. Then for any $i \geq 1$,*

$$||\boldsymbol{x}_i - \boldsymbol{y}_{i-1}||^2 \leq \frac{R^2}{i^2}. \tag{27}$$

Proof Using Lemma 4.1, the general proximal point method with $\boldsymbol{h}$ (25) satisfies

$$\frac{1}{R^2}||\boldsymbol{x}_N - \boldsymbol{y}_{N-1}||^2 \leq \mathcal{B}_D(\boldsymbol{h}) \leq \frac{1}{N^2}. \tag{28}$$

Since the iterates of the method are recursive and do not depend on a given N, the bound (28) generalizes to the intermediate iterates of the method. By Proposition 4.1, the proposed accelerated proximal point method also satisfies the bound (28), which concludes the proof.   □

The bound (8) of the proximal point method was found to be exact in [32] by specifying a certain operator $\boldsymbol{M}$ achieving the bound (8) exactly; that is, for given

---

[1] The convergence of the fixed-point residual does not guarantee the convergence of the sequence of the iterates $\{\boldsymbol{x}_i\}$. We leave analyzing the convergence of the sequence as future work, possibly based on the convergence analysis in [10] for Nesterov's fast gradient method [51,52] and FISTA [8] in convex minimization.
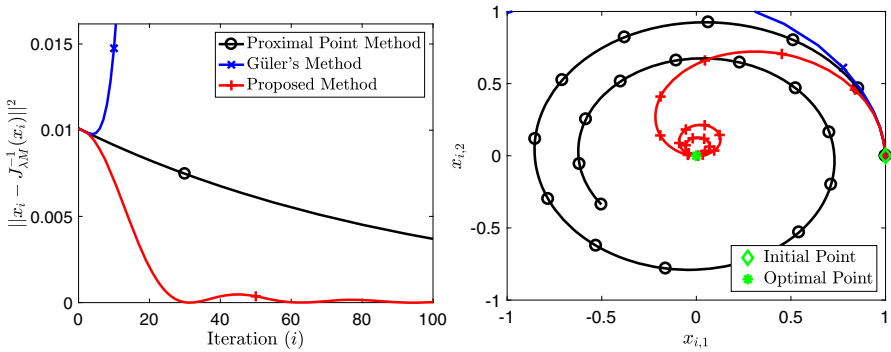
Fig. 1 Solving a worst-case monotone inclusion problem of the proximal point method with M (29) with $N = 100$; (left) the fixed-point residual vs. iteration, (right) the trajectory of the iterates $x_i = [x_{i,1}, \ x_{i,2}]^\top$ (markers are displayed every 5th iterations)

$N \geq 2$, the proximal point method exactly achieves the bound (8) for the operator

$$M \begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{\lambda \sqrt{N-1}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}, \tag{29}$$

with an initial point $x_0 = [1\ 0]^\top$. Such exact analysis is important since it reveals the worst-case behavior of the iterates of the method. However, we were not able to show that the bound (27) of the proposed method is exact, which we leave as future work. Instead, we compared the behavior of the iterates of the proximal point method and its accelerated variants on the operator M in (29). Figure 1 compares the proximal point method, Güler's first accelerated method with M instead of ∂f (i.e., an instance of the inertia method) and the proposed accelerated method, with an initial point $x_0 = [1\ 0]^\top$ and the optimal point $x_* = 0$. Note that the Güler's first method is almost equivalent to the proposed accelerated method without the correction term $-\frac{i}{i+2}(x_i - y_{i-1})$, and this exhibits diverging behavior in Fig. 1. The figure illustrates that the correction term greatly helps the iterates to rapidly converge by reducing the radius of the orbit of the iterates, compared to other methods.

We further investigate the behavior of the proposed method for a convex–concave saddle-point problem

$$\min_{u \in \mathcal{H}_1} \max_{v \in \mathcal{H}_2} \phi(u, v), \tag{30}$$

where $\mathcal{H}_1$ and $\mathcal{H}_2$ denote real Hilbert spaces equipped with inner product $\langle \cdot, \cdot \rangle$, and $\phi(\cdot, v) \in \mathcal{F}(\mathcal{H}_1)$, $-\phi(u, \cdot) \in \mathcal{F}(\mathcal{H}_2)$, which we further study in Sects. 5 and 6.1. The saddle subdifferential of ϕ,

$$\begin{bmatrix} \partial_u \phi(u, v) \\ \partial_v (-\phi(u, v)) \end{bmatrix}, \tag{31}$$

is monotone [57]. The proposed accelerated method applied to (31) with $\boldsymbol{x}_i := (\boldsymbol{u}_i, \boldsymbol{v}_i)$ and $\boldsymbol{x}_* := (\boldsymbol{u}_*, \boldsymbol{v}_*)$ (see Sect. 6.1 for details) satisfies

$$\phi(\boldsymbol{u}_i, \boldsymbol{v}_*) - \phi(\boldsymbol{u}_*, \boldsymbol{v}_i) \leq \frac{||\boldsymbol{u}_0 - \boldsymbol{u}_*||^2 + ||\boldsymbol{v}_0 - \boldsymbol{v}_*||^2}{4\lambda i} \tag{32}$$

for any $i \geq 1$. This is numerically conjectured by the PEP analysis in (19) with the objective function $\frac{1}{R}||\boldsymbol{x}_N - \boldsymbol{y}_{N-1}||^2$ and the inequality $\langle \boldsymbol{x}_N - \boldsymbol{x}_*, \boldsymbol{q}_N \rangle \geq 0$ in (19) replaced by $\phi(\boldsymbol{u}_N, \boldsymbol{v}_*) - \phi(\boldsymbol{u}_*, \boldsymbol{v}_N)$ and $\langle \boldsymbol{x}_N - \boldsymbol{x}_*, \boldsymbol{q}_N \rangle \geq \phi(\boldsymbol{u}_N, \boldsymbol{v}_*) - \phi(\boldsymbol{u}_*, \boldsymbol{v}_N)$, respectively.[2]

5 Restarting the accelerated proximal point method for strongly monotone operators

For strongly monotone operators, the proximal point method has a linear rate (9), whereas the proposed accelerated method is not guaranteed to have such a fast rate. Technically, one should be able to find an accelerated method for strong monotone operators via PEP, as we did for the monotone operators in the previous section. However, the resulting PEP problem, a reminiscent of (HD), is much more difficult to solve, and we leave it as future work. Instead, we consider a fixed restarting technique in [50, Section 11.4], [53, Section 5.1] that restarts an accelerated method with a sublinear rate every certain number of iterations to yield a fast linear rate, particularly for $\boldsymbol{M} \in \mathcal{M}_\mu(\mathcal{H})$ in this section.

Suppose one restarts the proposed method every k (inner) iterations by initializing the $(j + 1)$th outer iteration $\boldsymbol{x}_{j+1,0} = \boldsymbol{y}_{j+1,0} = \boldsymbol{y}_{j+1,-1}$ by $\boldsymbol{x}_{j,k}$, where $\boldsymbol{x}_{j,l}$ and $\boldsymbol{y}_{j,l}$ denote iterates at the jth outer iteration and lth inner iteration for $j = 0, 1, \dots$ and $l = -1, 0, 1, \dots, k$. Using the rate (27) (with $R = ||\boldsymbol{x}_{j,0} - \boldsymbol{x}_*||$) and the strong monotonicity condition (2), we have

$$||\boldsymbol{x}_{j,k} - \boldsymbol{y}_{j,k-1}||^2 \leq \frac{||\boldsymbol{x}_{j,0} - \boldsymbol{x}_*||^2}{k^2} \leq \frac{1}{\mu^2 k^2}||\boldsymbol{M}\boldsymbol{x}_{j,0}||^2 \tag{33}$$

for $j = 0, 1, \dots$. Since $\frac{1}{\lambda}(\boldsymbol{x}_{j-1,k} - \boldsymbol{y}_{j-1,k-1}) \in \boldsymbol{M}\boldsymbol{x}_{j,0}$, we have a linear rate

$$||\boldsymbol{x}_{j,k} - \boldsymbol{y}_{j,k-1}||^2 \leq \frac{1}{\lambda^2 \mu^2 k^2}||\boldsymbol{x}_{j-1,k} - \boldsymbol{y}_{j-1,k-1}||^2. \tag{34}$$

For a given $N = jk$ total number of steps, minimizing the overall rate with respect to k yields an optimal choice of the restarting interval given by $k_{\text{opt}} \approx \frac{e}{\lambda\mu}$, where e is Euler's number. The corresponding linear rate is $O((e^{\lambda\mu/e})^{-2N})$.

