A SYMMETRIC EXTRAPOLATED PROXIMAL ALTERNATING PREDICTOR-CORRECTOR METHOD FOR SADDLE-POINT PROBLEMS. *

LIHAN ZHOU † AND FENG MA‡

Abstract. The proximal alternating predictor-corrector (PAPC) method is a widely used first-order algorithm for solving convex-concave saddle-point problems involving both smooth and non-smooth components. Unlike the primal-dual hybrid gradient (PDHG) method, which incorporates an extrapolation step with parameter $\theta \in (0, 1]$ to improve convergence, the existing convergence analysis of PAPC has been limited to the case $\theta = 1$. As a result, the behavior of PAPC under general extrapolation parameters remains largely unexplored. Moreover, despite the intrinsic symmetry between primal and dual variables in saddle-point formulations, the classical PAPC employs an asymmetric update scheme that does not exploit this structure. In this paper, we introduce a simple yet effective extrapolation step into the PAPC framework, resulting in a new method termed the symmetric proximal alternating predictor-corrector (SPAPC) algorithm. The proposed scheme incorporates extrapolation symmetrically in both primal and dual updates, yielding a balanced and structurally coherent iteration. We demonstrate that this modification allows for relaxed step-size conditions compared to those required by the classical PAPC method. Moreover, we establish the global convergence of SPAPC and provide both ergodic and nonergodic convergence rate guarantees. Numerical experiments further validate the improved efficiency of the proposed method.

Key words. Saddle-point problem, Lipschitz continuity, Primal–dual algorithms, Proximal alternating predictor-corrector, Symmetry.

MSC codes. 68U10, 90C25, 65K10, 65J22

1. Introduction.

In this paper, we consider the convex-concave saddle-point problem of the form

$$\min_{x\in\mathcal{R}^n}\max_{s\in\mathcal{R}^m}\{\varphi(x,s)=f(x)+s^TAx-h^*(s)\}, \tag{1.1}$$

where $A \in \mathcal{R}^{m\times n}$ is a given matrix, and $f : \mathcal{R}^n \to \mathcal{R}$ and $h : \mathcal{R}^m \to \mathcal{R}$ are convex, proper, and lower semicontinuous functions. Additionally, we assume that f has an L_f-Lipschitz continuous gradient, and h^* represents the Fenchel conjugate of h. The function h^* is assumed to be proximal tractable, meaning that its proximal operator (see (2.1)) can be efficiently computed. Such saddle-point problems frequently arise in optimization tasks in machine learning and image processing, including regularized regression, CT reconstruction and so on. We refer readers to e.g., [4, 11, 20, 24] and the references therein for further discussion.

A classical algorithm for solving (1.1) is the primal-dual hybrid gradient (PDHG) method, introduced by Chmabolle and Pock [3, 5]. It iterates as follows:

$$\begin{cases} x^{k+1} = \mathrm{Prox}_{(rf)}[x^k - rA^Ts^k], & \text{(1.2a)}\\ \bar{x}^{k+1} = x^{k+1} + \theta(x^{k+1} - x^k), & \text{(1.2b)}\\ s^{k+1} = \mathrm{Prox}_{(\delta h^*)}[s^k + \sigma A\bar{x}^{k+1}], & \text{(1.2c)}\end{cases}$$

where (1.2b) denotes a extrapolation step, $\theta \in (0, 1]$ denotes the extrapolation parameter, and $\mathrm{Prox}_{(rf)}$ and $\mathrm{Prox}_{(\delta h^*)}$ denote the proximal operators of rf and δh^*,

*This research was partially supported by the NSFC Grant 12171481 and 12471298.

†214611022@csu.edu.cn, School of Automation, Central South University, Changsha, 410083, Hunan, China.

‡mafengnju@gmail.com, High-Tech Institute of Xi'an, Xi'an, 710025, Shaanxi, China.

respectively (see (2.1)). By reformulating the PDHG within a proximal point algorithmic framework [18], the convergence condition of the PDHG with $\theta = 1$ can be obtained by

$$r\delta \leq \frac{1}{\|A^T A\|}. \tag{1.3}$$

In the PDHG scheme, each iteration involves two subproblems: updating the primal variable x and the dual variable s alternately. These subproblems often have closed-form solutions or allow efficient computation via proximal operators, making PDHG straightforward to implement, especially in the image process. However, when the objective function f is differentiable, solving the subproblem (1.2a) involves a matrix inversion. Consequently, the solution for the subproblem (1.2a) may not be attainable. To address this difficulty, several algorithms have been developed, including proximal alternating predictor-corrector (PAPC) algorithm [10], primal–dual fixed point algorithm based on the proximity operator (PDFP^{2}O$_k$) [6], the adaptive PDHG variant [28], proximal forward-backward splitting (PFBS) algorithm [8]. For further discussion of the connections between these algorithms, we refer readers to e.g., [7, 4, 29, 9, 25, 27]. To motivate our study, we present two representative algorithms: the PAPC method and the PDFP^{2}O$_k$ method.

The Proximal Alternating Predictor-Corrector (PAPC) method, introduced in [10], employs a gradient mapping for updating the primal variable x^{k+1}, thereby avoiding matrix inversion. The iteration scheme of PAPC is given by

$$\begin{cases} \tilde{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T s^k, & \text{(1.4a)} \\ s^{k+1} = \mathrm{Prox}_{\delta h^*}\left(s^k + \delta A\tilde{x}^{k+1}\right), & \text{(1.4b)} \\ x^{k+1} = x^k - r\nabla f(x^k) - rA^T s^{k+1}. & \text{(1.4c)} \end{cases}$$

The convergence of PAPC is guaranteed under the condition

$$r\delta\|AA^T\| < 1, \quad 0 < r < \frac{1}{L_f}, \tag{1.5}$$

where L_f is the Lipschitz constant of ∇f. This algorithm was initially presented in [22] in the context of ℓ_1-penalized least squares problems. It has since been independently rediscovered and appears under various names, including the Primal-Dual Fixed-Point algorithm based on the Proximity Operator (PDFP^{2}O) [6] and the PAPC algorithm [10]. An interpretation of PAPC as a primal-dual forward–backward splitting method was given in [9]. Unlike the PDHG method, PAPC performs only a single evaluation of the gradient of the smooth component and one proximal operation for the nonsmooth component per iteration. This feature makes it particularly attractive for large-scale problems where computational efficiency is critical. However, the theoretical convergence analysis imposes a stricter constraint on the step-size parameter r, which may limit the flexibility of the method in practical settings.

The PDFP^{2}O$_k$ algorithm, proposed in [6], extends the PAPC method by introducing two under-relaxation steps to improve stability and flexibility. The iterative scheme is given by:

$$\begin{cases} \tilde{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T s^k, & \text{(1.6a)} \\ \tilde{s}^{k+1} = \mathrm{Prox}_{(\delta h^*)}[s^k + \delta A\tilde{x}^{k+1}], & \text{(1.6b)} \\ \bar{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T \tilde{s}^{k+1}, & \text{(1.6c)} \\ s^{k+1} = ks^k + (1-k)\tilde{s}^{k+1}, & \text{(1.6d)} \\ x^{k+1} = kx^k + (1-k)\bar{x}^{k+1}, & \text{(1.6e)} \end{cases}$$

where $k \in [0,1]$ denotes a relaxation parameter. When $k = 0$, the PDFP^{2}O$_k$ reduces to the PAPC. By reformulating the scheme (1.6) within the fixed point algorithmic framework, the convergence can be proved under the condition

$$r\delta\|AA^T\| < 1, \quad 0 < r < \frac{2}{L_f}. \tag{1.7}$$

Compared to the PAPC convergence condition, this result relaxes the upper bound on the primal step size r, potentially improving convergence speed. However, it introduces an under-relaxation parameter k, which may affect practical performance depending on the choice of k.

Both the PAPC and PDFP^{2}O$_k$ algorithms employ alternating update schemes for primal and dual variables while avoiding the matrix inversion typically required in the x-subproblem, which significantly reduces computational complexity. However, these algorithms do not fully exploit the intrinsic symmetry of saddle-point problems. In particular, within each iteration, extrapolation is applied only to either the primal or the dual variable, but not both. To highlight this asymmetry, we reformulate the PAPC algorithm in an equivalent form:

$$\begin{cases} \tilde{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T s^k, & \text{(1.8a)} \\ s^{k+1} = \mathrm{Prox}_{\delta h^*}\left[s^k + \delta A\tilde{x}^{k+1}\right], & \text{(1.8b)} \\ x^{k+1} = \tilde{x}^{k+1} - rA^T(s^{k+1} - s^k). & \text{(1.8c)} \end{cases}$$

This scheme can be further expressed as:

$$\begin{cases} \bar{x}^{k+1} = \tilde{x}^{k+1}, & \text{(1.9a)} \\ \hat{s}^{k+1} = s^k + \delta A\bar{x}^{k+1}, & \text{(1.9b)} \\ s^{k+1} = \arg\min_s \left\{ h^*(s) + \frac{1}{2\delta}\|s - \hat{s}^{k+1}\|^2 \right\}, & \text{(1.9c)} \\ \bar{s}^{k+1} = s^{k+1} + (s^{k+1} - s^k), & \text{(1.9d)} \\ \hat{x}^{k+2} = \bar{x}^{k+1} - r\nabla f(x^{k+1}) - rA^T\bar{s}^{k+1}, & \text{(1.9e)} \\ \tilde{x}^{k+2} = \arg\min_x \left\{ \frac{1}{2r}\|x - \hat{x}^{k+2}\|^2 \right\}. & \text{(1.9f)} \end{cases}$$

As evident from steps (1.9a) and (1.9d), the update of $\bar{x}^{k+1}$ relies solely on the most recent iterate $\tilde{x}^{k+1}$ and does not utilize the preceding iterate x^k, whereas $\bar{s}^{k+1}$ explicitly incorporates both s^k and s^{k+1}. Thus, the dual update involves an extrapolation with factor 1, while the primal update lacks any form of extrapolation. This structural imbalance renders the PAPC inherently asymmetric. A similar asymmetry is also present in the PDFP^{2}O$_k$ algorithm.

