Convergence Results for Primal-Dual Algorithms in the Presence of Adjoint Mismatch

Emilie Chouzenoux, Andrés Contreras, Jean-Christophe Pesquet, Marion Savanier

Convergence Results for Primal-Dual Algorithms in the Presence of Adjoint Mismatch

Emilie Chouzenoux, Andrés Contreras, Jean-Christophe Pesquet, and Marion Savanier

Abstract

Most optimization problems arising in imaging science involve high-dimensional linear operators and their adjoints. In the implementations of these operators, approximations may be introduced for various practical considerations (e.g., memory limitation, computational cost, convergence speed), leading to an *adjoint mismatch*. This occurs for the X-ray tomographic inverse problems found in Computed Tomography (CT), where the adjoint of the measurement operator (called projector) is often replaced by a surrogate operator. The resulting adjoint mismatch can jeopardize the convergence properties of iterative schemes used for image recovery. In this paper, we study the theoretical behavior of a panel of primal-dual proximal algorithms, which rely on forward-backward-(forward) splitting schemes, when an adjoint mismatch occurs. We analyze these algorithms by focusing on the resolution of possibly non-smooth convex penalized minimization problems in an infinite-dimensional setting. By using tools from fixed point theory, we show that they can solve monotone inclusions that go beyond minimization problems. Such findings indicate these algorithms can be seen as a generalization of classical primal-dual formulations. The applicability of our findings are also demonstrated through two numerical experiments in the context of CT image reconstruction.

Index Terms

Adjoint mismatch, convex optimization, primal-dual algorithms, convergence analysis, fixed point theory, CT reconstruction, forward-backward splitting, Tseng's splitting.

I. Introduction

Many imaging science applications [2] can be formulated as inverse problems where a high-dimensional variable has to be estimated from noisy indirect observations. Solving an inverse problem can often be recast as the problem of minimizing a cost function expressed as a sum of convex terms composed with linear operators [41]. Especially when these functions are non-smooth, first-order proximal splitting algorithms appear as appealing minimization strategies [19], [11]. In these algorithms, each cost function term is handled separately, either through its gradient or through its proximity operator. At the core of the convergence analysis of most proximal splitting methods is the interpretation of the optimization problem as an instance of a monotone inclusion problem [21]. For example, the forward-backward (also called proximal gradient [23] in the variational case) algorithm [24] solves an inclusion based on a sum of two monotone operators. Owing to its solid convergence guarantees, this scheme is widely used in inverse problems involving sparsity priors [27], [15]. However, when the cost function involves a non-smooth term composed with a linear operator, the proximal step in the forward-backward algorithm may not have a closed form, so it must be computed using an inner loop at the expense of practical efficiency. A way to avoid performing these sub-iterations is to rely on primal-dual proximal splitting algorithms [36], [9], [22], [6], [1], [31]. These algorithmic schemes are grounded on splitting strategies such as the forward-backward, the Douglas-Rachford [5], [33], or Tseng's forward-backward-forward [52] algorithms for finding a zero of a sum of maximally monotone operators. For instance, the Chambolle-Pock algorithm [14] is one of the most popular methods. More flexibility can be added in the considered formulation by taking into account the gradient of Lipschitz differentiable terms in the objective function. The Combettes-Pesquet algorithm [22] based on Tseng's forward-backward-forward splitting was the first algorithm offering this possibility. Then, Condat-Vũ algorithm [55], [25] was proposed, which constitutes an extension of the Chambolle-Pock one. The Loris-Verhoeven algorithm [40] is another powerful algorithm relying on the forward-backward splitting. It offers more freedom in choosing the algorithm parameters, but deals with a more restricted class of optimization problems. Primal-dual algorithms have numerous applications in the resolution of signal/image processing problems [51], [44], [17], [7]. These schemes increase the dimension of the original minimization variables by introducing auxiliary dual variables, yielding the resolution of a system of two decoupled monotone inclusions in the primal-dual product space.

In real-life imaging applications, an *adjoint mismatch* frequently occurs. This term corresponds to the situation when, in the iterative process, the adjoints of the linear operators in the cost function are replaced by surrogate operators. An adjoint mismatch typically occurs in Computed Tomography (CT) image reconstruction. CT involves the resolution of a linear inverse problem, where the forward operator is a discrete Radon transform that relates an image to a set of projections measured at several angles during the rotation of an X-ray imaging system [35]. During the reconstruction, the adjoint of the projector, called

This work was supported by the European Research Council Starting Grant MAJORIS ERC-2019-STG-850925, the ANRT CIFRE Convention 2018/1587, and the ANR Research and Teaching Chair in Artificial Intelligence BRIDGEABLE.

E. Chouzenoux, J.-C. Pesquet, A. Contreras and M. Savanier are with Univ. Paris-Saclay, CentraleSupélec, CVN, Inria, Gif-sur-Yvette, France (e-mail: firstname.lastname@centralesupelec.fr).

M. Savanier is with GE Healthcare, Buc, France (e-mail: firstname.lastname@ge.com).

backprojector, is often approximated to produce an estimate of the image more rapidly (by improving the conditioning [58] or reducing computational complexity per iteration [32]) or to improve the quality of the reconstruction [57]. An adjoint mismatch has also been advocated in SPECT (Single Photon Emission Computed Tomography) imaging [46], [58] and in microscopy imaging [45]. With such mathematical approximations, the convergence guarantees of classical minimization algorithms no longer hold. Hence there is a need for designing and adapting existing results and schemes to these situations.

The effect of an adjoint mismatch has first been studied for CT in the context of simultaneous or row-action algebraic algorithms [35] based on projections onto convex sets (POCS). In that case, the convergence can be analyzed with tools from linear algebra. Some works gave convergence conditions [30], [58], [28], [38] and focused on fixing the convergence of these schemes [28]. In contrast, others proposed to use more general optimization schemes that can directly deal with non-symmetric normal fixed-point equations [42], [50]. The resulting algorithms are semi-convergent; the number of iterations acts as a regularization parameter. The literature on the influence of an adjoint mismatch on more complex proximal splitting schemes has been expanding recently. The convergence of the proximal gradient algorithm (PGA) has been studied in [18] when an adjoint mismatch is present on the forward operator. In the context of microscopy imaging where the forward operator satisfies a specific orthogonality condition, [45] gave conditions for the Douglas-Rachford/ADMM iterations to converge in the Multi-Agent Consensus Equilibrium (MACE) framework [12]. They also highlighted that using an adjoint mismatch in a quadratic data fidelity term was equivalent to using the proximity operator (or agent) of the quadratic term with a different prior model for the image that depends on the mismatched adjoint. A similar observation was made in [49], where the authors showed that considering an adjoint mismatch in PGA could be seen as a problem of unmatched preconditioning where the metric used in the gradient step differs from the one used in the proximity step. Up to our knowledge, the first proposal to analyze a primal-dual proximal splitting method under an adjoint mismatch is [39] which studies two mismatched forms of the Chambolle-Pock algorithm with fixed and varying step sizes. These analyses are conducted under strong convexity assumptions. The authors give conditions on the strong convexity modulus of the involved functions and they derive update rules for the step sizes to recover similar convergence rates to those of the matched schemes. However, they do not investigate the existence and uniqueness of the fixed points of the mismatched methods.

The contribution of this paper is to analyze the behavior of some of the most popular primal-dual proximal splitting algorithms in the presence of adjoint mismatch by relying on fixed-point theory [21]. Our study leverages some of our previous results in [18]. Our analysis applies to a broad class of penalized least-squares problems. More specifically, we provide sufficient conditions for the existence and uniqueness of fixed point of the modified schemes, and we also provide conditions to ensure the convergence of the schemes to such a fixed point. Our methodology allows us to develop extensions of existing algorithms that are robust to adjoint mismatches. Their resulting fixed points satisfy general notions of equilibrium (similar to the one arising in neural network architectures/paradigms, plug-and-play methods [21], [43], [34]) instead of being the solution to a minimization problem. In this aspect, our work is in line with the recent efforts to design [49], [48] and learn [3], [10], [43], [12] more expressive variants of well-known optimization schemes while keeping convergence guarantees.

The paper is organized as follows. In section II, we present our notation and useful definitions underlying the class of minimization and monotone inclusion problems under the scope of our study. We also present the penalized least-squares cost function we aim to minimize, which involves two linear operators. The first linear operator appears in a quadratic term, while the second arises in a non-necessarily smooth convex term. Sections III, IV, and V gather our methodological contributions, namely the study of the properties of three main classes of primal-dual splitting algorithms in the presence of adjoint mismatch. Specifically, section III focuses on the Condat-Vũ algorithm [55], [25], and its mismatched version. We consider the case when an adjoint mismatch appears on the linear operator in the quadratic term and the second linear operator is perturbed. Note that this differs from the mismatch considered in [39], which appears on the second linear operator. This analysis is then extended to the projected form of the Condat-Vũ algorithm proposed by Briceño-Arias and López [8]. In section IV, we perform an analysis of the Loris-Verhoeven algorithm [40] in the case of an adjoint mismatch on the linear operator involved in the quadratic term. In section V, we analyze the Combettes-Pesquet algorithm [22] when the adjoints of both linear operators are approximated. We derive convergence results for all the above algorithms and characterize the resulting fixed points. Finally, in section VI, we illustrate our theoretical findings on examples involving two different inverse problems in CT reconstruction, with two types of regularization and noise modeling.

II. Preliminaries

A. Notation and background

Let us introduce some basic notation and definitions on convex analysis, proximity operators, and monotone operators. Most of our notation follow from the reference book [4].

We consider optimization problems over real Hilbert spaces. We denote by $\mathcal{H}$, $\mathcal{G}$, $\mathcal{L}$ some real Hilbert spaces, and $\mathcal{B}(\mathcal{H},\mathcal{G})$ the set of bounded and linear operators from $\mathcal{H}$ to $\mathcal{G}$. The norm of $\mathcal{H}$ is denoted by $\|\cdot\|_{\mathcal{H}}$ and the associated inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$. The operator norm of $D\in\mathcal{B}(\mathcal{H},\mathcal{G})$ is denoted by $\|D\|_{\mathcal{H},\mathcal{L}}=\sup\{\|Dx\|_{\mathcal{L}} \mid x\in\mathcal{H}, \|x\|_{\mathcal{H}}=1\}$. Given $C\subset\mathcal{H}$, $\operatorname{sri} C$ denotes its strong relative interior.