---

[2] A convex–concave function ϕ satisfies $\phi(\boldsymbol{u}_*, \boldsymbol{v}_N) \geq \phi(\boldsymbol{u}_N, \boldsymbol{v}_N) + \langle \boldsymbol{u}_* - \boldsymbol{u}_N, \boldsymbol{q}_{\boldsymbol{u},N} \rangle$ for $\boldsymbol{q}_{\boldsymbol{u},N} \in \partial_{\boldsymbol{u}}\phi(\boldsymbol{u}_N, \boldsymbol{v}_N)$ and $-\phi(\boldsymbol{u}_N, \boldsymbol{v}_*) \geq -\phi(\boldsymbol{u}_N, \boldsymbol{v}_N) + \langle \boldsymbol{v}_* - \boldsymbol{v}_N, -\boldsymbol{q}_{\boldsymbol{v},N} \rangle$ for $-\boldsymbol{q}_{\boldsymbol{v},N} \in \partial_{\boldsymbol{v}}(-\phi(\boldsymbol{u}_N, \boldsymbol{v}_N))$. Adding these two inequalities yields $\langle \boldsymbol{x}_N - \boldsymbol{x}_*, \boldsymbol{q}_N \rangle \geq \phi(\boldsymbol{u}_N, \boldsymbol{v}_*) - \phi(\boldsymbol{u}_*, \boldsymbol{v}_N)$, where $\boldsymbol{x}_N := (\boldsymbol{u}_N, \boldsymbol{v}_N)$ and $\boldsymbol{q}_N := (\boldsymbol{q}_{\boldsymbol{u},N}, -\boldsymbol{q}_{\boldsymbol{v},N})$.

We further investigate the behavior of the restarting technique for a saddle-point problem (30) with an assumption that ϕ is strongly-convex–strongly-concave, i.e., $\phi(\cdot, \boldsymbol{v}) - \frac{\mu}{2}||\cdot||^2 \in \mathcal{F}(\mathcal{H}_1)$ and $-\phi(\boldsymbol{u}, \cdot) - \frac{\mu}{2}||\cdot||^2 \in \mathcal{F}(\mathcal{H}_2)$. The associated saddle subdifferential (31) is μ-strongly monotone. For such case, using the rate (32), and the inequalities $\phi(\boldsymbol{u}_*, \boldsymbol{v}_*) + \frac{\mu}{2}||\boldsymbol{u} - \boldsymbol{u}_*||^2 \leq \phi(\boldsymbol{u}, \boldsymbol{v}_*)$ and $-\phi(\boldsymbol{u}_*, \boldsymbol{v}_*) + \frac{\mu}{2}||\boldsymbol{v} - \boldsymbol{v}_*||^2 \leq -\phi(\boldsymbol{u}_*, \boldsymbol{v})$, the proposed method with restarting every k iterations satisfies

$$\phi(\boldsymbol{u}_{j,k}, \boldsymbol{v}_*) - \phi(\boldsymbol{u}_*, \boldsymbol{v}_{j,k}) \leq \frac{1}{2\lambda\mu k}(\phi(\boldsymbol{u}_{j,0}, \boldsymbol{v}_*) - \phi(\boldsymbol{u}_*, \boldsymbol{v}_{j,0})) \qquad (35)$$

for $j = 0, 1, \ldots$. The associated optimal restarting interval is $k_{\text{opt}}^{\phi} \approx \frac{e}{2\lambda\mu}$, which is twice smaller than k_{opt}. The corresponding linear rate is also $O((e^{\lambda\mu/e})^{-2N})$, whereas the proximal point method has the rate $O((1 + \lambda\mu)^{-2N})$ in (9). For any given positive μ, there is no positive λ that satisfies both $k_{\text{opt}}^{\phi}(\lambda, \mu) \geq 1$ and $e^{\lambda\mu/e} > 1 + \lambda\mu$. This contrasts with the fact that the worst-case rate of optimally restarting the proposed method is not slower than that of the proximal point method. This implies that the bounds (34) and (35) are not exact, and we leave finding their tight bounds as future work. The numerical experiment below (and those in Sect. 6) suggests that the restarting technique can perform better than the proximal point method.

We consider a toy problem that is a combination of the worst-case problems in $\mathcal{M}(\mathcal{H})$ and $\mathcal{M}_{\mu}(\mathcal{H})$ for the proximal point method:

$$\boldsymbol{M}\begin{bmatrix} u \\ v \end{bmatrix} = \left(\frac{1}{\lambda\sqrt{N-1}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} \mu & 0 \\ 0 & \mu \end{bmatrix} \right) \begin{bmatrix} u \\ v \end{bmatrix}, \qquad (36)$$

which is the saddle subdifferential operator of $\phi(u, v) = \frac{\mu}{2}u^2 + \frac{1}{\lambda\sqrt{N-1}}uv - \frac{\mu}{2}v^2$. We choose $N = 100$, $\lambda = 1$ and $\mu = 0.02$. The optimal restarting intervals are $k_{\text{opt}} \approx 136$ and $k_{\text{opt}}^{\phi} \approx 68$, and we run 200 iterations in the experiment, where restarting intervals 17, 34, 68, and 136 are considered. Figure 2 compares the proximal point method, its accelerated variants, and the proposed accelerated method with restarting, with an initial point $\boldsymbol{x}_0 = [1 \ 0]^{\top}$ and the optimal point $\boldsymbol{x}_* = \boldsymbol{0}$. Figure 2 presents that the proximal point method has a linear rate that is faster than the proposed method (with a sublinear rate), while the restarting greatly accelerates the proposed method with a fast linear rate. Figure 2 also illustrates that the optimal restarting intervals k_{opt} and k_{opt}^{ϕ} for strongly monotone operators and strongly-convex–strongly-concave functions, respectively, are not optimal for this specific case. Examples in the next section also present that the restarting can be useful even without strong monotonicity (but possibly with local strong monotonicity).

6 Applications of the accelerated proximal point method

As mentioned earlier, the proximal point method for maximally monotone operators include various well-known convex optimization methods. These include the augmented Lagrangian (i.e., the method of multipliers), the proximal method of mul-
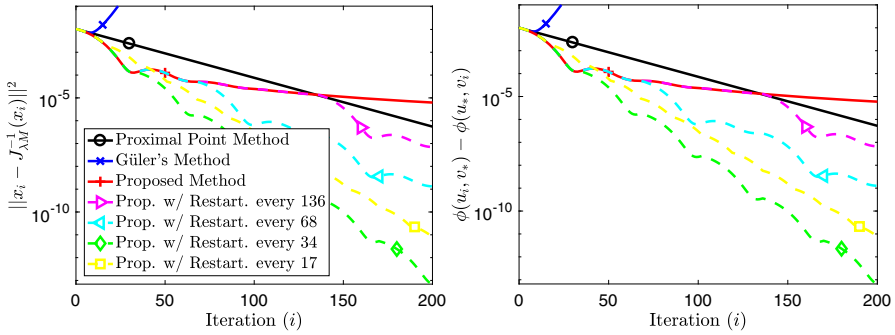
Fig. 2 Solving a strongly monotone inclusion problem with M (36); (left) the fixed-point residual versus iteration, (right) the function residual versus iteration

tipliers, and ADMM. The augmented Lagrangian method is equivalent to the proximal point method directly solving the dual convex minimization problem [58], so Güler's methods [34] already provide acceleration, whereas other instances of the proximal point method have no known accelerations yet. Thus, this section introduces accelerations to well-known instances of the proximal point method, which were not possible previously to the best of our knowledge (under this paper's setting).