However, several optimization algorithms for saddle-point problems have successfully exploited the inherent symmetry of such formulations by treating the primal and dual variables in a more balanced manner. Prominent examples include the Peaceman–Rachford splitting method [21, 26, 15] and various symmetric variants of the alternating direction method of multipliers (ADMM) [12, 17]. These methods leverage symmetric update structures, which have been shown to enhance convergence behavior and broaden the range of practical applicability. In particular, they often permit a wider range for the extrapolation parameter, and empirical studies suggest that such relaxed conditions yield improved numerical performance. Motivated by these observations, a natural question arises: *Can we design a symmetric* *optimization algorithm that applies extrapolation to both the primal and dual updates, while admitting a broad admissible range for the extrapolation parameter?*

In this work, we provide an affirmative answer to this question. Motivated by the symmetric PDHG algorithm introduced in [23], we extend this framework to address the problem (1.1). The resulting symmetric scheme incorporates extrapolation into both the primal and dual updates and is formally given by:

$$\begin{cases} \tilde{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T s^k, & \text{(1.10a)} \\ \bar{x}^{k+1} = \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - x^k), & \text{(1.10b)} \\ s^{k+1} = \mathrm{Prox}_{(\delta h^*)}\left[s^k + \delta A\bar{x}^{k+1}\right], & \text{(1.10c)} \\ x^{k+1} = x^k - (1+\theta)r\nabla f(x^k) - \theta r A^T s^k - rA^T s^{k+1}. & \text{(1.10d)} \end{cases}$$

This framework admits an equivalent form as follows

$$\text{(SPAPC)} \quad \begin{cases} \tilde{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T s^k, & \text{(1.11a)} \\ \bar{x}^{k+1} = \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - x^k), & \text{(1.11b)} \\ s^{k+1} = \mathrm{Prox}_{(\delta h^*)}\left[s^k + \delta A\bar{x}^{k+1}\right], & \text{(1.11c)} \\ x^{k+1} = \bar{x}^{k+1} - rA^T(s^{k+1} - s^k). & \text{(1.11d)} \end{cases}$$

For the detailed derivation process, we refer readers to Section 2.4. To further elucidate the structural properties of our method, we rewrite it in an alternative formulation:

$$\begin{cases} \bar{x}^{k+1} = \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - x^k), & \text{(1.12a)} \\ \hat{s}^{k+1} = s^k + \delta A\bar{x}^{k+1}, & \text{(1.12b)} \\ s^{k+1} = \arg\min_s \{h^*(s) + \frac{1}{2\delta}\|s - \hat{s}^{k+1}\|^2\}, & \text{(1.12c)} \\ \bar{s}^{k+1} = s^{k+1} + (s^{k+1} - s^k), & \text{(1.12d)} \\ \hat{x}^{k+2} = \bar{x}^{k+1} - r\nabla f(x^{k+1}) - rA^T\bar{s}^{k+1}, & \text{(1.12e)} \\ \tilde{x}^{k+2} = \arg\min_x \{\frac{1}{2r}\|x - \hat{x}^{k+2}\|^2\}. & \text{(1.12f)} \end{cases}$$

This algorithm applies coupled extrapolation to both the primal and dual variables in steps (1.12a) and (1.12d), resulting in a symmetric update scheme. The use of extrapolation on both sides not only accelerates the exchange of information between primal and dual iterates but also reflects the inherent symmetry of the saddle-point problem. As we will demonstrate through numerical experiments, this symmetry often leads to faster convergence in practice.

In this paper, we prove that the proposed symmetric proximal alternating predictor-corrector method (SPAPC), defined in (1.11), converges globally to a saddle point of (1.1) without requiring any additional assumptions. Specifically, our analysis demonstrates that the convergence is ensured when the following conditions are satisfied

$$r\delta\|AA^T\| < 1, \qquad 0 < r < \frac{4}{L_h}, \quad \theta \in (-1, 1 - \frac{rL_f}{2}). \tag{1.13}$$

Compared to PAPC and PDFP^{2}O$_k$, our method significantly enlarges the admissible range of the extrapolation parameter from $\theta = 0$ to $\theta \in \left(-1, 1 - \frac{rL_f}{2}\right)$, and relaxes the condition on r from $0 < r < \frac{2}{L_h}$ to $0 < r < \frac{4}{L_h}$, while retaining the same constraint on $r\delta$. This expanded parameter range offers two key advantages: (1) it enables more flexible adaptive tuning strategies that enhance both theoretical convergence guarantees and practical performance. (2) It preserves computational efficiency by eliminating matrix-vector multiplications and relying exclusively on vector addition operations, keeping the per-iteration complexity comparable to existing methods. Moreover, we establish both non-ergodic and ergodic convergence rates for SPAPC, providing a comprehensive theoretical analysis of its convergence behavior.

The rest of this paper is organized as follows: Section 2 gives some facts that are useful for further analysis. In particular, a prediction-correction interpretation for the SPAPC and a characterization of a solution point of the problem (1.1) are presented. Section 3 conducts the convergence analysis of the SPAPC, including its global convergence guarantee and convergence rate. Section 4 reports some results of numerical experiments to verify the theoretical results. Finally, Section 5 summarizes the conclusions.

2. Preliminaries.

In this section, we firstly establish the necessary notation and lemmas. Building upon them, we then derive the variational inequality formulation of the optimality conditions for the problem (1.1) and propose a prediction-correction framework for analyzing SPAPC. Finally, we give a characterization of a solution point of the problem (1.1) within the scheme (1.10).

2.1. Notation.

For a function f, its gradient and domain are, respectively, denoted by ∇f and $\mathbf{dom} f$. For vectors, $\|y\|_p$ represents the l_p norm of y, with $\|y\|$ reserved for the Euclidean norm (l_2) and $\|y\|_F$ reserved for the Frobenius norm. For matrices, $R \succ 0$ and $R \succeq 0$ indicate positive definiteness and positive semidefiniteness, respectively. For a nonsingular matrix R, $\lambda_{\min}(R)$ and $\lambda_{\max}(R)$ denote the maximum eigenvalue and the minimum eigenvalue of R, respectively. Given a symmetric matrix R and vector y, the weighted norm $\|y\|_R$ is defined as $\sqrt{y^T R y}$, while $\|R\|$ denotes the spectral norm (largest singular value). The proximal mapping of a convex, lower semicontinuous function is given by

$$\text{Prox}_{\tau f}(x) = \arg\min_u \{f(u) + \frac{1}{2\tau}\|u - x\|^2\}, \quad \forall x \in \Re^n. \tag{2.1}$$

Note that in this case, it always yields a unique solution.

In the following, we define some auxiliary variables and matrices, which facilitate the convergence analysis. The auxiliary variables are given by

$$v = \begin{pmatrix} x \\ s \end{pmatrix}, \quad \nabla\mathcal{F}(v) = \begin{pmatrix} A^T s \\ -Ax \end{pmatrix}, \quad \nabla\mathcal{G}(v) = \begin{pmatrix} A^T s + \nabla f(x) \\ -Ax \end{pmatrix}, \tag{2.2}$$

where v^{k+1} and $\tilde{v}^{k+1}$ are, respectively, defined as follows

$$v^{k+1} = \begin{pmatrix} x^{k+1} \\ s^{k+1} \end{pmatrix}, \quad \tilde{v}^{k+1} = \begin{pmatrix} \tilde{x}^{k+1} \\ \tilde{s}^{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{x}^{k+1} \\ s^{k+1} \end{pmatrix}. \tag{2.3}$$

The matrices are given by

$$Q = \begin{pmatrix} \frac{1}{r}I & -A^T \\ -\theta A & \frac{1}{\delta}I \end{pmatrix}, \tag{2.4}$$

$$M = \begin{pmatrix} (1+\theta)I & -rA^T \\ 0 & I \end{pmatrix}, \tag{2.5}$$

$$Q_f = \begin{pmatrix} \frac{L_f}{2} & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.6}$$

$$Q_f^{'} = \begin{pmatrix} L_f & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.7}$$

where L_f is a Lipschitz constant of the function f. Building upon the definition (2.4) and (2.5), we present some matrices

$$H = QM^{-1} = \begin{pmatrix} r\lambda I & -(1-\rho)A^T \\ -(1-\rho)A & \frac{1}{\delta}I - r(1-\rho)AA^T \end{pmatrix}, \tag{2.8}$$

$$G = Q + Q^T - M^T H M = \begin{pmatrix} \frac{1-\theta}{r}I & 0 \\ 0 & \frac{1}{\delta}I - rAA^T \end{pmatrix}, \tag{2.9}$$

$$G_f = G - Q_f = \begin{pmatrix} (\frac{1-\theta}{r} - \frac{L_f}{2})I & 0 \\ 0 & \frac{1}{\delta} - rAA^T \end{pmatrix}, \tag{2.10}$$

$$G_f^{'} = G - Q_f^{'} = \begin{pmatrix} (\frac{1-\theta}{r} - L_f)I & 0 \\ 0 & \frac{1}{\delta} - rAA^T \end{pmatrix}, \tag{2.11}$$

where $\rho = \frac{1}{1+\theta}$.

2.2. Some lemmas. In the section, we introduce some lemmas which are used to prove the convergence and derive the convergence rate.

LEMMA 2.1. *For any vector $x, y, z, p \in \mathcal{R}^n$, and a symmetric matrix $R \in \mathcal{R}^{n\times n}$, the following identity holds*

$$(x-y)^T R(z-p) = \frac{1}{2}\Big(\|x-p\|_R^2 - \|x-z\|_R^2 + \|y-z\|_R^2 - \|y-p\|_R^2\Big). \tag{2.12}$$

LEMMA 2.2. *For any vector $x, y \in \mathcal{R}^n$, and a symmetric matrix $R \in \mathcal{R}^{n\times n}$, the following identity holds*

$$\|x\|_R^2 - \|y\|_R^2 = 2x^T R(x-y) - \|x-y\|_R^2. \tag{2.13}$$

LEMMA 2.3. *[1] Let f be convex differentiable with L_f-Lipschitz gradient. Then $\forall x, y, z \in$ **dom**f:*

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geq \frac{1}{L_f}\|\nabla f(x) - \nabla f(y)\|^2. \tag{2.14}$$

LEMMA 2.4. *Let f be convex differentiable with L_f-Lipschitz gradient. Then $\forall x, y, z \in$ **dom**f:*

$$f(x) \leq f(y) + (x-y)^T \nabla f(z) + \frac{L_f}{2}\|x-z\|^2. \tag{2.15}$$

Proof. For any convex differentiable functions f with L_f-Lipschitz continuous gradient, we have

$$f(x) \leq f(z) + (x-z)^T \nabla f(z) + \frac{L_f}{2}\|x-z\|^2. \tag{2.16}$$

It follows from the property of the convex function that

$$f(z) \leq f(y) + (z-y)^T \nabla f(z). \tag{2.17}$$

Adding the inequalities (2.16) and (2.17) yields (2.15). ◻

LEMMA 2.5. *[16] Let $\mathbf{X} \in \mathcal{R}^n$ be a closed convex set. Given two convex functions $\mathcal{F}(x) : \mathbf{X} \to \mathcal{R}$ and $\mathcal{G}(x) : \mathbf{X} \to \mathcal{R}$ with $\nabla \mathcal{G}(x)$ differentiable. Suppose that the optimization problem $\min\{\mathcal{F}(x) + \mathcal{G}(x) | x \in \mathbf{X}\}$ has a non-empty solution set. Then, $x^* \in \mathbf{X}$ is optimal if and only if*

$$x^* \in \mathbf{X},\ \mathcal{F}(x) - \mathcal{F}(x^*) + (x-x^*)^T \nabla \mathcal{G}(x^*) \geq 0,\ \forall x \in \mathbf{X}. \tag{2.18}$$

2.3. Optimization condition in terms of variational inequality. In this section, we derive the variational inequality formulation of the optimality conditions for problem (1.1). We firstly revisit that (x^*, s^*) is called a saddle point for the problem (1.1), if it satisfies

$$\varphi(x^*, s) \leq \varphi(x^*, s^*) \leq \varphi(x, s^*).$$

Then, it follows from Lemma 2.5 that the first inequality is equivalent to

$$h^*(s) - h^*(s^*) + (s-s^*)^T(-Ax^*) \geq 0. \tag{2.19}$$

Similarly, the second inequality can be formulated as

$$f(x) - f(x^*) + (x-x^*)^T(A^T s^*) \geq 0. \tag{2.20}$$

Since f is differentiable, the inequality (2.20) can further reduce to

$$(x-x^*)^T(A^T s^* + \nabla f(x^*)) \geq 0. \tag{2.21}$$

Adding the inequalities (2.19) and (2.21), and using the notations in (2.2), we obtain the variational inequality of the optimization condition for the problem (1.1)

$$h^*(s) - h^*(s^*) + (v-v^*)^T \nabla \mathcal{G}(v^*) \geq 0. \tag{2.22}$$

2.4. A prediction-correction interpretation. In this section, we follow the approach outlined in, e.g., [14, 19], to reformulate the SPAPC scheme (1.10) within a prediction–correction framework. This reformulation reveals a structure akin to that of proximal point algorithms and facilitates a convenient convergence analysis.