$2^{\mathcal{H}}$ denotes the power set of $\mathcal{H}$. Given a set-valued operator $\mathcal{M}: \mathcal{H} \to 2^{\mathcal{H}}$, we define $\operatorname{dom}\mathcal{M} = \{x \in \mathcal{H} \mid \mathcal{M}x \neq \varnothing\}$ as its domain, and $\operatorname{ran}\mathcal{M}$ and $\operatorname{gra}\mathcal{M}$ as its range and graph, respectively.
$\mathcal{M}^{-1}: \mathcal{H} \to 2^{\mathcal{H}}$ denotes the inverse operator of $\mathcal{M}$, with domain $\operatorname{ran}\mathcal{M}$ and range $\operatorname{dom}\mathcal{M}$, and $\mathcal{M}^{-1}(0) = \{x \in \mathcal{H} \mid 0 \in \mathcal{M}x\}$ the set of zeros of $\mathcal{M}$.
$\mathcal{M}$ is said to be monotone when, for every $(x,y) \in \operatorname{gra}\mathcal{M}$ and $(x',y') \in \operatorname{gra}\mathcal{M}$, $\langle y'-y, x'-x\rangle_{\mathcal{H}} \geq 0$. $\mathcal{M}$ is said to be maximal when $\operatorname{gra}\mathcal{M}$ is maximal in the sense of the inclusion in $\mathcal{H}\times\mathcal{H}$ among the graphs of monotone operators. Moreover, $\mathcal{M}$ is strongly monotone with parameter $\rho \geq 0$ if for every $(x,y) \in \operatorname{gra}\mathcal{M}$ and $(x',y') \in \operatorname{gra}\mathcal{M}$, $\langle y'-y, x'-x\rangle_{\mathcal{H}} \geq \rho\|x'-x\|_{\mathcal{H}}^2$.
The resolvent of a maximally monotone operator $\mathcal{M}$ scaled with a parameter $\gamma > 0$ is the mapping $J_{\gamma\mathcal{M}}: \mathcal{H} \to \mathcal{H}: x \mapsto J_{\gamma\mathcal{M}}(x) = (\mathrm{Id}_{\mathcal{H}} + \gamma\mathcal{M})^{-1}x$, where $\mathrm{Id}_{\mathcal{H}}$ refers to the identity operator in $\mathcal{H}$.
An operator $\mathscr{C}: \mathcal{H} \to \mathcal{H}$ is cocoercive if there exists $\Theta > 0$ such that for every $(x,x') \in \mathcal{H}^2$, $\langle \mathscr{C}x' - \mathscr{C}x, x'-x\rangle_{\mathcal{H}} \geq \Theta\|\mathscr{C}x' - \mathscr{C}x\|_{\mathcal{H}}^2$.
The Fenchel-Legendre conjugate of a function $f: \mathcal{H} \to]-\infty,+\infty]$, is expressed as $f^*: \mathcal{H} \to [-\infty,+\infty]: u \mapsto \sup_{x\in\mathcal{H}} (\langle x,u\rangle_{\mathcal{H}} - f(x))$.
The domain of f is $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$ and this function is said to be proper when $\operatorname{dom} f \neq \varnothing$.
The class of convex, lower semicontinuous, and proper function defined on $\mathcal{H}$ is denoted by $\Gamma_0(\mathcal{H})$ and the Moreau subdifferential of any function f in $\Gamma_0(\mathcal{H})$ is denoted by ∂f, while the proximity operator of f is given by $\operatorname{prox}_f: \mathcal{H} \to \mathcal{H}: x \mapsto \operatorname{argmin}_{v\in\mathcal{H}} \left(f(v) + \frac{1}{2}\|v-x\|^2 \right)$.
From [4, Corollary 16.30] we have $(\partial f)^{-1} = \partial f^*$ and, if f is strongly convex with modulus $\sigma \geq 0$, then ∂f is σ-strongly monotone [4, Example 22.4 (iv)].
Finally, given $\alpha \in]0,1[$, an operator $T: \mathcal{H} \to \mathcal{H}$ is α-averaged if, for every $(x,y) \in \mathcal{H}^2$, there exists a nonexpansive operator $Q: \mathcal{H} \to \mathcal{H}$ such that $T = (1-\alpha)\mathrm{Id}_{\mathcal{H}} + \alpha Q$.

B. Problem statement

Let us now state the generic minimization problem considered in this paper. Let $H \in \mathcal{B}(\mathcal{H},\mathcal{G})$, $D \in \mathcal{B}(\mathcal{H},\mathcal{L})$, $f \in \Gamma_0(\mathcal{H})$, and $g \in \Gamma_0(\mathcal{L})$. Given $y \in \mathcal{G}$, we are interested in solving the primal optimization problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad \frac{1}{2}\|y - Hx\|_{\mathcal{G}}^2 + f(x) + g(Dx). \tag{II.1}$$

By using first-order optimality conditions [4, Theorems 16.47, 27.1], Problem (II.1) can be reformulated as finding a solution to the variational inclusion problem

$$0 \in \partial f(x) + D^*\partial g(Dx) + H^*(Hx - y). \tag{II.2}$$

A solution to (II.2) is a solution to (II.1), but the converse may not be true. From now on, the solution set of (II.2) is supposed to be nonempty, and then so is the solution set of (II.1). Under mild qualification constraints, the solutions to (II.1) and (II.2) are the same. One of such qualification conditions is

$$0 \in \operatorname{sri}\left(D(\operatorname{dom} f) - \operatorname{dom} g\right). \tag{II.3}$$

We refer the reader to [4, Proposition 27.5, Corollary 27.6] for other examples of qualification constraints.
The dual problem associated with (II.1) reads

$$\underset{u\in\mathcal{G}}{\text{minimize}} \quad (f+h)^*(-D^*u) + g^*(u), \tag{II.4}$$

where $h: \mathcal{H} \to \mathbb{R}: x \mapsto \frac{1}{2}\|y - Hx\|_{\mathcal{G}}^2$.
On the one hand, x is a solution to (II.1) and u is a solution to the dual problem associated with (II.1) if and only if $z = (x,u) \in \mathcal{Z} = \mathcal{H}\times\mathcal{L}$ is a solution to

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial f(x) + D^*u \\ -Dx + \partial g^*(u) \end{pmatrix}}_{\mathcal{A}z} + \underbrace{\begin{pmatrix} H^*Hx - H^*y \\ 0 \end{pmatrix}}_{Bz}. \tag{II.5}$$

The operator $\mathcal{A}: \mathcal{Z} \to 2^{\mathcal{Z}}: (x,u) \mapsto (\partial f(x) + D^*u, -Dx + \partial g^*(u))$ is maximally monotone [4, Proposition 26.32 (iii)]. Moreover, the operator $B: \mathcal{Z} \to \mathcal{Z}: (x,u) \mapsto (H^*Hx - H^*y, 0)$ is θ-cocoercive, with $\theta = 1/\|H\|_{\mathcal{H},\mathcal{G}}^2$. Based on this structure, the forward-backward splitting [24] can be applied to solve (II.5). Combined with preconditioning [26], it yields a simple form of iteration in terms of primal/dual updates.
On the other hand, (II.5) is also equivalent to the following splitting

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial f(x) \\ \partial g^*(u) \end{pmatrix}}_{\mathscr{M}z} + \underbrace{\begin{pmatrix} H^*Hx - H^*y + D^*u \\ -Dx \end{pmatrix}}_{Qz}, \tag{II.6}$$

where $\mathscr{M}$ is a maximally monotone operator (see [4, Theorem 26.32 (iii)]) and Q is a monotone operator which is Lipschitz continuous with constant $\vartheta \leq \|H\|_{\mathcal{H}}^2 + \|D\|_{\mathcal{H},\mathcal{L}}$ (see [22]). We can use Tseng's approach [4, Theorem 26.17], [9, Theorem 2.5 (ii)] to build primal-dual schemes delivering a solution to splitting (II.6).
Both (II.5) and (II.6) splitting strategies will be at the basis of the primal-dual algorithms under study. The Condat-Vũ (CV) algorithm (section III) and the Loris-Verhoeven (LV) algorithm (section IV) will be derived from (II.5) while the Combettes-Pesquet (CP) algorithm (section V) will rely on (II.6).

III. The mismatched Condat-Vũ algorithm

A. CV algorithm

In this section, we use the forward-backward splitting to solve (II.6). The resulting iteration by the implicit and preconditioned discretization of (II.5) reads: for every $n \in \mathbb{N}$,

$$0 \in \mathcal{A} z_{n+\frac{1}{2}} + B z_n + P(z_{n+\frac{1}{2}} - z_n), \tag{III.1}$$

where $z_n = (x_n, u_n) \in \mathcal{Z}$, and $z_{n+\frac{1}{2}} = (x_{n+\frac{1}{2}}, u_{n+\frac{1}{2}}) \in \mathcal{Z}$. The preconditioning metric P is defined as

$$(\forall z = (x,u) \in \mathcal{Z}) \quad Pz = \begin{pmatrix} \frac{1}{\tau} x - D^* u \\ -Dx + \frac{1}{\sigma} u \end{pmatrix}, \tag{III.2}$$

where τ and σ are two positive real parameters and initialization $x_0 \in \mathcal{H}$ and $u_0 \in \mathcal{L}$.
Let $(\Theta_n)_{n\in\mathbb{N}}$ be a sequence of relaxation parameters. The corresponding primal-dual forward-backward iteration [4], often called the Condat-Vũ (CV) iteration in the literature [55], [25], reads

CV iterations for solving (II.5):

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} x_{n+\frac{1}{2}} = \operatorname{prox}_{\tau f}\left(x_n - \tau(H^*(Hx_n - y) + D^* u_n)\right) \\ u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\left(u_n + \sigma D(2x_{n+\frac{1}{2}} - x_n)\right) \\ x_{n+1} = x_n + \Theta_n(x_{n+\frac{1}{2}} - x_n) \\ u_{n+1} = u_n + \Theta_n(u_{n+\frac{1}{2}} - u_n). \end{array} \right. \tag{III.3}$$

The convergence of (III.3) to a solution pair (x, u) to (II.5) is guaranteed by [25, Theorem 3.1] for step sizes $\tau > 0$, $\sigma > 0$ such that $\tau\left(\sigma\|D\|_{\mathcal{H},\mathcal{L}}^2 + \dfrac{1}{2\theta}\right) < 1$, and relaxation parameters $(\Theta_n)_{n\in\mathbb{N}} \subset [0, \delta]$, with $\delta = 2 - \dfrac{1}{2\theta}\left(\dfrac{1}{\tau} - \sigma\|D\|_{\mathcal{H},\mathcal{L}}^2\right)$ and θ the cocoercivity constant of operator B such that $\displaystyle\sum_{n\in\mathbb{N}} \Theta_n\,(\delta - \Theta_n) = +\infty$.

