Optimization Letters manuscript No.
(will be inserted by the editor)

The operator splitting schemes revisited: primal-dual gap and degeneracy reduction by a unified analysis

Feng Xue

Received: date / Accepted: date

Abstract We revisit the operator splitting schemes proposed in a recent work of [*Some extensions of the operator splitting schemes based on Lagrangian and primal-dual: A unified proximal point analysis*, Feng Xue, Optimization, 2022, doi: 10.1080/02331934.2022.2057309], and further analyze the convergence of the generalized Bregman distance and the primal-dual gap of these algorithms within a unified proximal point framework. The possibility of reduction to a simple resolvent is also discussed by exploiting the structure and possible degeneracy of the underlying metric.

Keywords Operator splitting · proximal point algorithms · primal-dual gap · generalized Bregman distance · degeneracy · resolvent

Mathematics Subject Classification (2010) 47H05 · 49M29 · 49M27 · 90C25

1 Introduction

The present paper discusses the operator splitting schemes for solving [41, Eq.(1)]

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{A} : \mathbb{R}^N \mapsto \mathbb{R}^M$ is a linear operator, $f : \mathbb{R}^N \mapsto \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^M \mapsto \mathbb{R} \cup \{+\infty\}$ are proper, lower semi-continuous (l.s.c.), convex (not necessarily smooth) and proximable[1] functions. As observed in [18, Sect. 2.1] and

Feng Xue
National key laboratory, Beijing, China
E-mail: fxue@link.cuhk.edu.hk

[1] We say a convex function f is *proximable*, if the proximity operator of f has a closed-form representation or at least can be solved efficiently up to high precision [17]. This property is also called '*simple*' [17] or '*with inexpensive proximity operator*' [36].

[46, Sect. 1.1], the problem (1) also covers the minimization of the sum of multiple functions g_i composed with linear operators $\mathbf{A}_i$, i.e. $\min_{\mathbf{x}} \sum_{i=1}^I g_i(\mathbf{A}_i\mathbf{x})$, if we define $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_I \end{bmatrix} : \mathbb{R}^N \mapsto \mathbb{R}^M$, with $\mathbf{A}_i : \mathbb{R}^N \mapsto \mathbb{R}^{M_i}$ and $M = \sum_{i=1}^I M_i$, $g : \mathbb{R}^{M_1} \times \cdots \times \mathbb{R}^{M_I} \mapsto \mathbb{R} \cup \{+\infty\} : (\mathbf{a}_1, \cdots, \mathbf{a}_I) \mapsto \sum_{i=1}^I g_i(\mathbf{a}_i)$. This problem can be solved by many classes of operator splitting algorithms, e.g., Douglas-Rachford splitting (DRS) [31], primal-dual splitting (PDS) [51,11], the alternating direction method of multipliers (ADMM) [22,21], Bregman methods [35,49,23,50], and so on[2].

In the recent work of [46], we gave a brief review of the typical ADMM-type and PDS algorithms, and presented a unified proximal point treatment. Following this work, we in this paper attempt to answer two important questions:

1. *What value do these operator splitting algorithms attempt to minimize? Is it possible to analyze the convergence within the unified proximal point framework?*
2. *Observing that some splitting strategies generate auxiliary variables that maybe redundant in the iterations, can they be reduced to a simpler form with smallest number of variables? Is it possible to detect and reduce the degeneracy or redundancy under the unified proximal point analysis?*

The contributions of this paper are in order.

– We show more possibilities of devising new algorithms than [46] in a more systematic way, according to the metric structure.
– We give an affirmative answer to the above first question: what the operator splitting algorithms try to minimize is the *Bregman distance* of some convex functional (and the associated *primal-dual gap* under additional conditions). It can be inferred by the proximal point framework (Sect. 3.4, 4.2 and 5.2), which enables us to perform a unified gap analysis of all the schemes developed in [46], that is much simpler than the existing case studies of specific algorithms, e.g. [7, Theorem 2.1] and [11, Theorem 1].
– The unified proximal point interpretation paves a way for expressing many algorithms proposed in [46] as a simple resolvent. More remarkably, by exploiting the metric degeneracy, some algorithms, particularly the standard ADMM/DRS, can be reduced to a simple resolvent involving only active variables and an explicit expression of the associated maximally monotone operator. This is also a case study of the degenerate analysis recently proposed in [8].

[2] Note that the well-known *proximal forward-backward splitting* (PFBS) algorithm may not in general be applied to solve (1), since neither f nor g is assumed to be differentiable with a Lipschitz continuous gradient.

2 Preliminaries

2.1 Notations and definitions

We use standard notations and concepts from convex analysis and variational analysis, which, unless otherwise specified, can all be found in the classical and recent monographs [4, 5, 38, 39].

A few more words about our notations are as follows. The class of symmetric and positive semi-definite/definite matrices is denoted by $\mathbb{S}_+$ or $\mathbb{S}_{++}$, respectively. We use the boldface uppercase to denote a matrix, e.g., $\mathbf{M}$, the calligraphic uppercase to denote a block-structured matrix or operator, e.g., $\mathcal{M}$. The identity operator and identity matrix of size $N \times N$ are denoted by $\mathcal{I}$ and $\mathbf{I}_N$. The $\mathbf{M}$-norm with $\mathbf{M} \in \mathbb{S}_+$ is defined as: $\|\mathbf{x}\|_{\mathbf{M}}^2 := \langle \mathbf{x}|\mathbf{M}\mathbf{x}\rangle$.

The generalized proximity operator, denoted by $\mathsf{prox}_f^{\mathbf{M}}$ is defined as $\mathsf{prox}_f^{\mathbf{M}}: \mathbf{x} \mapsto \arg\min_{\mathbf{u}} f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|_{\mathbf{M}}^2$, with $\mathbf{M} \in \mathbb{S}_+$ as an induced metric [42, Eq.(4)], [13, Definition 2.3]. If $\mathbf{M} = \frac{1}{\tau}\mathbf{I}$ (i.e. the scalar case), the generalized proximity operator reduces to ordinary one, denoted by $\mathsf{prox}_{\tau f}: \mathbf{x} \mapsto \arg\min_{\mathbf{u}} f(\mathbf{u}) + \frac{1}{2\tau}\|\mathbf{u}-\mathbf{x}\|^2$ [4, Definition 12.23], [16, Eq.(2.13)].

The classical Bregman distance associated with the function φ between $\mathbf{x}$ and $\mathbf{y}$ is defined as $D_\varphi(\mathbf{x},\mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \nabla\varphi(\mathbf{y})|\mathbf{x}-\mathbf{y}\rangle$, which requires the function φ to be differentiable and strictly convex. It was then extnded to the context of proper, l.s.c. and convex function φ along a direction of $\mathbf{v}$ within its subdifferential $\partial\varphi$ [30]:

$$D_\varphi^{\mathbf{v}}(\mathbf{x},\mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \mathbf{v}|\mathbf{x}-\mathbf{y}\rangle, \quad \mathbf{v} \in \partial\varphi(\mathbf{y}). \tag{2}$$

This plays a central role in various Bregman algorithms, e.g., [10, 23, 35, 50]. [33] further proposed two types of the generalized Bregman distance (2):

$$\begin{cases} D_\varphi^\sharp(\mathbf{x},\mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) + \sup_{\mathbf{v}\in\partial\mathbf{y}}\langle \mathbf{v}|\mathbf{y}-\mathbf{x}\rangle, \\ D_\varphi^\flat(\mathbf{x},\mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) + \inf_{\mathbf{v}\in\partial\mathbf{y}}\langle \mathbf{v}|\mathbf{y}-\mathbf{x}\rangle, \end{cases}$$

which are the upper and lower bounds of the Bregman distance generated by φ.

2.2 Some existing results of proximal point algorithm

The generalized proximal point algorithm (PPA) is given as

$$\left\lfloor \begin{array}{lll} \mathbf{0} & :\in \mathcal{A}\tilde{\mathbf{c}}^k + \mathcal{Q}(\tilde{\mathbf{c}}^k - \mathbf{c}^k), & \text{(proximal step)} \\ \mathbf{c}^{k+1} & := \mathbf{c}^k + \mathcal{M}(\tilde{\mathbf{c}}^k - \mathbf{c}^k), & \text{(relaxation step)} \end{array} \right. \tag{3}$$

where $\mathcal{A}$ is a (possibly set-valued) monotone operator, $\mathcal{Q}$ is a metric, $\mathcal{M}$ is an *invertible* relaxation matrix. The convergence of (3) has been extensively studied in the contexts of DRS, PDHG and multi-block ADMM algorithms, e.g., [29, 25, 28], and recently revisited in our recent works of [44, 46].

Here, we restate the main results therein with more straightforward proofs. More than that, our analysis admits any choices of $\mathcal{A}$, $\mathcal{Q}$ and $\mathcal{M}$, not limited to any specific algorithms.

Lemma 1 *Let $\{\mathbf{c}^k\}_{k\in\mathbb{N}}$ be a sequence generated by* (3) *and $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$. Denote $\mathcal{S} := \mathcal{Q}\mathcal{M}^{-1}$, $\mathcal{G} := \mathcal{Q} + \mathcal{Q}^\top - \mathcal{M}^\top\mathcal{Q}$. Denote the operator $\mathcal{T} := (\mathcal{A}+\mathcal{Q})^{-1}\mathcal{Q}$, and $\mathcal{R} := \mathcal{I} - \mathcal{T}$. If $\mathcal{A}$ is maximally monotone and $\mathcal{S}, \mathcal{G} \in \mathbb{S}_{++}$, then, the following hold.*

(i) $\|\mathbf{c}^{k+1} - \mathbf{c}^\star\|_{\mathcal{S}}^2 \le \|\mathbf{c}^k - \mathbf{c}^\star\|_{\mathcal{S}}^2 - \|\mathbf{c}^k - \mathbf{c}^{k+1}\|_{\mathcal{M}^{-\top}\mathcal{G}\mathcal{M}^{-1}}^2$;

(ii) $\langle \mathcal{M}^\top\mathcal{Q}\mathcal{R}\mathbf{c}^k | \mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\rangle \ge \frac{1}{2}\|\mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\|_{\mathcal{Q}+\mathcal{Q}^\top}^2$;

(iii) $\|\mathbf{c}^k - \mathbf{c}^{k+1}\|_{\mathcal{S}}^2 - \|\mathbf{c}^{k+1} - \mathbf{c}^{k+2}\|_{\mathcal{S}}^2 \ge \|\mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\|_{\mathcal{G}}^2$.

Proof (i) Based on (3), we have:

$$
\begin{aligned}
0 &\le \langle \mathcal{A}\tilde{\mathbf{c}}^k - \mathcal{A}\mathbf{c}^\star | \tilde{\mathbf{c}}^k - \mathbf{c}^\star\rangle \quad \text{by monotonicity of } \mathcal{A} \\
&= \langle \mathcal{Q}(\mathbf{c}^k - \tilde{\mathbf{c}}^k) | \tilde{\mathbf{c}}^k - \mathbf{c}^\star\rangle \quad \text{by (3) and } \mathbf{0} \in \mathcal{A}\mathbf{b}^\star \\
&= \langle \mathcal{Q}\mathcal{M}^{-1}(\mathbf{c}^k - \mathbf{c}^{k+1}) | \mathbf{c}^k + \mathcal{M}^{-1}(\mathbf{c}^{k+1} - \mathbf{c}^k) - \mathbf{c}^\star\rangle \quad \text{by (3)} \\
&= \langle \mathcal{S}(\mathbf{c}^k - \mathbf{c}^{k+1}) | \mathbf{c}^k - \mathbf{c}^\star\rangle - \frac{1}{2}\|\mathbf{c}^k - \mathbf{c}^{k+1}\|_{\mathcal{M}^{-\top}\mathcal{S}+\mathcal{S}\mathcal{M}^{-1}}^2 \\
&= \frac{1}{2}\|\mathbf{c}^k - \mathbf{c}^\star\|_{\mathcal{S}}^2 - \frac{1}{2}\|\mathbf{c}^{k+1} - \mathbf{c}^\star\|_{\mathcal{S}}^2 - \frac{1}{2}\|\mathbf{c}^k - \mathbf{c}^{k+1}\|_{\mathcal{M}^{-\top}\mathcal{G}\mathcal{M}^{-1}}^2, \quad \text{by } \mathcal{S} \in \mathbb{S}_{++}.
\end{aligned}
$$

(ii) By [45, Lemma 2.6], we obtain:

$$
\langle \mathbf{c}^k - \mathbf{c}^{k+1} | \mathcal{Q}\mathcal{R}\mathbf{c}^k - \mathcal{Q}\mathcal{R}\mathbf{c}^{k+1}\rangle \ge \|\mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\|_{\mathcal{Q}}^2 = \frac{1}{2}\|\mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\|_{\mathcal{Q}+\mathcal{Q}^\top}^2.
$$

Then, (ii) follows from $\mathbf{c}^k - \mathbf{c}^{k+1} = \mathcal{M}\mathcal{R}\mathbf{c}^k$.

(iii) We develop:

$$
\begin{aligned}
&\|\mathbf{c}^k - \mathbf{c}^{k+1}\|_{\mathcal{S}}^2 - \|\mathbf{c}^{k+1} - \mathbf{c}^{k+2}\|_{\mathcal{S}}^2 \\
&= \|\mathcal{M}\mathcal{R}\mathbf{c}^k\|_{\mathcal{S}}^2 - \|\mathcal{M}\mathcal{R}\mathbf{c}^{k+1}\|_{\mathcal{S}}^2 \quad \text{by (3)} \\
&= 2\langle \mathcal{M}^\top\mathcal{S}\mathcal{M}\mathcal{R}\mathbf{b}^k | \mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\rangle - \|\mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\|_{\mathcal{M}^\top\mathcal{S}\mathcal{M}}^2 \\
&\ge \|\mathcal{R}\mathbf{c}^k - \mathcal{R}\mathbf{c}^{k+1}\|_{\mathcal{G}}^2, \qquad \text{by Lemma 1–(ii) and } \mathcal{Q} = \mathcal{S}\mathcal{M}.
\end{aligned}
$$

Remark 1 Lemma 1–(i) can be found in [29, Theorem 1], [28, Theorem 4.1] and [25, Theorem 3.2]. The item (ii) is same as [29, Lemma 3], [28, Lemma 5.3] and [25, Lemma 5.4]. The item (iii) is a restatement of [29, Theorem 5] and [28, Theorem 5.1]. The proof presented here outlines the key ingredients only. Refer to [44, Lemma 5.2] for more details.

Then, the convergence properties of (3) are given as follows.

Theorem 1 (Convergence in terms of metric distance) *Under the notations and assumptions of Lemma 1, then the following hold.*

(i) [Basic convergence] *There exists* $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$*, such that* $\mathbf{c}^k \to \mathbf{c}^\star$*, as* $k \to \infty$.

(ii) [Asymptotic regularity] $\|\mathbf{c}^k - \mathbf{c}^{k+1}\|_{\mathcal{S}}$ *has the pointwise convergence rate of* $\mathcal{O}(1/\sqrt{k})$*, i.e.,*

$$\big\|\mathbf{c}^{k+1} - \mathbf{c}^k\big\|_{\mathcal{S}} \le \frac{1}{\sqrt{k+1}} \sqrt{\frac{\lambda_{\max}(\mathcal{S})}{\lambda_{\min}(\mathcal{M}^{-\top}\mathcal{G}\mathcal{M}^{-1})}} \big\|\mathbf{c}^0 - \mathbf{c}^\star\big\|_{\mathcal{S}}, \quad \forall k \in \mathbb{N},$$

where $\lambda_{\max}$ *and* $\lambda_{\max}$ *denote the largest and smallest eigenvalues of a matrix.*

Proof (i) The basic convergence is established based on Lemma 1–(i), invoking Opial's lemma [4, Lemma 2.47].

(ii) In view of Lemma 1–(i), we have:

$$\big\|\mathbf{c}^{i+1} - \mathbf{c}^\star\big\|_{\mathcal{S}}^2 \le \big\|\mathbf{c}^{i} - \mathbf{c}^\star\big\|_{\mathcal{S}}^2 - \frac{\lambda_{\min}(\mathcal{M}^{-\top}\mathcal{G}\mathcal{M}^{-1})}{\lambda_{\max}(\mathcal{S})} \big\|\mathbf{c}^{i} - \mathbf{c}^{i+1}\big\|_{\mathcal{S}}^2. \quad (4)$$

Finally, (ii) is obtained, by summing up (4) from $i = 0$ to k and noting that the sequence $\{\|\mathbf{c}^i - \mathbf{c}^{i+1}\|_{\mathcal{S}}\}_{i\in\mathbb{N}}$ is non-increasing (by Lemma 1–(iii)).

Remark 2 Refer to [44, Theorem 5.3] for more details. The non-ergodic rate of asymptotic regularity (ii) has also been established in [29, Theorem 6], [28, Theorem 6.1] and [25, Theorem 5.5].

In particular, if $\mathcal{M} = \mathcal{I}$, the scheme (3) reduces to a standard PPA: $\mathbf{0} \in \mathcal{A}\mathbf{c}^{k+1} + \mathcal{Q}(\mathbf{c}^{k+1} - \mathbf{c}^k)$, which can be rewritten as

$$\mathbf{c}^{k+1} := (\mathcal{A} + \mathcal{Q})^{-1}\mathcal{Q}\mathbf{c}^k, \quad (5)$$

whose convergence is given below.

Corollary 1 (Convergence of standard PPA) *Given the scheme* (5) *with maximally monotone* $\mathcal{A}$ *and metric* $\mathcal{Q} \in \mathbb{S}_{++}$*, the following hold.*

(i) [Basic convergence] *There exists* $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$*, such that* $\mathbf{c}^k \to \mathbf{c}^\star$*, as* $k \to \infty$.

(ii) [Asymptotic regularity] $\|\mathbf{c}^k - \mathbf{c}^{k+1}\|_{\mathcal{Q}}$ *has a pointwise convergence rate of* $\mathcal{O}(1/\sqrt{k})$*, i.e.*

$$\big\|\mathbf{c}^{k+1} - \mathbf{c}^k\big\|_{\mathcal{Q}} \le \frac{1}{\sqrt{k+1}} \big\|\mathbf{c}^0 - \mathbf{c}^\star\big\|_{\mathcal{Q}}, \quad \forall k \in \mathbb{N},$$

(iii) [Resolvent] *The scheme* (5) *can be rewritten as a resolvent form:*

$$\mathbf{c}^{k+1} := (\mathcal{I} + \mathcal{Q}^{-1} \circ \mathcal{A})^{-1}\mathbf{c}^k = \mathcal{Q}^{-\frac{1}{2}}\big(\mathcal{I} + \mathcal{Q}^{-\frac{1}{2}} \circ \mathcal{A} \circ \mathcal{Q}^{-\frac{1}{2}}\big)^{-1}\mathcal{Q}^{\frac{1}{2}}\mathbf{c}^k.$$

Proof (i) and (ii) follow from Theorem 1.

(iii) [44, Lemma 2.1-(iii)].

All the results presented in Sect. 2.2 require the associated metrics $\mathcal{Q}$ (or $\mathcal{S}, \mathcal{G}$) to be *strictly* positive definite. However, based on a recent analysis of [8], this condition can be sometimes loosened to positive *semi*-definite in the applications to operator splitting algorithms, which leads to some interesting reductions by removing redundant variables. See Sect. 3.5 and 4.3 for detailed discussions.

2.3 An extension of Moreau's decomposition identity

The following result extends the classical Moreau's decomposition identity (see, for instance, [16, Eq.(2.21)]) to arbitrary linear operator $\mathbf{A}$, and links the proximity operator of the infimal postcomposition of f by $\mathbf{A}$ to that of the conjugate f^*. The the notion of 'infimal postcomposition' was recently studied in [2] in details.

Lemma 2 *Given a proper, l.s.c. and convex function $f : \mathbb{R}^N \mapsto \mathbb{R} \cup \{+\infty\}$ and arbitrary matrix $\mathbf{A} : \mathbb{R}^N \mapsto \mathbb{R}^M$, the following holds:*

$$\mathsf{prox}_{\mathbf{A} \triangleright f} + \mathsf{prox}_{f^* \circ \mathbf{A}^\top} = \mathbf{I}_M,$$

where $\mathbf{A} \triangleright f$ denotes the infimal postcomposition of f by $\mathbf{A}$, defined as $\mathbf{A} \triangleright f : \mathbb{R}^M \mapsto \mathbb{R} : \mathbf{t} \mapsto \min_{\mathbf{Ax}=\mathbf{t}} f(\mathbf{x})$.

Proof First, incorporating a hard constraint of $\mathbf{t} = \mathbf{Ax}$ [48, Sect. 3], we have:

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) + \frac{1}{2}\|\mathbf{Ax} - \mathbf{u}\|^2 &= \min_{\mathbf{x},\mathbf{t}} f(\mathbf{x}) + \frac{1}{2}\|\mathbf{t} - \mathbf{u}\|^2 + \iota_{\{\mathbf{x}:\mathbf{Ax}=\mathbf{t}\}}(\mathbf{t}) \\ &= \min_{\mathbf{t}} F(\mathbf{t}) + \frac{1}{2}\|\mathbf{t} - \mathbf{u}\|^2, \end{aligned} \tag{6}$$

where ι_C is an indicator function of a set C, the function $F(\mathbf{t}) := \min_{\mathbf{x}} f(\mathbf{x}) + \iota_{\{\mathbf{x}:\mathbf{Ax}=\mathbf{t}\}}(\mathbf{x}) = \min_{\mathbf{Ax}=\mathbf{t}} f(\mathbf{x})$ is the so-called infimal postcomposition of f by $\mathbf{A}$, simply denoted as $F = \mathbf{A} \triangleright f$ [4, Definition 12.34]. (6) implies that $\mathbf{t}^\star = \mathsf{prox}_F(\mathbf{u}) = \mathbf{Ax}^\star$, where $\mathbf{x}^\star = \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{1}{2}\|\mathbf{Ax} - \mathbf{u}\|^2$.