6.1 Accelerating the proximal point method for convex–concave saddle-point problem

This section considers a convex–concave saddle-point problem (30), where the associated saddle subdifferential operator (31) is monotone. [59] applied the proximal point method on such operator to solve the convex–concave saddle-point problem, and this section further applies the proposed acceleration to such proximal point method as below.

Accelerated Proximal Point Method for Convex–Concave Saddle-Point Problem

Input: $\phi(\cdot, \boldsymbol{v}) \in \mathcal{F}(\mathcal{H}_1)$, $-\phi(\boldsymbol{u}, \cdot) \in \mathcal{F}(\mathcal{H}_2)$, $\hat{\boldsymbol{u}}_0 \in \mathcal{H}_1$, $\hat{\boldsymbol{v}}_0 \in \mathcal{H}_2$,

$\quad \boldsymbol{x}_0 = \boldsymbol{y}_0 = \boldsymbol{y}_{-1} = (\hat{\boldsymbol{u}}_0, \hat{\boldsymbol{v}}_0)$, $\lambda \in \mathbb{R}_{++}$.

For $i = 0, 1, \ldots$

$$\boldsymbol{x}_{i+1} = (\boldsymbol{u}_{i+1}, \boldsymbol{v}_{i+1}) = \arg \min_{\boldsymbol{u} \in \mathcal{H}_1} \max_{\boldsymbol{v} \in \mathcal{H}_2} \left\{ \phi(\boldsymbol{u}, \boldsymbol{v}) + \frac{1}{2\lambda} ||\boldsymbol{u} - \hat{\boldsymbol{u}}_i||^2 - \frac{1}{2\lambda} ||\boldsymbol{v} - \hat{\boldsymbol{v}}_i||^2 \right\},$$

$$\boldsymbol{y}_{i+1} = (\hat{\boldsymbol{u}}_{i+1}, \hat{\boldsymbol{v}}_{i+1}) = \boldsymbol{x}_{i+1} + \frac{i}{i+2} (\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) - \frac{i}{i+2} (\boldsymbol{x}_i - \boldsymbol{y}_{i-1}).$$

One primary use of this accelerated method is the following convex–concave Lagrangian problem

$$\min_{\boldsymbol{u} \in \mathcal{H}_1} \max_{\boldsymbol{v} \in \mathcal{H}_2} \{ L(\boldsymbol{u}, \boldsymbol{v}) := f(\boldsymbol{u}) + \langle \boldsymbol{v}, \, A\boldsymbol{u} - \boldsymbol{b} \rangle \}, \tag{37}$$

associated with the linearly constrained problem

$$\min_{\boldsymbol{u} \in \mathcal{H}_1} \ f(\boldsymbol{u})$$
$$\text{subject to} \ \ \boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}, \tag{38}$$

where $\boldsymbol{A} \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ and $\boldsymbol{b} \in \mathcal{H}_2$. The resulting method is called the proximal method of multipliers in [58], and applying the proposed acceleration to this method leads to below.

Accelerated Proximal Method of Multipliers

Input: $\phi(\cdot, \boldsymbol{v}) \in \mathcal{F}(\mathcal{H}_1)$, $-\phi(\boldsymbol{u}, \cdot) \in \mathcal{F}(\mathcal{H}_2)$, $\hat{\boldsymbol{u}}_0 \in \mathcal{H}_1$, $\hat{\boldsymbol{v}}_0 \in \mathcal{H}_2$,

$\quad\quad \boldsymbol{x}_0 = \boldsymbol{y}_0 = \boldsymbol{y}_{-1} = (\hat{\boldsymbol{u}}_0, \hat{\boldsymbol{v}}_0)$, $\lambda \in \mathbb{R}_{++}$.

For $i = 0, 1, \dots$

$$\boldsymbol{u}_{i+1} = \arg \min_{\boldsymbol{u} \in \mathcal{H}_1} \left\{ L(\boldsymbol{u}, \hat{\boldsymbol{v}}_i) + \frac{\lambda}{2} ||\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}||^2 + \frac{1}{2\lambda} ||\boldsymbol{u} - \hat{\boldsymbol{u}}_i||^2 \right\},$$

$$\boldsymbol{x}_{i+1} = (\boldsymbol{u}_{i+1}, \hat{\boldsymbol{v}}_i + \lambda(\boldsymbol{A}\boldsymbol{u}_{i+1} - \boldsymbol{b})),$$

$$\boldsymbol{y}_{i+1} = (\hat{\boldsymbol{u}}_{i+1}, \hat{\boldsymbol{v}}_{i+1}) = \boldsymbol{x}_{i+1} + \frac{i}{i+2}(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i) - \frac{i}{i+2}(\boldsymbol{x}_i - \boldsymbol{y}_{i-1}).$$

Note that this method without the acceleration and the term $\frac{1}{2\lambda} ||\boldsymbol{u} - \boldsymbol{u}_i||^2$ reduces to the augmented Lagrangian method. This method has an advantage over the augmented Lagrangian method and its accelerated variants; the primal iterate $\boldsymbol{u}_{i+1}$ is uniquely defined with a better conditioning.

Example 6.1 We apply the accelerated proximal method of multipliers to a basis pursuit problem

$$\min_{\boldsymbol{u} \in \mathbb{R}^{d_1}} \ ||\boldsymbol{u}||_1$$
$$\text{subject to} \ \ \boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}, \tag{39}$$

where $\boldsymbol{A} \in \mathbb{R}^{d_2 \times d_1}$ and $\boldsymbol{b} \in \mathbb{R}^{d_2}$. In the experiment, we choose $d_1 = 100$, $d_2 = 20$, and randomly generated $\boldsymbol{A}$. A true sparse $\boldsymbol{u}_{\text{true}}$ is randomly generated followed by a thresholding to sparsify nonzero elements, and $\boldsymbol{b}$ is then given by $\boldsymbol{A}\boldsymbol{u}_{\text{true}}$. We run 100 iterations of the proximal method of multipliers and its variants with $\lambda = 0.01$ and initial $\boldsymbol{x}_0 = \boldsymbol{0}$. Since the $\boldsymbol{u}_{i+1}$-update does not have a closed form, we used a sufficient number of iterations to solve the $\boldsymbol{u}_{i+1}$-update using the strongly convex version of FISTA [8] in [12, Theorem 4.10].

Figure 3 compares the proximal method of multipliers and its accelerated variants. Similar to Fig. 1, Güler's first accelerated version diverges, while the proposed method has accelerating behavior, compared to the non-accelerated version. The proposed method exhibits an oscillation in Fig. 3 (and a subtle oscillation in Fig. 1). This might be due to *high momentum*, owing from the acceleration, discussed in [54]. So in Fig. 3
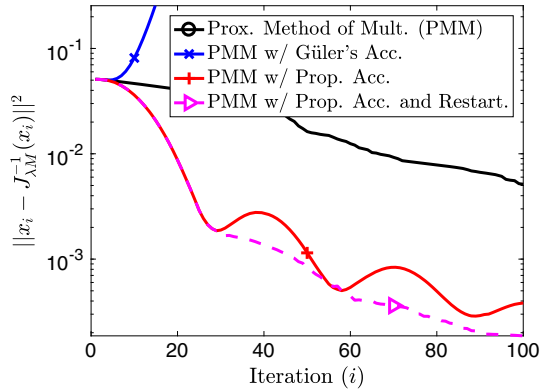
we heuristically restarted the method every 30 iterations to avoid such oscillation and accelerate, as suggested in [54]. Developing an approach to appropriately choosing a restarting interval or adaptively restarting the method as in [54] for such problem are left as future work.[3]