Remark 3.1: The condition on the step sizes (τ, σ) implies that $\tau\sigma\|D\|_{\mathcal{H},\mathcal{L}}^2 < 1$, which allows us to conclude that operator P present in (III.1) is strongly positive (bounded from below) and therefore invertible. Hereafter, we will consider the Hilbert space $\mathcal{Z}_P$ obtained by equipping $\mathcal{Z}$ with the inner product $\langle\cdot,\cdot\rangle_P \colon (z, z') \mapsto \langle z, z'\rangle_P = \langle z, Pz'\rangle_{\mathcal{Z}}$. Hence, operators $P^{-1}\mathcal{A}$ and $P^{-1}B$ are maximally monotone [4, Proposition 20.24] and cocoercive [25, Theorem 3.1] in $\mathcal{Z}_P$, respectively.

B. Convergence properties of the mismatched CV algorithm

In this section, we will be interested in studying the impact of replacing operator H^* in the n-th iteration of Algorithm (III.3) by a surrogate $K_n \in \mathcal{B}(\mathcal{G}, \mathcal{H})$. This yields the following modified algorithm:

Mismatched CV iterations:

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} x_{n+\frac{1}{2}} = \operatorname{prox}_{\tau f}\left(x_n - \tau(K_n(Hx_n - y) + D^* u_n)\right) \\ u_{n+\frac{1}{2}} = \operatorname{prox}_{\sigma g^*}\left(u_n + \sigma D(2x_{n+\frac{1}{2}} - x_n)\right) \\ x_{n+1} = x_n + \Theta_n(x_{n+\frac{1}{2}} - x_n) \\ u_{n+1} = u_n + \Theta_n(u_{n+\frac{1}{2}} - u_n), \end{array} \right. \tag{III.4}$$

The sequence of surrogates $(K_n)_{n\in\mathbb{N}}$ is related to a constant linear operator $\overline{K}$ through the following assumption, which is similar to [18, Assumption 3.1],

Assumption 3.2: There exists $\overline{K} \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ and $\{\omega_n\}_{n\in\mathbb{N}} \subset [0, +\infty[$ such that $\sum_{n\in\mathbb{N}} \omega_n < +\infty$ and

$$\overline{K} H \neq 0 \tag{III.5}$$

$$(\forall n \in \mathbb{N}) \quad \|K_n - \overline{K}\|_{\mathcal{G},\mathcal{H}} \leq \omega_n. \tag{III.6}$$

To perform the convergence analysis of (III.4), we introduce notation involved in the characterization of the spectra of the linear operators in the cost function.

Notation 3.3: For every $n \in \mathbb{N}$, we define the following quantities:

(i) $L = \overline{K}H$
(ii) $\lambda_{\min} = \inf\{\langle x, Lx\rangle_{\mathcal{H}} \mid x \in \mathcal{H}, \|x\|_{\mathcal{H}} = 1\}$
(iii) $\widetilde{L} = \Pi L$, $\widetilde{L}_n = \Pi K_n H$, $\widetilde{K} = \Pi K$, and $\widetilde{K}_n = \Pi K_n$, with $\Pi\colon \mathcal{Z} \to \mathcal{Z}\colon (x,u) \mapsto (x,0)$.
(iv) $T_n\colon \mathcal{Z} \to \mathcal{Z}\colon z \mapsto J_{P^{-1}\mathcal{A}}\left(z - P^{-1}\left(\widetilde{L}_n z - \widetilde{K}_n y\right)\right)$ and
$\overline{T}\colon \mathcal{Z} \to \mathcal{Z}\colon z \mapsto J_{P^{-1}\mathcal{A}}\left(z - P^{-1}\left(\widetilde{L} z - \widetilde{K} y\right)\right)$.

Under this notation, Algorithm (III.4) can be rewritten more concisely in terms of a simple update rule on the pair $z_n = (x_n, u_n)$:

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = z_n + \Theta_n\left(T_n\left(z_n\right) - z_n\right). \tag{III.7}$$

Algorithm (III.7) can be viewed as a mismatched form of a forward-backward algorithm for finding a zero of a sum of maximally monotone operators in $\mathcal{Z}_P$. Similarly to [18], we rely on the cocoercivity properties of operator $\widetilde{L}$ to study the convergence of (III.7).

Proposition 3.4: Assume that $(\tau,\sigma) \in]0,+\infty[^2$ are such that $\tau\sigma\|D\|^2_{\mathcal{H},\mathcal{L}} < 1$. We have the following properties.

(i) $P^{-1}\widetilde{L}$ is cocoercive in $\mathcal{Z}_P$ if and only if L is cocoercive.
(ii) Suppose that $\mathrm{Ran}(L + L^*)$ is closed. Then $P^{-1}\widetilde{L}$ is cocoercive in $\mathcal{Z}_P$ with constant $\widetilde{\eta} > 0$ if and only if $\lambda_{\min} \geq 0$, $\mathrm{Ker}(L + L^*) = \mathrm{Ker}\, L$, and

$$\widetilde{\eta} \leq \widetilde{\eta}_{\max} = \frac{2}{\|P^{-1/2}\Pi M\|^2_{\mathcal{H},\mathcal{Z}}}, \tag{III.8}$$

where

$$M = (\mathrm{Id}_{\mathcal{H}} + (L - L^*)(L + L^*)^{\#})(L + L^*)^{1/2} \tag{III.9}$$

and $(L + L^*)^{\#}$ denotes the pseudo-inverse of $L + L^*$. In addition,

$$\widetilde{\eta}_{\max} \geq \tau^{-1}(1 - \tau\sigma\|D\|^2_{\mathcal{H},\mathcal{L}})\eta_{\max} \tag{III.10}$$

where $\eta_{\max} = 2/\|M\|^2_{\mathcal{H},\mathcal{H}}$ is the largest cocoercivity constant of L.

Proof:

(i) $P^{-1}\widetilde{L}$ is cocoercive in $\mathcal{Z}_P$ if and only if there exists $\widetilde{\eta} > 0$ such that, for every $z \in \mathcal{Z}$,

$$\begin{aligned} & (\forall z \in \mathcal{Z}) \quad \langle z, P^{-1}\widetilde{L}z\rangle_{\mathcal{Z}_P} \geq \widetilde{\eta}\|P^{-1}\widetilde{L}z\|^2_{\mathcal{Z}_P} \\ \Leftrightarrow \quad & (\forall z' \in \mathcal{Z}) \quad \langle z', P^{-1/2}\widetilde{L}P^{-1/2}z'\rangle_{\mathcal{Z}} \geq \widetilde{\eta}\|P^{-1/2}\widetilde{L}P^{-1/2}z'\|^2_{\mathcal{Z}}. \end{aligned} \tag{III.11}$$

Therefore $P^{-1}\widetilde{L}$ is cocoercive in $\mathcal{Z}_P$ if and only if $P^{-1/2}\widetilde{L}P^{-1/2}$ is cocoercive in $\mathcal{Z}$. In turn, it follows from [4, Proposition 4.12] that, if $\widetilde{L}$ is cocoercive, then $P^{-1/2}\widetilde{L}P^{-1/2}$ is cocoercive. Conversely, if $P^{-1/2}\widetilde{L}P^{-1/2}$ is cocoercive, then $P^{1/2}(P^{-1/2}\widetilde{L}P^{-1/2})P^{1/2} = \widetilde{L}$ is cocoercive, that is

$$\begin{aligned} & (\forall z \in \mathcal{Z}) \quad \langle z, \widetilde{L}z\rangle_{\mathcal{Z}} \geq \eta\|\widetilde{L}z\|^2_{\mathcal{Z}} \\ \Leftrightarrow \quad & (\forall x \in \mathcal{H}) \quad \langle x, Lx\rangle_{\mathcal{H}} \geq \eta\|Lx\|^2_{\mathcal{H}}, \end{aligned} \tag{III.12}$$

for some $\eta > 0$.

(ii) Let $\widetilde{L}^*_P$ denote the adjoint of $P^{-1}\widetilde{L}$ in $\mathcal{Z}_P$ and let

$$\widetilde{\lambda}_{\min} = \inf\left\{\langle z, P^{-1}\widetilde{L}z\rangle_{\mathcal{Z}_P} \mid z \in \mathcal{Z}_P, \|z\|_{\mathcal{Z}_P} = 1\right\}. \tag{III.13}$$

According to [18, Proposition 3.4(ii)], provided that $\mathrm{Ran}(P^{-1}\widetilde{L} + \widetilde{L}^*_P)$ is closed, $P^{-1}\widetilde{L}$ is cocoercive in $\mathcal{Z}_P$ with constant $\widetilde{\eta}$ if and only if $\widetilde{\lambda}_{\min} \geq 0$, $\mathrm{Ker}(P^{-1}\widetilde{L} + \widetilde{L}^*_P) = \mathrm{Ker}(P^{-1}\widetilde{L})$, and

$$\widetilde{\eta} \leq \frac{2}{\|(\mathrm{Id}_{\mathcal{Z}} + (P^{-1}\widetilde{L} - \widetilde{L}^*_P)(P^{-1}\widetilde{L} + \widetilde{L}^*_P)^{\#})(P^{-1}\widetilde{L} + \widetilde{L}^*_P)^{1/2}\|^2_{\mathcal{Z}_P,\mathcal{Z}_P}}. \tag{III.14}$$

Since $\widetilde{L}^*_P = P^{-1}\widetilde{L}^*$, $\mathrm{Ker}(P^{-1}\widetilde{L} + \widetilde{L}^*_P) = \mathrm{Ker}(\widetilde{L} + \widetilde{L}^*) = \mathrm{Ker}(L + L^*) \times \mathcal{L}$ and, similarly, $\mathrm{Ker}(P^{-1}\widetilde{L}) = \mathrm{Ker}\, L \times \mathcal{L}$. Therefore $\mathrm{Ker}(P^{-1}\widetilde{L} + \widetilde{L}_P) = \mathrm{Ker}(P^{-1}\widetilde{L})$ if and only if $\mathrm{Ker}(L + L^*) = \mathrm{Ker}\, L$. Similarly, $\mathrm{Ran}(P^{-1}\widetilde{L} + \widetilde{L}^*_P) = P^{-1}\,\mathrm{Ran}(\widetilde{L} + \widetilde{L}^*) = P^{-1}(\mathrm{Ran}(L + L^*) \times \{0\})$. Thus, $\mathrm{Ran}(P^{-1}\widetilde{L} + \widetilde{L}^*_P)$ is closed if and only if $\mathrm{Ran}(L + L^*)$ is closed.