Then, by Fenchel duality, the above is equivalent to a saddle-point problem:

$$\min_{\mathbf{t}} \max_{\mathbf{s}} \langle \mathbf{s}|\mathbf{t}\rangle - F^*(\mathbf{s}) + \frac{1}{2}\|\mathbf{t} - \mathbf{u}\|^2.$$

Exchanging the order of min and max, we have:

$$\max_{\mathbf{s}} \min_{\mathbf{t}} -F^*(\mathbf{s}) + \frac{1}{2}\|\mathbf{t} - \mathbf{u} + \mathbf{s}\|^2 - \frac{1}{2}\|\mathbf{s}\|^2 + \langle \mathbf{u}|\mathbf{s}\rangle,$$

which yields that $\mathbf{t}^\star = \mathbf{u} - \mathbf{s}^\star$, where $\mathbf{s}^\star = \arg\min_{\mathbf{s}} F^*(\mathbf{s}) + \frac{1}{2}\|\mathbf{s} - \mathbf{u}\|^2 = \mathsf{prox}_{F^*}(\mathbf{u})$. Finally, the proof is completed by noting that $F^* = f^* \circ \mathbf{A}^\top$ [48].

3 Operator splitting based on Lagrangian

3.1 The Lagrangian schemes and their PPA interpretations

First, we consider the Lagrangian of (1) [46, Eq.(13)]:

$$\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{p}) := f(\mathbf{x}) + g(\mathbf{a}) + \mathbf{p}^\top(\mathbf{A}\mathbf{x}-\mathbf{a}), \tag{7}$$

or generalized augmented Lagrangian:

$$\mathcal{L}_{\mathbf{\Gamma}}(\mathbf{x},\mathbf{a},\mathbf{p}) := f(\mathbf{x}) + g(\mathbf{a}) + \mathbf{p}^\top(\mathbf{A}\mathbf{x}-\mathbf{a}) + \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{a}\|_{\mathbf{\Gamma}}^2, \tag{8}$$

which extends the standard augmented Lagrangian [46, Eq.(3)] from the scalar penalty parameter γ to the matrix metric $\mathbf{\Gamma}$.

Then, similar to [13,20,42,24], defining the proximal metrics by $\mathbf{M}:\mathbb{R}^N\mapsto\mathbb{R}^N$, $\mathbf{\Omega}:\mathbb{R}^M\mapsto\mathbb{R}^M$, $\mathbf{\Gamma}:\mathbb{R}^M\mapsto\mathbb{R}^M$, the alternating optimization of (7) or (8) yields the algorithms listed in Table 1. `LAG-I,II,V,VI` and `VII` can be found in [46, Sect. 3 and 4], and are extended to general proximal metrics here. Table 2 shows the PPA reinterpretations of the schemes. One can check the PPA fitting by the similar procedure with [9,32,3,26], verify the convergence condition for each algorithm (shown in Table 3) by computing the corresponding $\mathcal{S}$ and $\mathcal{G}$ by Theorem 1, and further write down the specific convergence property of asymptotic regularity, which are omitted here. Also note that:

- `LAG-I` and `LAG-II` correspond to symmetric $\mathcal{Q}$ (without relaxation);
- `LAG-III` and `LAG-IV` correspond to upper triangular $\mathcal{Q}$;
- `LAG-V` and `LAG-VI` correspond to lower triangular $\mathcal{Q}$;
- `LAG-VII` corresponds to skew-symmetric $\mathcal{Q}$.

These algorithms can be interpreted by alternating optimization of some cost function. For instance, `LAG-I` and `LAG-VII` stem from the alternating optimization of non-augmented Lagrangian $\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{p})$. For example, both $\mathbf{x}$- and $\mathbf{a}$-updates of `LAG-I` come from

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x},\mathbf{a}^k,\mathbf{p}^k) + \frac{1}{2}\|\mathbf{x}-\mathbf{x}^k\|_{\mathbf{M}}^2, \\ \mathbf{a}^{k+1} = \arg\min_{\mathbf{a}} \mathcal{L}(\mathbf{x}^k,\mathbf{a},\mathbf{p}^k) + \frac{1}{2}\|\mathbf{a}-\mathbf{a}^k\|_{\mathbf{\Omega}}^2. \end{array}\right.$$

`LAG-V` and `LAG-VI` are based on the augmented Lagrangian $\mathcal{L}_{\mathbf{\Gamma}}(\mathbf{x},\mathbf{a},\mathbf{p})$. For example, the $\mathbf{x}$- and $\mathbf{a}$-updates of `LAG-V` are obtained by

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathcal{L}_{\mathbf{\Gamma}}(\mathbf{x},\mathbf{a}^k,\mathbf{p}^k) + \frac{1}{2}\|\mathbf{x}-\mathbf{x}^k\|_{\mathbf{M}}^2, \\ \mathbf{a}^{k+1} = \arg\min_{\mathbf{a}} \mathcal{L}_{\mathbf{\Gamma}}(\mathbf{x}^k,\mathbf{a},\mathbf{p}^k) + \frac{1}{2}\|\mathbf{a}-\mathbf{a}^k\|_{\mathbf{\Omega}}^2. \end{array}\right.$$

The $\mathbf{x}$- and $\mathbf{a}$-updates of `LAG-II`, `LAG-III` and `LAG-IV` are the hybrid optimizations of both non-augmented and augmented forms. For example, the $\mathbf{x}$-update of `LAG-IV` is from non-augmented, while the $\mathbf{a}$-update is from augmented, i.e.,

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x},\mathbf{a}^k,\mathbf{p}^k) + \frac{1}{2}\|\mathbf{x}-\mathbf{x}^k\|_{\mathbf{M}}^2, \\ \mathbf{a}^{k+1} = \arg\min_{\mathbf{a}} \mathcal{L}_{\mathbf{\Gamma}}(\mathbf{x}^{k+1},\mathbf{a},\mathbf{p}^k) + \frac{1}{2}\|\mathbf{a}-\mathbf{a}^k\|_{\mathbf{\Omega}}^2. \end{array}\right.$$

Table 1 The proposed Lagrangian-based algorithms

name	iterative scheme
`LAG-I` [46, Eq.(14)]	$\left\lfloor \begin{aligned} \mathbf{x}^{k+1} &:= \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^k\big) \\ \mathbf{a}^{k+1} &:= \mathsf{prox}_g^{\mathbf{\Omega}}(\mathbf{a}^k + \mathbf{\Omega}^{-1}\mathbf{p}^k) \\ \mathbf{p}^{k+1} &:= \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k) - (2\mathbf{a}^{k+1} - \mathbf{a}^k)\big) \end{aligned} \right.$
`LAG-II` [46, Eq.(24)]	$\left\lfloor \begin{aligned} \mathbf{x}^{k+1} &:= \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A}}\big((\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k + \mathbf{A}^\top\mathbf{\Gamma}\mathbf{a}^k - \mathbf{A}^\top\mathbf{p}^k)\big) \\ \mathbf{p}^{k+1} &= \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^k\big) \\ \mathbf{a}^{k+1} &= \mathsf{prox}_g^{\mathbf{\Omega}}\big(\mathbf{a}^k + \mathbf{\Omega}^{-1}(2\mathbf{p}^{k+1} - \mathbf{p}^k)\big) \end{aligned} \right.$
`LAG-III`	$\left\lfloor \begin{aligned} \mathbf{a}^{k+1} &:= \mathsf{prox}_g^{\mathbf{\Omega}}(\mathbf{a}^k + \mathbf{\Omega}^{-1}\mathbf{p}^k) \\ \mathbf{x}^{k+1} &:= \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A}}\big((\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k - \mathbf{A}^\top\mathbf{p}^k + \mathbf{A}^\top\mathbf{\Gamma}\mathbf{a}^{k+1})\big) \\ \mathbf{p}^{k+1} &:= \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{a}^k - 2\mathbf{a}^{k+1}\big) \end{aligned} \right.$
`LAG-IV`	$\left\lfloor \begin{aligned} \mathbf{x}^{k+1} &:= \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^k\big) \\ \mathbf{a}^{k+1} &:= \mathsf{prox}_g^{\mathbf{\Omega}+\mathbf{\Gamma}}\big((\mathbf{\Omega}+\mathbf{\Gamma})^{-1}(\mathbf{\Omega}\mathbf{a}^k + \mathbf{\Gamma}\mathbf{A}\mathbf{x}^{k+1} + \mathbf{p}^k)\big) \\ \mathbf{p}^{k+1} &:= \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k) - \mathbf{a}^{k+1}\big) \end{aligned} \right.$
`LAG-V` [46, Eq.(19)]	$\left\lfloor \begin{aligned} \mathbf{x}^{k+1} &:= \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A}}\big((\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k + \mathbf{A}^\top\mathbf{\Gamma}\mathbf{a}^k - \mathbf{A}^\top\mathbf{p}^k)\big) \\ \mathbf{a}^{k+1} &:= \mathsf{prox}_g^{\mathbf{\Omega}+\mathbf{\Gamma}}\big((\mathbf{\Omega}+\mathbf{\Gamma})^{-1}(\mathbf{\Omega}\mathbf{a}^k + \mathbf{\Gamma}\mathbf{A}\mathbf{x}^k + \mathbf{p}^k)\big) \\ \mathbf{p}^{k+1} &= \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^{k+1}\big) \end{aligned} \right.$
`LAG-VI` [46, Eq.(20)]	$\left\lfloor \begin{aligned} \mathbf{x}^{k+1} &:= \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A}}\big((\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k + \mathbf{A}^\top\mathbf{\Gamma}\mathbf{a}^k - \mathbf{A}^\top\mathbf{p}^k)\big) \\ \mathbf{a}^{k+1} &:= \mathsf{prox}_g^{\mathbf{\Omega}+\mathbf{\Gamma}}\big((\mathbf{\Omega}+\mathbf{\Gamma})^{-1}(\mathbf{\Omega}\mathbf{a}^k + \mathbf{\Gamma}\mathbf{A}\mathbf{x}^{k+1} + \mathbf{p}^k)\big) \\ \mathbf{p}^{k+1} &:= \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^{k+1}\big) \end{aligned} \right.$
`LAG-VII` [46, Eq.(18)]	$\left\lfloor \begin{aligned} \tilde{\mathbf{x}}^k &:= \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^k\big) \\ \tilde{\mathbf{a}}^k &:= \mathsf{prox}_g^{\mathbf{\Omega}}\big(\mathbf{a}^k + \mathbf{\Omega}^{-1}\mathbf{p}^k\big) \\ \mathbf{x}^{k+1} &:= \tilde{\mathbf{x}}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{\Gamma}(\mathbf{A}\mathbf{x}^k - \mathbf{a}^k) \\ \mathbf{a}^{k+1} &:= \tilde{\mathbf{a}}^k + \mathbf{\Omega}^{-1}\mathbf{\Gamma}(\mathbf{A}\mathbf{x}^k - \mathbf{a}^k) \\ \mathbf{p}^{k+1} &:= \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}\tilde{\mathbf{x}}^k - \tilde{\mathbf{a}}^k\big) \end{aligned} \right.$

Table 2 The PPA reinterpretations of the Lagrangian-based schemes

schemes	$\mathbf{c}$	$\mathcal{A}$	$\mathcal{Q}$	$\mathcal{M}$
`LAG-I`			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\mathbf{I}_{2M+N}$
`LAG-II`			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & -\mathbf{I}_M \\ \mathbf{0} & -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	
`LAG-III`			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{I}_M \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Gamma} & \mathbf{I}_M \end{bmatrix}$
`LAG-IV`	$\begin{bmatrix} \mathbf{x} \\ \mathbf{a} \\ \mathbf{p} \end{bmatrix}$	$\begin{bmatrix} \partial f & \mathbf{0} & \mathbf{A}^\top \\ \mathbf{0} & \partial g & -\mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{\Gamma}\mathbf{A} & \mathbf{0} & \mathbf{I}_M \end{bmatrix}$
`LAG-V`			$\begin{bmatrix} \mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}+\mathbf{\Gamma} & \mathbf{0} \\ \mathbf{A} & -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{\Gamma}\mathbf{A} & -\mathbf{\Gamma} & \mathbf{I}_M \end{bmatrix}$
`LAG-VI`			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}+\mathbf{\Gamma} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Gamma} & \mathbf{I}_M \end{bmatrix}$
`LAG-VII`			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{I}_M \\ \mathbf{A} & -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & -\mathbf{M}^{-1}\mathbf{A}^\top \\ \mathbf{0} & \mathbf{I}_M & \mathbf{\Omega}^{-1} \\ \mathbf{\Gamma}\mathbf{A} & -\mathbf{\Gamma} & \mathbf{I}_M \end{bmatrix}$

The preconditioning technique [11, Sect. 4.3] can be applied to the $\mathbf{x}$-updates of `LAG-II,III,V,VI` and $\mathbf{a}$-updates of `LAG-IV,V,VI`, see [46, Sect. 4.1] for more details.

Table 3 The corresponding $\mathcal{S}$ and $\mathcal{G}$ of the Lagrangian-based schemes

schemes	$\mathcal{S}$	$\mathcal{G}$	convergence condition
`LAG-I`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\mathbf{M}, \mathbf{\Omega}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{\Gamma}^{-1} \succ \mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top + \mathbf{\Omega}^{-1}$
`LAG-II`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & -\mathbf{I}_M \\ \mathbf{0} & -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & -\mathbf{I}_M \\ \mathbf{0} & -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\mathbf{M} \in \mathbb{S}_{+}$ $\mathbf{\Omega}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{\Omega} \succ \mathbf{\Gamma}$
`LAG-III`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}+\mathbf{\Gamma} & \mathbf{I}_M \\ \mathbf{0} & \mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{I}_M \\ \mathbf{0} & \mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$	
`LAG-IV`	$\begin{bmatrix} \mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A} & \mathbf{0} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} \\ -\mathbf{A} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} \\ -\mathbf{A} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\mathbf{M}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{\Omega} \in \mathbb{S}_{+}$ $\mathbf{M} \succ \mathbf{A}^\top\mathbf{\Gamma}\mathbf{A}$
`LAG-V`	$\begin{bmatrix} \mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}+\mathbf{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{A}^\top\mathbf{\Gamma} & \mathbf{0} \\ \mathbf{\Gamma}\mathbf{A} & \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\mathbf{M}, \mathbf{\Omega}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{A}^\top\mathbf{\Gamma}\mathbf{\Omega}^{-1}\mathbf{\Gamma}\mathbf{A}$
`LAG-VI`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}+\mathbf{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\mathbf{M}, \mathbf{\Omega} \in \mathbb{S}_{+}$ $\mathbf{\Gamma} \in \mathbb{S}_{++}$
`LAG-VII`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M}-\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A} & \mathbf{A}^\top\mathbf{\Gamma} & \mathbf{0} \\ \mathbf{\Gamma}\mathbf{A} & \mathbf{\Omega}-\mathbf{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1}-\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top-\mathbf{\Omega}^{-1} \end{bmatrix}$	$\mathbf{M}, \mathbf{\Omega}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{\Gamma}^{-1} \succ \mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top + \mathbf{\Omega}^{-1}$

If $\mathbf{\Gamma} = \gamma \mathbf{I}_M$, `LAG-V` and `LAG-VI` reduce to [46, Eqs.(19) and (20)]. Their comparisons and connections to [40, Algorithms 1 and 2] have been discussed in [46, Sect. 4.1]. In addition, the convergence condition of [46, Eq.(19)], by [40, Proposition 5.2 and Theorem 5.1], is $\mathbf{M} \succ \gamma \mathbf{A}^\top \mathbf{A}$ and $\mathbf{\Omega} \succ \gamma \mathbf{I}_M$. Our analysis in Table 3 shows that this condition can be relaxed to $\mathbf{M} \succ \gamma^2 \mathbf{A}^\top \mathbf{\Omega}^{-1} \mathbf{A}$, which is obviously milder than $\mathbf{M} \succ \gamma \mathbf{A}^\top \mathbf{A}$ and $\mathbf{\Omega} \succ \gamma \mathbf{I}_M$.

Finally, note that the monotone operator $\mathcal{A}$ represents the optimality condition of (7). Indeed, $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$ implies the KKT conditions: $-\mathbf{A}^\top \mathbf{p}^\star \in \partial f(\mathbf{x}^\star)$, $\mathbf{p}^\star \in \partial g(\mathbf{a}^\star)$ and $\mathbf{a}^\star = \mathbf{A}\mathbf{x}^\star$, i.e. $\mathbf{0} \in \partial f(\mathbf{x}^\star) + \mathbf{A}^\top \partial g(\mathbf{A}\mathbf{x}^\star)$. This is the reason for why all the Lagrangian-based schemes in Table 1 share the same $\mathcal{A}$.

Another important observation is that $\mathcal{A}$ bears a typical (diagonal) monotone + (off-diagonal) skew-symmetric structure:

$$\begin{bmatrix} \partial f & \mathbf{0} & \mathbf{A}^\top \\ \mathbf{0} & \partial g & -\mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{0} \end{bmatrix} = \underbrace{\begin{bmatrix} \partial f & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \partial g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\text{monotone}} + \underbrace{\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{A}^\top \\ \mathbf{0} & \mathbf{0} & -\mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{0} \end{bmatrix}}_{\text{skew}},$$

which has also been noticed in [1,9,15]. This remark also applies to other classes of algorithms, see Sect. 4 and 5.

3.2 Connections to existing algorithms

[46, Sect. 3.1] discussed the connection of a special case of `LAG-I` to PDHG. We here show more connections.

3.2.1 `LAG-I`: two forms of PDHG

Letting $\mathbf{u} := \begin{bmatrix}\mathbf{x}\\ \mathbf{a}\end{bmatrix}$, $q(\mathbf{u}) := f(\mathbf{x}) + g(\mathbf{a})$, $\mathbf{U} := \begin{bmatrix}\mathbf{A} & -\mathbf{I}_M\end{bmatrix}$, the Lagrangian (7) is compactly given as

$$\mathcal{L}(\mathbf{u}, \mathbf{p}) = q(\mathbf{u}) + \langle \mathbf{p} | \mathbf{U}\mathbf{u} \rangle. \tag{9}$$

This Lagrangian objective consists of the primal part of $q(\mathbf{u})$, the dual part of 0, and their interplay represented by $\langle \mathbf{p}|\mathbf{U}\mathbf{u}\rangle$. `LAG-I` is equivalently written as

$$\left\lfloor \begin{array}{l} \mathbf{u}^{k+1} := \mathsf{prox}_q^{\mathbf{R}}\big(\mathbf{u}^k - \mathbf{R}^{-1}\mathbf{U}^\top \mathbf{p}^k\big), \\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Gamma}\mathbf{U}\big(2\mathbf{u}^{k+1} - \mathbf{u}^k\big), \end{array} \right.$$

where $\mathbf{R} = \begin{bmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}\end{bmatrix}$. This is essentially a special case of `PDS-I` in Sect. 4.1, where $\mathbf{u}$-step is a primal update, $\mathbf{p}$-step is a dual update. Compared to the commonly used primal-dual form (27), `LAG-I` associated with the Lagrangian (7) or (9) adopts a different splitting strategy, which treats $\mathbf{u} = (\mathbf{x}, \mathbf{a})$ as primal variable and $\mathbf{p}$ as dual, whereas (27) treats $f(\mathbf{x})$ as primal and $g^*(\mathbf{p})$ as dual.

The following proposition shows that under a certain condition, `LAG-I` can be simplified to the alternating updates between f and g^*, which coincides with the splitting strategy of (27). This result also extends the discussion in [46, Sect. 3.1] to general proximal metrics, and thus, the proof is omitted.

Proposition 1 *Given `LAG-I`, then, the following hold.*

(i) *`LAG-I` is equivalent to*

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top \mathbf{p}^k\big), \\ \mathbf{s}^{k+1} := \mathsf{prox}_{g^*}^{\boldsymbol{\Omega}^{-1}}(\boldsymbol{\Omega}\mathbf{a}^k + \mathbf{p}^k), \\ \mathbf{a}^{k+1} := \mathbf{a}^k + \boldsymbol{\Omega}^{-1}(\mathbf{p}^k - \mathbf{s}^{k+1}), \\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Gamma}\big(\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k) - (2\mathbf{a}^{k+1} - \mathbf{a}^k)\big). \end{array} \right. \tag{10}$$

(ii) *If $\boldsymbol{\Omega} = 2\boldsymbol{\Gamma}$, $\mathbf{s}^k = \mathbf{p}^k$,* (10) *reduces to*

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top \mathbf{p}^k\big), \\ \mathbf{p}^{k+2} := \mathsf{prox}_{g^*}^{\boldsymbol{\Omega}^{-1}}\big(\mathbf{p}^k + \boldsymbol{\Omega}\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k)\big). \end{array} \right. \tag{11}$$

Observe that the scheme (11) is essentially `PDS-I`—a generalized version of PDHG [46, Eq. (8)], which will be discussed in Sect. 4.

If one chooses $\boldsymbol{\Omega} = \boldsymbol{\Gamma}$ in `LAG-I` (which violates the convergence condition), and $\mathbf{s}^k = \mathbf{p}^k$, then combining the updates of $\mathbf{s}$, $\mathbf{a}$ and $\mathbf{p}$ in (10), we obtain $\mathbf{p}^{k+2} = \mathsf{prox}_{g^*}^{\boldsymbol{\Gamma}^{-1}}\big(\mathbf{p}^{k+1} + \boldsymbol{\Gamma}\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k)\big)$. Thus, `LAG-I` becomes

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top \mathbf{p}^k\big), \\ \mathbf{p}^{k+1} := \mathsf{prox}_{g^*}^{\boldsymbol{\Gamma}^{-1}}\big(\mathbf{p}^k + \boldsymbol{\Gamma}\mathbf{A}(2\mathbf{x}^k - \mathbf{x}^{k-1})\big). \end{array} \right.$$

This is a PDHG-like algorithm, but with illogical and weird update (noting that $\mathbf{p}^{k+1}$ is obtained without using $\mathbf{x}^{k+1}$). It is not guaranteed to converge, due to the unreasonable assumption $\boldsymbol{\Omega} = \boldsymbol{\Gamma}$.