6.2 Accelerating the primal–dual hybrid gradient method

This section considers a linearly coupled convex–concave saddle-point problem

$$\min_{\boldsymbol{u} \in \mathcal{H}_1} \max_{\boldsymbol{v} \in \mathcal{H}_2} \left\{ \phi(\boldsymbol{u}, \boldsymbol{v}) \equiv f(\boldsymbol{u}) + \langle \boldsymbol{K}\boldsymbol{u}, \boldsymbol{v} \rangle - g(\boldsymbol{v}) \right\}, \tag{40}$$

where $f \in \mathcal{F}(\mathcal{H}_1)$, $g \in \mathcal{F}(\mathcal{H}_2)$ and $\boldsymbol{K} \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$. One widely known method for such problem is the primal–dual hybrid gradient (PDHG) method [11,25], which is a *preconditioned* proximal point method (with $\lambda = 1$) for the saddle subdifferential operator of ϕ (31) [13,35]. The associated preconditioner is

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\tau}\boldsymbol{I} & -\boldsymbol{K}^* \\ -\boldsymbol{K} & \frac{1}{\sigma}\boldsymbol{I} \end{bmatrix}, \tag{41}$$

which is positive definite when $\tau\sigma\|\boldsymbol{K}\|^2 < 1$, where $\|\boldsymbol{K}\| = \sup_{\|\boldsymbol{x}\| \leq 1} \|\boldsymbol{K}\boldsymbol{x}\|$. As mentioned in remark 2.1, we can directly apply our results to the PDHG method as below.

---

[3] We found that adaptively restarting the method when the fixed-point residual increases seems to be a good option in practice.

Accelerated PDHG Method

Input: $f \in \mathcal{F}(\mathcal{H}_1),\ g \in \mathcal{F}(\mathcal{H}_2),\ K \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2),\ \hat{u}_0 \in \mathcal{H}_1,\ \hat{v}_0 \in \mathcal{H}_2,\ \tau\sigma||K||^2 < 1,$
$\quad x_0 = y_0 = y_{-1} = (\hat{u}_0, \hat{v}_0).$

For $i = 0, 1, \ldots$

$$u_{i+1} = \underset{u \in \mathcal{H}_1}{\arg\min} \left\{ f(u) + \langle Ku, \hat{v}_i \rangle + \frac{1}{2\tau} ||u - \hat{u}_i||^2 \right\}$$

$$v_{i+1} = \underset{v \in \mathcal{H}_2}{\arg\min} \left\{ g(v) - \langle K(2u_{i+1} - \hat{u}_i), v \rangle + \frac{1}{2\sigma} ||v - \hat{v}_i||^2 \right\}$$

$$x_{i+1} = (u_{i+1}, v_{i+1})$$

$$y_{i+1} = (\hat{u}_{i+1}, \hat{v}_{i+1}) = x_{i+1} + \frac{i}{i+2}(x_{i+1} - x_i) - \frac{i}{i+2}(x_i - y_{i-1})$$

Corollary 6.1 *Assume that* $\langle P(x_0 - x_*),\ x_0 - x_* \rangle \leq R^2$ *for some* $x_* \in X_*(M)$. *The PDHG method satisfies*

$$\langle P(x_i - x_{i-1}),\ x_i - x_{i-1} \rangle \leq \left(1 - \frac{1}{i} \right)^{i-1} \frac{R^2}{i},$$

and the proposed accelerated PDHG method satisfies

$$\langle P(x_i - y_{i-1}),\ x_i - y_{i-1} \rangle \leq \frac{R^2}{i^2}.$$

Example 6.2 We apply the accelerated PDHG method to the bilinear game problem

$$\min_{u \in \mathbb{R}^{d_1}}\ \max_{v \in \mathbb{R}^{d_2}}\ \langle a, u \rangle + \langle Ku, v \rangle - \langle b, v \rangle, \tag{42}$$

where $K \in \mathbb{R}^{d_2 \times d_1}, a \in \mathbb{R}^{d_1}$ and $b \in \mathbb{R}^{d_2}$. The main part of the corresponding method is as below:

$$u_{i+1} = \hat{u}_i - \tau(K^*\hat{v}_i + a)$$
$$v_{i+1} = \hat{v}_i + \sigma(K(2u_{i+1} - \hat{u}_i) - b). \tag{43}$$

In the experiment, we choose $d_1 = 1000, d_2 = 500$, and a matrix K and vectors a, b are randomly generated. We run 100 iterations of the PDHG method and its variants with initial $\hat{u}_0 = [10\ \cdots\ 10]^\top$, $\hat{v}_0 = [10\ \cdots\ 10]^\top$ and $\tau = \sigma = \frac{0.99}{||K||}$. Figure 4 plots the *preconditioned* fixed-point residual, where Güler's first accelerated method diverges. The PDHG method and its proposed accelerated variant are comparable in this experiment, and heuristically restarting the accelerated method every 10 iterations yields a big acceleration. While [11] found restarting (reinitializing) a relaxed PDHG method not useful, our experiment suggests that restarting can be effective in some practical cases.

Fig. 4 Solving a bilinear game problem (42) by a *preconditioned* proximal point method; the *preconditioned* fixed-point residual versus iteration. $\tilde{\boldsymbol{J}}_{\boldsymbol{M},\boldsymbol{P}}:=(\boldsymbol{P}+\boldsymbol{M})^{-1}\boldsymbol{P}$ denotes the *preconditioned* resolvent operator
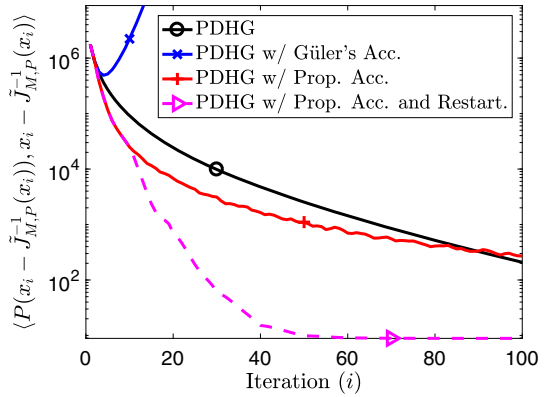
6.3 Accelerating the Douglas–Rachford splitting method

This section considers a monotone inclusion problem in a form

$$\text{Find } \boldsymbol{x} \in \mathcal{H} \text{ subject to } \boldsymbol{0} \in (\boldsymbol{M}_1 + \boldsymbol{M}_2)\boldsymbol{x} \tag{44}$$

for $\boldsymbol{M}_1, \boldsymbol{M}_2 \in \mathcal{M}(\mathcal{H})$, where $\boldsymbol{J}_{\rho \boldsymbol{M}_1}$ and $\boldsymbol{J}_{\rho \boldsymbol{M}_2}$ are more efficient than $\boldsymbol{J}_{\rho(\boldsymbol{M}_1+\boldsymbol{M}_2)}$ for a positive real number ρ. For such problem, the Douglas–Rachford splitting method [19,44] that iteratively applies the operator

$$\boldsymbol{G}_{\rho,\boldsymbol{M}_1,\boldsymbol{M}_2}:=\boldsymbol{J}_{\rho \boldsymbol{M}_1}\circ(2\boldsymbol{J}_{\rho \boldsymbol{M}_2} - \boldsymbol{I}) + (\boldsymbol{I} - \boldsymbol{J}_{\rho \boldsymbol{M}_2}) \tag{45}$$

has been found to be effective in many applications including ADMM, which we discuss in the next section.

In [24, Theorem 4], the Douglas–Rachford operator (45) was found to be a resolvent $\boldsymbol{J}_{\boldsymbol{M}_{\rho,\boldsymbol{M}_1,\boldsymbol{M}_2}}$ of a maximally monotone operator

$$\boldsymbol{M}_{\rho,\boldsymbol{M}_1,\boldsymbol{M}_2}:=\boldsymbol{G}_{\rho,\boldsymbol{M}_1,\boldsymbol{M}_2}^{-1} - \boldsymbol{I}. \tag{46}$$

In other words, the Douglas–Rachford splitting method is an instance of the proximal point method (with $\lambda = 1$) as

$$\boldsymbol{v}_{i+1} = \boldsymbol{J}_{\boldsymbol{M}_{\rho,\boldsymbol{M}_1,\boldsymbol{M}_2}}(\boldsymbol{v}_i) = \boldsymbol{G}_{\rho,\boldsymbol{M}_1,\boldsymbol{M}_2}(\boldsymbol{v}_i) \tag{47}$$

for $i = 0, 1, \ldots$. Therefore, we can apply the proposed acceleration to the Douglas–Rachford splitting method as below.