For every $z = (x,u) \in \mathcal{Z}$,

$$\begin{aligned} & \langle z, P^{-1}\widetilde{L}z\rangle_{\mathcal{Z}_P} \geq \widetilde{\lambda}_{\min}\|z\|^2_{\mathcal{Z}_P} \\ \Leftrightarrow \quad & \langle z, \widetilde{L}z\rangle_{\mathcal{Z}} \geq \widetilde{\lambda}_{\min}\|z\|^2_{\mathcal{Z}_P} \\ \Leftrightarrow \quad & \langle x, Lx\rangle_{\mathcal{H}} \geq \widetilde{\lambda}_{\min}\|(x,u)\|^2_{\mathcal{Z}_P}. \end{aligned} \tag{III.15}$$

So, $\widetilde{\lambda}_{\min} \geq 0$ if and only if, for every $x \in \mathcal{H}$, $\langle x, Lx\rangle_{\mathcal{H}} \geq 0$, that is $\lambda_{\min} \geq 0$. In addition, when condition $\lambda_{\min} = 0$ is met, $\widetilde{\lambda}_{\min} = 0$.

Since $\|\cdot\|_{\mathcal{Z}_P,\mathcal{Z}_P} = \|P^{1/2} \cdot P^{-1/2}\|_{\mathcal{Z},\mathcal{Z}}$, we have

$$\begin{aligned}
&\|(\mathrm{Id}_{\mathcal{Z}} + (P^{-1}\widetilde{L} - \widetilde{L}_P^*)(P^{-1}\widetilde{L} + \widetilde{L}_P^*)^{\#})(P^{-1}\widetilde{L} + \widetilde{L}_P^*)^{1/2}\|^2_{\mathcal{Z}_P,\mathcal{Z}_P} \\
&= \|(\mathrm{Id}_{\mathcal{Z}} + (P^{-1}\widetilde{L} - \widetilde{L}_P^*)(P^{-1}\widetilde{L} + \widetilde{L}_P^*)^{\#})(P^{-1}\widetilde{L} + \widetilde{L}_P^*)(\mathrm{Id}_{\mathcal{Z}} + (P^{-1}\widetilde{L} + \widetilde{L}_P^*)^{\#}(\widetilde{L}_P^* - P^{-1}\widetilde{L})\|_{\mathcal{Z}_P,\mathcal{Z}_P} \\
&= \|P^{1/2}(\mathrm{Id}_{\mathcal{Z}} + P^{-1}(\widetilde{L} - \widetilde{L}^*)(\widetilde{L} + \widetilde{L}^*)^{\#}P)P^{-1}(\widetilde{L} + \widetilde{L}^*)(\mathrm{Id}_{\mathcal{Z}} + (\widetilde{L} + \widetilde{L}^*)^{\#}PP^{-1}(\widetilde{L}^* - \widetilde{L}))P^{-1/2}\|_{\mathcal{Z},\mathcal{Z}} \\
&= \|P^{-1/2}(\mathrm{Id}_{\mathcal{Z}} + (\widetilde{L} - \widetilde{L}^*)(\widetilde{L} + \widetilde{L}^*)^{\#})(\widetilde{L} + \widetilde{L}^*)(\mathrm{Id}_{\mathcal{Z}} + (\widetilde{L} + \widetilde{L}^*)^{\#}(\widetilde{L}^* - \widetilde{L}))P^{-1/2}\|_{\mathcal{Z},\mathcal{Z}} \\
&= \|P^{-1/2}(\mathrm{Id}_{\mathcal{Z}} + (\widetilde{L} - \widetilde{L}^*)(\widetilde{L} + \widetilde{L}^*)^{\#})(\widetilde{L} + \widetilde{L}^*)^{1/2}\|^2_{\mathcal{Z},\mathcal{Z}}.
\end{aligned}$$

By using the specific form of $\widetilde{L}$, we deduce that

$$\|(\mathrm{Id}_{\mathcal{Z}} + (P^{-1}\widetilde{L} - \widetilde{L}_P^*)(P^{-1}\widetilde{L} + \widetilde{L}_P^*)^{\#})(P^{-1}\widetilde{L} + \widetilde{L}_P^*)^{1/2}\|^2_{\mathcal{Z}_P,\mathcal{Z}_P} = \|P^{-1/2}\Pi M\|^2_{\mathcal{H},\mathcal{Z}}. \tag{III.16}$$

Altogether with (III.14), this yields (III.8).

In addition,

$$\|P^{-1/2}\Pi M\|_{\mathcal{H},\mathcal{Z}} \leq \|P^{-1/2}\Pi\|_{\mathcal{Z},\mathcal{Z}}\|M\|_{\mathcal{H},\mathcal{H}} = \|P^{-1/2}\Pi\|_{\mathcal{Z},\mathcal{Z}}\sqrt{\frac{2}{\eta_{\max}}}, \tag{III.17}$$

where

$$\begin{aligned}
\|P^{-1/2}\Pi\|_{\mathcal{Z},\mathcal{Z}} = \|\Pi^* P^{-1/2}\|_{\mathcal{Z},\mathcal{Z}} &= \|\Pi P^{-1/2}\|_{\mathcal{Z},\mathcal{Z}} \\
&= \sup_{z' \in \mathcal{Z}\setminus\{0\}} \frac{\|\Pi P^{-1/2} z'\|_{\mathcal{Z}}}{\|z'\|_{\mathcal{Z}}} \\
&= \sup_{z=(x,u)\in\mathcal{Z}\setminus\{0\}} \frac{\|x\|_{\mathcal{H}}}{\sqrt{\langle z, Pz\rangle_{\mathcal{Z}}}}.
\end{aligned} \tag{III.18}$$

For every $z = (x, u) \in \mathcal{Z}$,

$$\begin{aligned}
\langle z, Pz\rangle_{\mathcal{Z}} &= \frac{1}{\tau}\|x\|^2_{\mathcal{H}} - 2\langle Dx, u\rangle_{\mathcal{L}} + \frac{1}{\sigma}\|u\|^2_{\mathcal{L}} \\
&= \frac{1}{\tau}\|x\|^2_{\mathcal{H}} + \frac{1}{\sigma}\|u - \sigma Dx\|^2_{\mathcal{L}} - \sigma\|Dx\|^2_{\mathcal{L}} \\
&\geq \frac{1}{\tau}(1 - \tau\sigma\|D\|^2_{\mathcal{H},\mathcal{L}})\|x\|^2_{\mathcal{H}}.
\end{aligned} \tag{III.19}$$

We deduce from (III.18) and (III.19) that

$$\|P^{-1/2}\Pi\|^2_{\mathcal{Z},\mathcal{Z}} \leq \frac{\tau}{1 - \tau\sigma\|D\|^2_{\mathcal{H},\mathcal{L}}}, \tag{III.20}$$

which, combined with (III.17), yields (III.10).

□

The following results provides a characterization of the fixed point set of $\overline{T}$.

Proposition 3.5: Let $(\widetilde{x}, \widetilde{u}) \in \mathcal{Z}$. Then $(\widetilde{x}, \widetilde{u}) \in \mathrm{Fix}\overline{T}$ if and only if $(\widetilde{x}, \widetilde{u})$ belongs to

$$\overline{\mathcal{F}} = \left\{(x, u) \in \mathcal{Z} \mid \overline{K}y \in \partial f(x) + Lx + D^*u,\ u \in \partial g(Dx)\right\}, \tag{III.21}$$

which is nonempty if $L + \partial f + D^* \circ \partial g \circ D$ is surjective.

(i) If $\lambda_{\min} \geq 0$, then $\overline{\mathcal{F}}$ is closed and convex.

(ii) Let $\overline{\mathcal{F}}_1 = \{x \in \mathcal{H} \mid (\exists u \in \mathcal{L})\ (x, u) \in \overline{\mathcal{F}}\}$.

$\overline{\mathcal{F}}_1$ has at most one element if one of the following conditions holds:

(a) $L + \partial f + D^* \circ \partial g \circ D$ is strictly monotone.

(b) $\lambda_{\min} \geq 0$ and $g \circ D + f$ is strictly convex.

$\overline{\mathcal{F}}_1$ is a singleton if (II.3) is satisfied and one of the following conditions holds:

(c) $\lambda_{\min} \geq 0$ and $L + \partial f + D^* \circ \partial g \circ D$ is strongly monotone.

(d) $\lambda_{\min} > 0$.

(e) $\lambda_{\min} \geq 0$, and f is strongly convex or [g is strongly convex and D^*D is strongly positive].

(f) L is cocoercive, and f is strongly convex or [g is strongly convex and D^*D is strongly positive].