3.2.2 `LAG-V`: semi-implicit Arrow-Hurwicz scheme

We now show that `LAG-V` is essentially an instance of the classical semi-implicit Arrow-Hurwicz scheme.

Let $\mathbf{u} := \begin{bmatrix}\mathbf{x}\\ \mathbf{a}\end{bmatrix}$, $q(\mathbf{u}) := f(\mathbf{x}) + g(\mathbf{a})$, $\mathbf{U} := \begin{bmatrix}\mathbf{A} & -\mathbf{I}_M\end{bmatrix}$, the augmented Lagrangian (8) is compactly written as

$$\mathcal{L}_{\boldsymbol{\Gamma}}(\mathbf{u},\mathbf{p}) = q(\mathbf{u}) + \langle \mathbf{p}|\mathbf{U}\mathbf{u}\rangle + \frac{1}{2}\|\mathbf{U}\mathbf{u}\|_{\boldsymbol{\Gamma}}^2.$$

With the variable metrics $\mathbf{R}$ and $\boldsymbol{\Gamma}$, the *semi-implicit* Arrow-Hurwicz scheme is given by

$$\left\lfloor \begin{array}{l} \mathbf{u}^{k+1} :\in \mathbf{u}^k - \mathbf{R}^{-1}\big(\partial q(\mathbf{u}^{k+1}) + \mathbf{U}^\top\mathbf{p}^k + \mathbf{U}^\top\boldsymbol{\Gamma}\mathbf{U}\mathbf{u}^{k+1}\big),\\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Gamma}\mathbf{U}\mathbf{u}^{k+1}, \end{array}\right.$$

i.e.,

$$\left\lfloor \begin{array}{l} \mathbf{u}^{k+1} := \mathsf{prox}_q^{\mathbf{R}+\mathbf{U}^\top\boldsymbol{\Gamma}\mathbf{U}}\big((\mathbf{R}+\mathbf{U}^\top\boldsymbol{\Gamma}\mathbf{U})^{-1}(\mathbf{R}\mathbf{u}^k - \mathbf{U}^\top\mathbf{p}^k)\big),\\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Gamma}\mathbf{U}\mathbf{u}^{k+1}. \end{array}\right. \tag{12}$$

The equivalent PPA form is given as

$$\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix} \in \begin{bmatrix}\partial q & \mathbf{U}^\top\\ -\mathbf{U} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{u}^{k+1}\\ \mathbf{p}^{k+1}\end{bmatrix} + \begin{bmatrix}\mathbf{R} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Gamma}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{u}^{k+1}-\mathbf{u}^k\\ \mathbf{p}^{k+1}-\mathbf{p}^k\end{bmatrix},$$

for which it is easy to show the convergence.

Furthermore, if one chooses $\mathbf{R} = \begin{bmatrix}\mathbf{M} & \mathbf{A}^\top\boldsymbol{\Gamma}\\ \boldsymbol{\Gamma}\mathbf{A} & \boldsymbol{\Omega}\end{bmatrix}$, such that $\mathbf{R} + \mathbf{U}^\top\boldsymbol{\Gamma}\mathbf{U} = \begin{bmatrix}\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Omega}+\boldsymbol{\Gamma}\end{bmatrix}$, which makes $\mathbf{x}$ and $\mathbf{a}$ to be fully decoupled, the Arrow-Hurwicz scheme (12) can be split into $(\mathbf{x},\mathbf{a},\mathbf{p})$:

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A}}\big(\mathbf{x}^k - (\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A})^{-1}(\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A}\mathbf{x}^k - \mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{a}^k + \mathbf{A}^\top\mathbf{p}^k)\big),\\ \mathbf{a}^{k+1} := \mathsf{prox}_g^{\boldsymbol{\Omega}+\boldsymbol{\Gamma}}\big(\mathbf{a}^k - (\boldsymbol{\Omega}+\boldsymbol{\Gamma})^{-1}(-\boldsymbol{\Gamma}\mathbf{A}\mathbf{x}^k + \boldsymbol{\Omega}\mathbf{a}^k - \mathbf{p}^k)\big),\\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Gamma}(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^{k+1}), \end{array}\right.$$

which is exactly `LAG-V`. The convergence condition (as shown in Table 3) follows from $\mathbf{R} \succ \mathbf{0}$.

3.2.3 `LAG-VI`: ADMM and PDHG

`LAG-VI` is essentially a proximal ADMM with proximal metrics $\mathbf{M}$ and $\boldsymbol{\Omega}$. We now show the connection of `LAG-VI` to PDHG.

Proposition 2 *Given `LAG-VI`, the following hold:*

(i) *If $\boldsymbol{\Omega} = \mathbf{0}$, `LAG-VI` is equivalent to*

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A}}\big(\mathbf{x}^k - (\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A})^{-1}\mathbf{A}^\top(2\mathbf{p}^k - \mathbf{p}^{k-1})\big),\\ \mathbf{p}^{k+1} := \mathsf{prox}_{g^*}^{\boldsymbol{\Gamma}^{-1}}\big(\boldsymbol{\Gamma}\mathbf{A}\mathbf{x}^{k+1} + \mathbf{p}^k\big), \end{array}\right. \tag{13}$$

(ii) *If* $\mathbf{M} = \frac{1}{\tau}\mathbf{I}_N - \gamma\mathbf{A}^\top\mathbf{A}$, $\boldsymbol{\Omega} = \mathbf{0}$, $\boldsymbol{\Gamma} = \gamma\mathbf{I}_M$, `LAG-VI` *reduces to the PDHG [46, Eq.(9)]:*

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_{\tau f}\big(\mathbf{x}^k - \tau\mathbf{A}^\top(2\mathbf{p}^k - \mathbf{p}^{k-1})\big), \\ \mathbf{p}^{k+1} := \mathsf{prox}_{\gamma g^*}\big(\mathbf{p}^k + \gamma\mathbf{A}\mathbf{x}^{k+1}\big), \end{array} \right.$$

Proof (i) If $\boldsymbol{\Omega} = \mathbf{0}$, similar to Proposition 1–(i), `LAG-VI` is equivalent to

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A}}\big((\mathbf{M} + \mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k + \mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{a}^k - \mathbf{A}^\top\mathbf{p}^k)\big), \\ \mathbf{s}^{k+1} := \mathsf{prox}_{g^*}^{\boldsymbol{\Gamma}^{-1}}\big(\boldsymbol{\Gamma}\mathbf{A}\mathbf{x}^{k+1} + \mathbf{p}^k)\big), \\ \mathbf{a}^{k+1} := \mathbf{A}\mathbf{x}^{k+1} + \boldsymbol{\Gamma}^{-1}(\mathbf{p}^k - \mathbf{s}^{k+1}), \\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Gamma}\big(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^{k+1}\big), \end{array} \right.$$

which yields that $\mathbf{s}^k = \mathbf{p}^k$. Substituting $\mathbf{a}^k = \mathbf{A}\mathbf{x}^k + \boldsymbol{\Gamma}^{-1}(\mathbf{p}^{k-1} - \mathbf{p}^k)$ into $\mathbf{x}$-update completes the proof.

(ii) clear.

Observe that the scheme (13) is essentially a generalized version of PDHG [46, Eq.(9)], the corresponding PPA form is given as

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \begin{bmatrix} \partial f & \mathbf{A}^\top \\ -\mathbf{A} & \partial g^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^{k+1} \\ \mathbf{p}^k \end{bmatrix} + \begin{bmatrix} \mathbf{M} + \mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A} & \mathbf{A}^\top \\ \mathbf{A} & \boldsymbol{\Gamma}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}^{k+1} - \mathbf{x}^k \\ \mathbf{p}^k - \mathbf{p}^{k-1} \end{bmatrix}.$$

We consider `LAG-V` as a comparison with `LAG-VI`. If one chooses $\boldsymbol{\Omega} = \mathbf{0}$ (which violates the convergence condition), following the similar steps of Proposition 2, `LAG-V` becomes

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A}}\big(\mathbf{x}^k - (\mathbf{M} + \mathbf{A}^\top\boldsymbol{\Gamma}\mathbf{A})^{-1}\mathbf{A}^\top(2\mathbf{p}^k - \mathbf{p}^{k-1})\big), \\ \mathbf{p}^{k+1} := \mathsf{prox}_{g^*}^{\boldsymbol{\Gamma}^{-1}}\big(\boldsymbol{\Gamma}\mathbf{A}\mathbf{x}^k + \mathbf{p}^k\big), \end{array} \right.$$

where the $\mathbf{p}$-update is illogical and weird (noting that $\mathbf{p}^{k+1}$ is computed *without* using $\mathbf{x}^{k+1}$). It is not guaranteed to converge, due to the unreasonable assumption $\boldsymbol{\Omega} = \mathbf{0}$.

3.3 The generalized Bregman distance

We will use the PPA interpretations to show that *the objective value that the Lagrangian schemes in Table 1 try to minimize is essentially an instance of generalized Bregman distance associated with* $f(\mathbf{x}) + g(\mathbf{a})$.

First, we define a quantity[3]:

$$\Pi(\mathbf{c}, \mathbf{c}') := \mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{p}') - \mathcal{L}(\mathbf{x}', \mathbf{a}', \mathbf{p}),$$

where $\mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{p})$ is given by (7). Given the schemes in Table 1, the following lemma presents a key inequality, which directly connects $\Pi(\tilde{\mathbf{c}}^k, \mathbf{c})$ to the metric $\mathcal{Q}$.

[3] For any pair of $(\mathbf{c}, \mathbf{c}')$, the quantity of $\Pi(\mathbf{c}, \mathbf{c}')$ is generally only a difference, but not a distance, since it is not guaranteed to be non-negative.

Lemma 3 *Given the Lagrangian $\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{p})$ as (7), consider all the Lagrangian-based schemes listed in Table 1, where $\tilde{\mathbf{c}}^k=(\tilde{\mathbf{x}}^k,\tilde{\mathbf{a}}^k,\tilde{\mathbf{p}}^k)$ denotes the proximal output, when the schemes are interpreted by the PPA (shown in Table 2). Then, the following holds, $\forall \mathbf{c}=(\mathbf{x},\mathbf{a},\mathbf{p})\in\mathbb{R}^N\times\mathbb{R}^M\times\mathbb{R}^M$:*
(i) $\Pi(\tilde{\mathbf{c}}^k,\mathbf{c})\le\big\langle\mathcal{Q}(\tilde{\mathbf{c}}^k-\mathbf{c}^k)\big|\mathbf{c}-\tilde{\mathbf{c}}^k\big\rangle$,
(ii) $\Pi\big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i,\mathbf{c}\big)\le\frac{1}{2k}\big\|\mathbf{c}^0-\mathbf{c}\big\|_{\mathcal{S}}^2$.

Proof (i) First, note that the proximal step of all the Lagrangian-based schemes listed in Table 1 can be written as:

$$\begin{bmatrix}\mathbf{0}\\\mathbf{0}\\\mathbf{0}\end{bmatrix}\in\begin{bmatrix}\partial f & \mathbf{0} & \mathbf{A}^\top\\ \mathbf{0} & \partial g & -\mathbf{I}_M\\ -\mathbf{A} & \mathbf{I}_M & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{x}}^k\\\tilde{\mathbf{a}}^k\\\tilde{\mathbf{p}}^k\end{bmatrix}+\begin{bmatrix}-\mathbf{Q}_1-\\-\mathbf{Q}_2-\\-\mathbf{Q}_3-\end{bmatrix}(\tilde{\mathbf{c}}^k-\mathbf{c}^k),$$

which is:

$$\left\lfloor\begin{array}{l}\mathbf{0}\in\partial f(\tilde{\mathbf{x}}^k)+\mathbf{A}^\top\tilde{\mathbf{p}}^k+\mathbf{Q}_1(\tilde{\mathbf{c}}^k-\mathbf{c}^k),\\ \mathbf{0}\in\partial g(\tilde{\mathbf{a}}^k)-\tilde{\mathbf{p}}^k+\mathbf{Q}_2(\tilde{\mathbf{c}}^k-\mathbf{c}^k),\\ \mathbf{0}=-\mathbf{A}\tilde{\mathbf{x}}^k+\tilde{\mathbf{a}}^k+\mathbf{Q}_3(\tilde{\mathbf{c}}^k-\mathbf{c}^k).\end{array}\right. \tag{14}$$

Then, by convexity of f and g, we develop:

$$\begin{aligned}f(\mathbf{x})&\ge f(\tilde{\mathbf{x}}^k)+\langle\partial f(\tilde{\mathbf{x}}^k)|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle\\&=f(\tilde{\mathbf{x}}^k)-\langle\mathbf{A}^\top\tilde{\mathbf{p}}^k|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle-\langle\mathbf{Q}_1(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle,\quad\text{by (14)}\end{aligned}$$

and

$$\begin{aligned}g(\mathbf{a})&\ge g(\tilde{\mathbf{a}}^k)+\langle\partial g(\tilde{\mathbf{a}}^k)|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle\\&=g(\tilde{\mathbf{a}}^k)+\langle\tilde{\mathbf{p}}^k|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle-\langle\mathbf{Q}_2(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle.\quad\text{by (14)}\end{aligned}$$

Summing up both inequalities yields

$$\begin{aligned}f(\mathbf{x})+g(\mathbf{a})-f(\tilde{\mathbf{x}}^k)-g(\tilde{\mathbf{a}}^k)\ge&-\langle\mathbf{A}^\top\tilde{\mathbf{p}}^k|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle-\langle\mathbf{Q}_1(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle\\&+\langle\tilde{\mathbf{p}}^k|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle-\langle\mathbf{Q}_2(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle.\end{aligned}$$

Finally, we obtain

$$\begin{aligned}\Pi(\tilde{\mathbf{c}}^k,\mathbf{c})&=\mathcal{L}(\tilde{\mathbf{x}}^k,\tilde{\mathbf{a}}^k,\mathbf{p})-\mathcal{L}(\mathbf{x},\mathbf{a},\tilde{\mathbf{p}}^k)\\&=f(\tilde{\mathbf{x}}^k)+g(\tilde{\mathbf{a}}^k)-f(\mathbf{x})-g(\mathbf{a})+\langle\mathbf{p}|\mathbf{A}\tilde{\mathbf{x}}^k-\tilde{\mathbf{a}}^k\rangle-\langle\tilde{\mathbf{p}}^k|\mathbf{A}\mathbf{x}-\mathbf{a}\rangle\\&\le\langle\mathbf{Q}_1(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle+\langle\mathbf{Q}_2(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle\\&-\langle\tilde{\mathbf{p}}^k|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle+\langle\mathbf{A}^\top\tilde{\mathbf{p}}^k|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle+\langle\mathbf{p}|\mathbf{A}\tilde{\mathbf{x}}^k-\tilde{\mathbf{a}}^k\rangle-\langle\tilde{\mathbf{p}}^k|\mathbf{A}\mathbf{x}-\mathbf{a}\rangle\\&=\langle\mathbf{Q}_1(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle+\langle\mathbf{Q}_2(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle+\langle\mathbf{A}\tilde{\mathbf{x}}^k-\tilde{\mathbf{a}}^k|\mathbf{p}-\tilde{\mathbf{p}}^k\rangle\\&=\langle\mathbf{Q}_1(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{x}-\tilde{\mathbf{x}}^k\rangle+\langle\mathbf{Q}_2(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{a}-\tilde{\mathbf{a}}^k\rangle+\langle\mathbf{Q}_3(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{p}-\tilde{\mathbf{p}}^k\rangle\ \text{by (14)}\\&=\langle\mathcal{Q}(\tilde{\mathbf{c}}^k-\mathbf{c}^k)|\mathbf{c}-\tilde{\mathbf{c}}^k\rangle.\end{aligned}$$

(ii) By Lemma 3-(i) and Lemma 1-(i), we further develop:

$$\begin{aligned}\Pi(\tilde{\mathbf{c}}^i,\mathbf{c})&\le\big\langle\mathcal{Q}(\tilde{\mathbf{c}}^i-\mathbf{c}^i)\big|\mathbf{c}-\tilde{\mathbf{c}}^i\big\rangle\\&=\frac{1}{2}\big\|\mathbf{c}^i-\mathbf{c}\big\|_{\mathcal{S}}^2-\frac{1}{2}\big\|\mathbf{c}^{i+1}-\mathbf{c}\big\|_{\mathcal{S}}^2-\frac{1}{2}\big\|\mathbf{c}^i-\mathbf{c}^{i+1}\big\|_{\mathcal{M}^{-\top}\mathcal{G}\mathcal{M}^{-1}}^2.\end{aligned}$$

Then, summing up from $i=0$ to $k-1$, we obtain $\sum_{i=0}^{k-1}\Pi(\tilde{\mathbf{c}}^i,\mathbf{c}) \le \frac{1}{2}\big\|\mathbf{c}^0-\mathbf{c}\big\|_{\mathcal{S}}^2$. Since $\Pi(\tilde{\mathbf{c}}^i,\mathbf{c})$ is a convex function w.r.t. $\tilde{\mathbf{c}}^i$ (by its definition), it yields that $\frac{1}{k}\sum_{i=0}^{k-1}\Pi(\tilde{\mathbf{c}}^i,\mathbf{c}) \ge \Pi\big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i,\mathbf{c}\big)$, which completes the proof.

Noting that Lemma 3 is valid for any $\mathbf{c}\in\mathbb{R}^N\times\mathbb{R}^M\times\mathbb{R}^M$, $\Pi(\tilde{\mathbf{c}}^k,\mathbf{c})$ is not a distance, since it may be negative. However, $\Pi(\mathbf{c},\mathbf{c}^\star)$ with $\mathbf{c}^\star\in\mathsf{zer}\mathcal{A}$ is essentially a particular instance of the generalized Bregman distance generated by $q(\mathbf{u}):=f(\mathbf{x})+g(\mathbf{a})$ between any point $\mathbf{u}=(\mathbf{x},\mathbf{a})$ and a saddle point $\mathbf{u}^\star=(\mathbf{x}^\star,\mathbf{a}^\star)$. More specifically, $0\le D_q^\flat(\mathbf{u},\mathbf{u}^\star)\le\Pi(\mathbf{c},\mathbf{c}^\star)\le D_q^\sharp(\mathbf{u},\mathbf{u}^\star)$. Indeed, the generalized Bregman distance is given as

$$\begin{aligned}
0 &\le D_q^\flat(\mathbf{u},\mathbf{u}^\star) = q(\mathbf{u})-q(\mathbf{u}^\star)+\inf_{\mathbf{v}\in\partial q(\mathbf{u}^\star)}\langle\mathbf{v}|\mathbf{u}^\star-\mathbf{u}\rangle\\
&= f(\mathbf{x})-f(\mathbf{x}^\star)+g(\mathbf{a})-g(\mathbf{a}^\star)+\inf_{\mathbf{v}\in\partial f(\mathbf{x}^\star)}\langle\mathbf{v}|\mathbf{x}^\star-\mathbf{x}\rangle+\inf_{\mathbf{t}\in\partial g(\mathbf{a}^\star)}\langle\mathbf{t}|\mathbf{a}^\star-\mathbf{a}\rangle\\
&\le f(\mathbf{x})-f(\mathbf{x}^\star)+g(\mathbf{a})-g(\mathbf{a}^\star)+\langle\mathbf{A}^\top\mathbf{p}^\star|\mathbf{x}-\mathbf{x}^\star\rangle-\langle\mathbf{p}^\star|\mathbf{a}-\mathbf{a}^\star\rangle\\
&= f(\mathbf{x})-f(\mathbf{x}^\star)+g(\mathbf{a})-g(\mathbf{a}^\star)+\langle\mathbf{p}^\star|\mathbf{A}\mathbf{x}-\mathbf{a}\rangle\ \text{(by } \mathbf{A}\mathbf{x}^\star=\mathbf{a}^\star)\\
&= f(\mathbf{x})-f(\mathbf{x}^\star)+g(\mathbf{a})-g(\mathbf{a}^\star)+\langle\mathbf{p}^\star|\mathbf{A}\mathbf{x}-\mathbf{a}\rangle-\langle\mathbf{p}|\underbrace{\mathbf{A}\mathbf{x}^\star-\mathbf{a}^\star}_{=\mathbf{0}}\rangle\\
&= \Pi(\mathbf{c},\mathbf{c}^\star)\le D_q^\sharp(\mathbf{u},\mathbf{u}^\star),
\end{aligned}\tag{15}$$

where the first inequality, i.e., the non-negativity of $D_q^\flat(\mathbf{u},\mathbf{u}^\star)$, is due to the convexity of q.

Then, we obtain the convergence rate of $\Pi(\mathbf{c}^k,\mathbf{c}^\star)$ in an ergodic sense.

Theorem 2 *For all the Lagrangian-based algorithms shown in Table 1, the generalized Bregman distance associated with $f(\mathbf{x})+g(\mathbf{a})$ between the ergodic point $\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i$ and a saddle point $\mathbf{c}^\star\in\mathsf{zer}\mathcal{A}$ has a rate of $\mathcal{O}(1/k)$:*

$$0\le\Pi\bigg(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i,\mathbf{c}^\star\bigg)\le\frac{1}{2k}\big\|\mathbf{c}^0-\mathbf{c}^\star\big\|_{\mathcal{S}}^2,$$

where $\{\tilde{\mathbf{c}}^i\}_{i\in\mathbb{N}}$ and $\mathcal{S}$ are defined in Lemma 3 and 1.