First, substituting the step (1.10a) into the step (1.10d), we obtain an equivalent update of x^{k+1} as follows

$$\begin{aligned} x^{k+1} &= x^k - (1+\theta) r \nabla f(x^k) - \theta r A^T s^k - r A^T s^{k+1} \\ &= (1+\theta)\left(x^k - r\nabla f(x^k) - rA^T s^k\right) - \theta x^k - rA^T(s^{k+1} - s^k) \\ &= \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - x^k) - rA^T(s^{k+1} - s^k) \\ &= x^k + (1+\theta)(\tilde{x}^{k+1} - x^k) - rA^T(s^{k+1} - s^k). \end{aligned} \tag{2.23}$$

Combining the step (2.23) with the scheme (1.10), we get the equivalent form of the S-PAPC (1.11), i.e.,

$$\begin{cases} \tilde{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T s^k, \\ \bar{x}^{k+1} = \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - x^k), \\ s^{k+1} = \mathrm{Prox}_{(\delta h^*)}\big[s^k + \delta A\bar{x}^{k+1}\big], \\ x^{k+1} = x^k + (1+\theta)(\tilde{x}^{k+1} - x^k) - rA^T(s^{k+1} - s^k). \end{cases}$$

Note that in the following, we mainly study the properties of the scheme (1.11) rather than the scheme (1.10) for the sake of analysis convenience. The scheme (1.11) can be further split into two parts: prediction part and correction part. The prediction part is given by

$$\text{(Prediction)} \quad \begin{cases} \tilde{x}^{k+1} = x^k - r\nabla f(x^k) - rA^T s^k, \\ s^{k+1} = \mathrm{Prox}_{(\delta h^*)}\big[s^k + \delta A(\bar{x}^{k+1} + \theta(\tilde{x}^{k+1} - x^k)\big]. \end{cases}$$

The corresponding correction part is given by

$$\text{(Correction)} \quad \begin{cases} x^{k+1} = x^k + (1+\theta)(\tilde{x}^{k+1} - x^k) - rA^T(s^{k+1} - s^k), \\ s^{k+1} = s^k + (s^{k+1} - s^k). \end{cases}$$

We further simplify the prediction part and correction part. Applying Lemma 2.5 to the prediction part yields

$$\begin{cases} (x - \tilde{x}^{k+1})^T \frac{1}{r}(\tilde{x}^{k+1} - x^k + rA^T s^{k+1} + rp^k) \geq 0, & \text{(2.24a)} \\ h^*(s) - h^*(s^{k+1}) + (s - s^{k+1})^T \frac{1}{\delta}(s^{k+1} - s^k - \delta A\bar{x}^{k+1}) \geq 0, & \text{(2.24b)} \end{cases}$$

where $p^k = \nabla f(x^k) - A^T(s^{k+1} - s^k)$. Combining the inequalities (2.24a) and (2.24b) yields

$$\begin{aligned} h^*(s) - h^*(\tilde{s}^{k+1}) + (v - \tilde{v}^{k+1})^T \nabla\mathcal{F}(\tilde{v}^{k+1}) + (x - \tilde{x}^{k+1})^T \nabla f(x^k) \\ + (v - \tilde{v}^{k+1})^T Q(\tilde{v}^{k+1} - v^k) \geq 0, \end{aligned} \tag{2.25}$$

where Q is defined in (2.4), $\nabla\mathcal{F}(\cdot)$ is defined in (2.2) and $\tilde{v}^{k+1}$ is given by (2.3). This inequality can be reformulated as

$$\begin{aligned} (v - \tilde{v}^{k+1})^T Q(\tilde{v}^{k+1} - v^k) &\geq h^*(\tilde{s}^{k+1}) - h^*(s) + (\tilde{v}^{k+1} - v)^T \nabla\mathcal{F}(\tilde{v}^{k+1}) \\ &\quad + (\tilde{x}^{k+1} - x)^T \nabla f(x^k) \\ &= h^*(\tilde{s}^{k+1}) - h^*(s) + (\tilde{v}^{k+1} - v)^T \nabla\mathcal{F}(v) \\ &\quad + (\tilde{x}^{k+1} - x)^T \nabla f(x^k), \end{aligned} \tag{2.26}$$

where the last identity follows from the fact that $(\tilde{v}^{k+1} - v)^T \nabla\mathcal{F}(\tilde{v}^{k+1}) = (\tilde{v}^{k+1} - v)^T \nabla\mathcal{F}(v)$. For the last term of the inequality (2.26), we have

$$\begin{aligned} (\tilde{x}^{k+1} - x)^T \nabla f(x^k) =& (\tilde{x}^{k+1} - x)^T \nabla f(x) + (\tilde{x}^{k+1} - x)^T (\nabla f(x^k) - \nabla f(x)), \\ =& (\tilde{x}^{k+1} - x)^T \nabla f(x) + (\tilde{x}^{k+1} - x^k)^T (\nabla f(x^k) - \nabla f(x)) \\ &+ (x^k - x)^T (\nabla f(x^k) - \nabla f(x)). \end{aligned} \tag{2.27}$$

The inequality (2.27) can further reduce to

$$\begin{aligned}(\tilde{x}^{k+1}-x)^T\nabla f(x^k) \geq& (\tilde{x}^{k+1}-x)^T\nabla f(x) + (\tilde{x}^{k+1}-x^k)^T(\nabla f(x^k)-\nabla f(x))\\ &+\frac{1}{L_f}\|\nabla f(x^k)-\nabla f(x)\|^2\\ \geq& (\tilde{x}^{k+1}-x)^T\nabla f(x) + \frac{1}{L_f}\|\nabla f(x^k)-\nabla f(x)\|^2\\ &-\frac{L_f}{4}\|\tilde{x}^{k+1}-x^k\|^2 - \frac{1}{L_f}\|\nabla f(x^k)-\nabla f(x)\|^2,\end{aligned} \tag{2.28}$$

where the first inequality is derived from the Lemma 2.3 and the second inequality follows from the Cauchy-Schwarz inequality. Substituting (2.28) into the inequality (2.26), we obtain the simplified prediction part as follows

$$\begin{aligned}&\text{(Prediction)}\\ &(v-\tilde{v}^{k+1})^T Q(\tilde{v}^{k+1}-v^k) \geq h^*(\tilde{s}^{k+1}) - h^*(s) + (\tilde{v}^{k+1}-v)^T\nabla\mathcal{G}(v)\\ &\qquad\qquad -\frac{L_f}{4}\|\tilde{x}^{k+1}-x^k\|^2,\end{aligned} \tag{2.29}$$

where $\nabla\mathcal{G}$ is defined in (2.2).

For the correction part, we have

$$x^{k+1} = x^k + (1+\theta)(\tilde{x}^{k+1}-x^k) - rA^T(\tilde{s}^{k+1}-s^k), \tag{2.30}$$

where $\tilde{s}^{k+1} = s^{k+1}$, (see (2.2)). Combining the update of x^{k+1} and s^{k+1} in the correction part, and using the notation (2.2) yields

$$\text{(Correction)}\quad v^{k+1} = v^k - M(v^k - \tilde{v}^{k+1}), \tag{2.31}$$

where the matrix M is defined in (2.5).

The prediction-correction framework, defined by steps (2.29) and (2.31), provides a structured approach for analyzing the convergence of the SPAPC. This framework is a modified version in [23] by replacing the proximal mapping in the prediction step with a gradient mapping. Such modification leverages the Lipschitz continuity of f, which not only preserves convergence guarantees but also improves the algorithmic efficiency.

2.5. Characterization of a solution point of (1.1)**.** In this section, we conducts a rigorous analysis of solution point characterization for the problem (1.1) by the prediction-correction framework.

THEOREM 2.6. *Let* (s^{k+1}, x^{k+1}) *be the iterate generated by the scheme* (1.11) *from the current iterate* (x^k, s^k)*. Then* (x^{k+1}, s^{k+1}) *is a solution point of the problem* (1.1) *if the following identity holds*

$$M(\tilde{v}^{k+1}-v^k) = 0. \tag{2.32}$$

Proof. Since the matrix M is nonsingular (see (2.5)), $\tilde{v}^{k+1} = v^k$ holds when the condition (2.32) is satisfied. Substituting this result into the correction step (2.31), we get $v^{k+1} = v^k$. Combining the two results, we obtain $\tilde{v}^{k+1} = v^k = v^{k+1}$. By the relation $H = QM^{-1}$ (see (2.8)), the inequality (2.25) can be rewritten as

$$\begin{aligned}h^*(s) - h^*(\tilde{s}^{k+1}) + (v-\tilde{v}^{k+1})^T\nabla\mathcal{F}(\tilde{v}^{k+1}) + (x-\tilde{x}^{k+1})^T\nabla f(x^k)\\ + (v-\tilde{v}^{k+1})^T HM(\tilde{v}^{k+1}-v^k) \geq 0,\end{aligned}$$

Substituting $\tilde{v}^{k+1} = v^k = v^{k+1}$ into this inequality yields

$$h^*(s) - h^*(s^{k+1}) + (v - v^{k+1})^T \nabla \mathcal{F}(v^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1}) \geq 0$$

Using the definition (2.2), we have

$$h^*(s) - h^*(s^{k+1}) + (v - \tilde{v}^{k+1})^T \nabla \mathcal{G}(\tilde{v}^{k+1}) \geq 0 \tag{2.33}$$

By the optimization condition (2.22), we have that $v^{k+1} = (x^{k+1}, s^{k+1})$ is a optimal solution of the problem (1.1). ◻

Theorem 2.6 shows the term $\|M(\tilde{v}^{k+1} - v)\|_H^2$ can be adopted as a measurement that characterizes the solution precision of $\tilde{v}^{k+1}$ for the problem (1.1).

3. Global Convergence of the SPAPC.

In the section, we firstly analyze the contractive property of the sequence generated by SPAPC (1.11). Then, building upon the contractive property, we prove the convergence and derive the convergence rate of the SPAPC.

3.1. Contractive properties.

For the sequence $\{v^k\}$ within the prediction-correction framework defined by (2.29) and (2.31), we have the main theoretical results as follows:

THEOREM 3.1. *Let $\{v^k\}$ be the sequence generated by* (1.11). *Then, there holds*

$$\|v^k - v^*\|_H^2 \geq \|v^{k+1} - v^*\|_H^2 + \|\tilde{v}^{k+1} - v^k\|_{G_f}^2, \tag{3.1}$$

where $\tilde{v}^{k+1}$ is defined in (2.3)*, and the matrices H and G_f are characterized by* (2.8) *and* (2.10)*, respectively.*

Proof. Setting $v = v^*$ in (2.29), we get

$$(v^* - \tilde{v}^{k+1})^T Q(\tilde{v}^{k+1} - v^k) \geq h^*(\tilde{s}^{k+1}) - h^*(s^*) + (\tilde{v}^{k+1} - v^*)^T \nabla \mathcal{G}(v^*) - \frac{L_f}{4}\|\tilde{x}^{k+1} - x^k\|^2,$$

where v^* is the optimal solution to the problem (1.1). By the optimality condition (2.22), the above inequality can reduce to

$$(v^* - \tilde{v}^{k+1})^T Q(\tilde{v}^{k+1} - v^k) \geq -\frac{L_f}{4}\|\tilde{x}^{k+1} - x^k\|^2.$$

Substituting (2.31) into this inequality yields

$$(v^* - \tilde{v}^{k+1})^T Q M^{-1}(v^{k+1} - v^k) \geq -\frac{L_f}{4}\|\tilde{x}^{k+1} - x^k\|^2.$$

By the relation $H = QM^{-1}$, we have

$$(v^* - \tilde{v}^{k+1})^T H(v^{k+1} - v^k) \geq -\frac{L_f}{4}\|\tilde{x}^{k+1} - x^k\|^2. \tag{3.2}$$