Proof:

$$\widetilde{z} = (\widetilde{x}, \widetilde{u}) \in \mathrm{Fix}\overline{T} \Leftrightarrow \widetilde{z} = \overline{T}\widetilde{z} \Leftrightarrow 0 \in P^{-1}\left(\widetilde{L}\widetilde{z} - \widetilde{K}y\right) + P^{-1}\mathcal{A}\widetilde{z} \Leftrightarrow \overline{K}y \in \partial f(\widetilde{x}) + L\widetilde{x} + D^*\widetilde{u},\ D\widetilde{x} \in \partial g^*(\widetilde{u}),$$

that is $\widetilde{z} \in \overline{\mathcal{F}}$. In addition,

$$(\exists \widetilde{u} \in \mathcal{L}) \quad \begin{cases} \overline{K}y \in \partial f(\widetilde{x}) + L\widetilde{x} + D^*\widetilde{u} \\ D\widetilde{x} \in \partial g^*(\widetilde{u}) \end{cases} \quad \Leftrightarrow \quad \overline{K}y \in \partial f(\widetilde{x}) + L\widetilde{x} + D^*\partial g(D\widetilde{x}), \tag{III.22}$$

and the latter condition is satisfied if $\partial f + L + D^* \circ \partial g \circ D$ is surjective.

(i) If $\lambda_{\min} \geq 0$, then, for every $z = (x, u) \in \mathcal{Z}$, $\langle \widetilde{L}z, z\rangle_{\mathcal{Z}} = \langle Lx, x\rangle_{\mathcal{H}} \geq 0$ and, since $\widetilde{L}$ is continuous, $\widetilde{L}$ is maximally monotone on $\mathcal{Z}$. Operator $\mathcal{A}$ being maximally monotone, by [4, Proposition 23.39], we conclude that $\overline{\mathcal{F}} = \operatorname{zer}(\widetilde{L} - \widetilde{K}y + \mathcal{A})$ is closed and convex.

(ii) According to (III.22), $\overline{\mathcal{F}}_1 = \operatorname{zer}(L - \overline{K}y + \partial f + D^* \circ \partial g \circ D)$.

(a) Follows from [4, Proposition 23.35].

(b) From standard subdifferential calculus rules, $\overline{\mathcal{F}}_1 \subset \operatorname{zer}(L - Ky + \partial(f + g \circ D))$. $\lambda_{\min} \geq 0 \Leftrightarrow L + L^*$ positive $\Leftrightarrow L$ is monotone. In addition, $f + g \circ D$ is strictly convex if and only if $\partial(f + g \circ D)$ is strictly monotone. Thus, under the stated condition, $L - Ky + \partial(f + g \circ D)$ is strictly monotone, and $\overline{\mathcal{F}}_1$ has at most one element.

(c) Since ∂f and $D^* \circ \partial g \circ D$ are maximally monotone and (II.3) holds, $\partial f + D^* \circ \partial g \circ D$ is maximally monotone [4, Corollary 25.4]. Since L is maximally monotone and has a full domain, $L + \partial f + D^* \circ \partial g \circ D$ is thus maximally monotone. Then we deduce the result from [4, Proposition 23.37].

(d) If $\lambda_{\min} > 0$, L is strongly monotone and (c) is satisfied.

(e) If g is ρ_{G}-strongly convex with $\rho_{\mathrm{G}} > 0$, there exists $h \in \Gamma_0(\mathcal{L})$ such that $g = h + \rho_{\mathrm{G}}\|\cdot\|_{\mathcal{L}}^2/2$ and $D^* \circ \partial g \circ D = D^* \circ \partial h \circ D + \rho_{\mathrm{G}} D^*D$ is strongly monotone as D^*D is strongly positive. If f is strongly convex, then ∂f is strongly monotone. The result then follows from (c).

(f) Since L is cocoercive, from [18, Lemma 3.3(i)], $\lambda_{\min} \geq 0$ and the result follows from (e).

□

We will now turn our attention to the convergence of the mismatched CV algorithm by establishing a first result regarding the averagedness properties of operator $\overline{T}$.

Lemma 3.6: Let $\widetilde{\eta} \in]\frac{1}{2}, +\infty[$, $\overline{\alpha} = \dfrac{1}{2 - \frac{1}{2\widetilde{\eta}}}$, and $\overline{W} = \mathrm{Id}_{\mathcal{Z}_P} - P^{-1}\widetilde{L}$. Then, if $P^{-1}\widetilde{L}$ is $\widetilde{\eta}$-cocoercive in $\mathcal{Z}_P$, then $\overline{T}$ is $\overline{\alpha}$-averaged in $\mathcal{Z}_P$.

Proof: If $P^{-1}\widetilde{L}$ is $\widetilde{\eta}$-cocoercive in $\mathcal{Z}_P$ and $\widetilde{\eta} > 1/2$, according to [4, Proposition 4.39], $\overline{W}$ is $\frac{1}{2\widetilde{\eta}}$-averaged in $\mathcal{Z}_P$. Similarly to the proof of [18, Lemma 3.14], we deduce that

$$(\forall z \in \mathcal{Z}_P) \quad \|\overline{W}z - 2(1 - \overline{\alpha})z\|_{\mathcal{Z}_P} + \|\overline{W}z\|_{\mathcal{Z}_P} \leq 2\overline{\alpha}\|z\|_{\mathcal{Z}_P}. \tag{III.23}$$

Since $P^{-1}\mathcal{A}$ is maximally monotone on $\mathcal{Z}_P$, then $J_{P^{-1}\mathcal{A}}$ is firmly nonexpansive [4, Corollary 23.9]. Finally, it follows from (III.23) and [20, Theorem 3.8], that $\overline{T} = J_{P^{-1}\mathcal{A}}(\overline{W} \cdot + P^{-1}\widetilde{K}y)$ is $\overline{\alpha}$-averaged. □ The following theorem provides conditions under which iteration (III.4) converges to a fixed point of $\overline{T}$.

Theorem 3.7: Let $(\tau, \sigma) \in]0, +\infty[^2$ be such that $\tau\sigma\|D\|_{\mathcal{H},\mathcal{L}}^2 < 1$. Assume that $\widetilde{\eta} \in]1/2, +\infty[$ is a cocoercivity constant of $P^{-1}\widetilde{L}$ in $\mathcal{Z}_P$. For $\delta = 2 - 1/(2\widetilde{\eta})$, let $\{\Theta_n\}_{n\in\mathbb{N}} \subset [0, \delta]$ be a sequence such that $\sum_{n\in\mathbb{N}} \Theta_n(\delta - \Theta_n) = +\infty$, and suppose that $\overline{\mathcal{F}} \neq \varnothing$. Then the sequence $((x_n, u_n))_{n\in\mathbb{N}}$ given by (III.4) converges weakly to some point in $\overline{\mathcal{F}}$.

Proof: Let $z_0 = r_0 \in \mathcal{Z}$, let $(z_n)_{n\geq 1}$ be given by (III.7), and let $(r_n)_{n\geq 1}$ be defined as

$$(\forall n \in \mathbb{N}) \quad r_{n+1} = r_n + \Theta_n\left(\overline{T}(r_n) - r_n\right). \tag{III.24}$$

By applying Theorem 3.6, operator $\overline{T}$ is $\overline{\alpha}$-averaged in $\mathcal{Z}_P$ with $\overline{\alpha} = 1/\delta = 1/(2 - \frac{1}{2\widetilde{\eta}})$. Thus, the sequence $(r_n)_{n\in\mathbb{N}}$ converges weakly to some $\overline{r} \in \operatorname{Fix}\left(\overline{T}\right)$ [4, Proposition 5.16], which implies that $\varrho = \sup_{n\in\mathbb{N}} \|r_n\|_{\mathcal{Z}_P} < +\infty$.

For every $n \in \mathbb{N}$, let us bound $z_{n+1} - r_{n+1}$ as follows

$$\begin{aligned} \|z_{n+1} - r_{n+1}\|_{\mathcal{Z}_P} &= \left\|z_n - r_n + \Theta_n(r_n - z_n) + \Theta_n\left(T_n(z_n) - \overline{T}(r_n)\right)\right\|_{\mathcal{Z}_P} \\ &\leq \left\|z_n - r_n + \Theta_n\left(\overline{T}(z_n) - \overline{T}(r_n) - z_n + r_n\right)\right\|_{\mathcal{Z}_P} + \Theta_n\left\|T_n(z_n) - \overline{T}(z_n)\right\|_{\mathcal{Z}_P}. \end{aligned} \tag{III.25}$$

$\overline{T}$ being $\overline{\alpha}$-averaged in $\mathcal{Z}_P$, there exist a nonexpansive operator $W\colon \mathcal{Z}_P \to \mathcal{Z}_P$ such that $\overline{T} = (1 - \overline{\alpha})\mathrm{Id} + \overline{\alpha}W$. Since $\{\Theta_n\}_{n\in\mathbb{N}} \subset [0, 1/\overline{\alpha}]$, we have

$$\begin{aligned} \left\|z_n - r_n + \Theta_n\left(\overline{T}(z_n) - \overline{T}(r_n) - z_n + r_n\right)\right\|_{\mathcal{Z}_P} &= \left\|(1 - \overline{\alpha}\Theta_n)(z_n - r_n) + \overline{\alpha}\Theta_n(W(z_n) - W(r_n))\right\|_{\mathcal{Z}_P} \\ &\leq (1 - \overline{\alpha}\Theta_n)\|z_n - r_n\|_{\mathcal{Z}_P} + \overline{\alpha}\Theta_n\|W(z_n) - W(r_n)\|_{\mathcal{Z}_P} \\ &\leq \|z_n - r_n\|_{\mathcal{Z}_P}. \end{aligned} \tag{III.26}$$