Proof The first inequality (i.e. non-negativity) follows from (15). The second inequality is concluded by simply taking $\mathbf{c}=\mathbf{c}^\star\in\mathsf{zer}\mathcal{A}$ in Lemma 3-(ii).

One can write down the specific form of $\|\mathbf{c}^0-\mathbf{c}^\star\|_{\mathcal{S}}$ for each Lagrangian algorithm, according to the associated $\mathcal{S}$. In particular, for `LAG-I` and `LAG-II`, we have: $\Pi\big(\frac{1}{k}\sum_{i=1}^{k}\mathbf{c}^i,\mathbf{c}^\star\big)\le\frac{1}{2k}\big\|\mathbf{c}^0-\mathbf{c}^\star\big\|_{\mathcal{Q}}^2$ (noting that $\tilde{\mathbf{c}}^i=\mathbf{c}^{i+1}$ due to $\mathcal{M}=\mathcal{I}$).

Remark 3 (Degenerate case of Bregman distance) The generalized Bregman distance $D_q^\flat(\mathbf{u},\mathbf{u}^\star)$ or $D_q^\sharp(\mathbf{u},\mathbf{u}^\star)$ does not always reflect or control the distance between any point $\mathbf{u}=(\mathbf{x},\mathbf{a})$ and a saddle point $\mathbf{u}^\star=(\mathbf{x}^\star,\mathbf{a}^\star)$, particularly for the non-strictly convex case of q. Consider a degenerate case of linear functional, when $f(\mathbf{x})=\langle\mathbf{x}|\mathbf{v}\rangle$ with a constant vector $\mathbf{v}$ and $g(\mathbf{a})=0$. Then, the Bregman distance between any two points $\mathbf{u}$ and $\mathbf{u}'$ is always 0, since $D_q^\flat(\mathbf{u},\mathbf{u}') = D_q^\sharp(\mathbf{u},\mathbf{u}') = D_f(\mathbf{x},\mathbf{x}') = \langle \mathbf{x}|\mathbf{v}\rangle - \langle \mathbf{x}'|\mathbf{v}\rangle - \langle \mathbf{v}|\mathbf{x}-\mathbf{x}'\rangle = 0$. It implies that the Bregman distance loses the control of the distance between $(\mathbf{x},\mathbf{a})$ and $(\mathbf{x}',\mathbf{a}')$. Theorem 2 becomes more informative, when the involved functions f and g are strictly convex.

3.4 The ergodic primal-dual gap

Considering $\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{p})$ given as (7), for given sets $B_1 \subset \mathbb{R}^N$, $B_2 \subset \mathbb{R}^M$ and $B_3 \subset \mathbb{R}^M$, we introduce a *primal-dual gap* function restricted to $B_1 \times B_2 \times B_3$ [6, Eq.(2.6)], [11, Sect. 3.1], [34, Eq.(2.14)]:

$$\Psi_{B_1\times B_2\times B_3}(\mathbf{c}) = \sup_{\mathbf{p}'\in B_3} \mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{p}') - \inf_{(\mathbf{x}',\mathbf{a}')\in B_1\times B_2} \mathcal{L}(\mathbf{x}',\mathbf{a}',\mathbf{p}), \tag{16}$$

from which also follows that $\Psi_{B_1\times B_2\times B_3}(\mathbf{c}) = \sup_{\mathbf{c}'\in B_1\times B_2\times B_3} \Pi(\mathbf{c},\mathbf{c}')$.

Corollary 2 *Under the conditions of Theorem 2, if the set $B_1 \times B_2 \times B_3$ is bounded, the primal-dual gap defined as* (16) *has the upper bound:*

$$\Psi_{B_1\times B_2\times B_3}\bigg(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i\bigg) \le \frac{1}{2k} \sup_{\mathbf{c}\in B_1\times B_2\times B_3} \big\|\mathbf{c}^0-\mathbf{c}\big\|_{\mathcal{S}}^2. \tag{17}$$

Furthermore, $\Psi_{B_1\times B_2\times B_3}(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i) \ge 0$, if the set $B_1 \times B_2 \times B_3$ contains a saddle point $\mathbf{c}^\star = (\mathbf{x}^\star,\mathbf{a}^\star,\mathbf{p}^\star) \in \mathsf{zer}\mathcal{A}$.

Proof Since Lemma 3-(ii) is valid for any $(\mathbf{x},\mathbf{a},\mathbf{p}) \in \mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^M$, passing to the supremum and infimum over $(\mathbf{x},\mathbf{a}) \in B_1 \times B_2$ and $\mathbf{p} \in B_3$ yields (17). The non-negativity of $\Psi_{B_1\times B_2\times B_3}$ follows from Theorem 2, since $\Psi_{B_1\times B_2\times B_3}$ $\big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i\big) \ge \Pi\big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i,\mathbf{c}^\star\big) \ge 0$, provided that $B_1 \times B_2 \times B_3$ contains a saddle point $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$.

Remark 4 The ergodic primal-dual gap for specific algorithms has been given in [11,12,34]. Our analysis of the primal-dual gap is general, easy and clear, compared to the original complicated case studies of specific algorithms, e.g. [11, Theorem 1-(b)], [6, Theorem 9-(b)] and [7, Theorem 2.1-(d)]. All the results presented in Sect. 3.3 and 3.4 are valid for all Lagrangian-based algorithms with the same monotone operator $\mathcal{A}$, not limited to the listed ones. More importantly, this observation also applies to other classes of algorithms, see Sect. 4.2 and 5.2.

Remark 5 By Theorem 1, the multiplier $\{\tilde{\mathbf{p}}^k\}_{k\in\mathbb{N}}$ converges and therefore lies in some (unknown) bounded set $B_3 \subset \mathbb{R}^M$. If $\mathsf{dom}f$ and $\mathsf{dom}g$ are bounded, Corollary 2 could lead to an interesting result: *the sequence of the objective value of dual to* (1) *taken at the ergodic averaging point $\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i$ converges at the rate of* $\mathcal{O}(1/k)$, namely, it holds that:

$$f^*\bigg(-\mathbf{A}^\top\Big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{p}}^i\Big)\bigg) + g^*\bigg(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{p}}^i\bigg) - f^*(-\mathbf{A}^\top\mathbf{p}^\star) - g^*(\mathbf{p}^\star) \le C/k, \tag{18}$$

for some constant C.

Indeed, if $\mathsf{dom}f$ and $\mathsf{dom}g$ are bounded, one can choose $B_1 = \mathsf{dom}f$ and $B_2 = \mathsf{dom}g$. Since the sequence $\{\tilde{\mathbf{p}}^k\}_{k\in\mathbb{N}}$ lies in B_3, and thus, $\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{p}}^i \in B_3$, $\mathbf{p}^\star \in B_3$. Denoting the ergodic averaging point by $\widehat{\mathbf{c}}^k = \frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i$ ($\widehat{\mathbf{x}}^k$, $\widehat{\mathbf{a}}^k$ and $\widehat{\mathbf{p}}^k$ are defined similarly), using Fenchel-Young inequality [4, Proposition 13.15], we develop

$$
\begin{aligned}
& \Psi_{B_1\times B_2\times B_3}\big(\widehat{\mathbf{c}}^k\big) \\
&= \sup_{\mathbf{p}'\in B_3} \mathcal{L}\big(\widehat{\mathbf{x}}^k, \widehat{\mathbf{a}}^k, \mathbf{p}'\big) - \inf_{(\mathbf{x}',\mathbf{a}')\in B_1\times B_2} \mathcal{L}\big(\mathbf{x}', \mathbf{a}', \widehat{\mathbf{p}}^k\big) \\
&= \sup_{\mathbf{p}'\in B_3} q\big(\widehat{\mathbf{u}}^k\big) + \big\langle \mathbf{p}' \big| \mathbf{U}\widehat{\mathbf{u}}^k \big\rangle - \inf_{\mathbf{u}'\in B_1\times B_2} \big(q(\mathbf{u}') + \big\langle \widehat{\mathbf{p}}^k \big| \mathbf{U}\mathbf{u}'\big\rangle\big) \quad \text{by (9)} \\
&\geq q\big(\widehat{\mathbf{u}}^k\big) + \big\langle \mathbf{U}^\top \mathbf{p}^\star \big| \widehat{\mathbf{u}}^k \big\rangle + q^*\big(-\mathbf{U}^\top \widehat{\mathbf{p}}^k\big) \\
&\geq -q^*(-\mathbf{U}^\top \mathbf{p}^\star) + q^*\big(-\mathbf{U}^\top \widehat{\mathbf{p}}^k\big),
\end{aligned}
$$

which, combining with Corollary 2, yields

$$
q^*\big(-\mathbf{U}^\top \widehat{\mathbf{p}}^k\big) - q^*(-\mathbf{U}^\top \mathbf{p}^\star) \leq \frac{1}{2k} \sup_{\mathbf{c}\in B_1\times B_2\times B_3} \big\|\mathbf{c}^0 - \mathbf{c}\big\|_{\mathcal{S}}^2.
$$

By the definitions of q and $\mathbf{U}$ of (9), $q^*(-\mathbf{U}^\top\mathbf{p}) = f^*(-\mathbf{A}^\top\mathbf{p}) + g^*(\mathbf{p})$, which exactly coincides with the dual of (1), which is given as (25). The conclusion (18) is reached.

On the other hand, note that $q^*(-\mathbf{U}^\top\mathbf{p}) = \sup_{\mathbf{u}\in B_1\times B_2}\big(-q(\mathbf{u}) - \langle \mathbf{p}|\mathbf{U}\mathbf{u}\rangle\big) = -\inf_{\mathbf{u}\in B_1\times B_2}\big(q(\mathbf{u}) + \langle \mathbf{p}|\mathbf{U}\mathbf{u}\rangle\big) = -\min_{\mathbf{u}} \mathcal{L}(\mathbf{u},\mathbf{p})$, and thus, the saddle-point problem of $\mathcal{L}(\mathbf{u},\mathbf{p})$ (9) becomes

$$
\max_{\mathbf{p}} \min_{\mathbf{u}} \mathcal{L}(\mathbf{u},\mathbf{p}) = \max_{\mathbf{p}} (-q^*(-\mathbf{U}^\top p)) = -\min_{\mathbf{p}} q^*(-\mathbf{U}^\top \mathbf{p}),
$$

which is essentially the minimization problem of the dual $q^*(-\mathbf{U}^\top\cdot)$.

Finally, we stress that the convergence rate of $\mathcal{O}(1/k)$ of the dual value holds for all the the Lagrangian-based algorithms shown in Table 1. However, as contrary to Remark 7, it is difficult to obtain an *a priori* estimate of the constant C, since the bounded set B_3 is unknown in practice.

3.5 Reductions of some Lagrangian schemes

3.5.1 `LAG-I` and `LAG-II`

Table 2 shows that `LAG-I` and `LAG-II` can be expressed as a standard PPA (5) with $\mathcal{M} = \mathcal{I}$. Both of them can be reduced to a simple resolvent by Corollary 1-(iii):

$$
\mathbf{v}^{k+1} := \big(\mathcal{I} + \mathcal{Q}^{-\frac{1}{2}} \circ \mathcal{A} \circ \mathcal{Q}^{-\frac{1}{2}}\big)^{-1}\mathbf{v}^k, \tag{19}
$$

where $\mathbf{v}^k = \mathcal{Q}^{\frac{1}{2}}\mathbf{c}^k$, $\mathcal{Q}$ is specified in Table 2 for `LAG-I` or `II`.

`LAG-II` deserves particular attention, since the corresponding metric $\mathcal{Q}$ is allowed to be degenerate.

Low degeneracy of `LAG-II` Notice that $\mathbf{M} = \mathbf{0}$ is allowed for `LAG-II`, which becomes

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{1}{2}\big\|\mathbf{A}\mathbf{x} - \mathbf{a}^k + \mathbf{\Gamma}^{-1}\mathbf{p}^k\big\|_{\mathbf{\Gamma}}^2, \\ \mathbf{p}^{k+1} = \mathbf{p}^k + \mathbf{\Gamma}\big(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^k\big), \\ \mathbf{a}^{k+1} = \mathsf{prox}_g^{\mathbf{\Omega}}\big(\mathbf{a}^k + \mathbf{\Omega}^{-1}(2\mathbf{p}^{k+1} - \mathbf{p}^k)\big). \end{array} \right.$$

Now, the metric $\mathcal{Q}$ is degenerate (i.e. positive *semi*-positive) with $\mathsf{rank}\mathcal{Q} = 2M < \mathsf{dim}(\mathbb{R}^N \times \mathbb{R}^M \times \mathbb{R}^M) = N + 2M$. The rank-deficiency of $\mathcal{Q}$ shows that the variable $\mathbf{x}$ is redundant that does not really take part in the iterations of `LAG-II`. We can reduce `LAG-II` based on the analysis of [8].

Proposition 3 *`LAG-II` with $\mathbf{M} = \mathbf{0}$ and $\mathbf{\Omega} \succ \mathbf{\Gamma}$ can be expressed as the following resolvent:*

$$\mathbf{v}^{k+1} = \big(\mathcal{I} + \mathcal{D}(\mathcal{L} + \mathcal{K}^\top \circ \partial f^* \circ \mathcal{K})\mathcal{D}\big)^{-1}\mathbf{v}^k,$$

where $\mathcal{D} = \begin{bmatrix} \mathbf{\Omega} & -\mathbf{I}_M \\ -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}^{-\frac{1}{2}}$, $\mathcal{L} = \begin{bmatrix} \partial g & -\mathbf{I}_M \\ \mathbf{I}_M & \mathbf{0} \end{bmatrix}$, $\mathcal{K} = \begin{bmatrix} \mathbf{0} & -\mathbf{A}^\top \end{bmatrix}$. *Here, the variable* $\mathbf{v}$ *is linked to* $(\mathbf{a}, \mathbf{p})$ *in `LAG-II` via:* $\mathbf{v}^k := \begin{bmatrix} \mathbf{\Omega} & -\mathbf{I}_M \\ -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} \mathbf{a}^k \\ \mathbf{p}^k \end{bmatrix} \in \mathbb{R}^M \times \mathbb{R}^M$.

Proof $\mathcal{Q}$ of `LAG-II` can be decomposed as: $\mathcal{Q} = \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} \begin{bmatrix} \mathbf{0} \; \mathcal{D}^\top \end{bmatrix}$, where $\mathcal{D}\mathcal{D}^\top = \begin{bmatrix} \mathbf{\Omega} & -\mathbf{I}_M \\ -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}$. For simplicity, one can choose $\mathcal{D} = \mathcal{D}^\top = \begin{bmatrix} \mathbf{\Omega} & -\mathbf{I}_M \\ -\mathbf{I}_M & \mathbf{\Gamma}^{-1} \end{bmatrix}^{\frac{1}{2}}$. The standard PPA form (5) becomes

$$\mathbf{c}^{k+1} = \left(\mathcal{A} + \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} \begin{bmatrix} \mathbf{0} \; \mathcal{D}^\top \end{bmatrix}\right)^{-1} \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} \begin{bmatrix} \mathbf{0} \; \mathcal{D}^\top \end{bmatrix} \mathbf{c}^k.$$

Let $\mathbf{v}^k := \begin{bmatrix} \mathbf{0} \; \mathcal{D}^\top \end{bmatrix} \mathbf{c}^k = \mathcal{D}^\top \begin{bmatrix} \mathbf{a}^k \\ \mathbf{p}^k \end{bmatrix} \in \mathbb{R}^M \times \mathbb{R}^M$. By [8, Theorem 2.13], we obtain the reduced PPA:

$$\begin{aligned} \mathbf{v}^{k+1} &= \begin{bmatrix} \mathbf{0} \; \mathcal{D}^\top \end{bmatrix} \left(\mathcal{A} + \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} \begin{bmatrix} \mathbf{0} \; \mathcal{D}^\top \end{bmatrix}\right)^{-1} \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} \mathbf{v}^k \\ &= \left(\mathcal{I} + \left(\begin{bmatrix} \mathbf{0} \; \mathcal{D}^\top \end{bmatrix} \mathcal{A}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix}\right)^{-1}\right)^{-1} \mathbf{v}^k := \big(\mathcal{I} + \tilde{\mathcal{A}}\big)^{-1}\mathbf{v}^k. \end{aligned}$$

To evaluate $\tilde{\mathcal{A}}$, we rewrite $\mathcal{A} = \begin{bmatrix} \partial f & -\mathcal{K} \\ \mathcal{K}^\top & \mathcal{L} \end{bmatrix}$, where $\mathcal{L} = \begin{bmatrix} \partial g & -\mathbf{I}_M \\ \mathbf{I}_M & \mathbf{0} \end{bmatrix}$, $\mathcal{K} = \begin{bmatrix} \mathbf{0} & -\mathbf{A}^\top \end{bmatrix}$. Then,

$$\mathcal{R} = \mathcal{A}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} = \begin{bmatrix} \partial f & -\mathcal{K} \\ \mathcal{K}^\top & \mathcal{L} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} = \begin{bmatrix} \mathcal{R}_1 \\ \mathcal{R}_2 \end{bmatrix},$$

which yields the solution: $\mathcal{R}_2 = (\mathcal{L} + \mathcal{K}^\top \circ \partial f^* \circ \mathcal{K})^{-1}\mathcal{D}$. Thus,

$$\tilde{\mathcal{A}} = \left(\begin{bmatrix} \mathbf{0} \ \mathcal{D}^\top \end{bmatrix} \mathcal{A}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathcal{D} \end{bmatrix} \right)^{-1} = \left(\begin{bmatrix} \mathbf{0} \ \mathcal{D}^\top \end{bmatrix} \mathcal{R} \right)^{-1} = (\mathcal{D}^\top \mathcal{R}_2)^{-1},$$

Substituting $\mathcal{R}_2$ into above concludes the proof.

High degeneracy of `LAG-II` Furthermore, if $\mathbf{M} = \mathbf{0}$ and $\mathbf{\Omega} = \mathbf{\Gamma} = \mathbf{I}_M$, `LAG-II` becomes

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{1}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{a}^k + \mathbf{p}^k\right\|^2, \\ \mathbf{p}^{k+1} = \ \mathbf{p}^k + \mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^k, \\ \mathbf{a}^{k+1} = \ \mathsf{prox}_g\big(\mathbf{a}^k + 2\mathbf{p}^{k+1} - \mathbf{p}^k\big). \end{array} \right. \tag{20}$$

Now, the corresponding metric $\mathcal{Q}$ is 'more' degenerate with $\mathsf{rank}\mathcal{Q} = M$. The following result shows that the active variable of `LAG-II` is actually $\mathbf{a}^k - \mathbf{p}^k$.

Proposition 4 *`LAG-II` with $\mathbf{M} = \mathbf{0}$ and $\mathbf{\Omega} = \mathbf{\Gamma} = \mathbf{I}_M$ can be expressed as*

$$\mathbf{v}^{k+1} = \left(\mathcal{I} + (\tilde{\mathcal{D}}^\top (\mathcal{L} + \mathcal{K}^\top \circ \partial f^* \circ \mathcal{K})^{-1} \tilde{\mathcal{D}})^{-1}\right)^{-1} \mathbf{v}^k,$$

where $\tilde{\mathcal{D}} = \begin{bmatrix} \mathbf{I}_M & -\mathbf{I}_M \end{bmatrix}^\top$, $\mathcal{L} = \begin{bmatrix} \partial g & -\mathbf{I}_M \\ \mathbf{I}_M & \mathbf{0} \end{bmatrix}$, $\mathcal{K} = \begin{bmatrix} \mathbf{0} & -\mathbf{A}^\top \end{bmatrix}$. *Here, the variable* $\mathbf{v}$ *is linked to* $(\mathbf{a}, \mathbf{p})$ *in `LAG-II` via:* $\mathbf{v}^k := \mathbf{a}^k - \mathbf{p}^k \in \mathbb{R}^M$.

Proof In this case, $\mathcal{Q}$ of `LAG-II` can be decomposed as: $\mathcal{Q} = \mathcal{D}\mathcal{D}^\top$ where $\mathcal{D} = \begin{bmatrix} \mathbf{0} \ \mathbf{I}_M \ -\mathbf{I}_M \end{bmatrix}^\top$. The standard PPA form (5) becomes: $\mathbf{c}^{k+1} = \big(\mathcal{A} + \mathcal{D}\mathcal{D}^\top\big)^{-1} \mathcal{D}\mathcal{D}^\top \mathbf{c}^k$. Finally, the proof is completed, by [8, Theorem 2.13] and the proof of Proposition 3.

The active variable of (20) can also be identified without the degenerate PPA analysis, as shown below.