By Lemma 2.1, the inequality (3.2) is equivalent to

$$\begin{aligned}\frac{1}{2}\Big(\|v^* - v^k\|_H^2 - \|v^* - v^{k+1}\|_H^2\Big) + \frac{1}{2}\Big(\|\tilde{v}^{k+1} - v^{k+1}\|_H^2 - \|\tilde{v}^{k+1} - v^k\|_H^2\Big) \\ \geq -\frac{L_f}{4}\|\tilde{x}^{k+1} - x^k\|^2.\end{aligned} \tag{3.3}$$

For the last two terms in the left side of the inequality (3.3), we have

$$
\begin{aligned}
&\|v^k - \tilde{v}^{k+1}\|_H^2 - \|v^{k+1} - \tilde{v}^{k+1}\|_H^2 \\
&\quad = \|v^k - \tilde{v}^{k+1}\|_H^2 - \|(v^k - \tilde{v}^{k+1}) - (v^k - v^{k+1})\|_H^2 \\
&\quad \overset{(2.31)}{=} \|v^k - \tilde{v}^{k+1}\|_H^2 - \|(v^k - \tilde{v}^{k+1}) - M(v^k - \tilde{v}^{k+1})\|_H^2 \\
&\quad = 2(v^k - \tilde{v}^{k+1})^T HM(v^k - \tilde{v}^{k+1}) - (v^k - \tilde{v}^{k+1})^T M^T HM(v^k - \tilde{v}^{k+1}) \\
&\quad \overset{(2.8)}{=} (v^k - \tilde{v}^{k+1})^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^{k+1}) \\
&\quad \overset{(2.9)}{=} \|v^k - \tilde{v}^{k+1}\|_G^2, \qquad (3.4)
\end{aligned}
$$

where the matrix G is characterized by (2.9). Furthermore, applying the definition (2.10), the expression $\|\tilde{v}^{k+1} - v^k\|_G^2 - \frac{L_f}{2}\|\tilde{x}^{k+1} - x^k\|^2$ can be reformulated as

$$
\|\tilde{v}^{k+1} - v^k\|_G^2 - \frac{L_f}{2}\|\tilde{x}^{k+1} - x^k\|^2 = \|\tilde{v}^{k+1} - v^k\|_{G_f}^2. \qquad (3.5)
$$

Substituting (3.4) and (3.5) into (3.3), we obtain the assertion (3.1). ∎

Theorem 3.1 establishes that the sequence $\{\|v^k - v^*\|_H^2\}$ exhibits a non-increasing behavior under the positive definiteness of H and G_f. This ensures the convergence of $\{v^k\}$ to the optimal solution v^*. To guarantee both $H > 0$ and $G_f > 0$, we establish the following sufficient conditions.

LEMMA 3.2. *Let H and G_f be the matrices defined in* (2.8) *and* (2.10)*, respectively. Suppose the parameters θ, r, δ in the scheme* (1.11) *satisfy the conditions* (1.13)*. Then, the matrices H and G_f are positive definite.*

Proof. Since G_f is a diagonal matrix, $G_f > 0$ holds if and only if its diagonal elements satisfy

$$
\theta < 1 - \frac{rL_f}{2}, \quad \frac{1}{\delta} - r\|AA^T\| > 0. \qquad (3.6)
$$

For the positive definiteness of the matrix H, we firstly define a nonsingular matrix W as follows

$$
W = \begin{pmatrix} I & r\theta A^T \\ 0 & I \end{pmatrix}.
$$

By multiplying H by W^T and W from the left and right, respectively, we obtain the matrix H_w:

$$
H_w = W^T HW = \begin{pmatrix} \frac{1}{r(1+\theta)}I & 0 \\ 0 & \frac{1}{\delta}I - r\theta AA^T \end{pmatrix}.
$$

Since the matrix W is nonsingular, $H > 0$ is equivalent to $H_w > 0$. Similar to the matrix G_f, $H_w > 0$ holds if and only if its diagonal elements satisfy

$$
\frac{1}{r(1+\theta)} > 0, \quad \frac{1}{\delta} - r\theta\|AA^T\| > 0. \qquad (3.7)
$$

Combining the inequalities (3.6) and (3.7), we conclude the range of θ as follows

$$
\theta \in (-1, 1 - \frac{rL_f}{2}),
$$

which implies that $\theta < 1$. Building the range of θ, we have

$$\frac{1}{\delta} > r\|AA^T\| > \theta r\|AA^T\|.$$

Thus, we obtain the condition (1.13). ∎

3.2. Global convergence. In this section, two main results are established: (1) a critical lemma for proving S-PAPC convergence, and (2) S-PAPC convergence under conditions (1.13).

LEMMA 3.3. *Let* $\{v^k\}$ *be the sequence generated by the scheme* (1.11) *with the condition* (1.13)*. Then, we have*

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0, \tag{3.8a}$$

$$\lim_{k\to\infty} \|s^{k+1} - s^k\| = 0. \tag{3.8b}$$

Proof. Theorem 3.1 establishes that the sequence $\{v^k\}$ generated by the scheme (1.11) satisfies the inequality

$$\|v^k - v^*\|_H^2 \geq \|v^{k+1} - v^*\|_H^2 + \|\tilde{v}^{k+1} - v^k\|_{G_f}^2.$$

Summing the above inequality over $k = 0, 1, \dots, \infty$ gives

$$\sum_{k=0}^{\infty} \|\tilde{v}^{k+1} - v^k\|_{G_f}^2 \leq \|v^0 - v^*\|_H^2,$$

which implies

$$\lim_{k\to\infty} \|\tilde{v}^{k+1} - v^k\|_{G_f}^2 = 0.$$

It follows from (2.3) that this identity is equivalent to

$$\lim_{k\to\infty} \|\tilde{v}^{k+1} - v^k\|_{G_f}^2 = \lim_{k\to\infty} \left\| \begin{pmatrix} \tilde{x}^{k+1} - x^k \\ s^{k+1} - s^k \end{pmatrix} \right\|_{G_f}^2.$$

Lemma 3.2 shows that under the conditions (1.13), the matrix G_f maintains positive definiteness. Therefore, under the conditions, we obtain

$$\lim_{k\to\infty} \|\tilde{x}^{k+1} - x^k\| = 0, \tag{3.9a}$$

$$\lim_{k\to\infty} \|s^{k+1} - s^k\| = 0. \tag{3.9b}$$

Applying the convergence result (3.9b) and taking the limit in iteration (1.11d), we establish that $\lim_{k\to\infty} x^{k+1} = x^k$. Combined with (3.9b), this yields the final convergence result (3.8). ∎

The following result establishes the global convergence of SPAPC.

THEOREM 3.4. *Let* $\{v^k\}$ *be the sequence generated by the scheme* (1.11) *with the conditions* (1.13)*. Then, the sequence* $\{v^k\}$ *converges to the optimal point* $v^\infty = v^*$ *for the problem* (1.1)*.*

Proof. Theorem 3.1 and Lemma 3.2 establish that under condition (1.13), the sequence $\{\|v^k - v^*\|_H^2\}$ generated by (1.11) is non-increasing. Consequently, $\{v^k\}$ is bounded. Consider a convergent subsequence $\{v_b^k\} \to v_b^\infty = (x_b^\infty, s_b^\infty)$. By substituting (x_b^∞, s_b^∞) into (1.11), we derive the subsequent iterates $(\tilde{x}_b^{\infty+}, s_b^{\infty+}, x_b^{\infty+})$. It follows from Lemma 3.3 that

$$
\begin{aligned}
x_b^{\infty+} &= \tilde{x}_b^{\infty+} = x_b^\infty, \\
s_b^{\infty+} &= s_b^\infty,
\end{aligned}
$$

which implies $v_b^\infty = v_b^{\infty+}$. By the definition (2.3), we have $\tilde{v}_b^{\infty+} = v_b^{\infty+}$. Thus, $\tilde{v}_b^{\infty+} = v_b^\infty$. Substituting $\tilde{v}_b^{\infty+} = v_b^\infty$ into (2.24) leads to

$$
\begin{cases}
(x - \tilde{x}_b^{\infty+})^T (A^T \tilde{s}_b^{\infty+} + \nabla f(\tilde{x}_b^{\infty+})) \geq 0, \\
h^*(s) - h^*(\tilde{s}_b^{\infty+}) - (y - \tilde{y}_b^{\infty+})^T (A\tilde{x}_b^{\infty+}) \geq 0.
\end{cases}
$$

By the notations (2.2), the above inequalities are equivalent to the form

$$
h^*(s) - h^*(\tilde{s}_b^{\infty+}) + (v - \tilde{v}_b^{\infty+})^T \nabla \mathcal{G}(\tilde{v}_b^{\infty+}) \geq 0.
$$

By *Lemma 2.5*, we obtain that $\tilde{v}_b^{\infty+} = v_b^\infty$ is an optimal point for (1.1). Furthermore, inequality (3.1) shows the contraction property, i.e., $\|v^{k+1} - v_b^\infty\|_H^2 \leq \|v^k - v_b^\infty\|_H^2$. This implies that $\{v^k\}$ has a unique cluster point. Thus, the sequence $\{v^k\}$ converges to the optimal point $v^* = v_b^{\infty+} = v_b^\infty$. ∎

3.3. Convergence rate. In this section, we present the non-ergodic convergence rate of the SPAPC, while establishing that the primal-dual gap generated by the SPAPC can achieve an ergodic convergence rate. Specifically, Theorem 2.6 shows that the term $M\|\tilde{v}^{k+1} - v^k\|_H^2$ can be used to measure the accuracy of iterates. Thus, we present the non-ergodic convergence rate with respect to this term. For the ergodic convergence rate, we firstly derive a stricter convergence condition than that in (1.13). Then, building upon this convergence condition, we derive the ergodic convergence rate.

In the following, we prove a lemma which is crucial in analyzing the non-ergodic convergence rate.