> **Accelerated Douglas–Rachford Splitting Method**
>   Input: $\boldsymbol{M}_1, \boldsymbol{M}_2 \in \mathcal{M}(\mathcal{H})$, $\boldsymbol{v}_0 = \boldsymbol{\eta}_0 = \boldsymbol{\eta}_{-1} \in \mathcal{H}$, $\rho \in \mathbb{R}_{++}$.
>   For $i = 0, 1, \ldots$
> $$\boldsymbol{v}_{i+1} = \boldsymbol{G}_{\rho, \boldsymbol{M}_1, \boldsymbol{M}_2}(\boldsymbol{\eta}_i)$$
> $$\boldsymbol{\eta}_{i+1} = \boldsymbol{v}_{i+1} + \frac{i}{i+2}(\boldsymbol{v}_{i+1} - \boldsymbol{v}_i) - \frac{i}{i+2}(\boldsymbol{v}_i - \boldsymbol{\eta}_{i-1})$$

Using (8) and (27), we have the following worst-case rates for the Douglas–Rachford splitting method and its accelerated variant. Finding exact bounds for the Douglas–Rachford splitting method and its variant is left as future work; [61] used PEP to analyze the exact worst-case rate of Douglas–Rachford splitting method under some additional conditions.

Corollary 6.2 *Assume that* $||\boldsymbol{v}_0 - \boldsymbol{v}_*|| \leq R$ *for some* $\boldsymbol{v}_* \in X_*(\boldsymbol{M}_{\rho, \boldsymbol{M}_1, \boldsymbol{M}_2})$. *The Douglas–Rachford splitting method satisfies*

$$||\boldsymbol{v}_i - \boldsymbol{\eta}_{i-1}||^2 \leq \left(1 - \frac{1}{i}\right)^{i-1} \frac{R^2}{i}, \tag{48}$$

and the proposed accelerated Douglas–Rachford splitting method satisfies

$$||\boldsymbol{v}_i - \boldsymbol{\eta}_{i-1}||^2 \leq \frac{R^2}{i^2}. \tag{49}$$

[23,24] illustrated that ADMM is equivalent to the Douglas–Rachford splitting method on the dual problem, so we naturally develop an accelerated ADMM in the next section and provide numerical experiment of the accelerated ADMM and thus the accelerated Douglas–Rachford splitting method.

6.4 Accelerating the alternating direction method of multipliers (ADMM)

Let $\mathcal{H}_1, \mathcal{H}_2, \mathcal{G}$ be real Hilbert spaces equipped with inner product $\langle \cdot, \cdot \rangle$. This section considers a linearly constrained convex problem

$$\min_{\boldsymbol{x} \in \mathcal{H}_1, \boldsymbol{z} \in \mathcal{H}_2} f(\boldsymbol{x}) + g(\boldsymbol{z})$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{c}, \tag{50}$$

where $f \in \mathcal{F}(\mathcal{H}_1)$, $g \in \mathcal{F}(\mathcal{H}_2)$, $\boldsymbol{A} \in \mathcal{B}(\mathcal{H}_1, \mathcal{G})$, $\boldsymbol{B} \in \mathcal{B}(\mathcal{H}_2, \mathcal{G})$ and $\boldsymbol{c} \in \mathcal{G}$. Its dual problem is

$$\max_{\boldsymbol{v} \in \mathcal{G}} \left\{ -f^*(-\boldsymbol{A}^*\boldsymbol{v}) - g^*(-\boldsymbol{B}^*\boldsymbol{v}) + \langle \boldsymbol{c}, \boldsymbol{v} \rangle \right\}, \tag{51}$$

where $f^*(y):=\sup_{x\in\mathcal{H}_1}\{\langle y, x\rangle - f(x)\}$ and $g^*(y):=\sup_{z\in\mathcal{H}_2}\{\langle y, z\rangle - g(z)\}$ are the conjugate functions of f and g, respectively. The dual problem (51) is equivalent to the following monotone inclusion problem

$$\text{Find} \quad v \in \mathcal{G} \quad \text{subject to} \quad \mathbf{0} \in -A\partial f^*(-A^*v) - B\partial g^*(-B^*v) - c. \qquad (52)$$

We next use the connection between ADMM for solving (50) and the Douglas–Rachford splitting method for solving (52) in [17, Proposition 9] [60] to develop an accelerated ADMM, using the accelerated Douglas–Rachford splitting method in the previous section.

Denoting

$$M_1:= -A\partial f^*(-A^*\cdot) - c \quad \text{and} \quad M_2:= -B\partial g^*(-B^*\cdot) \qquad (53)$$

converts the problem (52) into a form of the monotone inclusion problem (44). Then we use the following equivalent form of the accelerated Douglas–Rachford splitting method to solve (44) with (53):

$$\begin{aligned}
\boldsymbol{\zeta}_{i+1} &= J_{\rho M_2}(\boldsymbol{\eta}_i) \\
\boldsymbol{\xi}_{i+1} &= J_{\rho M_1}(2\boldsymbol{\zeta}_{i+1} - \boldsymbol{\eta}_i) \\
\boldsymbol{v}_{i+1} &= \boldsymbol{\eta}_i + (\boldsymbol{\xi}_{i+1} - \boldsymbol{\zeta}_i) \\
\boldsymbol{\eta}_{i+1} &= \boldsymbol{v}_{i+1} + \frac{i}{i+2}(\boldsymbol{v}_{i+1} - \boldsymbol{v}_i) - \frac{i}{i+2}(\boldsymbol{v}_i - \boldsymbol{\eta}_{i-1})
\end{aligned} \qquad (54)$$

for $i = 0, 1, \ldots$. Replacing the resolvent operators of M_1 and M_2 in (53) by minimization steps yields

$$\begin{aligned}
z_{i+1} &= \underset{z\in H_2}{\arg\min}\left\{g(z) + \langle\boldsymbol{\eta}_i, Bz\rangle + \frac{\rho}{2}||Bz||^2\right\} \\
\boldsymbol{\zeta}_{i+1} &= \boldsymbol{\eta}_i + \rho Bz_{i+1} \\
\tilde{x}_{i+1} &= \underset{x\in H_1}{\arg\min}\left\{f(x) + \langle\boldsymbol{\eta}_i + 2\rho Bz_{i+1}, Ax - c\rangle + \frac{\rho}{2}||Ax - c||^2\right\} \\
\boldsymbol{\xi}_{i+1} &= \boldsymbol{\eta}_i + \rho(A\tilde{x}_{i+1} - c) + 2\rho Bz_{i+1} \\
\boldsymbol{v}_{i+1} &= \boldsymbol{\eta}_i + \rho(A\tilde{x}_{i+1} + Bz_{i+1} - c) \\
\boldsymbol{\eta}_{i+1} &= \boldsymbol{v}_{i+1} + \frac{i}{i+2}(\boldsymbol{v}_{i+1} - \boldsymbol{v}_i) - \frac{i}{i+2}(\boldsymbol{v}_i - \boldsymbol{\eta}_{i-1}).
\end{aligned} \qquad (55)$$