From (20), we have

$$\mathbf{a}^{k+1} - \mathbf{p}^{k+1} = \mathsf{prox}_g\big(\mathbf{a}^k + 2\mathbf{p}^{k+1} - \mathbf{p}^k\big) - \mathbf{p}^k - \mathbf{A}\mathbf{x}^{k+1} + \mathbf{a}^k.$$

Denoting $\mathbf{v}^k := \mathbf{a}^k - \mathbf{p}^k$, it becomes

$$\begin{aligned}
\mathbf{v}^{k+1} &= \mathsf{prox}_g\big(\mathbf{v}^k + 2\mathbf{p}^{k+1}\big) + \mathbf{v}^k - \mathbf{A}\mathbf{x}^{k+1} \\
&= \mathsf{prox}_g\big(\mathbf{v}^k + 2\mathbf{p}^k + 2\mathbf{A}\mathbf{x}^{k+1} - 2\mathbf{a}^k\big) + \mathbf{v}^k - \mathbf{A}\mathbf{x}^{k+1} \\
&= \mathsf{prox}_g\big(2\mathbf{A}\mathbf{x}^{k+1} - \mathbf{v}^k\big) + \mathbf{v}^k - \mathbf{A}\mathbf{x}^{k+1} \\
&= \mathsf{prox}_g\big(2\mathsf{prox}_{\mathbf{A}\triangleright f}(\mathbf{v}^k) - \mathbf{v}^k\big) + \mathbf{v}^k - \mathsf{prox}_{\mathbf{A}\triangleright f}(\mathbf{v}^k) \\
&= \Big(\mathsf{prox}_g \circ \big(2\mathsf{prox}_{\mathbf{A}\triangleright f} - \mathcal{I}\big) + \mathcal{I} - \mathsf{prox}_{\mathbf{A}\triangleright f}\Big)(\mathbf{v}^k) \\
&= \Big(\mathsf{prox}_g \circ \big(2(\mathcal{I} - \mathsf{prox}_{f^* \circ \mathbf{A}^\top}) - \mathcal{I}\big) + \mathsf{prox}_{f^* \circ \mathbf{A}^\top}\Big)(\mathbf{v}^k) \quad \text{(by Lemma 2)} \\
&= \Big(\mathsf{prox}_g \circ \big(\mathcal{I} - 2\mathsf{prox}_{f^* \circ \mathbf{A}^\top}\big) + \mathsf{prox}_{f^* \circ \mathbf{A}^\top}\Big)(\mathbf{v}^k)
\end{aligned}$$

which shows that (20) is essentially a DRS algorithm (see Eq.(26)).

3.5.2 LAG-VI and the related standard ADMM/DRS

It seems more interesting to investigate the degenerate case of LAG-VI, which is a representative ADMM-type algorithm.

Table 2 shows that the corresponding PPA of LAG-VI has a non-trivial relaxation step (i.e. $\mathcal{M} \neq \mathcal{I}$). This also coincides with a pioneering work of [27, Sect. 3]. Due to the non-trivial relaxation, it is difficult to obtain the equivalent resolvent from this PPA interpretation. *Can LAG-VI be written in a standard PPA form without relaxation step?* To achieve this, by changing variable of $\mathbf{p}^k := \mathbf{z}^k - \mathbf{\Gamma}\mathbf{a}^k$, LAG-VI becomes (with a flipped update order of $\mathbf{x} \to \mathbf{z} \to \mathbf{a}$)

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A}}\big((\mathbf{M}+\mathbf{A}^\top\mathbf{\Gamma}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k + 2\mathbf{A}^\top\mathbf{\Gamma}\mathbf{a}^k - \mathbf{A}^\top\mathbf{z}^k)\big), \\ \mathbf{z}^{k+1} := \mathbf{z}^k + \mathbf{\Gamma}(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^k), \\ \mathbf{a}^{k+1} := \mathsf{prox}_g^{\mathbf{\Omega}+\mathbf{\Gamma}}\big((\mathbf{\Omega}+\mathbf{\Gamma})^{-1}((\mathbf{\Omega}-\mathbf{\Gamma})\mathbf{a}^k + \mathbf{\Gamma}\mathbf{A}\mathbf{x}^{k+1} + \mathbf{z}^k)\big). \end{array} \right. \tag{21}$$

It is easy to verify that (21) corresponds to the standard PPA form (i.e. $\mathcal{M} = \mathcal{I}$):

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \begin{bmatrix} \partial f & -\mathbf{A}^\top\mathbf{\Gamma} & \mathbf{A}^\top \\ \mathbf{\Gamma}\mathbf{A} & \partial g & -\mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}^{k+1} \\ \mathbf{a}^k \\ \mathbf{z}^{k+1} \end{bmatrix} + \begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Gamma}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}^{k+1} - \mathbf{x}^k \\ \mathbf{a}^k - \mathbf{a}^{k-1} \\ \mathbf{z}^{k+1} - \mathbf{z}^k \end{bmatrix}.$$

Non-degenerate case: proximal ADMM If $\mathbf{M} \succ \mathbf{0}$ and $\mathbf{\Omega} \succ \mathbf{0}$, $\mathcal{Q}$ is non-degenerate. By Corollary 1-(iii), the equivalent resolvent is given as (19), where $\mathcal{A}$ and $\mathcal{Q}$ are specified as above, $\mathbf{v}^k$ is related to $(\mathbf{x}, \mathbf{a}, \mathbf{p})$ of (22) and $(\mathbf{x}, \mathbf{a}, \mathbf{z})$ of (23) via: $\mathbf{v}^k = \mathcal{Q}^{\frac{1}{2}}\mathbf{c}^k = (\mathbf{M}^{\frac{1}{2}}\mathbf{x}^k, \mathbf{\Omega}^{\frac{1}{2}}\mathbf{a}^{k-1}, \mathbf{\Gamma}^{-\frac{1}{2}}\mathbf{z}^k) = (\mathbf{M}^{\frac{1}{2}}\mathbf{x}^k, \mathbf{\Omega}^{\frac{1}{2}}\mathbf{a}^{k-1}, \mathbf{\Gamma}^{-\frac{1}{2}}\mathbf{p}^k + \mathbf{\Gamma}^{\frac{1}{2}}\mathbf{a}^k)$.

Degenerate case: standard ADMM If $\mathbf{M} = \mathbf{0}$, $\mathbf{\Omega} = \mathbf{0}$ and $\mathbf{\Gamma} = \gamma\mathbf{I}_M$, LAG-VI boils down to a standard ADMM [46, Eq.(4)]:

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{\gamma}{2}\big\|\mathbf{A}\mathbf{x} - \mathbf{a}^k + \frac{1}{\gamma}\mathbf{p}^k\big\|^2, \\ \mathbf{a}^{k+1} := \mathsf{prox}_{g/\gamma}\big(\mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\gamma}\mathbf{p}^k\big), \\ \mathbf{p}^{k+1} := \mathbf{p}^k + \gamma(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^{k+1}). \end{array} \right. \tag{22}$$

By the variable changing of $\mathbf{p}^k := \mathbf{z}^k - \gamma\mathbf{a}^k$, (22) becomes (with a flipped update order of $\mathbf{x} \to \mathbf{z} \to \mathbf{a}$)

$$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{\gamma}{2}\big\|\mathbf{A}\mathbf{x} - 2\mathbf{a}^k + \frac{1}{\gamma}\mathbf{z}^k\big\|^2, \\ \mathbf{z}^{k+1} := \mathbf{z}^k + \gamma(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^k), \\ \mathbf{a}^{k+1} := \mathsf{prox}_{g/\gamma}\big(\frac{1}{\gamma}\mathbf{z}^{k+1}\big). \end{array} \right. \tag{23}$$

It is easy to verify that (23) corresponds to the standard PPA form (i.e., $\mathcal{M} = \mathcal{I}$):

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \begin{bmatrix} \partial f & -\gamma\mathbf{A}^\top & \mathbf{A}^\top \\ \gamma\mathbf{A} & \partial g & -\mathbf{I}_M \\ -\mathbf{A} & \mathbf{I}_M & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}^{k+1} \\ \mathbf{a}^k \\ \mathbf{z}^{k+1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\gamma}\mathbf{I}_M \end{bmatrix} \begin{bmatrix} \mathbf{x}^{k+1} - \mathbf{x}^k \\ \mathbf{a}^k - \mathbf{a}^{k-1} \\ \mathbf{z}^{k+1} - \mathbf{z}^k \end{bmatrix}. \tag{24}$$

The degenerate (i.e., positive *semi*-definite) metric $\mathcal{Q}$ indicates the redundancy of the variables $\mathbf{x}$ and $\mathbf{a}$. Based on the recent result of [8, Theorem 2.13], the standard ADMM (22) or (23) can be reduced to a simple resolvent.

Theorem 3 *The ADMM scheme* (23)*, being equivalent to* (22)*, can be expressed as*

$$\mathbf{v}^{k+1} = (\mathcal{I} + \gamma\mathcal{K}^\top\mathcal{L}^{-1}\mathcal{K})^{-1}\mathbf{v}^k,$$

where $\mathcal{L} = \begin{bmatrix} \partial f & -\gamma\mathbf{A}^\top \\ \gamma\mathbf{A} & \partial g \end{bmatrix}$, $\mathcal{K} = \begin{bmatrix} -\mathbf{A} \ \mathbf{I}_M \end{bmatrix}^\top$*. Here, the variable* $\mathbf{v}$ *is linked to* $\mathbf{z}$ *in* (23) *and* $(\mathbf{a}, \mathbf{p})$ *in* (22) *via* $\mathbf{v}^k = \frac{1}{\sqrt{\gamma}}\mathbf{z}^k = \frac{1}{\sqrt{\gamma}}\mathbf{p}^k + \sqrt{\gamma}\mathbf{a}^k \in \mathbb{R}^M$.

Proof $\mathcal{Q}$ in (24) can be decomposed as $\mathcal{Q} = \mathcal{D}\mathcal{D}^\top = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \frac{1}{\sqrt{\gamma}}\mathbf{I}_M \end{bmatrix} \begin{bmatrix} \mathbf{0} \ \mathbf{0} \ \frac{1}{\sqrt{\gamma}}\mathbf{I}_M \end{bmatrix}$. Then, the standard PPA form (24) becomes $\mathbf{c}^{k+1} = (\mathcal{A} + \mathcal{D}\mathcal{D}^\top)^{-1}\mathcal{D}\mathcal{D}^\top\mathbf{c}^k$. Let $\mathbf{v}^k := \mathcal{D}^\top\mathbf{c}^k = \frac{1}{\sqrt{\gamma}}\mathbf{z}^k$. Finally, the result can be obtained by [8, Theorem 2.13] and similar proof of Proposition 3.

Connection to standard DRS It is well known that the ADMM scheme (22) is equivalent to the standard DRS algorithm [31] applied to the dual problem of (1) (see [41, Eq.(2)] for example):

$$\min_{\mathbf{p}} f^*(-\mathbf{A}^\top\mathbf{p}) + g^*(\mathbf{p}), \tag{25}$$

which reads as

$$\mathbf{z}^{k+1} := \mathbf{z}^k - J_{\gamma\mathcal{B}_2}(\mathbf{z}^k) + J_{\gamma\mathcal{B}_1}\big(2J_{\gamma\mathcal{B}_2}(\mathbf{z}^k) - \mathbf{z}^k\big), \tag{26}$$

where $\mathcal{B}_1 = (-\mathbf{A}) \circ \partial f^* \circ (-\mathbf{A}^\top)$, $\mathcal{B}_2 = \partial g^*$, $J_\mathcal{B}$ denotes a resolvent of $\mathcal{B}$: $J_\mathcal{B} = (\mathcal{I} + \mathcal{B})^{-1}$. The solution to (25) is given as: $\mathbf{p}^\star = J_{\gamma\mathcal{B}_2}(\mathbf{z}^\star) = \mathsf{prox}_{\gamma g^*}(\mathbf{z}^\star)$.

Let us first examine the equivalence between ADMM (22) or (23) and DRS (26), though this fact has long been recognized. By the development of (26):

$$\left\lfloor \begin{array}{ll} \mathbf{w}^{k+1} & := J_{\gamma\mathcal{B}_1}\big(2\mathbf{p}^k - \mathbf{z}^k\big), \\ \mathbf{z}^{k+1} & := \mathbf{z}^k + \mathbf{w}^{k+1} - \mathbf{p}^k, \\ \mathbf{p}^{k+1} & := \mathsf{prox}_{\gamma\mathcal{B}_2}\big(\mathbf{z}^{k+1}\big). \end{array} \right.$$

By the similar technique of [11, Sect. 4.2], we obtain by duality that:

$$\left\lfloor \begin{array}{ll} \mathbf{x}^{k+1} & := \arg\min_\mathbf{x} f(\mathbf{x}) + \frac{\gamma}{2}\big\|\mathbf{A}\mathbf{x} + \frac{1}{\gamma}(2\mathbf{p}^k - \mathbf{z}^k)\big\|^2, \\ \mathbf{w}^{k+1} & := 2\mathbf{p}^k - \mathbf{z}^k + \gamma\mathbf{A}\mathbf{x}^{k+1}, \\ \mathbf{z}^{k+1} & := \mathbf{z}^k + \mathbf{w}^{k+1} - \mathbf{p}^k, \\ \mathbf{a}^{k+1} & := \arg\min_\mathbf{u} g(\mathbf{a}) + \frac{\gamma}{2}\big\|\mathbf{a} - \frac{1}{\gamma}\mathbf{z}^{k+1}\big\|^2, \\ \mathbf{p}^{k+1} & := \mathbf{z}^{k+1} - \gamma\mathbf{a}^{k+1}. \end{array} \right.$$

Finally, (22) can be obtained by keeping $(\mathbf{x}, \mathbf{a}, \mathbf{p})$ and removing $(\mathbf{w}, \mathbf{z})$; while (23) is from keeping $(\mathbf{x}, \mathbf{a}, \mathbf{z})$ and removing $(\mathbf{w}, \mathbf{p})$.

The above equivalence implies that Theorem 3 also applies to the DRS iteration (26). Recall that the equivalence between DRS (26) and PPA was discussed in an early seminal work of [19], where the DRS (26) was shown to be equivalent to a resolvent with an *implicit* expression of the associated maximally monotone operator (see [19, Sect. 4]). Here, Theorem 3 shows another equivalent resolvent of DRS (26), with an *explicit* form of the monotone operator. However, the equivalence or connection between both forms requires further study.

4 Operator splitting based on primal-dual form

4.1 The PDS algorithms and their PPA interpretations

We then consider the alternating optimization of the primal-dual form [46, Eq.(7)]:

$$\min_{\mathbf{x}} \max_{\mathbf{p}} \mathcal{L}(\mathbf{x}, \mathbf{p}) := f(\mathbf{x}) + \mathbf{p}^\top \mathbf{A}\mathbf{x} - g^*(\mathbf{p}), \tag{27}$$

which gives rise to the PDS algorithms shown in Table 4. `PDS-I,II,V,VI` and `VII` can be found in [46, Sect. 5]. Tables 5–6 show their equivalent PPA forms, by noting that:

- `PDS-I` and `PDS-II` correspond to symmetric $\mathcal{Q}$ (without relaxation): the off-diagonal parts of both $\mathcal{Q}$ have opposite signs, which results in the reverse update orders of $\mathbf{x}$ and $\mathbf{p}$;
- `PDS-III` and `PDS-V` correspond to lower triangular $\mathcal{Q}$;
- `PDS-IV` and `PDS-VI` correspond to upper triangular $\mathcal{Q}$;
- `PDS-VII` corresponds to skew-symmetric $\mathcal{Q}$.

The connections of the proposed PDS algorithms to the previous works, e.g. [17, Algorithms 5.1 and 5.2], [14, Theorems 3.1 and 4.2] and [7, Algorithm 2.1], have been discussed in [46, Sect. 5].

4.2 The generalized Bregman distance and ergodic primal-dual gap

Similar to Sect. 3.3 and 3.4, the unified PPA framework also facilitates the gap analysis for the PDS algorithms.

Lemma 4 *Given the primal-dual form* $\mathcal{L}(\mathbf{x}, \mathbf{p})$ *as* (27)*, consider all the PDS schemes listed in Table 4, where* $\tilde{\mathbf{c}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{p}}^k)$ *denotes the proximal output, when the schemes are interpreted by the PPA (shown in Table 5). Then, the following holds,* $\forall \mathbf{c} = (\mathbf{x}, \mathbf{p}) \in \mathbb{R}^N \times \mathbb{R}^M$*:*
(i) $\Pi(\tilde{\mathbf{c}}^k, \mathbf{c}) \le \big\langle \mathcal{Q}(\tilde{\mathbf{c}}^k - \mathbf{c}^k) \big| \mathbf{c} - \tilde{\mathbf{c}}^k \big\rangle$,
(ii) $\Pi\big(\frac{1}{k}\sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i, \mathbf{c}\big) \le \frac{1}{2k}\big\|\mathbf{c}^0 - \mathbf{c}\big\|_{\mathcal{S}}^2$.

Table 4 The proposed PDS algorithms

name	iterative scheme
PDS-I [46, Eq.(26)]	$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathrm{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^k\big) \\ \mathbf{p}^{k+1} := \mathrm{prox}_{g^*}^{\mathbf{\Gamma}}\big(\mathbf{p}^k + \mathbf{\Gamma}^{-1}\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k)\big) \end{array} \right.$
PDS-II [46, Eq.(29)]	$\left\lfloor \begin{array}{l} \mathbf{p}^{k+1} := \mathrm{prox}_{g^*}^{\mathbf{\Gamma}}(\mathbf{p}^k + \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{x}^k) \\ \mathbf{x}^{k+1} := \mathrm{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top(2\mathbf{p}^{k+1} - \mathbf{p}^k)\big) \end{array} \right.$
PDS-III	$\left\lfloor \begin{array}{lll} \mathbf{p}^{k+1} & := & \mathrm{prox}_{g^*}^{\mathbf{\Gamma}}(\mathbf{p}^k + \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{x}^k) \\ \tilde{\mathbf{x}}^k & := & \mathrm{prox}_f^{\mathbf{M}}(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^{k+1}) \\ \mathbf{x}^{k+1} & := & \tilde{\mathbf{x}}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{\Gamma}^{-1}\mathbf{A}(\tilde{\mathbf{x}}^k - \mathbf{x}^k) - \mathbf{M}^{-1}\mathbf{A}^\top(\mathbf{p}^{k+1} - \mathbf{p}^k) \end{array} \right.$
PDS-IV	$\left\lfloor \begin{array}{lll} \mathbf{x}^{k+1} & := & \mathrm{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^k\big) \\ \tilde{\mathbf{p}}^k & := & \mathrm{prox}_{g^*}^{\mathbf{\Gamma}}(\mathbf{p}^k + \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{x}^{k+1}) \\ \mathbf{p}^{k+1} & := & \tilde{\mathbf{p}}^k - \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top(\tilde{\mathbf{p}}^k - \mathbf{p}^k) + \mathbf{\Gamma}^{-1}\mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k) \end{array} \right.$
PDS-V [46, Eq.(30)]	$\left\lfloor \begin{array}{lll} \tilde{\mathbf{p}}^k & := & \mathrm{prox}_{g^*}^{\mathbf{\Gamma}}(\mathbf{p}^k + \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{x}^k) \\ \mathbf{x}^{k+1} & := & \mathrm{prox}_f^{\mathbf{M}}(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^{k+1}) \\ \mathbf{p}^{k+1} & := & \tilde{\mathbf{p}}^k - \mathbf{\Gamma}^{-1}\mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k) \end{array} \right.$
PDS-VI [46, Eq.(31)]	$\left\lfloor \begin{array}{lll} \tilde{\mathbf{x}}^k & := & \mathrm{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^k\big) \\ \mathbf{p}^{k+1} & := & \mathrm{prox}_{g^*}^{\mathbf{\Gamma}}(\mathbf{p}^k + \mathbf{\Gamma}^{-1}\mathbf{A}\tilde{\mathbf{x}}^k) \\ \mathbf{x}^{k+1} & := & \tilde{\mathbf{x}}^k - \mathbf{M}^{-1}\mathbf{A}^\top(\mathbf{p}^{k+1} - \mathbf{p}^k) \end{array} \right.$
PDS-VII [46, Eq.(32)]	$\left\lfloor \begin{array}{lll} \tilde{\mathbf{x}}^k & := & \mathrm{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{p}^k\big) \\ \tilde{\mathbf{p}}^k & := & \mathrm{prox}_{g^*}^{\mathbf{\Gamma}}(\mathbf{p}^k + \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{x}^k) \\ \mathbf{x}^{k+1} & := & \tilde{\mathbf{x}}^k - \mathbf{M}^{-1}\mathbf{A}^\top(\tilde{\mathbf{p}}^k - \mathbf{p}^k) \\ \mathbf{p}^{k+1} & := & \tilde{\mathbf{p}}^k + \mathbf{\Gamma}^{-1}\mathbf{A}(\tilde{\mathbf{x}}^k - \mathbf{x}^k) \end{array} \right.$

Table 5 The PPA reinterpretations of the proposed PDS algorithms

schemes	$\mathbf{c}$	$\mathcal{A}$	$\mathcal{Q}$	$\mathcal{M}$
PDS-I	$\begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix}$	$\begin{bmatrix} \partial f & \mathbf{A}^\top \\ -\mathbf{A} & \partial g^* \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^\top \\ -\mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	$\mathbf{I}_{M+N}$
PDS-II			$\begin{bmatrix} \mathbf{M} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	
PDS-III			$\begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N - \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{\Gamma}^{-1}\mathbf{A} & -\mathbf{M}^{-1}\mathbf{A}^\top \\ \mathbf{0} & \mathbf{I}_M \end{bmatrix}$
PDS-IV			$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{\Gamma}^{-1}\mathbf{A} & \mathbf{I}_M - \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top \end{bmatrix}$
PDS-V			$\begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{\Gamma}^{-1}\mathbf{A} & \mathbf{I}_M \end{bmatrix}$
PDS-VI			$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & -\mathbf{M}^{-1}\mathbf{A}^\top \\ \mathbf{0} & \mathbf{I}_M \end{bmatrix}$
PDS-VII			$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^\top \\ \mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & -\mathbf{M}^{-1}\mathbf{A}^\top \\ \mathbf{\Gamma}^{-1}\mathbf{A} & \mathbf{I}_M \end{bmatrix}$

Proof (i) First, note that the proximal step of all the PDS schemes listed in Table 4 can be written as:

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \begin{bmatrix} \partial f & \mathbf{A}^\top \\ -\mathbf{A} & \partial g^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}^k \\ \tilde{\mathbf{p}}^k \end{bmatrix} + \begin{bmatrix} \text{—}\mathbf{Q}_1\text{—} \\ \text{—}\mathbf{Q}_2\text{—} \end{bmatrix} (\tilde{\mathbf{c}}^k - \mathbf{c}^k),$$

which is

$$\left\lfloor \begin{array}{l} \mathbf{0} \in \partial f(\tilde{\mathbf{x}}^k) + \mathbf{A}^\top\tilde{\mathbf{p}}^k + \mathbf{Q}_1(\tilde{\mathbf{c}}^k - \mathbf{c}^k), \\ \mathbf{0} \in \partial g^*(\tilde{\mathbf{p}}^k) - \mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{Q}_2(\tilde{\mathbf{c}}^k - \mathbf{c}^k). \end{array} \right. \tag{28}$$

Table 6 The corresponding $\mathcal{S}$ and $\mathcal{G}$ of the proposed PDS algorithms

schemes	$\mathcal{S}$	$\mathcal{G}$	convergence condition
PDS-I	$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^\top \\ -\mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^\top \\ -\mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	$\mathbf{M}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{A}^\top \mathbf{\Gamma}^{-1} \mathbf{A}$ or $\mathbf{\Gamma} \succ \mathbf{A} \mathbf{M}^{-1} \mathbf{A}^\top$
PDS-II	$\begin{bmatrix} \mathbf{M} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	
PDS-III	$\begin{bmatrix} \mathbf{M}^{-1} & -\mathbf{M}^{-1}\mathbf{A}^\top\mathbf{\Gamma}^{-1} \\ -\mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{M}^{-1} & \mathbf{\Gamma}^{-1} \end{bmatrix}^{-1}$	$\begin{bmatrix} \mathbf{M} + \mathbf{A}^\top\mathbf{\Gamma}^{-1}\mathbf{A} & \mathbf{A}^\top \\ \mathbf{A} & \mathbf{\Gamma} \end{bmatrix}$	
PDS-IV	$\begin{bmatrix} \mathbf{M}^{-1} & \mathbf{M}^{-1}\mathbf{A}^\top\mathbf{\Gamma}^{-1} \\ \mathbf{\Gamma}^{-1}\mathbf{A}\mathbf{M}^{-1} & \mathbf{\Gamma}^{-1} \end{bmatrix}^{-1}$	$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^\top \\ -\mathbf{A} & \mathbf{\Gamma} + \mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top \end{bmatrix}$	
PDS-V	$\begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} - \mathbf{A}^\top\mathbf{\Gamma}^{-1}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} \end{bmatrix}$	$\mathbf{M}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{A}^\top \mathbf{\Gamma}^{-1} \mathbf{A}$
PDS-VI		$\begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} - \mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top \end{bmatrix}$	$\mathbf{M}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{\Gamma} \succ \mathbf{A} \mathbf{M}^{-1} \mathbf{A}^\top$
PDS-VII		$\begin{bmatrix} \mathbf{M} - \mathbf{A}^\top\mathbf{\Gamma}^{-1}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma} - \mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top \end{bmatrix}$	$\mathbf{M}, \mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{A}^\top \mathbf{\Gamma}^{-1} \mathbf{A}$ $\mathbf{\Gamma} \succ \mathbf{A} \mathbf{M}^{-1} \mathbf{A}^\top$

Then, by convexity of f and g^*, we develop:

$$
\begin{aligned}
f(\mathbf{x}) &\geq f(\tilde{\mathbf{x}}^k) + \langle \partial f(\tilde{\mathbf{x}}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle \\
&= f(\tilde{\mathbf{x}}^k) - \langle \mathbf{A}^\top \tilde{\mathbf{p}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle - \langle \mathbf{Q}_1 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle, \quad \text{by (28)}
\end{aligned}
$$

and

$$
\begin{aligned}
g^*(\mathbf{p}) &\geq g^*(\tilde{\mathbf{p}}^k) + \langle \partial g^*(\tilde{\mathbf{p}}^k) | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle \\
&= g^*(\tilde{\mathbf{a}}^k) + \langle \mathbf{A} \tilde{\mathbf{x}}^k | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle - \langle \mathbf{Q}_2 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle. \quad \text{by (28)}
\end{aligned}
$$

Summing up both inequalities yields

$$
\begin{aligned}
f(\mathbf{x}) + g^*(\mathbf{p}) - f(\tilde{\mathbf{x}}^k) - g^*(\tilde{\mathbf{p}}^k) \geq & -\langle \mathbf{A}^\top \tilde{\mathbf{p}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle - \langle \mathbf{Q}_1 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle \\
& + \langle \mathbf{A} \tilde{\mathbf{x}}^k | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle - \langle \mathbf{Q}_2 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle.
\end{aligned}
$$

Finally, we have

$$
\begin{aligned}
& \mathcal{L}(\tilde{\mathbf{x}}^k, \mathbf{p}) - \mathcal{L}(\mathbf{x}, \tilde{\mathbf{p}}^k) \\
&= f(\tilde{\mathbf{x}}^k) + g^*(\tilde{\mathbf{p}}^k) - f(\mathbf{x}) - g^*(\mathbf{p}) + \langle \mathbf{p} | \mathbf{A} \tilde{\mathbf{x}}^k \rangle - \langle \tilde{\mathbf{p}}^k | \mathbf{A} \mathbf{x} \rangle \\
&\leq \langle \mathbf{Q}_1 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle + \langle \mathbf{Q}_2 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle \\
&+ \langle \mathbf{p} | \mathbf{A} \tilde{\mathbf{x}}^k \rangle - \langle \tilde{\mathbf{p}}^k | \mathbf{A} \mathbf{x} \rangle + \langle \mathbf{A}^\top \tilde{\mathbf{p}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle - \langle \mathbf{A} \tilde{\mathbf{x}}^k | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle \\
&= \langle \mathbf{Q}_1 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k \rangle + \langle \mathbf{Q}_2 (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{p} - \tilde{\mathbf{p}}^k \rangle \\
&= \langle \mathcal{Q} (\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{c} - \tilde{\mathbf{c}}^k \rangle.
\end{aligned}
$$

(ii) similar to the proof of Lemma 3-(ii).

Similar to the Lagrangian schemes, for the PDS schemes, $\Pi(\mathbf{c}, \mathbf{c}^\star)$ with $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$ essentially belongs to the generalized Bregman distance associated with $q(\mathbf{c}) := f(\mathbf{x}) + g^*(\mathbf{p})$ between any point $\mathbf{c} = (\mathbf{x}, \mathbf{p})$ and a saddle point

$\mathbf{c}^\star = (\mathbf{x}^\star, \mathbf{p}^\star)$, which satisfies $0 \le D_q^\flat(\mathbf{c}, \mathbf{c}^\star) \le \Pi(\mathbf{c}, \mathbf{c}^\star) \le D_q^\sharp(\mathbf{c}, \mathbf{c}^\star)$. Indeed, the generalized Bregman distance is given as

$$\begin{aligned}
0 \le D_q^\flat(\mathbf{c}, \mathbf{c}^\star) &= \inf_{\mathbf{v} \in \partial q(\mathbf{c}^\star)} q(\mathbf{c}) - q(\mathbf{c}^\star) + \langle \mathbf{v} | \mathbf{c}^\star - \mathbf{c} \rangle \\
&= f(\mathbf{x}) - f(\mathbf{x}^\star) + \inf_{\mathbf{v} \in \partial f(\mathbf{x}^\star)} \langle \mathbf{v} | \mathbf{x}^\star - \mathbf{x} \rangle + g^*(\mathbf{p}) - g^*(\mathbf{p}^\star) + \inf_{\mathbf{t} \in \partial g^*(\mathbf{p}^\star)} \langle \mathbf{t} | \mathbf{p}^\star - \mathbf{p} \rangle \\
&\le f(\mathbf{x}) - f(\mathbf{x}^\star) + \langle \mathbf{A}^\top \mathbf{p}^\star | \mathbf{x} - \mathbf{x}^\star \rangle + g^*(\mathbf{p}) - g^*(\mathbf{p}^\star) - \langle \mathbf{A}\mathbf{x}^\star | \mathbf{p} - \mathbf{p}^\star \rangle \\
&= f(\mathbf{x}) - f(\mathbf{x}^\star) + g^*(\mathbf{p}) - g^*(\mathbf{p}^\star) + \langle \mathbf{A}\mathbf{x} | \mathbf{p}^\star \rangle - \langle \mathbf{A}\mathbf{x}^\star | \mathbf{p} \rangle \\
&= \mathcal{L}(\mathbf{x}, \mathbf{p}^\star) - \mathcal{L}(\mathbf{x}^\star, \mathbf{p}) \\
&= \Pi(\mathbf{c}, \mathbf{c}^\star) \le D_q^\sharp(\mathbf{c}, \mathbf{c}^\star).
\end{aligned}$$

Then, we obtain the convergence rate of $\Pi(\mathbf{c}^k, \mathbf{c}^\star)$ in an ergodic sense.

Theorem 4 *For all the PDS algorithms shown in Table 4, the generalized Bregman distance generated by $f(\mathbf{x})+g^*(\mathbf{p})$ between the ergodic point $\frac{1}{k}\sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i$ and a saddle point $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$ has a rate of $\mathcal{O}(1/k)$:*

$$0 \le \Pi\bigg(\frac{1}{k} \sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i, \mathbf{c}^\star \bigg) \le \frac{1}{2k} \big\| \mathbf{c}^0 - \mathbf{c}^\star \big\|_{\mathcal{S}}^2,$$

where $\{\tilde{\mathbf{c}}^i\}_{i\in\mathbb{N}}$ and $\mathcal{S}$ are defined in Lemma 4 and 1.

The proof is similar to Theorem 2.

Likewise, for the class of PDS algorithms, for given sets $B_1 \subset \mathbb{R}^N$ and $B_2 \subset \mathbb{R}^M$, the *primal-dual gap* function restricted to $B_1 \times B_2$ is defined as:

$$\Psi_{B_1 \times B_2}(\mathbf{c}) = \sup_{\mathbf{p}' \in B_2} \mathcal{L}(\mathbf{x}, \mathbf{p}') - \inf_{\mathbf{x}' \in B_1} \mathcal{L}(\mathbf{x}', \mathbf{p}), \tag{29}$$

which has the upper bound:

Corollary 3 *Under the conditions of Theorem 4, if the set $B_1 \times B_2$ is bounded, the ergodic primal-dual gap defined as* (29) *has the upper bound:*

$$\Psi_{B_1 \times B_2}\bigg(\frac{1}{k} \sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i \bigg) \le \frac{1}{2k} \sup_{\mathbf{c} \in B_1 \times B_2} \big\| \mathbf{c}^0 - \mathbf{c} \big\|_{\mathcal{S}}^2.$$

Furthermore, $\Psi_{B_1 \times B_2}(\frac{1}{k}\sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i) \ge 0$, if the set $B_1 \times B_2$ contains a saddle point $\mathbf{c}^\star = (\mathbf{x}^\star, \mathbf{p}^\star) \in \mathsf{zer}\mathcal{A}$.

The proof is similar to Corollary 2.

Remark 6 For `PDS-I` and `II` with corresponding $\mathcal{M} = \mathcal{I}$, Theorem 4 and Corollary 3 can be simplified as $\Pi\big(\frac{1}{k}\sum_{i=1}^{k} \mathbf{c}^i, \mathbf{c}^\star\big) \le \frac{1}{2k}\big\|\mathbf{c}^0 - \mathbf{c}^\star\big\|_{\mathcal{Q}}^2$ and $\Psi_{B_1 \times B_2}$ $\big(\frac{1}{k}\sum_{i=1}^{k} \mathbf{c}^i\big) \le \frac{1}{2k} \sup_{\mathbf{c} \in B_1 \times B_2} \big\|\mathbf{c}^0 - \mathbf{c}\big\|_{\mathcal{Q}}^2$.

Remark 7 Similarly to Remark 5, under additional conditions on f and g^*, one can obtain *the convergence rate of $\mathcal{O}(1/k)$ of the sequence of the primal value of* (1)*, evaluated at the ergodic averaging point* $\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i$, namely, it holds that:

$$f\Big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{x}}^i\Big)+g\bigg(\mathbf{A}\Big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{x}}^i\Big)\bigg)-f(\mathbf{x}^\star)-g(\mathbf{A}\mathbf{x}^\star)\le\frac{1}{2k}\sup_{\mathbf{c}\in B_1\times B_2}\big\|\mathbf{c}^0-\mathbf{c}\big\|_{\mathcal{S}}^2. \tag{30}$$

Indeed, if $\mathsf{dom}f$ and $\mathsf{dom}g^*$ are bounded, then we can simply take the sets $B_1=\mathsf{dom}f$ and $B_2=\mathsf{dom}g^*$. Denoting the ergodic averaging points by $\widehat{\mathbf{c}}^k=\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i$ ($\widehat{\mathbf{x}}^k$ and $\widehat{\mathbf{p}}^k$ are defined similarly), using Fenchel-Young inequality [4, Proposition 13.15], we develop

$$\begin{aligned}
&\Psi_{B_1\times B_2}\big(\widehat{\mathbf{c}}^k\big)\\
=&\sup_{\mathbf{p}'\in B_2}\mathcal{L}\big(\widehat{\mathbf{x}}^k,\mathbf{p}'\big)-\inf_{\mathbf{x}'\in B_1}\mathcal{L}\big(\mathbf{x}',\widehat{\mathbf{p}}^k\big)\\
=&\sup_{\mathbf{p}'\in B_2}f\big(\widehat{\mathbf{x}}^k\big)+\big\langle\mathbf{p}'\big|\mathbf{A}\widehat{\mathbf{x}}^k\big\rangle-g^*(\mathbf{p}')-\inf_{\mathbf{x}'\in B_1}\Big(f(\mathbf{x}')+\big\langle\widehat{\mathbf{p}}^k\big|\mathbf{A}\mathbf{x}'\big\rangle-g^*\big(\widehat{\mathbf{p}}^k\big)\Big)\\
\ge&f\big(\widehat{\mathbf{x}}^k\big)+g\big(\mathbf{A}\widehat{\mathbf{x}}^k\big)+g^*\big(\widehat{\mathbf{p}}^k\big)-f(\mathbf{x}^\star)-\big\langle\widehat{\mathbf{p}}^k\big|\mathbf{A}\mathbf{x}^\star\big\rangle\\
\ge&f\big(\widehat{\mathbf{x}}^k\big)+g\big(\mathbf{A}\widehat{\mathbf{x}}^k\big)-f(\mathbf{x}^\star)-g(\mathbf{A}\mathbf{x}^\star),
\end{aligned}$$

which, combining with Corollary 3, yields (30). Still, (30) holds for all the PDS algorithms shown in Table 4.

4.3 Reductions of some PDHG algorithms

By Corollary 1-(iii), `PDS-I` and `II` can be reduced to a simple resolvent (19), where $\mathbf{v}^k=\mathcal{Q}^{\frac{1}{2}}\mathbf{c}^k$, $\mathcal{Q}$ is specified in Table 5 for `PDS-I` or `II`.

A degenerate case In particular, if $\mathbf{A}=\mathbf{I}_N$, $\mathbf{M}=\mathbf{I}_N$, $\boldsymbol{\Gamma}=\mathbf{I}_N$, then, $\mathcal{Q}=\begin{bmatrix}\mathbf{I}_N & -\mathbf{I}_N\\ -\mathbf{I}_N & \mathbf{I}_N\end{bmatrix}$ for `PDS-I` or $\mathcal{Q}=\begin{bmatrix}\mathbf{I}_N & \mathbf{I}_N\\ \mathbf{I}_N & \mathbf{I}_N\end{bmatrix}$ for `PDS-II`, which becomes degenerate. The convergence of this case, which is not covered by Table 6, can be answered by the degenerate analysis.

As an example, let us consider `PDS-I`, which becomes

$$\left\lfloor\begin{aligned}\mathbf{x}^{k+1}&:=\mathsf{prox}_f\big(\mathbf{x}^k-\mathbf{p}^k\big),\\ \mathbf{p}^{k+1}&:=\mathsf{prox}_{g^*}\big(\mathbf{p}^k+2\mathbf{x}^{k+1}-\mathbf{x}^k\big).\end{aligned}\right. \tag{31}$$

The metric $\mathcal{Q}$ can be decomposed as $\mathcal{Q}=\mathcal{D}\mathcal{D}^\top=\begin{bmatrix}\mathbf{I}_N\\ -\mathbf{I}_N\end{bmatrix}\begin{bmatrix}\mathbf{I}_N & -\mathbf{I}_N\end{bmatrix}$. Then, following the procedure similar to Proposition 3, we obtain the reduced PPA as:

$$\mathbf{v}^{k+1}=\big(\mathcal{I}+(\mathcal{D}^\top\mathcal{A}^{-1}\mathcal{D})^{-1}\big)^{-1}\mathbf{v}^k,$$

where $\mathbf{v}^k=\mathcal{D}^\top\mathbf{c}^k=\mathbf{x}^k-\mathbf{p}^k$.

The active variable of (31) can also be identified without the degenerate PPA analysis. Indeed, from (31), we have

$$\mathbf{x}^{k+1} - \mathbf{p}^{k+1} = \mathsf{prox}_f\big(\mathbf{x}^k - \mathbf{p}^k\big) - \mathsf{prox}_{g^*}\big(\mathbf{p}^k + 2\mathbf{x}^{k+1} - \mathbf{x}^k\big).$$

Denoting $\mathbf{v}^k := \mathbf{x}^k - \mathbf{p}^k$, it becomes

$$\begin{aligned}
\mathbf{v}^{k+1} &= \mathsf{prox}_f\big(\mathbf{v}^k\big) - \mathsf{prox}_{g^*}\big(2\mathbf{x}^{k+1} - \mathbf{v}^k\big) \\
&= \mathsf{prox}_f\big(\mathbf{v}^k\big) - \mathsf{prox}_{g^*}\big(2\mathsf{prox}_f(\mathbf{v}^k) - \mathbf{v}^k\big) \\
&= \Big(\mathsf{prox}_f - \mathsf{prox}_{g^*} \circ (2\mathsf{prox}_f - \mathcal{I})\Big)(\mathbf{v}^k) \\
&= \Big(\mathsf{prox}_f - (\mathcal{I} - \mathsf{prox}_g) \circ (2\mathsf{prox}_f - \mathcal{I})\Big)(\mathbf{v}^k) \\
&= \Big(\mathcal{I} - \mathsf{prox}_f + \mathsf{prox}_g \circ (2\mathsf{prox}_f - \mathcal{I})\Big)(\mathbf{v}^k),
\end{aligned}$$

which shows that (31) is essentially a DRS algorithm.

5 Operator splitting based on mixed strategies

Consider the hybrid strategy proposed in [46, Sect. 6], which aims at minimizing [46, Eq.(34)]:

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}) + h(\mathbf{B}\mathbf{x}), \tag{32}$$

where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{A} : \mathbb{R}^N \mapsto \mathbb{R}^{M_1}$, $\mathbf{B} : \mathbb{R}^N \mapsto \mathbb{R}^{M_2}$, $f : \mathbb{R}^N \mapsto \mathbb{R} \cup \{+\infty\}$, $g : \mathbb{R}^{M_1} \mapsto \mathbb{R} \cup \{+\infty\}$, $h : \mathbb{R}^{M_2} \mapsto \mathbb{R} \cup \{+\infty\}$.

5.1 The hybrid schemes and their PPA interpretations

Taking Lagrangian of g, and applying primal-dual to h in (32) yields [46, Eq(35)]:

$$\mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{b}, \mathbf{p}) := f(\mathbf{x}) + g(\mathbf{a}) + \mathbf{p}^\top(\mathbf{A}\mathbf{x} - \mathbf{a}) + \mathbf{b}^\top \mathbf{B}\mathbf{x} - h^*(\mathbf{b}), \tag{33}$$

or

$$\mathcal{L}_{\mathbf{\Theta}}(\mathbf{x}, \mathbf{a}, \mathbf{b}, \mathbf{p}) := f(\mathbf{x}) + g(\mathbf{a}) + \mathbf{p}^\top(\mathbf{A}\mathbf{x} - \mathbf{a}) + \frac{1}{2}\big\|\mathbf{A} - \mathbf{a}\big\|_{\mathbf{\Theta}}^2 + \mathbf{b}^\top \mathbf{B}\mathbf{x} - h^*(\mathbf{b}), \tag{34}$$

where $\mathbf{p} \in \mathbb{R}^{M_1}$, $\mathbf{a} \in \mathbb{R}^{M_1}$, $\mathbf{b} \in \mathbb{R}^{M_2}$, we devise the hybrid schemes based on the alternating optimization of (33) or (34), shown in Table 7. `MIX-I,III,IV` and `V` can be found in [46, Sect. 6], and are extended to general proximal metrics here.