LEMMA 3.5. *Let* $\{v^k\}$ *be the sequence generated by the scheme* (1.11). *Then it holds that*

$$
\begin{aligned}
(v^{k-1} - \tilde{v}^k)^T M^T H M \Big(v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1})\Big) \geq {} & -\frac{L_f}{4} \|\tilde{x}^k - x^{k-1} - \big(\tilde{x}^{k+1} - x^k\big)\|^2 \\
& + \frac{1}{2} \|v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1})\|_{Q^T+Q}^2.
\end{aligned} \tag{3.11}
$$

Proof. The inequality (2.25) for the iterates $k+1$ and k are, respectively, given by

$$
\begin{aligned}
h^*(s) - h^*(\tilde{s}^{k+1}) + (v - \tilde{v}^{k+1})^T \nabla \mathcal{F}(\tilde{v}^{k+1}) + (x - \tilde{x}^{k+1})^T \nabla f(x^k) \\
+ (v - \tilde{v}^{k+1})^T Q(\tilde{v}^{k+1} - v^k) \geq 0,
\end{aligned}
$$

and

$$
\begin{aligned}
h^*(s) - h^*(\tilde{s}^k) + (v - \tilde{v}^k)^T \nabla \mathcal{F}(\tilde{v}^k) + (x - \tilde{x}^k)^T \nabla f(x^{k-1}) \\
+ (v - \tilde{v}^k)^T Q(\tilde{v}^k - v^{k-1}) \geq 0,
\end{aligned}
$$

Setting $v = \tilde{v}^k$ in the first inequality and $v = \tilde{v}^{k+1}$ in the second inequality yields, respectively,

$$h^*(\tilde{s}^k) - h^*(\tilde{s}^{k+1}) + (\tilde{v}^k - \tilde{v}^{k+1})^T \nabla\mathcal{F}(\tilde{v}^{k+1}) + (\tilde{x}^k - \tilde{x}^{k+1})^T \nabla f(x^k) + (\tilde{v}^k - \tilde{v}^{k+1})^T Q(\tilde{v}^{k+1} - v^k) \geq 0, \tag{3.12}$$

and

$$h^*(\tilde{s}^{k+1}) - h^*(\tilde{s}^k) + (\tilde{v}^{k+1} - \tilde{v}^k)^T \nabla\mathcal{F}(\tilde{v}^k) + (\tilde{x}^{k+1} - \tilde{x}^k)^T \nabla f(x^{k-1}) + (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(\tilde{v}^k - v^{k-1}) \geq 0, \tag{3.13}$$

Combining the inequalities (3.12) and (3.13), and using $(\tilde{v}^{k+1} - v)^T \nabla\mathcal{F}(\tilde{v}^{k+1}) = (\tilde{v}^{k+1} - v)^T \nabla\mathcal{F}(v)$ yields

$$\begin{aligned}(\tilde{v}^k - \tilde{v}^{k+1})^T Q\Big(v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1})\Big) &\geq (x^{k-1} - x^k)^T (\nabla f(x^{k-1}) - \nabla f(x^k)) \\ &\quad + \Big(\tilde{x}^k - x^{k-1} - (\tilde{x}^{k+1} - x^k)\Big)^T (\nabla f(x^{k-1}) - \nabla f(x^k)).\end{aligned} \tag{3.14}$$

This inequality can further reduce to

$$\begin{aligned}(\tilde{v}^k - \tilde{v}^{k+1})^T Q\Big(v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1})\Big) &\geq \frac{1}{L_f}\|\nabla f(x^{k-1}) - \nabla f(x^k)\|^2 \\ &\quad + \Big(\tilde{x}^k - x^{k-1} - (\tilde{x}^{k+1} - x^k)\Big)^T (\nabla f(x^{k-1}) - \nabla f(x^k)) \\ &\geq -\frac{L_f}{4}\|\tilde{x}^k - x^{k-1} - (\tilde{x}^{k+1} - x^k)\|^2,\end{aligned} \tag{3.15}$$

where the first inequality follows from Lemma 2.14 and the second inequality is derived by the Cauchy-Schwarz inequality. Adding the term

$$\Big(v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1}\Big)^T Q\Big(v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1}\Big)$$

to both sides of (3.15) and using the fact that $v^T Q v = \frac{1}{2} v^T (Q^T + Q) v$, we get

$$\begin{aligned}(v^{k-1} - v^k)^T Q\Big(v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1})\Big) &\geq -\frac{L_f}{4}\|\tilde{x}^k - x^{k-1} - \big(\tilde{x}^{k+1} - x^k\big)\|^2 \\ &\quad + \frac{1}{2}\|v^{k-1} - \tilde{v}^k - (v^k - \tilde{v}^{k+1})\|^2_{Q^T+Q}.\end{aligned}$$

Using the relation $(v^{k-1} - v^k)^T = M(v^{k-1} - \tilde{v}^k)$ (see (2.31)) and $Q = HM$, this inequality is reformulated as the inequality (3.11). ∎

Theorem 3.6. *Let $\{v^k\}$ be the sequence generated by the scheme* (1.11) *with the conditions* (1.13). *Then it holds that*

$$\|M(v^t - \tilde{v}^{t+1})\|^2_H \leq \frac{1}{tD_0}\|v^0 - v^*\|^2_H, \tag{3.16}$$

where t denotes the iterate and D_0 is a constant given by

$$D_0 = \frac{\lambda_{\min}(G_f)}{\lambda_{\max}(M^T HM)}. \tag{3.17}$$

Proof. It follows from the definition (2.8) that the matrix H is symmetric. By Lemma 2.2, we have

$$\begin{aligned}\|M(v^{k-1}-\tilde{v}^k)\|_H^2 - \|M(v^k-\tilde{v}^{k+1})\|_H^2 &= -\|v^{k-1}-\tilde{v}^k-(v^k-\tilde{v}^{k+1})\|_{M^THM}^2 \\ &+2(v^{k-1}-\tilde{v}^k)^TM^THM\Big(v^{k-1}-\tilde{v}^k-(v^k-\tilde{v}^{k+1})\Big)\end{aligned}$$

Substituting the inequality (3.11) into the above inequality, we get

$$\begin{aligned}\|M(v^{k-1}-\tilde{v}^k)\|_H^2 - \|M(v^k-\tilde{v}^{k+1})\|_H^2 &\geq -\|v^{k-1}-\tilde{v}^k-(v^k-\tilde{v}^{k+1})\|_{M^THM}^2 \\ &-\frac{L_f}{2}\|\tilde{x}^k-x^{k-1}-\left(\tilde{x}^{k+1}-x^k\right)\|^2+\|v^{k-1}-\tilde{v}^k-(v^k-\tilde{v}^{k+1})\|_{Q^T+Q}^2.\end{aligned}$$

By the definition (2.10), we have

$$\|M(v^{k-1}-\tilde{v}^k)\|_H^2 - \|M(v^k-\tilde{v}^{k+1})\|_H^2 \geq \|v^{k-1}-\tilde{v}^k-(v^k-\tilde{v}^{k+1})\|_{G_f}^2. \tag{3.18}$$

Under the condition (1.13), Lemma 3.2 establishes the positive definiteness of the matrix G_f. Consequently, we obtain the following inequality

$$\|M(v^{k-1}-\tilde{v}^k)\|_H^2 \geq \|M(v^k-\tilde{v}^{k+1})\|_H^2.$$

Summing the above inequality over $k=0,1,...,t-1$ yields

$$t\|M(v^t-\tilde{v}^{t+1})\|_H^2 \leq \sum_{k=0}^{t-1}\|M(v^k-\tilde{v}^{k+1})\|_H^2. \tag{3.19}$$

Furthermore, since the matrix G_f is positive definite under the condition (1.13), it follows from Theorem 3.1 and the definition (3.17) that the following inequality holds

$$\begin{aligned}\|v^k-v^*\|_H^2 &\geq \|v^{k+1}-v^*\|_H^2+\|\tilde{v}^{k+1}-v^k\|_{G_f}^2 \\ &\geq \|v^{k+1}-v^*\|_H^2+D_0\|M(\tilde{v}^{k+1}-v^k)\|_H^2\end{aligned}$$

Summing the above inequality over $k=0,1,...,t-1$, we get

$$\|v^t-v^*\|_H^2+D_0\sum_{k=0}^{t-1}\|M(\tilde{v}^{k+1}-v^k)\|_H^2 \leq \|v^0-v^*\|_H^2. \tag{3.20}$$

Combining the inequalities (3.19) and (3.20), we get the assertion (3.16). ◻

In the following, a refined convergence condition is presented.

LEMMA 3.7. *Let* $\{v^k\}$ *be the sequence generated by the scheme* (1.11). *Suppose the parameters* θ, r, δ *in the scheme* (1.11) *satisfy the conditions*

$$\theta \in (-1, 1-rL_f), \quad r\delta < \frac{1}{\|AA^T\|}. \tag{3.21}$$

Then, the sequence $\{\|v^k-v^*\|_H^2\}$ *is non-increasing and satisfies*

$$\|v^k-v^*\|_H^2 \geq \|v^{k+1}-v^*\|_H^2+\|\tilde{v}^{k+1}-v^k\|_{G_f'}^2, \tag{3.22}$$

where $\tilde{v}^{k+1}$ *is the auxiliary variable in* (2.3), *and* H *and* G_f' *are defined by* (2.8) *and* (2.11), *respectively.*

Proof. By Lemma 2.4, the inequality (2.26) can further reduce to

$$\begin{aligned}(v-\tilde{v}^{k+1})^T Q(\tilde{v}^{k+1}-v^k) &\geq h^*(\tilde{s}^{k+1})-h^*(s)+(\tilde{v}^{k+1}-v)^T\nabla\mathcal{F}(\tilde{v}^{k+1})\\ &\quad+(\tilde{x}^{k+1}-x)^T\nabla f(x^k)\\ &\geq h^*(\tilde{s}^{k+1})-h^*(s)+(\tilde{v}^{k+1}-v)^T\nabla\mathcal{F}(\tilde{v}^{k+1})\\ &\quad+f(\tilde{x}^{k+1})-f(x)-\frac{L_f}{2}\|\tilde{x}^{k+1}-x^k\|^2.\end{aligned} \tag{3.23}$$

It follows from $(\tilde{v}^{k+1}-v)^T\nabla\mathcal{F}(\tilde{v}^{k+1})=(\tilde{v}^{k+1}-v)^T\nabla\mathcal{F}(v)$ that the above inequality is equivalent to

$$\begin{aligned}(v-\tilde{v}^{k+1})^T Q(\tilde{v}^{k+1}-v^k) &\geq h^*(\tilde{s}^{k+1})-h(s)+(\tilde{v}^{k+1}-v)^T\nabla\mathcal{F}(u)\\ &\quad+f(\tilde{x}^{k+1})-f(x)-\frac{L_f}{2}\|\tilde{x}^{k+1}-x^k\|^2.\end{aligned} \tag{3.24}$$

Using the property of the convex function f, we have

$$f(\tilde{x}^{k+1})-f(x)\geq(\tilde{x}^{k+1}-x)^T\nabla f(x). \tag{3.25}$$

Combining the inequalities (3.24) and (3.25), and using the notation (2.2), we have

$$(v-\tilde{v}^{k+1})^T Q(\tilde{v}^{k+1}-v^k)\geq h^*(\tilde{s}^{k+1})-h^*(s)+(\tilde{v}^{k+1}-v)^T\nabla\mathcal{G}(v)-\frac{L_f}{2}\|\tilde{x}^{k+1}-x^k\|^2. \tag{3.26}$$

Setting $v=v^*$ and applying the analytical framework developed in Theorem 3.1, we derive the inequality (3.22).

The inequality (3.22) implies that if $H>0$ and $G_f^{'}>0$, the sequence $\{\|v^k-v^*\|\}$ is non-increasing. Since $G_f^{'}$ is a diagonal matrix, $G_f^{'}>0$ holds when the parameters θ, r, δ satisfy

$$\theta<1-rL_f,\quad \frac{1}{\delta}-r\|AA^T\|>0. \tag{3.27}$$

Furthermore, it follows from Lemma 3.2 that the matrix H is positive definite if the inequality (3.7) holds. Combining the inequalities (3.27) and (3.7), we obtain (3.21), which proves the assertion (3.22). ∎

Note that Lemma 3.7 gives more conservative parameter conditions than that in Lemma 3.2. Building upon Lemma 3.7, we present the ergodic convergence rate of S-PAPC.