Since $J_{P^{-1}\mathcal{A}}$ is firmly nonexpansive in $\mathcal{Z}_P$, for every $z \in \mathcal{Z}$,

$$\begin{aligned}
\big\|T_n(z) - \overline{T}(z)\big\|_{\mathcal{Z}_P} &\le \|P^{-1}(\widetilde{L}_n z - \widetilde{K}_n y - \widetilde{L} z + \widetilde{K} y)\|_{\mathcal{Z}_P} \\
&\le \|P^{-1}(\widetilde{L}_n - \widetilde{L})z\|_{\mathcal{Z}_P} + \|P^{-1}(\widetilde{K}_n - \widetilde{K})y\|_{\mathcal{Z}_P} \\
&\le \|P^{-1}(\widetilde{L}_n - \widetilde{L})\|_{\mathcal{Z}_P,\mathcal{Z}_P}\|z\|_{\mathcal{Z}_P} + \|P^{-1}(\widetilde{K}_n - \widetilde{K})\|_{\mathcal{Z}_P,\mathcal{Z}_P}\|y\|_{\mathcal{Z}_P} \\
&= \|P^{-1/2}(\widetilde{L}_n - \widetilde{L})P^{-1/2}\|_{\mathcal{Z},\mathcal{Z}}\|z\|_{\mathcal{Z}_P} + \|P^{-1/2}(\widetilde{K}_n - \widetilde{K})P^{-1/2}\|_{\mathcal{Z},\mathcal{Z}}\|y\|_{\mathcal{Z}_P} \\
&\le \|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}(\|\widetilde{L}_n - \widetilde{L}\|_{\mathcal{Z},\mathcal{Z}}\|z\|_{\mathcal{Z}_P} + \|\widetilde{K}_n - \widetilde{K}\|_{\mathcal{Z},\mathcal{Z}}\|y\|_{\mathcal{Z}_P}) \\
&= \|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}(\|L_n - L\|_{\mathcal{H},\mathcal{H}}\|z\|_{\mathcal{Z}_P} + \|K_n - K\|_{\mathcal{G},\mathcal{H}}\|y\|_{\mathcal{Z}_P}) \\
&= \|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}(\|(K_n - K)H\|_{\mathcal{H},\mathcal{H}}\|z\|_{\mathcal{Z}_P} + \|K_n - K\|_{\mathcal{G},\mathcal{H}}\|y\|_{\mathcal{Z}_P}) \\
&\le \|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}(\|H\|_{\mathcal{H},\mathcal{G}}\|z\|_{\mathcal{Z}_P} + \|y\|_{\mathcal{Z}_P})\|K_n - K\|_{\mathcal{G},\mathcal{H}} \\
&\le \|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}(\|H\|_{\mathcal{H},\mathcal{G}}\|z\|_{\mathcal{Z}_P} + \|y\|_{\mathcal{Z}_P})\,\omega_n, \qquad \text{(III.27)}
\end{aligned}$$

where the last inequality follows from Theorem 3.2. Altogether (III.25), (III.26), and (III.27) yield, for every $n \in \mathbb{N}$,

$$\begin{aligned}
\|z_{n+1} - r_{n+1}\|_{\mathcal{Z}_P} &\le \|z_n - r_n\|_{\mathcal{Z}_P} + \Theta_n\|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}(\|H\|_{\mathcal{H},\mathcal{G}}\|z_n\|_{\mathcal{Z}_P} + \|y\|_{\mathcal{Z}_P})\omega_n \\
&\le \|z_n - r_n\|_{\mathcal{Z}_P} + \Theta_n\|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}\big(\|H\|_{\mathcal{H},\mathcal{G}}(\|z_n - r_n\|_{\mathcal{Z}_P} + \|r_n\|_{\mathcal{Z}_P}\big) + \|y\|_{\mathcal{Z}_P})\omega_n \\
&\le (1 + \mu_n)\|z_n - r_n\|_{\mathcal{Z}_P} + \nu_n \qquad \text{(III.28)}
\end{aligned}$$

with

$$\mu_n = \delta\|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}\|H\|_{\mathcal{H},\mathcal{G}}\,\omega_n \qquad \text{(III.29)}$$

$$\nu_n = \delta\|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}(\varrho\|H\|_{\mathcal{H},\mathcal{G}} + \|y\|_{\mathcal{Z}_P})\,\omega_n. \qquad \text{(III.30)}$$

Since $(\mu_n) \in \ell^1_+$ and $(\nu_n) \in \ell^1_+$, according to [4, Lemma 5.31], $\|z_n - r_n\|_{\mathcal{Z}_P} < +\infty$. Consequently, $\delta' = \sup_{n\in\mathbb{N}}\|z_n\|_{\mathcal{Z}_P} < \varrho + \sup_{n\in\mathbb{N}}\|z_n - r_n\|_{\mathcal{Z}_P} < +\infty$.
Let us now define

$$(\forall n \in \mathbb{N}) \quad e_n = \frac{T_n(z_n) - \overline{T}(z_n)}{\overline{\alpha}}, \qquad \text{(III.31)}$$

and it follows from (III.27) that

$$\sum_{n\in\mathbb{N}}\|e_n\|_{\mathcal{Z}_P} \le \frac{\|P^{-1}\|_{\mathcal{Z},\mathcal{Z}}}{\overline{\alpha}}(\|H\|_{\mathcal{H},\mathcal{G}}\delta' + \|y\|_{\mathcal{Z}_P})\sum_{n\in\mathbb{N}}\omega_n < +\infty. \qquad \text{(III.32)}$$

Algorithm (III.7) can be re-expressed as

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = z_n + \Theta'_n\,(Qz_n + e_n - z_n) \quad \text{with} \quad \Theta'_n = \overline{\alpha}\Theta_n \in]0,1[. \qquad \text{(III.33)}$$

Therefore, the weak convergence of $(z_n)_{n\in\mathbb{N}}$ to some $\widetilde{z} \in \mathrm{Fix}(Q) = \mathrm{Fix}\left(\overline{T}\right) = \overline{\mathcal{F}}$ follows from [4, Theorem 5.5]. □ We deduce the following more restrictive convergence result which is an extension of standard convergence results for CV algorithm.

Corollary 3.8: Assume that $\mathrm{Ran}(L + L^*)$ is closed, $\lambda_{\min} \ge 0$, and $\mathrm{Ker}(L + L^*) = \mathrm{Ker}\,L$. Let $(\tau, \sigma) \in]0, +\infty[^2$ be such that $\tau^{-1} - \sigma\|D\|^2_{\mathcal{H},\mathcal{L}} > \|M\|^2_{\mathcal{H},\mathcal{H}}/4$, where M is given by (III.9). For

$$\delta = 2 - \frac{1}{4}\Big(\frac{1}{\tau} - \sigma\|D\|^2_{\mathcal{H},\mathcal{L}}\Big)^{-1}\|M\|^2_{\mathcal{H},\mathcal{H}}, \qquad \text{(III.34)}$$

let $\{\Theta_n\}_{n\in\mathbb{N}} \subset [0, \delta]$ be a sequence such that $\sum_{n\in\mathbb{N}}\Theta_n\,(\delta - \Theta_n) = +\infty$, and suppose that $\overline{\mathcal{F}} \neq \varnothing$. Then the sequence $(x_n, u_n)_{n\in\mathbb{N}}$ given by (III.4) converges weakly to some point in $\overline{\mathcal{F}}$.
Proof: Note that, if $\tau^{-1} - \sigma\|D\|^2_{\mathcal{H},\mathcal{L}} > \|M\|^2_{\mathcal{H},\mathcal{H}}/4$, then $\tau\sigma\|D\|^2_{\mathcal{H},\mathcal{L}} < 1$. In addition, it follows from Theorem 3.4 that, when $\mathrm{Ran}(L + L^*)$ is closed, $\lambda_{\min} \ge 0$, and $\mathrm{Ker}(L + L^*) = \mathrm{Ker}\,L$, $P^{-1}\widetilde{L}$ is $\widetilde{\eta}$ -cocoercive in $\mathcal{Z}_P$ with

$$\widetilde{\eta} = \tau^{-1}(1 - \tau\sigma\|D\|^2_{\mathcal{H},\mathcal{L}})\eta_{\max} \qquad \text{(III.35)}$$

and $\eta_{\max} = 2/\|M\|^2_{\mathcal{H},\mathcal{H}}$. The result then follows by applying Theorem 3.7. □ We now provide an estimate of the distance between a Kuhn-Tucker pair $(\widehat{x}, \widehat{u})$ of the original optimization problem and a fixed point $(\widetilde{x}, \widetilde{u})$ of $\overline{T}$.

Proposition 3.9: Assume that (II.3) holds and L is a cocoercive operator.
If $f + g \circ D$ is strongly convex with modulus $\rho > 0$, then there exists a unique vector $\widetilde{x}$ in $\overline{\mathcal{F}}_1$ and a unique solution $\widehat{x}$ to the primal problem (II.1), and we have

$$\|\widetilde{x} - \widehat{x}\|_{\mathcal{H}} \le \frac{1}{\rho}\left\|\left(\overline{K} - H^*\right)(H\widehat{x} - y)\right\|_{\mathcal{H}}. \qquad \text{(III.36)}$$

In addition, if g is β-Lipschitz differentiable with $\beta \in [0, +\infty[$, there exists a unique $(\widetilde{x}, \widetilde{u}) \in \overline{\mathcal{F}}$ and a unique solution $\widehat{u}$ to the dual problem, and we have

$$\|\widetilde{u} - \widehat{u}\|_{\mathcal{L}} \leq \frac{\beta}{\rho} \|D\|_{\mathcal{H},\mathcal{L}} \left\| \left(\overline{K} - H^*\right) (H\widehat{x} - y) \right\|_{\mathcal{H}}. \tag{III.37}$$

Proof: Since (II.3) is satisfied, $\partial f + D^* \circ \partial g \circ D = \partial(f + g \circ D)$ [4, Theorem 16.47 (i)]. As we have assumed that $f + g \circ D$ is ρ-strongly convex and L is cocoercive, $L + \partial f + D^* \circ \partial g \circ D$ is strongly monotone. If follows from Theorem 3.5(c) that there exists a single element $\widetilde{x}$ in $\overline{\mathcal{F}}_1$ which is such that

$$\overline{K}y \in L\widetilde{x} + \partial f(\widetilde{x}) + D^* \partial g(D\widetilde{x}). \tag{III.38}$$

For any $\gamma > 0$, (III.38) is equivalent to

$$\widetilde{x} = \operatorname{prox}_{\gamma(f+g\circ D)} \left(\widetilde{x} - \gamma\overline{K}(H\widetilde{x} - y)\right). \tag{III.39}$$

For similar reasons, there exists a unique solution $\widehat{x}$ to the primal problem, which satisfies the fixed point equation

$$\widehat{x} = \operatorname{prox}_{\gamma(f+g\circ D)} \left(\widehat{x} - \gamma H^*(H\widehat{x} - y)\right). \tag{III.40}$$