The hybrid schemes can be interpreted by alternating optimization. For instance, the $\mathbf{x}$-updates of `MIX-I`, `MIX-II`, `MIX-IV` and `MIX-V` are from non-augmented (33):

$$\mathbf{x}^{k+1} := \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{a}^k, \mathbf{b}^k, \mathbf{p}^k) + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{M}}^2.$$

Table 7 The proposed hybrid algorithms

name	iterative scheme
MIX-I [46, Eq.(36)]	$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} = \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}(\mathbf{A}^\top\mathbf{p}^k + \mathbf{B}^\top\mathbf{b}^k)\big) \\ \mathbf{a}^{k+1} = \mathsf{prox}_g^{\boldsymbol{\Omega}}(\mathbf{a}^k + \boldsymbol{\Omega}^{-1}\mathbf{p}^k) \\ \mathbf{b}^{k+1} = \mathsf{prox}_{h^*}^{\boldsymbol{\Gamma}^{-1}}(\mathbf{b}^k + \boldsymbol{\Gamma}\mathbf{B}(2\mathbf{x}^{k+1} - \mathbf{x}^k)) \\ \mathbf{p}^{k+1} = \mathbf{p}^k + \boldsymbol{\Theta}\big(\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k) - (2\mathbf{a}^{k+1} - \mathbf{a}^k)\big) \end{array} \right.$
MIX-II	$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} = \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}(\mathbf{B}^\top\mathbf{b}^k + \mathbf{A}^\top\mathbf{p}^k)\big) \\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Theta}\big(\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k) - \mathbf{a}^k\big) \\ \mathbf{a}^{k+1} = \mathsf{prox}_g^{\boldsymbol{\Omega}}\big(\mathbf{a}^k + \boldsymbol{\Omega}^{-1}(2\mathbf{p}^{k+1} - \mathbf{p}^k)\big) \\ \mathbf{b}^{k+1} = \mathsf{prox}_{h^*}^{\boldsymbol{\Gamma}^{-1}}\big(\mathbf{b}^k + \boldsymbol{\Gamma}\mathbf{B}(2\mathbf{x}^{k+1} - \mathbf{x}^k)\big) \end{array} \right.$
MIX-III [46, Eq.(39)]	$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Theta}\mathbf{A}}\big((\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Theta}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k + \mathbf{A}^\top\boldsymbol{\Theta}\mathbf{a}^k - \mathbf{B}^\top\mathbf{b}^k - \mathbf{A}^\top\mathbf{p}^k)\big) \\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Theta}\big(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^k\big) \\ \mathbf{a}^{k+1} = \mathsf{prox}_g^{\boldsymbol{\Omega}}(\mathbf{a}^k + \boldsymbol{\Omega}^{-1}(2\mathbf{p}^{k+1} - \mathbf{p}^k)) \\ \mathbf{b}^{k+1} = \mathsf{prox}_{h^*}^{\boldsymbol{\Gamma}^{-1}}(\mathbf{b}^k + \boldsymbol{\Gamma}\mathbf{B}(2\mathbf{x}^{k+1} - \mathbf{x}^k)) \end{array} \right.$
MIX-IV [46, Eq.(40)]	$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} = \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}(\mathbf{A}^\top\mathbf{p}^k + \mathbf{B}^\top\mathbf{b}^k)\big) \\ \mathbf{a}^{k+1} = \mathsf{prox}_g^{\boldsymbol{\Omega}+\boldsymbol{\Theta}}\big((\boldsymbol{\Omega}+\boldsymbol{\Theta})^{-1}(\boldsymbol{\Omega}\mathbf{a}^k + \mathbf{p}^k + \boldsymbol{\Theta}\mathbf{A}\mathbf{x}^{k+1})\big) \\ \mathbf{b}^{k+1} = \mathsf{prox}_{h^*}^{\boldsymbol{\Gamma}^{-1}}(\mathbf{b}^k + \boldsymbol{\Gamma}\mathbf{B}\mathbf{x}^{k+1}) + \boldsymbol{\Gamma}\mathbf{B}(\mathbf{x}^{k+1} - \mathbf{x}^k) \\ \mathbf{p}^{k+1} = \mathbf{p}^k + \boldsymbol{\Theta}(2\mathbf{A}\mathbf{x}^{k+1} - \mathbf{A}\mathbf{x}^k - \mathbf{a}^{k+1}) \end{array} \right.$
MIX-V [46, Eq.(37)]	$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} = \mathsf{prox}_f^{\mathbf{M}}\big(\mathbf{x}^k - \mathbf{M}^{-1}(\mathbf{A}^\top\mathbf{p}^k + \mathbf{B}^\top\mathbf{b}^k)\big) \\ \mathbf{a}^{k+1} = \mathsf{prox}_g^{\boldsymbol{\Omega}}(\mathbf{a}^k + \boldsymbol{\Omega}^{-1}\mathbf{p}^k) \\ \mathbf{b}^{k+1} = \mathsf{prox}_{h^*}^{\boldsymbol{\Gamma}^{-1}}(\mathbf{b}^k + \boldsymbol{\Gamma}\mathbf{B}\mathbf{x}^{k+1}) + \boldsymbol{\Gamma}\mathbf{B}(\mathbf{x}^{k+1} - \mathbf{x}^k) \\ \mathbf{p}^{k+1} = \mathbf{p}^k + \boldsymbol{\Theta}\big(\mathbf{A}(2\mathbf{x}^{k+1} - \mathbf{x}^k) - (2\mathbf{a}^{k+1} - \mathbf{a}^k)\big) \end{array} \right.$
MIX-VI	$\left\lfloor \begin{array}{l} \mathbf{x}^{k+1} := \mathsf{prox}_f^{\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Theta}\mathbf{A}}\big((\mathbf{M}+\mathbf{A}^\top\boldsymbol{\Theta}\mathbf{A})^{-1}(\mathbf{M}\mathbf{x}^k + \mathbf{A}^\top\boldsymbol{\Theta}\mathbf{a}^k - \mathbf{B}^\top\mathbf{b}^k - \mathbf{A}^\top\mathbf{p}^k)\big) \\ \mathbf{a}^{k+1} := \mathsf{prox}_g^{\boldsymbol{\Omega}}\big(\mathbf{a}^k + \boldsymbol{\Omega}^{-1}(\boldsymbol{\Theta}\mathbf{A}\mathbf{x}^{k+1} - \boldsymbol{\Theta}\mathbf{a}^k + \mathbf{p}^k)\big) \\ \mathbf{b}^{k+1} := \mathsf{prox}_{h^*}^{\boldsymbol{\Gamma}^{-1}}\big(\mathbf{b}^k + \boldsymbol{\Gamma}\mathbf{B}(2\mathbf{x}^{k+1} - \mathbf{x}^k)\big) \\ \mathbf{p}^{k+1} := \mathbf{p}^k + \boldsymbol{\Theta}\big(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{a}^{k+1}\big) \end{array} \right.$

Table 8 The PPA reinterpretations of the hybrid algorithms

schemes	$\mathbf{c}$	$\mathcal{A}$	$\mathcal{Q}$	$\mathcal{M}$
MIX-I	$\begin{bmatrix} \mathbf{x} \\ \mathbf{a} \\ \mathbf{b} \\ \mathbf{p} \end{bmatrix}$	$\begin{bmatrix} \partial f & \mathbf{0} & \mathbf{B}^\top & \mathbf{A}^\top \\ \mathbf{0} & \partial g & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \partial h^* & \mathbf{0} \\ -\mathbf{A} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \boldsymbol{\Omega} & \mathbf{0} & \mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \boldsymbol{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & \mathbf{I}_{M_1} & \mathbf{0} & \boldsymbol{\Theta}^{-1} \end{bmatrix}$	$\mathbf{I}_{N+2M_1+M_2}$
MIX-II			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \boldsymbol{\Omega} & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \boldsymbol{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & -\mathbf{I}_{M_1} & \mathbf{0} & \boldsymbol{\Theta}^{-1} \end{bmatrix}$	
MIX-III			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \boldsymbol{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{M_1} & \mathbf{0} & \boldsymbol{\Theta}^{-1} \end{bmatrix}$	
MIX-IV			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \boldsymbol{\Omega} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Gamma}\mathbf{B} & \mathbf{0} & \mathbf{I}_{M_2} & \mathbf{0} \\ \boldsymbol{\Theta}\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{M_1} \end{bmatrix}$
MIX-V			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \boldsymbol{\Omega} & \mathbf{0} & \mathbf{I}_{M_1} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Gamma}\mathbf{B} & \mathbf{0} & \mathbf{I}_{M_2} & \mathbf{0} \\ \boldsymbol{\Theta}\mathbf{A} & -\boldsymbol{\Theta} & \mathbf{0} & \mathbf{I}_{M_1} \end{bmatrix}$
MIX-VI			$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} & \boldsymbol{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{M_1} & \mathbf{0} & \boldsymbol{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{M_2} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\Theta} & \mathbf{0} & \mathbf{I}_{M_1} \end{bmatrix}$

The $\mathbf{a}$-updates of `MIX-I`, `MIX-II`, `MIX-III` and `MIX-V` are from non-augmented (33). The $\mathbf{a}$-update of `MIX-I`, for instance, is

$$\mathbf{a}^{k+1} := \arg\min_{\mathbf{a}} \mathcal{L}(\mathbf{x}^k, \mathbf{a}, \mathbf{b}^k, \mathbf{p}^k) + \frac{1}{2}\|\mathbf{a} - \mathbf{a}^k\|_{\boldsymbol{\Omega}}^2.$$

Table 9 The corresponding $\mathcal{S}$ and $\mathcal{G}$ of the hybrid algorithms

schemes	$\mathcal{S}$	$\mathcal{G}$	convergence condition
`MIX-I`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & \mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & \mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\mathbf{M},\mathbf{\Omega},\mathbf{\Theta},\mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{A}^\top\mathbf{\Theta}\mathbf{A} + \mathbf{B}^\top\mathbf{\Gamma}\mathbf{B}$ $\mathbf{\Omega} \succ \mathbf{\Theta}$
`MIX-II`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & -\mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & -\mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	
`MIX-III`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\mathbf{M},\mathbf{\Omega},\mathbf{\Gamma},\mathbf{\Theta} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{B}^\top\mathbf{\Gamma}\mathbf{B}$ $\mathbf{\Omega} \succ \mathbf{\Theta}$
`MIX-IV`	$\begin{bmatrix} \mathbf{M}+\mathbf{A}^\top\mathbf{\Theta}\mathbf{A}+\mathbf{B}^\top\mathbf{\Gamma}\mathbf{B} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\mathbf{M},\mathbf{\Theta},\mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{\Omega} \in \mathbb{S}_{+}$ $\mathbf{M} \succ \mathbf{A}^\top\mathbf{\Theta}\mathbf{A} + \mathbf{B}^\top\mathbf{\Gamma}\mathbf{B}$
`MIX-V`	$\begin{bmatrix} \mathbf{M}+\mathbf{A}^\top\mathbf{\Theta}\mathbf{A}+\mathbf{B}^\top\mathbf{\Gamma}\mathbf{B} & -\mathbf{A}^\top\mathbf{\Theta} & -\mathbf{B}^\top & -\mathbf{A}^\top \\ -\mathbf{\Theta}\mathbf{A} & \mathbf{\Omega}+\mathbf{\Theta} & \mathbf{0} & \mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ -\mathbf{A} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & \mathbf{B}^\top & \mathbf{A}^\top \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & \mathbf{I}_{M_1} \\ \mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{A} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\mathbf{M},\mathbf{\Omega},\mathbf{\Theta},\mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{A}^\top\mathbf{\Theta}\mathbf{A} + \mathbf{B}^\top\mathbf{\Gamma}\mathbf{B}$ $\mathbf{\Omega} \succ \mathbf{\Theta}$
`MIX-VI`	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\begin{bmatrix} \mathbf{M} & \mathbf{0} & -\mathbf{B}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}-\mathbf{\Theta} & \mathbf{0} & \mathbf{0} \\ -\mathbf{B} & \mathbf{0} & \mathbf{\Gamma}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Theta}^{-1} \end{bmatrix}$	$\mathbf{M},\mathbf{\Omega},\mathbf{\Theta},\mathbf{\Gamma} \in \mathbb{S}_{++}$ $\mathbf{M} \succ \mathbf{B}^\top\mathbf{\Gamma}\mathbf{B}$ $\mathbf{\Omega} \succ \mathbf{\Theta}$

The $\mathbf{x}$-updates of `MIX-III` and `MIX-VI` are from augmented (34):

$$\mathbf{x}^{k+1} := \arg\min_{\mathbf{x}} \mathcal{L}_{\mathbf{\Theta}}(\mathbf{x}, \mathbf{a}^k, \mathbf{b}^k, \mathbf{p}^k) + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|^2_{\mathbf{M}}.$$

The $\mathbf{a}$-update of `MIX-VI` is from

$$\mathbf{a}^{k+1} := \arg\min_{\mathbf{a}} \mathcal{L}_{\mathbf{\Theta}}(\mathbf{x}^k, \mathbf{a}, \mathbf{b}^k, \mathbf{p}^k) + \frac{1}{2}\|\mathbf{a} - \mathbf{a}^k\|^2_{\mathbf{\Omega}}.$$

Following the discussion of [46, Sect. 6], it is easy to verify that `MIX-I` and `III` can be reduced to `PDS-I` under a certain conditions.

Again, the preconditioning can be applied to the $\mathbf{x}$-updates of `MIX-III,IV` or $\mathbf{a}$-update of `MIX-IV`.

5.2 The generalized Bregman distance and ergodic primal-dual gap

Similar to Sect. 3.4 and 4.2, the PPA framework could also provide a unified treatment of the primal-dual gap of the hybrid algorithms.

Lemma 5 *Given the hybrid form $\mathcal{L}(\mathbf{x}, \mathbf{a}, \mathbf{b}, \mathbf{p})$ as (33), consider all the hybrid schemes listed in Table 7, where $\tilde{\mathbf{c}}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{a}}^k, \tilde{\mathbf{b}}^k, \tilde{\mathbf{p}}^k)$ denotes the proximal output, when the schemes are interpreted by the PPA (shown in Table 8). Then, the following holds, $\forall \mathbf{c} = (\mathbf{x}, \mathbf{a}, \mathbf{b}, \mathbf{p}) \in \mathbb{R}^N \times \mathbb{R}^{M_1} \times \mathbb{R}^{M_2} \times \mathbb{R}^{M_1}$:*

(i) $\Pi(\tilde{\mathbf{c}}^k, \mathbf{c}) \le \big\langle \mathcal{Q}(\tilde{\mathbf{c}}^k - \mathbf{c}^k) \big| \mathbf{c} - \tilde{\mathbf{c}}^k \big\rangle$,

(ii) $\Pi\big(\frac{1}{k}\sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i, \mathbf{c}\big) \le \frac{1}{2k}\big\|\mathbf{c}^0 - \mathbf{c}\big\|^2_{\mathcal{S}}$.

Proof (i) First, note that the proximal step of all the hybrid schemes listed in Table 7 can be written as:

$$\begin{bmatrix}\mathbf{0}\\\mathbf{0}\\\mathbf{0}\\\mathbf{0}\end{bmatrix} \in \begin{bmatrix}\partial f & \mathbf{0} & \mathbf{B}^\top & \mathbf{A}^\top \\ \mathbf{0} & \partial g & \mathbf{0} & -\mathbf{I}_{M_1} \\ -\mathbf{B} & \mathbf{0} & \partial h^* & \mathbf{0} \\ -\mathbf{A} & \mathbf{I}_{M_1} & \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\tilde{\mathbf{x}}^k\\ \tilde{\mathbf{a}}^k \\ \tilde{\mathbf{b}}^k \\ \tilde{\mathbf{p}}^k\end{bmatrix} + \begin{bmatrix}—\mathbf{Q}_1— \\ —\mathbf{Q}_2— \\ —\mathbf{Q}_3— \\ —\mathbf{Q}_4—\end{bmatrix}(\tilde{\mathbf{c}}^k - \mathbf{c}^k),$$

which is

$$\begin{cases}\mathbf{0} \in \partial f(\tilde{\mathbf{x}}^k) + \mathbf{B}^\top \tilde{\mathbf{b}}^k + \mathbf{A}^\top \tilde{\mathbf{p}}^k + \mathbf{Q}_1(\tilde{\mathbf{c}}^k - \mathbf{c}^k), \\ \mathbf{0} \in \partial g(\tilde{\mathbf{a}}^k) - \tilde{\mathbf{p}}^k + \mathbf{Q}_2(\tilde{\mathbf{c}}^k - \mathbf{c}^k), \\ \mathbf{0} \in \partial h^*(\tilde{\mathbf{b}}^k) - \mathbf{B}\tilde{\mathbf{x}}^k + \mathbf{Q}_3(\tilde{\mathbf{c}}^k - \mathbf{c}^k), \\ \mathbf{0} = -\mathbf{A}\tilde{\mathbf{x}}^k + \tilde{\mathbf{a}}^k + \mathbf{Q}_4(\tilde{\mathbf{c}}^k - \mathbf{c}^k).\end{cases} \tag{35}$$

Then, by convexity of f, g and h^*, we develop:

$$\begin{aligned} f(\mathbf{x}) &\geq f(\tilde{\mathbf{x}}^k) + \langle \partial f(\tilde{\mathbf{x}}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle \\ &= f(\tilde{\mathbf{x}}^k) - \langle \mathbf{B}^\top \tilde{\mathbf{b}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle - \langle \mathbf{A}^\top \tilde{\mathbf{p}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle - \langle \mathbf{Q}_1(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle, \quad \text{by (35)}\end{aligned}$$

$$\begin{aligned} g(\mathbf{a}) &\geq g(\tilde{\mathbf{a}}^k) + \langle \partial g(\tilde{\mathbf{a}}^k) | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle \\ &= g(\tilde{\mathbf{a}}^k) + \langle \tilde{\mathbf{p}}^k | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle - \langle \mathbf{Q}_2(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle, \quad \text{by (35)}\end{aligned}$$

$$\begin{aligned} h^*(\mathbf{b}) &\geq h^*(\tilde{\mathbf{b}}^k) + \langle \partial h^*(\tilde{\mathbf{b}}^k) | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle \\ &= h^*(\tilde{\mathbf{b}}^k) + \langle \mathbf{B}\tilde{\mathbf{x}}^k | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle - \langle \mathbf{Q}_3(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle. \quad \text{by (35)}\end{aligned}$$

Summing up the above three inequalities yields

$$\begin{aligned} & f(\mathbf{x}) + g(\mathbf{a}) - f(\tilde{\mathbf{x}}^k) - g(\tilde{\mathbf{a}}^k) + h^*(\mathbf{b}) - h^*(\tilde{\mathbf{b}}^k) \\ \geq\ & -\langle \mathbf{B}^\top \tilde{\mathbf{b}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle - \langle \mathbf{A}^\top \tilde{\mathbf{p}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle - \langle \mathbf{Q}_1(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle \\ & + \langle \tilde{\mathbf{p}}^k | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle - \langle \mathbf{Q}_2(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle \\ & + \langle \mathbf{B}\tilde{\mathbf{x}}^k | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle - \langle \mathbf{Q}_3(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle.\end{aligned}$$

Finally, we have

$$\begin{aligned} & \mathcal{L}(\tilde{\mathbf{x}}^k, \tilde{\mathbf{a}}^k, \mathbf{b}, \mathbf{p}) - \mathcal{L}(\mathbf{x}, \mathbf{a}, \tilde{\mathbf{b}}^k, \tilde{\mathbf{p}}^k) \\ =\ & f(\tilde{\mathbf{x}}^k) + g(\tilde{\mathbf{a}}^k) + h^*(\tilde{\mathbf{b}}^k) - f(\mathbf{x}) - g(\mathbf{a}) - h^*(\mathbf{b}) \\ & + \langle \mathbf{p} | \mathbf{A}\tilde{\mathbf{x}}^k - \tilde{\mathbf{a}}^k\rangle - \langle \tilde{\mathbf{p}}^k | \mathbf{A}\mathbf{x} - \mathbf{a}\rangle - \langle \tilde{\mathbf{b}}^k | \mathbf{B}\mathbf{x}\rangle + \langle \mathbf{b} | \mathbf{B}\tilde{\mathbf{x}}^k\rangle \\ \leq\ & \langle \mathbf{Q}_1(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle + \langle \mathbf{Q}_2(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle + \langle \mathbf{Q}_3(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle \\ & + \langle \mathbf{B}^\top \tilde{\mathbf{b}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle + \langle \mathbf{A}^\top \tilde{\mathbf{p}}^k | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle - \langle \tilde{\mathbf{p}}^k | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle - \langle \mathbf{B}\tilde{\mathbf{x}}^k | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle \\ & + \langle \mathbf{p} | \mathbf{A}\tilde{\mathbf{x}}^k - \tilde{\mathbf{a}}^k\rangle - \langle \tilde{\mathbf{p}}^k | \mathbf{A}\mathbf{x} - \mathbf{a}\rangle - \langle \tilde{\mathbf{b}}^k | \mathbf{B}\mathbf{x}\rangle + \langle \mathbf{b} | \mathbf{B}\tilde{\mathbf{x}}^k\rangle \\ =\ & \langle \mathbf{Q}_1(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle + \langle \mathbf{Q}_2(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle + \langle \mathbf{Q}_3(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle \\ & + \langle \mathbf{A}\tilde{\mathbf{x}}^k - \tilde{\mathbf{a}}^k | \mathbf{p} - \tilde{\mathbf{p}}^k\rangle \\ =\ & \langle \mathbf{Q}_1(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{x} - \tilde{\mathbf{x}}^k\rangle + \langle \mathbf{Q}_2(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{a} - \tilde{\mathbf{a}}^k\rangle + \langle \mathbf{Q}_3(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{b} - \tilde{\mathbf{b}}^k\rangle \\ & + \langle \mathbf{Q}_4(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{p} - \tilde{\mathbf{p}}^k\rangle \quad \text{by (35)} \\ =\ & \langle \mathcal{Q}(\tilde{\mathbf{c}}^k - \mathbf{c}^k) | \mathbf{c} - \tilde{\mathbf{c}}^k\rangle.\end{aligned}$$

(ii) similar to Lemma 3-(ii) and Lemma 4-(ii).