THEOREM 3.8. *Let $\{v^k\}$ be the sequence generated by the scheme (1.11) with the conditions (3.21), and let $\tilde{v}_t=(\tilde{x}_t,\tilde{s}_t)$ be the averaged sequence given by*

$$\tilde{v}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{v}^k.$$

Then, there holds

$$\varphi(\tilde{x}_t,s)-\varphi(x,\tilde{s}_t)\leq\frac{1}{2(t+1)}\|v-v^0\|_H^2. \tag{3.28}$$

Proof. For the function $\varphi(\tilde{x}, s)$, it holds that

$$\begin{aligned}\varphi(\tilde{x}^{k+1}, s) - \varphi(x,\tilde{s}^{k+1}) &= \varphi(\tilde{x}^{k+1}, s) - \varphi(\tilde{x}^{k+1}, \tilde{s}^{k+1}) + \varphi(\tilde{x}^{k+1}, \tilde{s}^{k+1}) - \varphi(x, \tilde{s}^{k+1}) \\ &= h^*(\tilde{s}^{k+1}) - h^*(s) + (\tilde{v}^{k+1} - v)^T \nabla \mathcal{F}(\tilde{v}^{k+1}) + f(\tilde{x}^{k+1}) - f(x).\end{aligned}$$

It follows from $(\tilde{v}^{k+1} - v)^T \nabla \mathcal{F}(\tilde{v}^{k+1}) = (\tilde{v}^{k+1} - v)^T \nabla \mathcal{F}(v)$ that the above inequality can further be expressed as

$$\begin{aligned}\varphi(\tilde{x}^{k+1}, s) - \varphi(x, \tilde{s}^{k+1}) = h^*(\tilde{s}^{k+1}) - h^*(s) + (\tilde{v}^{k+1} - v)^T \nabla \mathcal{F}(v) \\ + f(\tilde{x}^{k+1}) - f(x).\end{aligned}$$

By (3.24), the following inequality holds

$$\begin{aligned}\varphi(\tilde{x}^{k+1}, s) - \varphi(x, \tilde{s}^{k+1}) &= h^*(\tilde{s}^{k+1}) - h^*(s) + (\tilde{v}^{k+1} - v)^T \nabla \mathcal{F}(v) + f(\tilde{x}^{k+1}) - f(x) \\ &\leq (v - \tilde{v}^{k+1})^T Q(\tilde{v}^{k+1} - v^k) + \frac{L_f}{2} \|\tilde{x}^{k+1} - x^k\|^2.\end{aligned}$$

By the correction step (2.31), this inequality can be rewritten as

$$\varphi(\tilde{x}^{k+1}, s) - \varphi(x, \tilde{s}^{k+1}) \leq (v - \tilde{v}^{k+1})^T H(v^{k+1} - v^k) + \frac{L_f}{2} \|\tilde{x}^{k+1} - x^k\|^2. \tag{3.29}$$

Using Lemma 2.1, we have

$$\begin{aligned}\varphi(\tilde{x}^{k+1}, s) - \varphi(x, \tilde{s}^{k+1}) \;&\leq\; \frac{1}{2}\Big(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2 + \|\tilde{v}^{k+1} - v^{k+1}\|_H^2 \\ &\qquad -\|\tilde{v}^{k+1} - v^k\|_H^2 + L_f \|\tilde{x}^{k+1} - x^k\|^2\Big) \\ &\overset{(3.4)}{=} \frac{1}{2}\Big(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2 - \|v^k - \tilde{v}^{k+1}\|_G^2 \\ &\qquad +L_f \|\tilde{x}^{k+1} - x^k\|^2\Big).\end{aligned} \tag{3.30}$$

Lemma 3.7 establishes that $G_f^{'} > 0$ holds under the conditions (3.21). Therefore, the inequality (3.30) can further simplified to

$$\varphi(\tilde{x}^{k+1}, s) - \varphi(x, \tilde{s}^{k+1}) \leq \frac{1}{2}\Big(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2\Big). \tag{3.31}$$

Summing the above inequality over $k = -1, 0, \ldots, t-1$ and using the convexity of the function φ gives (3.28). ∎

4. Numerical experiments.

In the section, we present some numerical experiments to verify the efficiency of the symmetric proximal alternating predictor-corrector (SPAPC) algorithm. We firstly examine the convergence condition (1.13) and identify the optimal parameters. Then, we compare the SPAPC algorithm with other algorithms, including the PDFP^{2}O$_k$ in [6], the PAPC in [10], and the PDHG in [3].

4.1. Lasso problem.

To investigate the impact of the convergence condition (1.13) on accelerating the convergence, we construct a general model as follows

$$\min_x \varphi(x) = \frac{1}{2} \|Kx - b\|_2^2 + \mu \|Ax\|_1, \tag{4.1}$$

where $K \in \mathcal{R}^{m\times n}$ is a random matrix, $b = Ka + \rho$ is a given vector and μ is a regularization parameter. We define $a \in \mathcal{R}^n$ as a random vector with exactly N non-zero components, and $\rho \in \mathcal{R}^m$ as a white gaussian noise of mean 0 and variance 1. The matrix $A \in \mathcal{R}^{n-1\times n}$ is given by

$$A = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & \dots & \\ & & & -1 & 1 \end{bmatrix}. \tag{4.2}$$

Assume that $y = Ax$ and s is a dual variable, then the problem (4.1) is equivalent to

$$\min_x \max_s \quad \frac{1}{2}\|Kx - b\|_2^2 + s^T Ax - I_\mu(\|s\|_\infty),$$

where $I_\mu(\|s\|_\infty)$ is a indicator function defined by

$$I_\mu(\|s\|_\infty) = \begin{cases} 0, & \text{if } \|s\|_\infty \le \mu, & (4.3a)\\ \infty, & \text{otherwise .} & (4.3b)\end{cases}$$

Note that $I_\mu(\|s\|_\infty)$ is the conjugate function of the function $\mu\|s\|_1$. For the detail derivation, we refer readers to [2]. By the definition (4.3), the update (1.11c) can further reduce to

$$s^{k+1} = \min(\max(s^k + \delta A\bar{x}^{k+1}, -\mu), \mu). \tag{4.4}$$

In the numerical experiments, we adopt the parameter configurations: $m = 500$, $n = 10000$, and $\mu = 200$. All optimization variables $(\tilde{x}^0, \bar{x}^0, s^0, x^0)$ are initialized to zero. The Lipschitz constant L_f, computed as the spectral norm $\|K^TK\|$, is 1.4838×10^4. The norm $\|AA^T\|$ evaluates to 3.9655. For the SPAPC algorithm (1.11), the step size parameter $r\delta$ is denoted by λ, and the maximum iteration number is 8000. To examine the convergence condition (1.13), we show the numerical performance of the SPAPC under the different parameters θ, λ and τ, focusing on the difference in the objective gap $\frac{\varphi^k - \varphi^*}{\varphi^*}$ and solution error $\|x^k - x^*\|_2^2$, where φ^* denotes the optimal objective function value. By applying the CVX solver in [13] to the problem (4.1), we obtain the optimal objective function value $\varphi^* \approx 2.47896 \times 10^4$ and the optimal solution x^*.

In the following, we firstly investigate the impact of the extrapolation parameter θ on accelerating the convergence. We fix the parameters λ and r at $\frac{1}{4}$ and $\frac{1}{2L_f}$, respectively. The parameter θ is chosen from the set $\{0.7, 0.5, 0.2, 0, -0.2, -0.5, -0.8\}$. It is clear that all the parameters θ, r, δ satisfy the convergence condition (1.13). The numerical results $\frac{\varphi^k - \varphi^*}{\varphi^*}$ and $\|x^k - x^*\|_2^2$ are, respectively, shown in Figure 1 and Table 1, where CPU denotes the running time of CPU. The numerical results demonstrate that as θ approaches 1, both the objective gap $\varphi - \varphi^*$ and solution error $\|x^k - x^*\|_2^2$ decrease. Furthermore, it is shown in Table 1 that the running time also reduce as $\theta \to 1$. These results imply that the convergence condition (1.11) can accelerate the convergence.

Building upon the above results, we identify the optimal extrapolation parameter as $\theta = 0.7$. With θ fixed along with $r = \frac{1}{2L_f}$, we analyze the performance of the SPAPC under the different parameter $\lambda = r\delta$. The numerical results $\frac{\varphi^k - \varphi^*}{\varphi^*}$ and $\|x^k - x^*\|_2^2$ are shown in Table 2. Note that since the CPU running time is similar

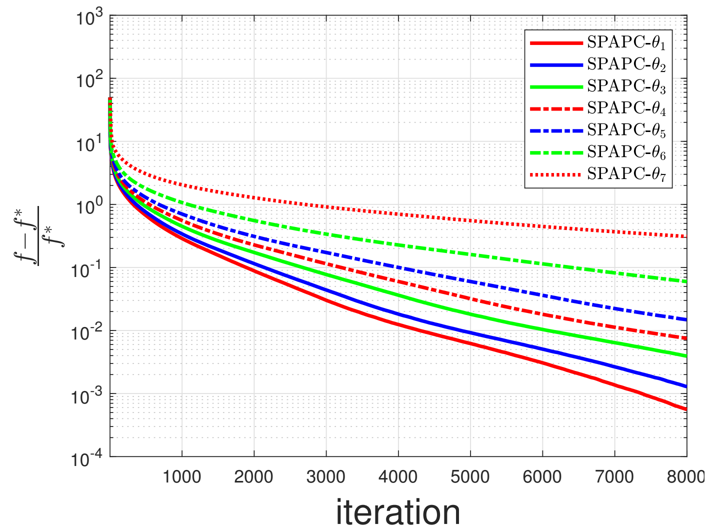


Fig. 1. *Numerical result* $\frac{\varphi^k-\varphi^*}{\varphi^*}$ *of* (1.11) *for* (4.1) *under the different extrapolation parameter:* $\theta_1 = 0.7$, $\theta_2 = 0.5$, $\theta_3 = 0.2$, $\theta_4 = 0$, $\theta_5 = -0.2$, $\theta_6 = -0.5$, $\theta_7 = -0.8$

Table 1
Numerical result $\|x^k - x^*\|_2^2$ *of* (1.11) *for* (4.1) *under the different extrapolation parameter* θ

θ	$\|x^k - x^*\|_2^2$	CPU	θ	$\|x^k - x^*\|_2^2$	CPU
0.7	1.6418	73.5394	0.5	2.5586	80.4803
0.2	4.6493	81.2929	0	6.5862	81.5051
−0.2	9.1304	75.5268	−0.5	15.8486	75.0197
−0.8	31.5168	74.5071			

in each case, we have ignored the CPU running time. From Table 2, it can be found that decreasing λ has little effect on the solution error $\|x^k - x^*\|_2^2$, while notably affecting the objective gap, i.e., as λ decrease, the objective gap increases. Thus, in the practical, selecting a moderately large λ is preferable.

Table 2
Numerical result of (1.11) *for* (4.1) *under the different step-size parameter* λ

λ	$\|x^k - x^*\|_2^2$	$\frac{\varphi^k-\varphi^*}{\varphi^*}$	λ	$\|x^k - x^*\|_2^2$	$\frac{\varphi^k-\varphi^*}{\varphi^*}$
$\frac{1}{4}$	1.6418	5.5964×10^{-4}	$\frac{1}{8}$	1.6367	6.1752×10^{-4}
$\frac{1}{16}$	1.6254	7.5074×10^{-4}	$\frac{1}{32}$	1.6098	1.0400×10^{-3}
$\frac{1}{64}$	1.5746	1.5634×10^{-3}			

Finally, we compare the numerical performance of the four optimization algorithms, including the SPAPC (1.11), the PAPC in [10], the PDHG in [3] and the PDFP^{2}O$_k$ in [6]. We fix the parameter $\lambda = 0.25$ across all methods. The step size r is chosen from a predefined set $\{\frac{1}{2L_f}, \frac{1}{L_f}, \frac{3}{2L_f}, \frac{19}{10L_f}\}$. To ensure that the convergence condition (1.13) holds, we set the extrapolation parameter $\theta \in \{0.7, 0.4, 0.2, 0.04\}$. The rest parameters remain unchanged. It should be noted that the PAPC algorithm's applicability is inherently limited by its convergence condition (1.5), which restricts its use to larger step sizes r. For the PDHG implementation, there exists a inversion problem about the matrix $(I + rK^TK)$ in the update (1.2a). To solve it, we apply the conjugate gradient method in [2] to the quadratic subproblem. The numerical results $\frac{\varphi^k - \varphi^*}{\varphi^*}$ and $\|x^k - x^*\|_2^2$ are presented in Figure 2 and Figure 3, respectively. The results indicate that the SPAPC presents superior numerical performance than other algorithms. Furthermore, it is clear that increasing the step size r leads to the fast convergence of the SPAPC, which guides us in choosing the step size parameter in practices.