By discarding $\boldsymbol{\zeta}_i$ and $\boldsymbol{\xi}_i$, and defining

$$\hat{v}_i:=v_i - \rho(A\tilde{x}_i - c) \quad \text{and} \quad \hat{\boldsymbol{\eta}}_i:=\boldsymbol{\eta}_i - \rho(A\tilde{x}_i - c), \qquad (56)$$

for $i = 0, 1, \ldots$, we have

$$z_{i+1} = \underset{z\in\mathcal{H}_2}{\arg\min}\left\{g(z) + \langle\hat{\boldsymbol{\eta}}_i + \rho(A\tilde{x}_i - c), Bz\rangle + \frac{\rho}{2}||Bz||^2\right\}$$

$$= \arg\min_{z \in \mathcal{H}_2} \left\{ g(z) + \langle \hat{\eta}_i, \ A\tilde{x}_i + Bz - c \rangle + \frac{\rho}{2} ||A\tilde{x}_i + Bz - c||^2 \right\}$$

$$\tilde{x}_{i+1} = \arg\min_{x \in \mathcal{H}_1} \left\{ f(x) + \langle \hat{v}_{i+1} + \rho Bz_{i+1}, \ Ax - c \rangle + \frac{\rho}{2} ||Ax - c||^2 \right\}$$

$$= \arg\min_{x \in \mathcal{H}_1} \left\{ f(x) + \langle \hat{v}_{i+1}, \ Ax + Bz_{i+1} - c \rangle + \frac{\rho}{2} ||Ax + Bz_{i+1} - c||^2 \right\}$$

$$\hat{v}_{i+1} = \hat{\eta}_i + \rho(A\tilde{x}_i + Bz_{i+1} - c)$$

$$\hat{\eta}_{i+1} = \hat{v}_{i+1} + \frac{i}{i+2}(\hat{v}_{i+1} - \hat{v}_i + \rho A(\tilde{x}_{i+1} - \tilde{x}_i))$$

$$\qquad - \frac{i}{i+2}(\hat{v}_i - \hat{\eta}_{i-1} + \rho A(\tilde{x}_i - \tilde{x}_{i-1})). \tag{57}$$

Then, replacing $\tilde{x}_i$ by x_{i+1} and reordering steps appropriately yield the following accelerated version of ADMM, which reduces to the standard ADMM when we let $\hat{\eta}_i = \hat{v}_i$ for $i = 0, 1, \ldots$.

Accelerated Alternating Direction Method of Multipliers

Input: $f \in \mathcal{F}(\mathcal{H}_1)$, $g \in \mathcal{F}(\mathcal{H}_2)$, $A \in \mathcal{B}(\mathcal{H}_1, \mathcal{G})$, $B \in \mathcal{B}(\mathcal{H}_2, \mathcal{G})$, $x_0 \in \mathcal{H}_1$, $z_0 \in \mathcal{H}_2$, $\hat{v}_0 \in \mathcal{G}$, $\rho \in \mathbb{R}_{++}$.

For $i = 0, 1, \ldots$

$$x_{i+1} = \arg\min_{x \in \mathcal{H}_1} \left\{ f(x) + \langle \hat{v}_i, \ Ax + Bz_i - c \rangle + \frac{\rho}{2} ||Ax + Bz_i - c||^2 \right\}$$

$$\hat{\eta}_i = \begin{cases} \hat{v}_i & i = 0, 1, \\ \hat{v}_i + \frac{i-1}{i+1}(\hat{v}_i - \hat{v}_{i-1} + \rho A(x_{i+1} - x_i)) - \frac{i-1}{i+1}(\hat{v}_{i-1} - \hat{\eta}_{i-2} + \rho A(x_i - x_{i-1})), & i = 2, 3, \ldots \end{cases}$$

$$z_{i+1} = \arg\min_{z \in \mathcal{H}_2} \left\{ g(z) + \langle \hat{\eta}_i, \ Ax_{i+1} + Bz - c \rangle + \frac{\rho}{2} ||Ax_{i+1} + Bz - c||^2 \right\}$$

$$\hat{v}_{i+1} = \hat{\eta}_i + \rho(Ax_{i+1} + Bz_{i+1} - c)$$

Since

$$v_i - \eta_{i-1} = \hat{v}_i - \hat{\eta}_{i-1} - \rho(Ax_{i+1} - Ax_i) = \rho(Ax_{i+1} + Bz_i - c), \tag{58}$$

we have the following worst-case rates with respect to the infeasibility for ADMM and its accelerated version, using (8) and (27).

Corollary 6.3 *Assume that* $||\hat{v}_0 + \rho A(x_0 - c) - v_*|| \leq R$ *for some* $v_* \in X_*(M_{\rho, -A\partial f^*(-A^* \cdot) - c, -B\partial g^*(-B^* \cdot)})$. *Alternating direction method of multipliers satisfies*

$$||Ax_{i+1} + Bz_i - c||^2 \leq \left(1 - \frac{1}{i}\right)^{i-1} \frac{R^2}{\rho^2 i}, \tag{59}$$

and the proposed accelerated alternating direction method of multipliers satisfies

$$||Ax_{i+1} + Bz_i - c||^2 \leq \frac{R^2}{\rho^2 i^2}. \tag{60}$$

The bound (59) is e-times asymptotically smaller than the known rate for ADMM in [17, Theorem 15], which originated from the bound (7). Finding exact bounds for the ADMM and its proposed variant is yet left as future work.

Remark 6.1 Many existing rates for the (preconditioned) ADMM consider the *ergodic* sequences $\{\bar{x}_i\}$ and $\{\bar{z}_i\}$, where $\bar{x}_i := \frac{1}{i} \sum_{l=1}^{i} x_l$ and $\bar{z}_i := \frac{1}{i} \sum_{l=1}^{i} z_l$ (see e.g., [11,13, 17,18]). In particular, in [17, Theorem 15], ADMM is found to satisfy

$$||A\bar{x}_{i+1} + B\bar{z}_i - c||^2 \leq \frac{16R^2}{\rho^2 i^2}, \tag{61}$$

which is faster than the rate of the *nonergodic* sequence $\{x_i, z_i\}$ of ADMM in (59) and is comparable to the rate of the proposed accelerated ADMM in (60). One should note that the feasibility convergence of the *ergodic* sequence, as in (61), does not necessarily imply the convergence of the fixed-point residual of the *ergodic* sequence, unlike (59) and (60) for the *nonergodic* sequence. In addition, some numerical experiments in [13] illustrate that the performance of the *nonergodic* sequence can be faster than that of the *ergodic* sequence. We leave further understanding the rates of the *ergodic* and *nonergodic* sequences of (preconditioned) ADMM and their relationship as future work.

Remark 6.2 [11,13,28] proposed accelerated variants of (preconditioned) ADMM under some additional conditions, while the proposed method does not require such conditions.

Example 6.3 We apply the accelerated ADMM to the problem

$$\min_{x \in \mathbb{R}^{d_1}, z \in \mathbb{R}^{d_2}} \frac{1}{2}||Hx - b||^2 + \gamma||z||_1$$
$$\text{subject to } Dx - z = 0, \tag{62}$$

with a positive real number γ, associated with the total-variation-regularized least-squares problem

$$\min_{x \in \mathbb{R}^{d_1}} \frac{1}{2}||Hx - b||^2 + \gamma||Dx||_1, \tag{63}$$

where $H \in \mathbb{R}^{p \times d_1}$, $b \in \mathbb{R}^p$, and a matrix $D \in \mathbb{R}^{d_2 \times d_1}$ is given as

$$D = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & 1 & -1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -1 \end{bmatrix}. \tag{64}$$

By letting $f(x) = \frac{1}{2}||Hx - b||^2$, $g(z) = \gamma||z||_1$, $A = D$, $B = -I$ and $c = 0$, we have the following accelerated ADMM method:

$$x_{i+1} = \arg\min_{x \in R^{d_1}} \left\{ \frac{1}{2}||Hx - b||^2 + \langle \hat{v}_i, Dx - z_i \rangle + \frac{\rho}{2}||Dx - z_i||^2 \right\}$$

$$= (H^\top H + \rho D^\top D)^{-1}(D^\top(\rho z_i - \hat{v}_i) + H^\top b)$$

$$\hat{\eta}_i = \begin{cases} \hat{v}_i, & i = 0, 1, \\ \hat{v}_i + \frac{i-1}{i+1}(\hat{v}_i - \hat{v}_{i-1} + \rho D(x_{i+1} - x_i)) - \frac{i-1}{i+1}(\hat{v}_{i-1} - \hat{\eta}_{i-2} + \rho D(x_i - x_{i-1})), & i = 2, 3, \dots \end{cases}$$

$$z_{i+1} = \arg\min_{z \in R^{d_2}} \left\{ \gamma||z||_1 + \langle \hat{\eta}_i, Dx_{i+1} - z \rangle + \frac{\rho}{2}||Dx_{i+1} - z||^2 \right\}$$

$$= S_{\frac{\gamma}{\rho}} \left(Dx_{i+1} + \frac{1}{\rho}\hat{\eta}_i \right)$$

$$\hat{v}_{i+1} = \hat{\eta}_i + \rho(Dx_{i+1} - z_{i+1}), \tag{65}$$

where the soft-thresholding operator is defined as $S_\tau(z) := \max\{|z| - \tau, 0\} \odot \text{sign}(z)$ with the element-wise absolute value, maximum and multiplication operators, $|\cdot|$, $\max\{\cdot, \cdot\}$ and $\odot$, respectively.