Because of the ρ-strong convexity of $f + g \circ D$, $\operatorname{prox}_{\gamma(f+g\circ D)}$ is $(1 + \gamma\rho)^{-1}$-Lipschitzian [4, Proposition 23.13]. The error bound in (III.36) is thus derived by the same arguments as in the proof of [18, Theorem 3.11].
In addition, if g is Gâteaux differentiable, there exists a unique $\widetilde{u} \in \mathcal{L}$ such that $(\widetilde{x}, \widetilde{u}) \in \overline{\mathcal{F}}$, which is given by

$$\widetilde{u} = \nabla g(D\widetilde{x}), \tag{III.41}$$

where ∇g is the gradient of g. Similarly, there exists a unique solution $\widehat{u}$ to the dual problem, given by

$$\widehat{u} = \nabla g(D\widehat{x}). \tag{III.42}$$

By using the fact that ∇g is β-Lipschitzian, we deduce that

$$\begin{aligned} \|\widetilde{u} - \widehat{u}\|_{\mathcal{L}} &\leq \beta \|D(\widetilde{x} - \widehat{x})\|_{\mathcal{L}} \\ &\leq \beta \|D\|_{\mathcal{H},\mathcal{L}} \|\widetilde{x} - \widehat{x}\|_{\mathcal{H}}. \end{aligned} \tag{III.43}$$

The upper error bound in (III.37) then follows from (III.36). □

C. Convergence properties of the mismatched CV algorithm with modified D

We now provide a brief analysis of the convergence of Condat-Vũ algorithm when D is replaced by an operator $V \in \mathcal{B}(\mathcal{H}, \mathcal{L})$ such that $\|V - D\|_{\mathcal{H},\mathcal{L}} \leq \epsilon$ for some $\epsilon > 0$. Let

(1) $\widetilde{\mathcal{A}} : \mathcal{Z} \to 2^{\mathcal{Z}} : z = (x, u) \mapsto (\partial f(x) + V^* u, -Vx + \partial g^*(u))$
(2) $\widetilde{P} : \mathcal{Z} \to \mathcal{Z} : z = (x, u) \mapsto \left(\frac{x}{\tau} - V^* u, -Vx + \frac{u}{\sigma}\right)$

with $(\tau, \sigma) \in]0, +\infty[^2$.
Here again $\widetilde{\mathcal{A}}$ is maximally monotone on $\mathcal{Z}$ and $\widetilde{P}$ is self adjoint and strongly positive for all $(\tau, \sigma) \in]0, +\infty]^2$ satisfying $\tau\sigma \left(\|D\|_{\mathcal{H},\mathcal{L}} + \epsilon\right)^2 < 1$, hence $\widetilde{P}^{-1}\widetilde{\mathcal{A}}$ is maximally monotone in $\mathcal{Z}_{\mathcal{P}}$ [4, Proposition 20.24]. Let the operators L, $\widetilde{L}$, $\widetilde{K}$, $\widetilde{K}_n$, and $\widetilde{L}_n$ be defined by Theorem 3.3 and suppose that $(K_n, \overline{K}) \in (\mathcal{B}(\mathcal{G}, \mathcal{H}))^2$ satisfy Theorem 3.2.
For every $n \in \mathbb{N}$, we define the operators

$$\widetilde{T} = J_{\widetilde{P}^{-1}\widetilde{\mathcal{A}}} \circ \left(\mathrm{Id}_{\mathcal{Z}} - \widetilde{P}^{-1}(\widetilde{L} - \widetilde{K})\right), \quad \widetilde{T}_n = J_{\widetilde{P}^{-1}\widetilde{\mathcal{A}}} \circ \left(\mathrm{Id}_{\mathcal{Z}} - \widetilde{P}^{-1}(\widetilde{L}_n - \widetilde{K}_n)\right). \tag{III.44}$$

Then the mismatched form of CV algorithm can be rewritten as

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = z_n + \Theta_n \left(\widetilde{T}_n (z_n) - z_n\right), \tag{III.45}$$

with $z_0 \in \mathcal{Z}$. Under conditions similar to those in Theorem 3.7, the weak convergence of sequence $(z_n)_{n\in\mathbb{N}}$ to some $\widetilde{z} = (\widetilde{x}, \widetilde{u}) \in \operatorname{Fix}\widetilde{T}$ can be readily established.
The following result provides bounds on the errors incurred in this additional change.

Proposition 3.10: Assume that L is a cocoercive operator, f is strongly convex with modulus $\rho > 0$, and g is Lipschitz-differentiable with constant $\beta > 0$. Then, there exists a unique fixed point $(\widetilde{x}, \widetilde{u})$ of $\widetilde{T}$ and unique solution $(\widehat{x}, \widehat{u})$ to the primal-dual problem(II.5). In addition,

$$\|\widetilde{x} - \widehat{x}\|_{\mathcal{H}} \leq \frac{1}{\rho} \left\|(\overline{K} - H^*)(H\widehat{x} - y) + V^*\nabla g(V\widehat{x}) - D^*\nabla g(D\widehat{x})\right\|_{\mathcal{H}} \tag{III.46}$$

$$\|\widetilde{u} - \widehat{u}\|_{\mathcal{L}} \leq \beta(\|V\|_{\mathcal{H},\mathcal{L}} \|\widetilde{x} - \widehat{x}\|_{\mathcal{H}} + \|(V - D)\widehat{x}\|_{\mathcal{L}}). \tag{III.47}$$

Proof: Since $\operatorname{dom} g = \mathcal{L}$, (II.3) holds. The existence and uniqueness of $(\widetilde{x}, \widetilde{u})$ and $(\widehat{x}, \widehat{u})$ follows from arguments similar to those in the proof of Theorem 3.9.
In addition, it follows from (III.38) that, for every $\gamma > 0$,

$$\widetilde{x} = \operatorname{prox}_{\gamma f}\big(\widetilde{x} - \gamma(\overline{K}(H\widetilde{x} - y) + V^*\nabla g(V\widetilde{x}))\big) \tag{III.48}$$

Similarly,

$$\widehat{x} = \operatorname{prox}_{\gamma f}\big(\widehat{x} - \gamma(H^*(H\widehat{x} - y) + D^*\nabla g(D\widehat{x}))\big). \tag{III.49}$$

Since f is ρ-strongly convex, $\operatorname{prox}_{\gamma f}$ is Lipschitz-continuous with constant $(1+\gamma\rho)^{-1}$ [4, Proposition 23.13]. Thus, using the triangle inequality and the Lipschitz property of $\operatorname{prox}_{\gamma f}$ yields

$$\begin{aligned}\|\widetilde{x} - \widehat{x}\|_{\mathcal{H}} &\leq \frac{1}{1+\gamma\rho}\big\|\widetilde{x} - \widehat{x} - \gamma(\overline{K}(H\widetilde{x} - y) + V^*\nabla g(V\widetilde{x}) - H^*(H\widehat{x} - y) - D^*\nabla g(D\widehat{x}))\big\|_{\mathcal{H}}\\ &\leq \frac{1}{1+\gamma\rho}\Big(\|(\mathrm{Id}_{\mathcal{H}} - \gamma L)(\widetilde{x} - \widehat{x}) - \gamma(V^*\nabla g(V\widetilde{x}) - V^*\nabla g(V\widehat{x}))\|_{\mathcal{H}}\\ &\quad + \gamma\big\|(\overline{K} - H^*)(H\widehat{x} - y) + V^*\nabla g(V\widehat{x}) - D^*\nabla g(D\widehat{x})\big\|_{\mathcal{H}}\Big).\end{aligned} \tag{III.50}$$

Since L is cocoercive with constant $\eta_{\max}$ and ∇g is β-Lipschitzian, hence cocoercive with constant $1/\beta$, $L + V^* \circ \nabla g \circ V$ is cocoercive with constant $\zeta = \big((\eta_{\max})^{-1} + \beta\|V\|^2_{\mathcal{H},\mathcal{L}}\big)^{-1}$ [4, Proposition 4.12]. Consequently, by choosing $\gamma \in [0, 2\zeta]$, $\mathrm{Id}_{\mathcal{H}} - \gamma(L + V^* \circ \nabla g \circ V)$ is nonexpansive [4, Proposition 4.39]. We then deduce from (III.50) that

$$\|\widetilde{x} - \widehat{x}\|_{\mathcal{H}} \leq \frac{1}{1+\gamma\rho}\Big(\|\widetilde{x} - \widehat{x}\|_{\mathcal{H}} + \gamma\big\|(\overline{K} - H^*)(H\widehat{x} - y) + V^*\nabla g(V\widehat{x}) - D^*\nabla g(D\widehat{x})\big\|_{\mathcal{H}}\Big), \tag{III.51}$$

which results in (III.46).
According to the first-order characterization of $\operatorname{Fix}\widetilde{T}$,

$$\widetilde{u} = \nabla g(V\widetilde{x}). \tag{III.52}$$

Using this equation and (III.42) allows us to derive the following inequality:

$$\begin{aligned}\|\widetilde{u} - \widehat{u}\|_{\mathcal{L}} &= \|\nabla g(V\widetilde{x}) - \nabla g(D\widehat{x})\|_{\mathcal{L}}\\ &\leq \beta\|V\widetilde{x} - D\widehat{x}\|_{\mathcal{L}}\\ &\leq \beta(\|V(\widetilde{x} - \widehat{x})\|_{\mathcal{L}} + \|(V - D)\widehat{x}\|_{\mathcal{L}}),\end{aligned} \tag{III.53}$$

so yielding (III.47). □

D. Remarks on the mismatched projected primal-dual splitting algorithm

When an additional constraint is added to our initial primal problem (II.1), the optimization problem becomes

$$\text{Find } \widehat{x} \in C \cap \operatorname*{Argmin}_{x\in\mathcal{H}} \frac{1}{2}\|Hx - y\|^2_{\mathcal{G}} + f(x) + g(Dx), \tag{III.54}$$

where C is a closed and convex nonempty subset of $\mathcal{H}$, and H, D, f and g are defined as previously. The dual problem reads

$$\text{Find } \widehat{u} \in E \cap \operatorname*{Argmin}_{u\in\mathcal{L}} (f + h)^*(-D^*u) + g^*(u), \tag{III.55}$$

where E is a closed vector subspace of $\mathcal{L}$ such that $\operatorname{ran} D \subset E$. Such a problem can be solved by the projected form of CV algorithm proposed by Briceño-Arias and López in [8]. We will be interested in the mismatched form of this algorithm with a fixed operator $\overline{K}$: Given $(x_0, u_0) \in \mathcal{Z}$ and $(\tau, \sigma) \in]0, +\infty[^2$,

Projected mismatched primal-dual algorithm:

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} p_n = \operatorname{prox}_{\tau f}\big(x_n - \tau\big(\overline{K}\,(Hx_n - y) + D^*u_n\big)\big)\\ x_{n+1} = \operatorname{proj}_C(p_n)\\ \overline{x}_n = x_{n+1} + p_n - x_n\\ q_n = \operatorname{prox}_{\sigma g^*}(u_n + \sigma D\overline{x}_n)\\ u_{n+1} = \operatorname{proj}_E(q_n)\end{array}\right. . \tag{III.56}$$

In particular, if $\overline{K} = H^*$, $C = \mathcal{H}$, and $E = \mathcal{L}$, we recover (III.3) in the case when, for every $n \in \mathbb{N}$, $\Theta_n = 1$.
We have then the following convergence result where L and $\overline{\mathcal{F}}$ are defined as previously.