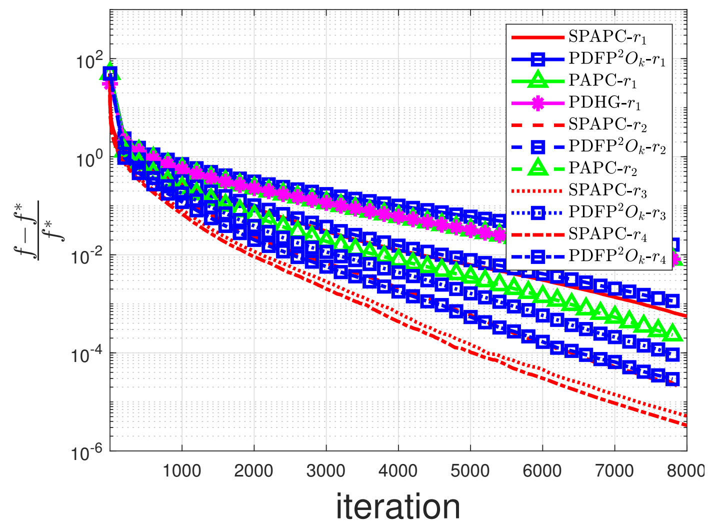


FIG. 2. *Numerical result* $\frac{\varphi^k - \varphi^*}{\varphi^*}$ *of* (1.11) *for* (4.1) *under different methods and the different primal step-size parameter* r*:* $r_1 = \frac{1}{2L_f}, r_2 = \frac{1}{L_f}, r_3 = \frac{3}{2L_f}, r_4 = \frac{19}{10L_f}$

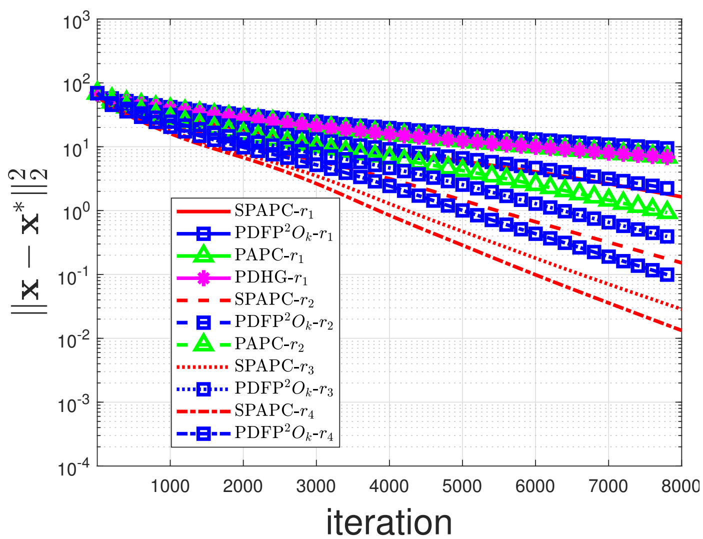


FIG. 3. *Numerical result* $\|x^k - x^*\|_2^2$ *of* (1.11) *for* (4.1) *under different methods and the different step-size parameter* r*:* $r_1 = \frac{1}{2L_f}, r_2 = \frac{1}{L_f}, r_3 = \frac{3}{2L_f}, r_4 = \frac{19}{10L_f}$

4.2. Total variation based image denoising. Consider an optimization problem for image denoising as follows

$$\min_{x\in[0,1]} \varphi(x) = \frac{1}{2}\|x-b\|_2^2 + \mu\|Ax\|_1, \tag{4.5}$$

where $x \in \mathcal{R}^{m\times n}$ denotes the denoised image, $b \in \mathcal{R}^{m\times n}$ is a noisy input image with noise level 0.1, and $\mu = 0.1$ is a regularization parameter. A denotes the 2D discrete gradient operator defined as

$$(Ax)_{i,j} = \begin{pmatrix} (Ax)^1_{i,j} \\ (Ax)^2_{i,j} \end{pmatrix} \tag{4.6}$$

where

$$(Ax)^1_{i,j} = \begin{cases} \frac{1}{g}(x_{i+1,j} - x_{i,j}), & \text{if } i < m, \quad (4.7a) \\ 0, & \text{if } i = m. \quad (4.7b) \end{cases}$$

and

$$(Ax)^2_{i,j} = \begin{cases} \frac{1}{g}(x_{i,j+1} - x_{i,j}), & \text{if } j < n, \quad (4.8a) \\ 0, & \text{if } j = n, \quad (4.8b) \end{cases}$$

and g denotes the gird step size. Let $y = Ax$ and s denote the dual variable. Then, the optimization problem (4.5) can be reformulated as

$$\min_{x\in[0,1]} \max_{s} \quad \frac{1}{2}\|x-b\|_2^2 + s^T Ax - I_\mu(\|s\|_\infty).$$

Similar to the update of s in the Lasso problem, the update (1.11c) in the image denoising problem can further reduce to (4.4).

For the numerical experiments, the parameter configuration is set as follows: the spectral norm of AA^T is estimated to be approximately 8, while the Lipschitz constant L_f is fixed at 1. The step size parameter $r\delta$ is defined as λ and the maximum iteration number is 200. Furthermore, to evaluate the convergence behavior of the SPAPC algorithm, we employ two quantitative metrics: the objective gap $\frac{\varphi^k - \varphi^*}{\varphi^*}$ and the Signal-to-Noise Ratio (SNR) defined as

$$SNR = 20\log_{10}\frac{\|x_o\|_F}{\|x_o - x^k\|_F},$$

where x_o denotes the original image without noisy and the optimal objective function value φ^* is estimated to be 1.0892×10^3.

We firstly compare the performance of the SPAPC under the different extrapolation parameters. With the fixed parameters $r = \frac{1}{2L_f}$ and $\lambda = \frac{1}{8}$, we choose the extrapolation parameter from a predefined set $\{0.7, 0.5, 0.2, 0, -0.2, -0.5, -0.8\}$, all of which rigorously satisfy the convergence condition (1.13). The numerical results are plotted in Figure 4 and Figure 5. The origin image, the image with noise and the denoised image are shown in Fig. 6. Since the output pages under the different extrapolation parameters θ look similar, we only present the image for $\theta = 0.7$. The result in Fig. 6 indicates that the SPAPC method can deal with the image denoising problem. The numerical results in Figure 4 and Figure 5 shows that as the extrapolation parameter θ approaches -1, there exists a significant reduction in convergence speed of the SPAPC. In contrast, when θ approaches the theoretical bound $1 - \frac{rL_f}{2}$, the convergence speed remains relatively stable. These observations suggest that the extrapolation parameter close to -1 should be avoided in image denoising problem.

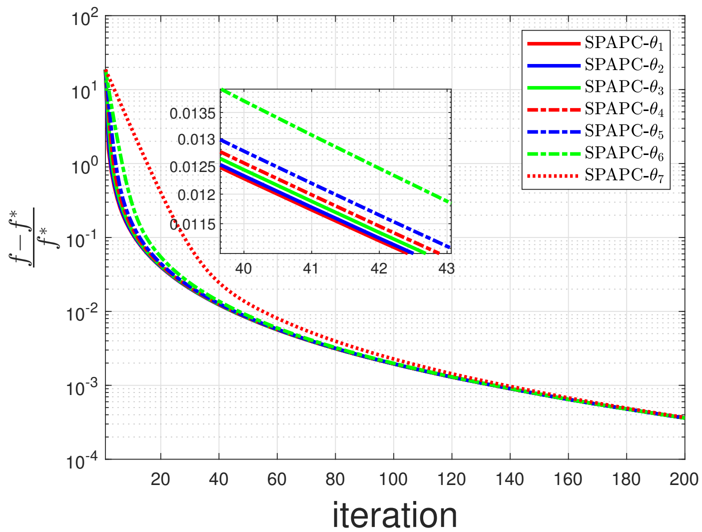


FIG. 4. *Numerical result* $\frac{\varphi^k-\varphi^*}{\varphi^*}$ *of* (1.11) *for* (4.5) *under the different extrapolation parameter:* $\theta_1 = 0.7$, $\theta_2 = 0.5$, $\theta_3 = 0.2$, $\theta_4 = 0$, $\theta_5 = -0.2$, $\theta_6 = -0.5$, $\theta_7 = -0.8$

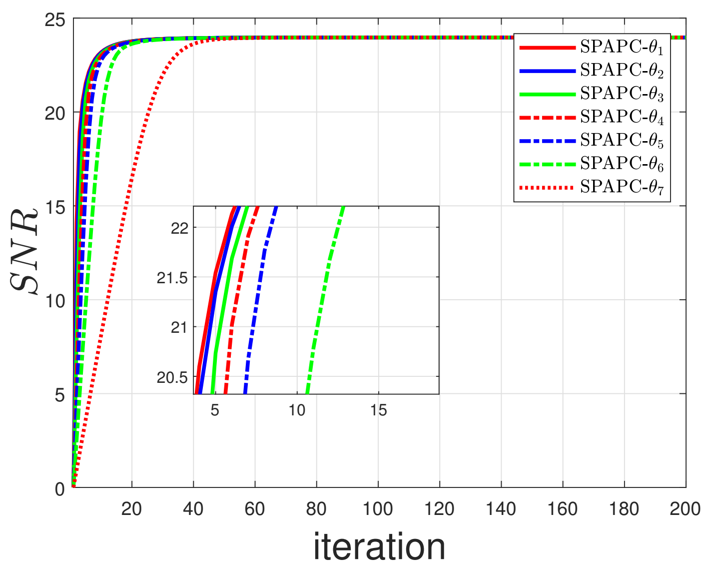
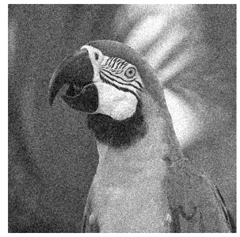


FIG. 5. *Numerical result SNR of* (1.11) *for* (4.5) *under the different extrapolation parameter:* $\theta_1 = 0.7$, $\theta_2 = 0.5$, $\theta_3 = 0.2$, $\theta_4 = 0$, $\theta_5 = -0.2$, $\theta_6 = -0.5$, $\theta_7 = -0.8$

Building upon the above results, we chose the extrapolation parameter θ as 0.7. Then, with the fixed parameter $r = \frac{1}{2L_f}$, we compare the performance of the SPAPC

(a)



(b)

(c)

Fig. 6. *Image.(a) Original image. (b) Image with noise. (c) Denoising image*

under the different parameter $\lambda = r\delta$. The numerical results $\frac{\varphi^k - \varphi^*}{\varphi^*}$ and SNR are presented in Figure 7 and Figure 8, respectively. The results indicate that increasing λ can accelerate the convergence of the SPAPC. Thus, in practices, we usually chose the large parameter λ.