In the experiment, we choose $d_1 = 100$, $d_2 = 99$, $p = 5$, and a true vector x_{true} is constructed such that a vector Dx_{true} has few nonzero elements. A matrix H is randomly generated and a noisy vector b is generated by adding randomly generated (noise) vector to Hx_{true}. We choose the parameters $\gamma = 3$ and $\rho = 0.05$ in the experiment.

Figure 5 illustrates the fixed-point residual of ADMM and its accelerated variants. Interestingly, ADMM has a rate comparable to the $O(1/i^2)$ rate of the proposed method. This does not contradict with the theory, and we leave further investigating the worst-case rate of ADMM under the Lipschitz continuity condition of ∇f; similar analysis but under different conditions can be found in [17,18]. Noticing the oscillating behavior of the proposed ADMM in Fig. 5, we heuristically restarted the proposed method every 20 iterations, yielding a linear rate, without a strong monotonicity condition.[4] Restarting has been previously found useful for a different accelerated ADMM in [28].

7 Accelerated forward method for cocoercive operators

This section applies the proposed acceleration to a forward method, such as a gradient method, for cocoercive operators. A single-valued operator $M : \mathcal{H} \to \mathcal{H}$ is β-cocoercive for $\beta \in \mathbb{R}_{++}$ if

$$\langle x - y, Mx - My \rangle \geq \beta||Mx - My||^2 \text{ for all } x, y \in \mathcal{H}. \tag{66}$$

---

[4] Since ∇f is Lipschitz continuous, the operator $M_1 = -D\partial f^*(-D^\top \cdot)$ in (53) for the problem (62) is strongly monotone, but this is insufficient to guarantee a strong monotonicity of M_{ρ, M_1, M_2} (46) for the problem (62).

Fig. 5 Solving a total-variation-regularized least-squares problem (63); the fixed-point residual versus iteration
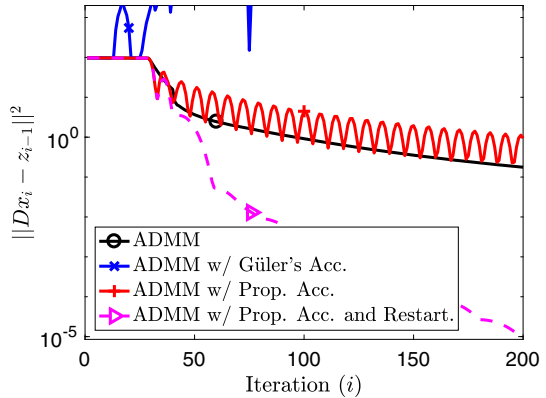
Let $\mathcal{C}_\beta(\mathcal{H})$ be the class of β-cocoercive operators on $\mathcal{H}$. For the β-cocoercive operator, the following forward method (that iteratively applies the forward operator $I - \beta M$) is guaranteed to converge weakly to a solution [7, Theorem 26.14].

Forward Method

Input: $M \in \mathcal{C}_\beta(\mathcal{H}), \ y_0 \in \mathcal{H}.$

For $i = 0, 1, \ldots$

$\quad y_{i+1} = (I - \beta M) y_i.$

An operator $T : \mathcal{H} \to \mathcal{H}$ is λ-cocoercive if and only if it is the Yosida approximation of index λ [7, Proposition 23.21]:

$$M_\lambda := \frac{1}{\lambda}(I - J_{\lambda M}) \tag{67}$$

of a maximally monotone operator $M : \mathcal{H} \to 2^{\mathcal{H}}$. We thus have the following equivalence between the resolvent (backward) operator of a maximally monotone operator M and a forward operator of the corresponding cocoercive operator M_λ:

$$J_{\lambda M} = (I + \lambda M)^{-1} = I - \lambda M_\lambda. \tag{68}$$

Therefore, the results on the proximal point method and its accelerated variant for monotone operators directly apply to the forward method and its accelerated variant for cocoercive operators.

8 Conclusion

This paper developed an accelerated proximal point method for maximally monotone operators, with respect to the fixed-point residual, using the computer-assisted performance estimation problem approach. Restarting technique was further employed under

the strong monotonicity condition. The proposed acceleration was applied to various instances of the proximal point method such as the proximal method of multipliers, the primal–dual hybrid gradient method, the Douglas–Rachford splitting method, and the alternating direction method of multipliers, yielding accelerations both theoretically and practically. The acceleration was also applied to a forward method for cocoercive operators.

We leave developing accelerations for more general or more specific classes of problems or methods as future work, possibly via the performance estimation problem approach; a comprehensive understanding of accelerations for the alternating direction method of multipliers with respect to various performance measures under various conditions are yet remain open.

References

1. Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set-Valued Anal. **9**(1–2), 3–11 (2001). https://doi.org/10.1023/A:1011253113155
2. Attouch, H., Cabot, A.: Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions. Appl. Math. Optim. **80**(3), 547–598 (2019). https://doi.org/10.1007/s00245-019-09584-z
3. Attouch, H., Cabot, A.: Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. Math. Program. **184**(1–2), 243–287 (2020). https://doi.org/10.1007/s10107-019-01412-0
4. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with Hessian driven damping. Math. Program. (2020). https://doi.org/10.1007/s10107-020-01591-1
5. Attouch, H., Chbani, Z., Riahi, H.: Fast proximal methods via time scaling of damped inertial dynamics. SIAM J. Optim. **29**(3), 2227–2256 (2019). https://doi.org/10.1137/18M1230207
6. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators. Math. Program. **174**(1–2), 391–432 (2019). https://doi.org/10.1007/s10107-018-1252-x
7. Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. Springer, Berlin (2011). https://doi.org/10.1007/978-1-4419-9467-7
8. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009). https://doi.org/10.1137/080716542
9. Brezis, H., Lions, P.L.: Produits infinis de resolvantes. Isr. J. Math. **29**(4), 329–345 (1978). https://doi.org/10.1007/BF02761171
10. Chambolle, A., Dossal, C.: On the convergence of the iterates of the "Fast iterative shrinkage/thresholding algorithm". J. Optim. Theory Appl. **166**(3), 968–82 (2015). https://doi.org/10.1007/s10957-015-0746-4
11. Chambolle, A., Pock, T.: A first-order primal–dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011). https://doi.org/10.1007/s10851-010-0251-1
12. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. Acta Numer. **25**, 161–319 (2016). https://doi.org/10.1017/S096249291600009X
13. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. Math. Program. **159**(1), 253–87 (2016). https://doi.org/10.1007/s10107-015-0957-3