Similar to the Lagrangian and PDS schemes, for the hybrid algorithms, $\Pi(\mathbf{c},\mathbf{c}^\star)$ with $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$ is essentially a special instance of generalized Bregman distance generated by $q(\mathbf{u}) := f(\mathbf{x}) + g(\mathbf{a}) + h^*(\mathbf{b})$ between any point $\mathbf{u} = (\mathbf{x},\mathbf{a},\mathbf{b})$ and a saddle point $\mathbf{u}^\star = (\mathbf{x}^\star,\mathbf{a}^\star,\mathbf{b}^\star)$. More specifically, $0 \le D_q^\flat(\mathbf{u},\mathbf{u}^\star) \le \Pi(\mathbf{c},\mathbf{c}^\star) \le D_q^\sharp(\mathbf{u},\mathbf{u}^\star)$. Indeed, the generalized Bregman distance is given as

$$\begin{aligned}
0 \le D_q^\flat(\mathbf{u},\mathbf{u}^\star) &= \inf_{\mathbf{v}\in\partial q(\mathbf{u}^\star)} q(\mathbf{u}) - q(\mathbf{u}^\star) + \langle \mathbf{v}|\mathbf{u}^\star - \mathbf{u}\rangle \\
&= f(\mathbf{x}) - f(\mathbf{x}^\star) + \inf_{\mathbf{v}\in\partial f(\mathbf{x}^\star)} \langle \mathbf{v}|\mathbf{x}^\star - \mathbf{x}\rangle + g(\mathbf{a}) - g(\mathbf{a}^\star) + \inf_{\mathbf{t}\in\partial g(\mathbf{a}^\star)} \langle \mathbf{t}|\mathbf{a}^\star - \mathbf{a}\rangle \\
&\quad + h^*(\mathbf{b}) - h^*(\mathbf{b}^\star) + \inf_{\mathbf{s}\in\partial h^*(\mathbf{b}^\star)} \langle \mathbf{s}|\mathbf{b}^\star - \mathbf{b}\rangle \\
&\le f(\mathbf{x}) - f(\mathbf{x}^\star) + \langle \mathbf{B}^\top\mathbf{b}^\star + \mathbf{A}^\top\mathbf{p}^\star|\mathbf{x} - \mathbf{x}^\star\rangle + g(\mathbf{a}) - g(\mathbf{a}^\star) - \langle \mathbf{p}^\star|\mathbf{a} - \mathbf{a}^\star\rangle \\
&\quad + h^*(\mathbf{b}) - h^*(\mathbf{b}^\star) - \langle \mathbf{B}\mathbf{x}^\star|\mathbf{b} - \mathbf{b}^\star\rangle \\
&= f(\mathbf{x}) - f(\mathbf{x}^\star) + g(\mathbf{a}) - g(\mathbf{a}^\star) + h^*(\mathbf{b}) - h^*(\mathbf{b}^\star) \\
&\quad + \langle \mathbf{B}\mathbf{x}|\mathbf{b}^\star\rangle - \langle \mathbf{B}\mathbf{x}^\star|\mathbf{b}\rangle + \langle \mathbf{p}^\star|\mathbf{A}\mathbf{x} - \mathbf{a}\rangle \\
&= f(\mathbf{x}) - f(\mathbf{x}^\star) + g(\mathbf{a}) - g(\mathbf{a}^\star) + h^*(\mathbf{b}) - h^*(\mathbf{b}^\star) \\
&\quad + \langle \mathbf{B}\mathbf{x}|\mathbf{b}^\star\rangle - \langle \mathbf{B}\mathbf{x}^\star|\mathbf{b}\rangle + \langle \mathbf{p}^\star|\mathbf{A}\mathbf{x} - \mathbf{a}\rangle - \langle \mathbf{p}^\star|\underbrace{\mathbf{A}\mathbf{x}^\star - \mathbf{a}^\star}_{=\mathbf{0}}\rangle \\
&= \mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b}^\star,\mathbf{p}^\star) - \mathcal{L}(\mathbf{x}^\star,\mathbf{a}^\star,\mathbf{b},\mathbf{p}) \\
&= \Pi(\mathbf{c},\mathbf{c}^\star) \le D_q^\sharp(\mathbf{u},\mathbf{u}^\star).
\end{aligned}$$

Then, we obtain the convergence rate of the generalized Bregman distance $\Pi(\mathbf{c}^k,\mathbf{c}^\star)$ in an ergodic sense.

Theorem 5 *For all the hybrid algorithms shown in Table 7, the generalized Bregman distance generated by $f(\mathbf{x}) + g(\mathbf{a}) + h^*(\mathbf{b})$ between the ergodic point $\frac{1}{k}\sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i$ and a saddle point $\mathbf{c}^\star \in \mathsf{zer}\mathcal{A}$ has a rate of $\mathcal{O}(1/k)$:*

$$0 \le \Pi\bigg(\frac{1}{k}\sum_{i=0}^{k-1} \tilde{\mathbf{c}}^i, \mathbf{c}^\star\bigg) \le \frac{1}{2k}\big\|\mathbf{c}^0 - \mathbf{c}^\star\big\|_{\mathcal{S}}^2,$$

where $\{\tilde{\mathbf{c}}^i\}_{i\in\mathbb{N}}$ and $\mathcal{S}$ are defined in Lemma 5 and 1.

The proof is similar to Theorem 2 or 4.

Likewise, for the class of hybrid algorithms, for given sets $B_1 \subset \mathbb{R}^N$, $B_2 \subset \mathbb{R}^{M_1}$, $B_3 \subset \mathbb{R}^{M_2}$ and $B_4 \subset \mathbb{R}^{M_1}$, the *primal-dual gap* function restricted to $B_1 \times B_2 \times B_3 \times B_4$ is defined as

$$\Psi_{B_1\times B_2\times B_3\times B_4}(\mathbf{c}) = \sup_{(\mathbf{b}',\mathbf{p}')\in B_3\times B_4} \mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b}',\mathbf{p}') - \inf_{(\mathbf{x}',\mathbf{a}')\in B_1\times B_2} \mathcal{L}(\mathbf{x}',\mathbf{a}',\mathbf{b},\mathbf{p}), \tag{36}$$

which has the upper bound:

Corollary 4 *Under the conditions of Theorem 5, if the set $B_1 \times B_2 \times B_3 \times B_4$ is bounded, the primal-dual gap defined as* (37) *has the upper bound:*

$$\Psi_{B_1\times B_2\times B_3\times B_4}\left(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i\right) \le \frac{1}{2k}\sup_{\mathbf{c}\in B_1\times B_2\times B_3\times B_4}\left\|\mathbf{c}^0-\mathbf{c}\right\|_{\mathcal{S}}^2.$$

Furthermore, $\Psi_{B_1\times B_2\times B_3\times B_4}\big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i\big) \ge 0$*, if the set* $B_1 \times B_2 \times B_3 \times B_4$ *contains a saddle point* $\mathbf{c}^\star = (\mathbf{x}^\star, \mathbf{a}^\star, \mathbf{b}^\star, \mathbf{p}^\star) \in \mathsf{zer}\mathcal{A}$*.*

The proof is similar to Corollary 2 or 3.

Remark 8 For `MIX-I, II` and `III` with corresponding $\mathcal{M} = \mathcal{I}$, Theorem 5 and Corollary 4 can be simplified as $\Pi\big(\frac{1}{k}\sum_{i=1}^{k}\mathbf{c}^i, \mathbf{c}^\star\big) \le \frac{1}{2k}\left\|\mathbf{c}^0-\mathbf{c}^\star\right\|_{\mathcal{Q}}^2$ and $\Psi_{B_1\times B_2\times B_3\times B_4}\ \big(\frac{1}{k}\sum_{i=1}^{k}\mathbf{c}^i\big) \le \frac{1}{2k}\sup_{\mathbf{c}\in B_1\times B_2}\left\|\mathbf{c}^0-\mathbf{c}\right\|_{\mathcal{Q}}^2$.

Remark 9 Similarly to Remarks 5 and 7, under additional conditions on f, g and h^*, one can obtain *the convergence rate of* $\mathcal{O}(1/k)$ *of the sequence of* $f(\mathbf{x})+g(\mathbf{a})+h(\mathbf{B}\mathbf{x})$*, evaluated at the ergodic averaging point* $\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i$, namely, it holds that:

$$f\left(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{x}}^i\right)+g\left(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{a}}^i\right)+h\left(\mathbf{B}\Big(\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{x}}^i\Big)\right)-f(\mathbf{x}^\star)-g(\mathbf{a}^\star)-h(\mathbf{B}\mathbf{x}^\star) \le C/k, \tag{37}$$

for some constant C.

To show this, we first rewrite $\mathcal{L}(\mathbf{x},\mathbf{a},\mathbf{b},\mathbf{p})$ in (33) as

$$\mathcal{L}(\mathbf{u},\mathbf{v}) := q(\mathbf{u}) + \langle\mathbf{v}|\mathbf{U}\mathbf{u}\rangle - l^*(\mathbf{v}),$$

where $\mathbf{u} = \begin{bmatrix}\mathbf{x}\\ \mathbf{a}\end{bmatrix}$, $\mathbf{v} = \begin{bmatrix}\mathbf{b}\\ \mathbf{p}\end{bmatrix}$, $\mathbf{U} = \begin{bmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{A} & -\mathbf{I}\end{bmatrix}$, $q : (\mathbf{x},\mathbf{a}) \mapsto f(\mathbf{x}) + g(\mathbf{a})$, $l : (\mathbf{b},\mathbf{p}) \mapsto h(\mathbf{b})$. Since the sequence $\{\tilde{\mathbf{p}}^k\}_{k\in\mathbb{N}}$ converges by Theorem 1, and thus lies in an unknown bounded set $B_4 \subset \mathbb{R}^{M_1}$. Obviously, $\frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{p}}^i \in B_4$, $\mathbf{p}^\star \in B_4$. On the other hand, if $\mathsf{dom}f$, $\mathsf{dom}g$ and $\mathsf{dom}h^*$ are bounded, then we can simply take the sets $B_1 = \mathsf{dom}f$, $B_2 = \mathsf{dom}g$ and $B_3 = \mathsf{dom}h^*$. Denoting the ergodic averaging point by $\widehat{\mathbf{c}}^k = \frac{1}{k}\sum_{i=0}^{k-1}\tilde{\mathbf{c}}^i$ ($\widehat{\mathbf{x}}^k$, $\widehat{\mathbf{a}}^k$, $\widehat{\mathbf{b}}^k$ and $\widehat{\mathbf{p}}^k$ are defined similarly), using Fenchel-Young inequality [4, Proposition 13.15], we develop

$$\begin{aligned}
&\Psi_{B_1\times B_2\times B_3\times B_4}\big(\widehat{\mathbf{c}}^k\big)\\
&= \sup_{\mathbf{v}'\in B_3\times B_4}\mathcal{L}\big(\widehat{\mathbf{u}}^k,\mathbf{v}'\big) - \inf_{\mathbf{u}'\in B_1\times B_2}\mathcal{L}\big(\mathbf{u}',\widehat{\mathbf{v}}^k\big)\\
&= \sup_{\mathbf{v}'\in B_3\times B_4} q\big(\widehat{\mathbf{u}}^k\big) + \big\langle\mathbf{v}'\big|\mathbf{U}\widehat{\mathbf{u}}^k\big\rangle - l^*(\mathbf{v}') - \inf_{\mathbf{u}'\in B_1\times B_2}\big(q(\mathbf{u}') + \big\langle\widehat{\mathbf{v}}^k\big|\mathbf{U}\mathbf{u}'\big\rangle - l^*\big(\widehat{\mathbf{v}}^k\big)\big)\\
&= q\big(\widehat{\mathbf{u}}^k\big) + l\big(\mathbf{U}\widehat{\mathbf{u}}^k\big) + l^*\big(\widehat{\mathbf{v}}^k\big) - \inf_{\mathbf{u}'\in B_1\times B_2}\big(q(\mathbf{u}') + \big\langle\widehat{\mathbf{v}}^k\big|\mathbf{U}\mathbf{u}'\big\rangle\big)\\
&\ge q\big(\widehat{\mathbf{u}}^k\big) + l\big(\mathbf{U}\widehat{\mathbf{u}}^k\big) + l^*\big(\widehat{\mathbf{v}}^k\big) - q(\mathbf{u}^\star) - \big\langle\widehat{\mathbf{v}}^k\big|\mathbf{U}\mathbf{u}^\star\big\rangle\\
&\ge q\big(\widehat{\mathbf{u}}^k\big) + l\big(\mathbf{U}\widehat{\mathbf{u}}^k\big) - q(\mathbf{u}^\star) - l(\mathbf{U}\mathbf{u}^\star)\\
&= q\big(\widehat{\mathbf{u}}^k\big) + h\big(\mathbf{B}\widehat{\mathbf{x}}^k\big) - q(\mathbf{u}^\star) - h(\mathbf{B}\mathbf{x}^\star),
\end{aligned}$$

which, combining with Corollary 4, yields (37). Still, (37) holds for all the hybrid algorithms listed in Table 7. Similarly with Remark 5, the constant C cannot be easily estimated, since the bounded set B_4 is not *a priori* known.

5.3 Reductions of some hybrid algorithms

As reported in Table 8, `MIX-I,II` and `III` correspond to a standard PPA with $\mathcal{M} = \mathcal{I}$, and thus, they can be readily expressed as a resolvent (19), by Corollary 1-(iii).

Observing that neither of $\mathbf{M}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ is allowed to be $\mathbf{0}$ for convergence. It is impossible to reduce any variables (i.e. all the variables are active) in `MIX-I, II` and `III`.

6 Concluding remarks

The numerical performance of these splitting algorithms has been reported in [46, Sect. 7], which is not discussed here.

The proximal point analysis is shown to be able to (i) provide a unified treatment of the generalized Bregman distance and ergodic primal-dual gap; (ii) identify the active variables and reduce the algorithmic dimensionality. The degeneracy reduction in this paper is essentially an application of the degenerate analysis of [8] to the operator splitting algorithms. An important implication of the degeneracy reduction is that it is possible to loosen the strict convergence results (e.g., Theorem 1 and Corollary 1) to positive *semi*-definite metric [45] under a certain conditions, which needs further careful study.

Despite of the success of interpretations using the proximal point analysis demonstrated in [46] and this paper, an evident limitation is that it cannot deal with, for example, the case of [11, Sect. 5], where only one function is assumed to be strongly convex. It may need to exploit the inner structure of $\mathcal{A}$ and $\mathcal{Q}$ based on a subspace analysis, e.g., partially strongly convex operator [43], which will further enrich the degenerate theory poineered in [8].

Acknowledgements I am gratefully indebted to the anonymous reviewers and the editor for helpful discussions, particularly related to the convergence analysis of PPA (Lemma 1), the notion of infimal postcomposition (Lemma 2), the generalized Bregman distance and primal-dual gap (Sect. 3.3, 3.4, 4.2 and 5.2), and for bringing references [34,37,47] to my attention.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Briceño Arias, L., Combettes, P.: A monotone+skew splitting model for composite monotone inclusions in duality. SIAM J. Control Optim. **21**(4), 1230–1250 (2011)
2. Briceño Arias, L., Roldán, F.: Resolvent of the parallel composition and the proximity operator of the infimal postcomposition. Optimization Letters, DOI: 10.1007/s11590-022-01906-5 (2022)
3. Bai, J., Zhang, H., Li, J.: A parameterized proximal point algorithm for separable convex optimization. Optimization Letters **12**, 1589–1608 (2018)
4. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Second Edition, CMS Books in Mathematics, Springer, New York, NY (2017)
5. Beck, A.: First-Order Methods in Optimization. SIAM-Society for Industrial and Applied Mathematics (2017)
6. Boţ, R., Csetnek, E.: On the convergence rate of a forward-backward type primal-dual primal-dual splitting algorithm for convex optimization problems. Optimization **64**(1), 5–23 (2014)
7. Boţ, R.I., Hendrich, C.: Convergence analysis for a primal-dual monotone+skew splitting algorithm with applications to total variation minimization. J. Math. Imaging Vis. **49**, 551–568 (2014)
8. Bredies, K., Chenchene, E., Lorenz, D.A., Naldi, E.: Degenerate preconditioned proximal point algorithms. SIAM Journal on Optimization **32**(3), 2376–2401 (2022)
9. Bredies, K., Sun, H.: A proximal point analysis of the preconditioned alternating direction method of multipliers. J. Optim. Theory Appl. **173**, 878–907 (2017)
10. Cai, J., Osher, S., Shen, Z.: Linearized bregman iterations for compressed sensing. Mathematics of Computation **78**, 1515–1536 (2009)
11. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imag. Vis. **40**(1), 120–145 (2011)
12. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal–dual algorithm. Math. Program., Ser. A **159**(1–2), 253–287 (2016)
13. Chouzenoux, E., Pesquet, J.C., Repetti, A.: A block coordinate variable metric forward-backward algorithm. Journal of Global Optimization **66**, 457–485 (2016)
14. Combettes, P., Pesquet, J.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued Var. Anal. **20**(2), 307–330 (2012)
15. Combettes, P., Pesquet, J.: Fixed point strategies in data science. IEEE Transactions on Signal Processing **69**, 3878–3905 (2021)
16. Combettes, P., Wajs, V.: Signal recovery by proximal forward-backward splitting. Multiscale Modeling and Simulation **4**(4), 1168–1200 (2005)
17. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable, and linear composite terms. J. Optim. Theory Appl. **158**(2), 460–479 (2013)
18. Drori, Y., Sabach, S., Teboulle, M.: A simple algorithm for a class of nonsmooth convex-concave saddle-point problems. Operations Research Letters **43**(2), 209–214 (2015)
19. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming **55**(1), 293–318 (1992)
20. Frankel, P., Garrigos, G., Peypouquet, J.: Splitting methods with variable metric for Kurdyka-Łojasiewicz Functions and General Convergence Rates. Journal of Optimization Theory and Applications **165**(3), 874–900 (2015)
21. Glowinski, R.: Numerical Methods for Nonlinear Variational Problems. Springer, New York (1984)
22. Glowinski, R., Marrocco, A.: Sur l'approximation par éléments finis d'ordure un et la résolution par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. Revue Fr. Autom. Inf. Rech. Opér. Anal. Numér. **2**, 41–76 (1975)
23. Goldstein, T., Osher, S.: The split Bregman method for ℓ_1-regularized problems. SIAM J. Imaging Sciences **2**(2), 323–343 (2009)

24. Gonçalves, M.L.N., Marques, A.M., Melo, J.G.: Pointwise and ergodic convergence rates of a variable metric proximal alternating direction method of multipliers. Journal of Optimization Theory and Applications **177**, 448–478 (2018)
25. He, B., Ma, F., Yuan, X.: An algorithmic framework of generalized primal-dual hybrid gradient methods for saddle point problems. Journal of Mathematical Imaging and Vision **58**(2), 279–293 (2017)
26. He, B., Xu, M., Yuan, X.: Block-wise ADMM with a relaxation factor for multiple-block convex programming. J. Oper. Res. Soc. China **6**, 485–505 (2018)
27. He, B., Yuan, X.: On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J. Numerical Analysis **50**(2), 700–709 (2012)
28. He, B., Yuan, X.: On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. Numerische Mathematik **130**(3), 567–577 (2015)
29. He, B., Yuan, X.: A class of ADMM-based algorithms for three-block separable convex programming. Comput. Optim. Appl. **70**, 791–826 (2018)
30. Kiwiel, K.: Proximal minimization methods with generalized bregman functions. SIAM journal on control and optimization **35**(4), 1142–1168 (1997)
31. Lions, P., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis **16**(6), 964–979 (1979)
32. Ma, F., Ni, M.: A class of customized proximal point algorithms for linearly constrained convex optimization. Comp. Appl. Math. **37**, 896–911 (2018)
33. Martínez-Legaz, R.S.B..J.E.: On bregman-type distances for convex functions and maximally monotone operators. Set-Valued and Variational Analysis **26**, 369–384 (2018)
34. Nemirovski, A.: Prox-method with rate of convergence $\mathcal{O}(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. Optim. **15**(1), 229–251 (2004)
35. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. Multiscale Model. Simul. **4**(2), 460–489 (2005)
36. O'Connor, D., Vandenberghe, L.: Primal-dual decomposition by operator splitting and applications to image deblurring. SIAM J. Imaging Sciences **7**(3), 1724–1754 (2014)
37. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the Mumford-Shah functional. In: IEEE Int. Conf. on Computer Vision, pp. 1133–1140 (2009)
38. Rockafellar, R.T.: Convex analysis. Princeton Landmarks in Mathematics and Physics, Princeton University Press (1996)
39. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Grundlehren der Mathematischen Wissenschaft, vol. 317 (2004)
40. Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. SIAM J. Optim. **24**(1), 269–297 (2014)
41. Tran-Dinh, Q., Fercoq, O., Cevher, V.: A smooth primal-dual optimization framework for nonsmooth composite convex minimization. SIAM J. Optim. **28**(1), 96–134 (2018)
42. Vũ, B.: A splitting algorithm for coupled system of primal-dual monotone inclusions. Journal of Optimization Theory and Applications **164**, 993–1025 (2015)
43. Valkonen, T., Pock, T.: Acceleration of the PDHGM on partially strongly convex functions. Journal of Mathematical Imaging and Vision **59**(3), 394–414 (2017)
44. Xue, F.: On the metric resolvent: nonexpansiveness, convergence rates and applications. arXiv preprint: arXiv:2108.06502 (2021)
45. Xue, F.: On the nonexpansive operators based on arbitrary metric: A degenerate analysis. Results in Mathematics, DOI: 10.1007/s00025-022-01766-6 (2022)
46. Xue, F.: Some extensions of the operator splitting schemes based on Lagrangian and primal-dual: A unified proximal point analysis. Optimization, DOI: 10.1080/02331934.2022.2057309 (2022)
47. Yan, M.: A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. Journal of Scientific Computing **76**, 1698–1717 (2018)
48. Yan, M., Yin, W.: Self Equivalence of the Alternating Direction Method of Multipliers, pp. 165–194. Springer, Cham. (2016)

49. Yin, W., Osher, S., Goldfarb, D., Darbon, J.: Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. SIAM J. Imaging Sciences **1**(1), 143–168 (2008)
50. Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. Journal of Scientific Computing **46**(1), 20–46 (2011)
51. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, UCLA (2008)