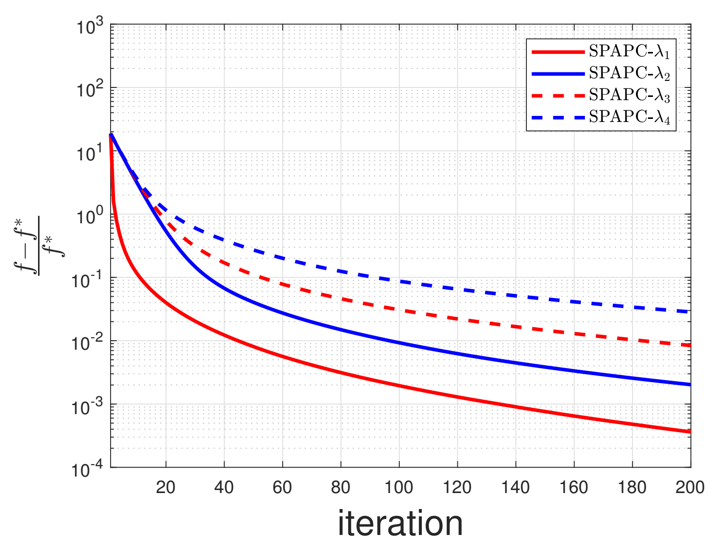


Fig. 7. *Numerical result* $\frac{\varphi^k - \varphi^*}{\varphi^*}$ *of* (1.11) *for* (4.5) *under the different parameter* λ*:* $\lambda_1 = \frac{1}{8}, \lambda_2 = \frac{1}{16}, \lambda_3 = \frac{1}{32}, \lambda_4 = \frac{1}{64}$

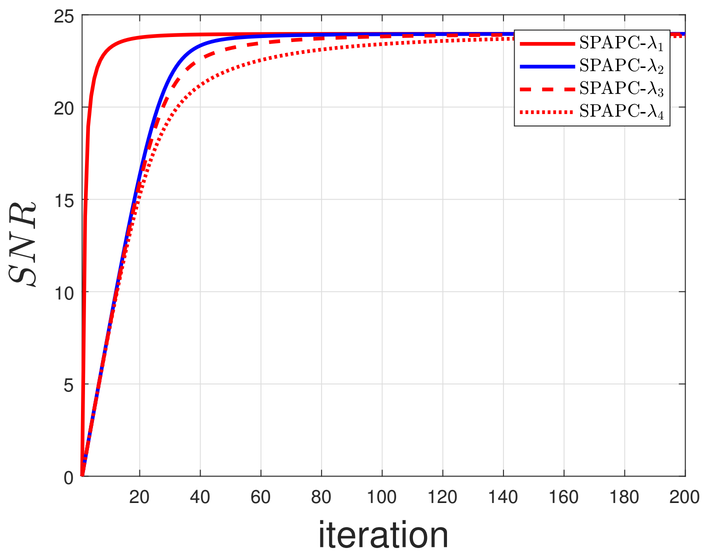


FIG. 8. *Numerical result SNR of* (1.11) *for* (4.5) *under the different parameter* λ*:* $\lambda_1 = \frac{1}{8}, \lambda_2 = \frac{1}{16}, \lambda_3 = \frac{1}{32}, \lambda_4 = \frac{1}{64}$

Finally, we compare the numerical performance of the SPAPC with the other methods, including the SPAPC (1.11), the PAPC in [10], the PDHG in [3] and the PDFP^{2}O$_k$ in [6]. We maintain the parameter $\lambda = 0.25$ and $\theta = 0.7$ across all methods. The step size r is chosen from a predefined set $\{\frac{1}{2L_f}, \frac{1}{L_f}, \frac{3}{2L_f}, \frac{\sqrt{2}}{4L_f}\}$. The numerical results $\frac{\varphi^k - \varphi^*}{\varphi^*}$ and SNR are presented in Figure 9 and Figure 10, respectively. It can be found that when the step size parameter r remains the same, the SPAPC present superior numerical performance than other algorithms. Furthermore, when the step size satisfies $r = \delta$, the SPAPC shows the fastest convergence compared with other cases. This indicates that, for the image denoising problem, setting equal step sizes for the primal and dual variables improves the numerical performance of the SPAPC.

5. Conclusion. In this work, we proposed a symmetric proximal alternating predictor-corrector (SPAPC) algorithm for solving convex-concave saddle-point problems, where one of the objective components is assumed to be differentiable. By introducing a coupled extrapolation strategy, the proposed method performs symmetric updates for both the primal and dual variables, with each update depending on both current and previous iterates. This symmetric design not only reflects the intrinsic structure of the saddle-point formulation but also facilitates improved convergence behavior. We establishes global convergence of the method under relaxed step-size conditions and provides both ergodic and nonergodic convergence rate guarantees. These results highlight the benefits of exploiting algorithmic symmetry and offer a new perspective on the design of first-order methods for structured optimization problems.

REFERENCES

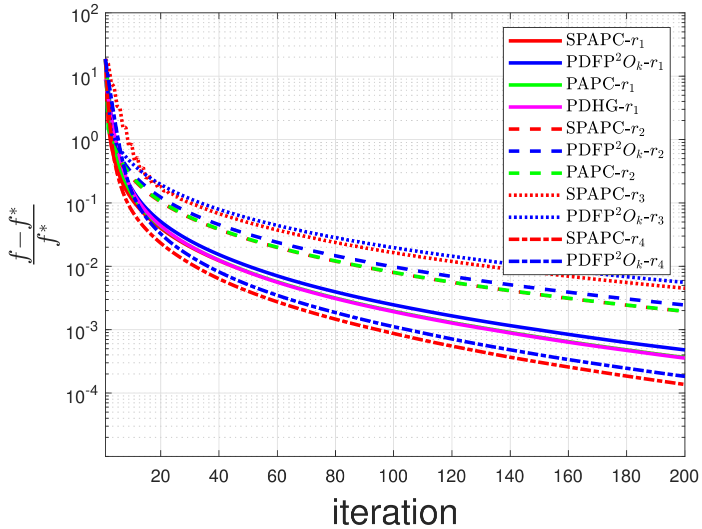


FIG. 9. *Numerical result* $\frac{\varphi^k-\varphi^*}{\varphi^*}$ *of* (1.11) *for* (4.5) *under different methods and the different primal step-size parameter* r*:* $r_1 = \frac{1}{2L_f}, r_2 = \frac{1}{L_f}, r_3 = \frac{3}{2L_f}, r_4 = \frac{\sqrt{2}}{4L_f}$

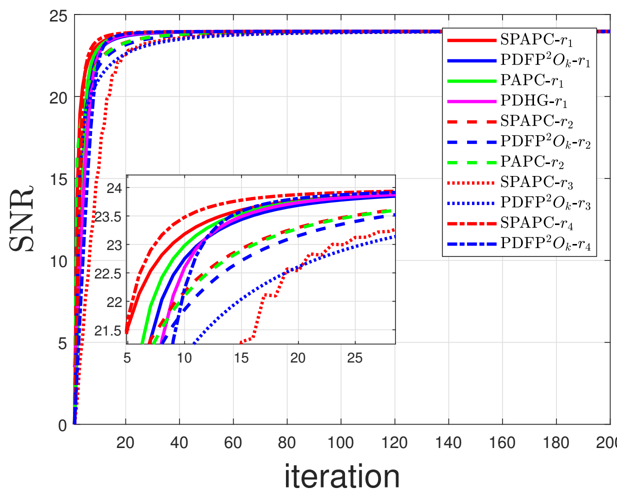


FIG. 10. *Numerical result* SNR *of* (1.11) *for* (4.5) *under different methods and the different step-size parameter* r*:* $r_1 = \frac{1}{2L_f}, r_2 = \frac{1}{L_f}, r_3 = \frac{3}{2L_f}, r_4 = \frac{\sqrt{2}}{4L_f}$

[1] H. BAUSCHKE AND P. L. COMBETTES, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2017.
[2] S. P. BOYD AND L. VANDENBERGHE, *Convex optimization*, Cambridge university press, 2004.
[3] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of mathematical imaging and vision, 40 (2011), pp. 120–145.
[4] A. CHAMBOLLE AND T. POCK, *An introduction to continuous optimization for imaging*, Acta Numerica, 25 (2016), pp. 161–319.
[5] A. CHAMBOLLE AND T. POCK, *On the ergodic convergence rates of a first-order primal–dual algorithm*, Mathematical Programming, 159 (2016), pp. 253–287.

[6] P. Chen, J. Huang, and X. Zhang, *A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration*, Inverse Problems, 29 (2013), p. 025011.

[7] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. Vũ, *A forward-backward view of some primal-dual optimization methods in image recovery*, in 2014 IEEE International Conference on Image Processing (ICIP), IEEE, 2014, pp. 4141–4145.

[8] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale modeling & simulation, 4 (2005), pp. 1168–1200.

[9] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, *Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists*, SIAM Review, 65 (2023), pp. 375–435.

[10] Y. Drori, S. Sabach, and M. Teboulle, *A simple algorithm for a class of nonsmooth convex–concave saddle-point problems*, Operations Research Letters, 43 (2015), pp. 209–214.

[11] E. Esser, X. Zhang, and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1015–1046.

[12] D. Goldfarb, S. Ma, and K. Scheinberg, *Fast alternating linearization methods for minimizing the sum of two convex functions*, Mathematical Programming, 141 (2013), pp. 349–382.

[13] M. Grant and S. P. Boyd, *Cvx: Matlab software for disciplined convex programming, version 2.1*, 2014, https://cvxr.com/cvx.

[14] B. He, *PPA-like contraction methods for convex optimization: a framework using variational inequality approach*, Journal of the Operations Research Society of China, 3 (2015), pp. 391–420.

[15] B. He, H. Liu, Z. Wang, and X. Yuan, *A strictly contractive peaceman–rachford splitting method for convex programming*, SIAM Journal on Optimization, 24 (2014), pp. 1011–1040.

[16] B. He, F. Ma, S. Xu, and X. Yuan, *A generalized primal-dual algorithm with improved convergence condition for saddle point problems*, SIAM Journal on Imaging Sciences, 15 (2022), pp. 1157–1183.

[17] B. He, F. Ma, and X. Yuan, *Convergence study on the symmetric version of ADMM with larger step sizes*, SIAM journal on imaging sciences, 9 (2016), pp. 1467–1501.

[18] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 119–149.

[19] B. He and X. Yuan, *A class of admm-based algorithms for three-block separable convex programming*, Computational Optimization and Applications, 70 (2018), pp. 791–826.

[20] N. Komodakis and J.-C. Pesquet, *Playing with duality: An overview of recent primal? dual approaches for solving large-scale optimization problems*, IEEE Signal Processing Magazine, 32 (2015), pp. 31–54.

[21] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, 16 (1979), pp. 964–979.

[22] I. Loris and C. Verhoeven, *On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty*, Inverse Problems, 27 (2011), p. 125007.

[23] F. Ma, S. Li, and X. Zhang, *A symmetric version of the generalized chambolle-pock-he-yuan method for saddle point problems*, Computational Optimization and Applications, (2025), pp. 1–26.

[24] C. A. Micchelli, L. Shen, and Y. Xu, *Proximity algorithms for image models: denoising*, Inverse Problems, 27 (2011), p. 045009.

[25] D. O'Connor and L. Vandenberghe, *On the equivalence of the primal-dual hybrid gradient method and douglas–rachford splitting*, Mathematical Programming, 179 (2020), pp. 85–108.

[26] D. W. Peaceman and H. H. Rachford, Jr, *The numerical solution of parabolic and elliptic differential equations*, Journal of the Society for industrial and Applied Mathematics, 3 (1955), pp. 28–41.

[27] T. Valkonen, *First-order primal–dual methods for nonsmooth non-convex optimization*, in Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, Springer, 2022, pp. 1–42.

[28] M.-L. Vladarean, Y. Malitsky, and V. Cevher, *A first-order primal-dual method with adaptivity to local smoothness*, Advances in neural information processing systems, 34 (2021), pp. 6171–6182.

[29] M. Yan, *A new primal–dual algorithm for minimizing the sum of three functions with a linear operator*, Journal of Scientific Computing, 76 (2018), pp. 1698–1717.