14. Combettes, P.L.: Monotone operator theory in convex optimization. Math. Program. **170**(1), 177–206 (2018)
15. Corman, E., Yuan, X.: A generalized proximal point algorithm and its convergence rate. SIAM J. Optim. **24**(4), 1614–38 (2014). https://doi.org/10.1137/130940402
16. CVX Research Inc.: CVX: Matlab software for disciplined convex programming, version 2.0. http://cvxr.com/cvx (2012)
17. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. In: Glowinski, R., Osher, S., Yin, W. (eds.) Splitting Methods in Communication, Imaging, Science, and Engineering. Springer, Berlin (2016)
18. Davis, D., Yin, W.: Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions. Math. Oper. Res. **42**(3), 783–805 (2017). https://doi.org/10.1287/moor.2016.0827
19. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. **82**(2), 421–39 (1956)
20. Drori, Y., Taylor, A.B.: Efficient first-order methods for convex minimization: a constructive approach. Math. Program. **184**(1–2), 183–220 (2020). https://doi.org/10.1007/s10107-019-01410-2
21. Drori, Y., Teboulle, M.: Performance of first-order methods for smooth convex minimization: a novel approach. Math. Program. **145**(1–2), 451–82 (2014). https://doi.org/10.1007/s10107-013-0653-0
22. Drori, Y., Teboulle, M.: An optimal variant of Kelley's cutting-plane method. Math. Program. **160**(1), 321–51 (2016). https://doi.org/10.1007/s10107-016-0985-7
23. Eckstein, J.: The Lions–Mercier splitting algorithm and the alternating direction method are instances of the proximal point method (1988). Technical report LIDS-P-1769
24. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(1–3), 293–318 (1992). https://doi.org/10.1007/BF01581204
25. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**(4), 1015–46 (2010). https://doi.org/10.1137/09076934X
26. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. Comput. Math. Appl. **2**(1), 17–40 (1976). https://doi.org/10.1016/0898-1221(76)90003-1
27. Glowinski, R., Marrocco, A.: Sur lapproximation par elements nis dordre un, et la resolution par penalisation-dualite dune classe de problemes de dirichlet nonlineaires, rev. francaise daut. Inf. Rech. Oper. **R–2**, 41–76 (1975)
28. Goldstein, T., O'Donoghue, B., Setzer, S., Baraniuk, R.: Fast alternating direction optimization methods. SIAM J. Imaging Sci. **7**(3), 1588–623 (2014). https://doi.org/10.1137/120896219
29. Gol'shtein, E.G., Tret'yakov, N.V.: Modified Lagrangians in convex programming and their generalizations. In: Huard, P. (ed.) Point-to-Set Maps and Mathematical Programming, Mathematical Programming Studies 10. Springer, Berlin (1979)
30. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, pp. 95–110. Springer, Berlin (2008)
31. Gu, G., Yang, J.: On the optimal ergodic sublinear convergence rate of the relaxed proximal point algorithm for variational inequalities (2019). arXiv:1905.06030
32. Gu, G., Yang, J.: Tight sublinear convergence rate of the proximal point algorithm for maximal monotone inclusion problems. SIAM J. Optim. **30**(3), 1905–1921 (2020). https://doi.org/10.1137/19M1299049
33. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**(2), 403–19 (1991). https://doi.org/10.1137/0329022
34. Güler, O.: New proximal point algorithms for convex minimization. SIAM J. Optim. **2**(4), 649–64 (1992). https://doi.org/10.1137/0802032
35. He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. **5**(1), 119–49 (2012). https://doi.org/10.1137/100814494
36. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**(5), 303–20 (1969). https://doi.org/10.1007/BF00927673
37. Kim, D., Fessler, J.A.: Optimized first-order methods for smooth convex minimization. Math. Program. **159**(1), 81–107 (2016). https://doi.org/10.1007/s10107-015-0949-3

38. Kim, D., Fessler, J.A.: Another look at the Fast Iterative Shrinkage/Thresholding Algorithm (FISTA). SIAM J. Optim. **28**(1), 223–50 (2018). https://doi.org/10.1137/16M108940X

39. Kim, D., Fessler, J.A.: Generalizing the optimized gradient method for smooth convex minimization. SIAM J. Optim. **28**(2), 1920–50 (2018). https://doi.org/10.1137/17m112124x

40. Kim, D., Fessler, J.A.: Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions. J. Optim. Theory Appl. (2020). https://doi.org/10.1007/s10957-020-01770-2

41. Lieder, F.: On the convergence rate of the Halpern-iteration. Optim. Lett. **15**, 405–18 (2020). https://doi.org/10.1007/s11590-020-01617-9

42. Lieder, F.: Projection based methods for conic linear programming—optimal first order complexities and norm constrained quasi newton methods (Doctoral dissertation, Universitäts-und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf) (2018). https://docserv.uniduesseldorf.de/servlets/DerivateServlet/Derivate-49971/Dissertation.pdf

43. Lin, H., Mairal, J., Harchaoui, Z.: Catalyst acceleration for first-order convex optimization: from theory to practice. J. Mach. Learn. Res. **18**(212), 1–54 (2018)

44. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–79 (1979). https://doi.org/10.1137/0716071

45. Lorenz, D., Pock, T.: An inertial forward–backward algorithm for monotone inclusions. J. Math. Imaging Vis. **51**(2), 311–25 (2015). https://doi.org/10.1007/s10851-014-0523-2

46. Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Rev. Française Informat. Recherche Opérationnelle **4**, 154–8 (1970)

47. Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**(3), 341–6 (1962). https://doi.org/10.1215/S0012-7094-62-02933-2

48. Minty, G.J.: On the monotonicity of the gradient of a convex function. Pac. J. Math. **14**, 243–7 (1964)

49. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bulletin de la Société Mathématique de France **93**, 273–99 (1965)

50. Nemirovski, A.: Efficient methods in convex programming (1994). http://www2.isye.gatech.edu/~nemirovs/Lect_EMCO.pdf . visited on 05/2019

51. Nesterov, Y.: A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. Dokl. Akad. Nauk. USSR **269**(3), 543–7 (1983)

52. Nesterov, Y.: On an approach to the construction of optimal methods of minimization of smooth convex functions. Ekonomika i Mateaticheskie Metody **24**, 509–517 (1988). in Russian

53. Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1), 125–61 (2013). https://doi.org/10.1007/s10107-012-0629-5

54. O'Donoghue, B., Candes, E.: Adaptive restart for accelerated gradient schemes. Found. Comput. Math. **15**(3), 715–32 (2015). https://doi.org/10.1007/s10208-013-9150-3

55. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. **4**(5), 1–17 (1964). https://doi.org/10.1016/0041-5553(64)90137-5

56. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)

57. Rockafellar, R.T.: Monotone operators associated with saddle functions and minimax problems. In: Browder, F.E. (ed.) Nonlinear Functional Analysis, Part 1, vol. 18, pp. 397–407. American Mathematical Society, New York (1970)

58. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976). https://doi.org/10.1287/moor.1.2.97

59. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–98 (1976). https://doi.org/10.1137/0314056

60. Ryu, E.K., Boyd, S.: A primer on monotone operator methods. Appl. Comput. Math. **15**(1), 3–43 (2016)

61. Ryu, E.K., Taylor, A.B., Bergeling, C., Giselsson, P.: Operator splitting performance estimation: tight contraction factors and optimal parameter selection. SIAM J. Optim. **30**(3), 2251–2271 (2020). https://doi.org/10.1137/19M1304854

62. Ryu, E.K., Yin, W.: Large-scale convex optimization via monotone operators (2020). https://large-scale-book.mathopt.com/LSCOMO.pdf . visited on 03/2021

63. Shi, B., Du, S.S., Jordan, M.I., Su, W.J.: Understanding the acceleration phenomenon via high-resolution differential equations (2018). arXiv:1810.08907

64. Su, W., Boyd, S., Candes, E.J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. J. Mach. Learn. Res. **17**(153), 1–43 (2016)
65. Taylor, A.B., Bach, F.: Stochastic first-order methods: non-asymptotic and computer-aided analyses via potential functions. In: Proceedings of the Conference on Learning Theory, pp. 2934–2992 (2019)
66. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Exact worst-case performance of first-order methods for composite convex optimization. SIAM J. Optim. **27**(3), 1283–313 (2017). https://doi.org/10.1137/16m108104x
67. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Smooth strongly convex interpolation and exact worst-case performance of first-order methods. Math. Program. **161**(1), 307–45 (2017). https://doi.org/10.1007/s10107-016-1009-3