Proposition 3.11: Assume that $\operatorname{Ran}(L + L^*)$ is closed, $\lambda_{\min} \geq 0$, and $\operatorname{Ker}(L + L^*) = \operatorname{Ker} L$. Let $(\tau, \sigma) \in]0, +\infty[^2$ be such that $\tau^{-1} - \sigma\|D\|^2_{\mathcal{H},\mathcal{L}} > \|M\|^2_{\mathcal{H},\mathcal{H}}/4$, where M is given by (III.9). Suppose that $\overline{\mathcal{F}} \cap (C \times E) \neq \varnothing$. Then the sequence

$(x_n, u_n)_{n\in\mathbb{N}}$ generated by (III.56) converges weakly to some point in $\overline{\mathcal{F}}\cap(C\times E)$. In addition $(p_n - x_n)_{n\in\mathbb{N}}$ and $(q_n - u_n)_{n\in\mathbb{N}}$ converge strongly to 0.

Proof: On the one hand, under the assumptions made on L, we have seen that $x \mapsto \overline{K}(Hx - y)$ is $\eta_{\max}$-cocoercive with $\eta_{\max} = 2/\|M\|^2_{\mathcal{H},\mathcal{H}}$. On the other hand, [8, Theorem 3.2 (ii)] guarantees the weak convergence of $(x_n, u_n)_{n\in\mathbb{N}}$ under the condition

$$\|D\|^2_{\mathcal{H},\mathcal{L}} < \frac{1}{\sigma}\left(\frac{1}{\tau} - \frac{1}{2\eta_{\max}}\right). \tag{III.57}$$

The strong convergence properties of $(p_n - x_n)_{n\in\mathbb{N}}$ and $(q_n - u_n)_{n\in\mathbb{N}}$ are also stated in the proof of [8, Theorem 3.2]. □ Note that the error related to the mismatch can still be quantified by Theorem 3.9.

IV. The mismatched Loris-Verhoeven algorithm

In this section, we consider a specific instance of our original template (II.1) where $f = \frac{\kappa}{2}\|\cdot\|^2_{\mathcal{H}}$ with $\kappa \in]0, +\infty[$. This problem can be solved by the primal-dual algorithm proposed by Loris and Verhoeven [40], which also relies on the forward-backward splitting. This algorithm, that will be designated as LV algorithm, also appears under the name of *Primal-Dual Fixed-Point algorithm based on the Proximity Operator* (PDFP2O) [16] and *Proximal Alternating Predictor-Corrector* (PAPC) algorithm [29].

The LV iterations can still be described by means of the implicit inclusion (III.1) where $\mathcal{A}$, B, and P are now given by

$$(\forall z = (x,u) \in \mathcal{Z}) \quad \mathcal{A}z = \begin{pmatrix} D^*u \\ -Dx + \partial g^*(u)\end{pmatrix} \tag{IV.1}$$

$$Bz = \begin{pmatrix} (H^*H + \kappa\,\mathrm{Id}_{\mathcal{H}})x - H^*y \\ 0\end{pmatrix} \tag{IV.2}$$

$$Pz = \begin{pmatrix} \frac{1}{\tau}x \\ (\frac{1}{\sigma}\,\mathrm{Id}_{\mathcal{L}} - \tau D^*D)u\end{pmatrix} \tag{IV.3}$$

with $(\tau, \sigma) \in]0, +\infty[^2$. The iterations of LV algorithm are then given by

LV iterations:

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} t_n = H^*(Hx_n - y) + \kappa x_n \\ u_{n+\frac{1}{2}} = \mathrm{prox}_{\sigma g^*}\Big(u_n + \sigma D\Big(x_n - \tau(t_n + D^*u_n)\Big)\Big) \\ x_{n+1} = x_n - \Theta_n\tau\Big(t_n + D^*u_{n+\frac{1}{2}}\Big) \\ u_{n+1} = u_n + \Theta_n\Big(u_{n+\frac{1}{2}} - u_n\Big) \end{array}\right. \tag{IV.4}$$

where $\{\Theta_n\}_{n\in\mathbb{N}}$ is a sequence of relaxation parameters and $(x_0, u_0) \in \mathcal{Z}$.

When, at iteration $n \in \mathbb{N}$, H^* is replaced by an operator $K_n \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ satisfying Theorem 3.2, the mismatched form of LV algorithm reads

Mismatched LV iterations:

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} t_n = K_n(Hx_n - y) + \kappa x_n \\ u_{n+\frac{1}{2}} = \mathrm{prox}_{\sigma g^*}\Big(u_n + \sigma D\Big(x_n - \tau(t_n + D^*u_n)\Big)\Big) \\ x_{n+1} = x_n - \Theta_n\tau\Big(t_n + D^*u_{n+\frac{1}{2}}\Big) \\ u_{n+1} = u_n + \Theta_n\Big(u_{n+\frac{1}{2}} - u_n\Big) \end{array}\right. \tag{IV.5}$$

This iteration can be reexpressed as (III.7) where Theorem 3.3(ii)-(iv) holds, but L is now defined as

$$L = \overline{K}H + \kappa\mathrm{Id}_{\mathcal{H}}. \tag{IV.6}$$

It follows that all the results in subsection III-B can be extended to the mismatched LV algorithm.

Theorem 4.1 is a straightforward adaptation of Theorem 3.5 to characterize the fixed point set of the nonlinear mapping $\overline{T}$ (see Theorem 3.3(iv)).

Proposition 4.1: Let $(\widetilde{x}, \widetilde{u}) \in \mathcal{Z}$. Then $(\widetilde{x}, \widetilde{u}) \in \mathrm{Fix}\overline{T}$ if and only if $(\widetilde{x}, \widetilde{u})$ belongs to

$$\overline{\mathcal{F}} = \left\{(x,u) \in \mathcal{Z} \mid \overline{K}y \in Lx + D^*u,\ u \in \partial g(Dx)\right\}, \tag{IV.7}$$

which is nonempty if $L + D^* \circ \partial g \circ D$ is surjective.

(i) If $\lambda_{\min} \geq 0$, then $\overline{\mathcal{F}}$ is closed and convex.

(ii) Let $\overline{\mathcal{F}}_1 = \{x \in \mathcal{H} \mid (\exists u \in \mathcal{L})\ (x,u) \in \overline{\mathcal{F}}\}$.

$\overline{\mathcal{F}}_1$ has at most one element if one of the following conditions holds:

(a) $L + D^* \circ \partial g \circ D$ is strictly monotone.
(b) $\lambda_{\min} \geq 0$ and $g \circ D$ is strictly convex.

$\overline{\mathcal{F}}_1$ is a singleton if (II.3) is satisfied and one of the following conditions holds:

(c) $\lambda_{\min} \geq 0$ and $L + D^* \circ \partial g \circ D$ is strongly monotone.
(d) $\lambda_{\min} > 0$.
(e) $\lambda_{\min} \geq 0$, g is strongly convex, and D^*D is strongly positive.
(f) L is cocoercive, g is strongly convex, and D^*D is strongly positive.

Similarly, we derive an equivalent of Theorem 3.8 concerning the convergence of the mismatched LV.

Proposition 4.2: Assume that $\mathrm{Ran}(L + L^*)$ is closed, $\lambda_{\min} \geq 0$, and $\mathrm{Ker}(L + L^*) = \mathrm{Ker}\, L$. Let $(\tau, \sigma) \in]0, +\infty[^2$ be such that $\tau < 4/\|M\|^2_{\mathcal{H},\mathcal{H}}$ and $\tau\sigma\|D\|^2_{\mathcal{H},\mathcal{L}} < 1$, where M is given by (III.9). For

$$\delta = 2 - \frac{\tau}{4}\|M\|^2_{\mathcal{H},\mathcal{H}}, \tag{IV.8}$$

let $\{\Theta_n\}_{n\in\mathbb{N}} \subset [0, \delta]$ be a sequence such that $\sum_{n\in\mathbb{N}} \Theta_n\,(\delta - \Theta_n) = +\infty$, and suppose that $\overline{\mathcal{F}} \neq \varnothing$. Then the sequence $(x_n, u_n)_{n\in\mathbb{N}}$ given by (IV.5) converges weakly to some point in $\overline{\mathcal{F}}$.

Proof: The result from Theorem 3.4 stating that, when $\mathrm{Ran}(L + L^*)$ is closed, $\lambda_{\min} \geq 0$, and $\mathrm{Ker}(L + L^*) = \mathrm{Ker}\, L$, $P^{-1}\widetilde{L}$ is cocoercive in $\mathcal{Z}_P$ with constant $\widetilde{\eta}_{\max} = 2/\|P^{-1/2}\Pi M\|^2_{\mathcal{H},\mathcal{Z}}$ still holds for LV algorithm. We also have

$$\widetilde{\eta}_{\max} \geq \frac{2}{\|P^{-1/2}\Pi\|^2_{\mathcal{H},\mathcal{Z}}\|M\|^2_{\mathcal{H},\mathcal{H}}} = \frac{2}{\tau\|M\|^2_{\mathcal{H},\mathcal{H}}} = \widetilde{\eta}, \tag{IV.9}$$

which shows that $P^{-1}\widetilde{L}$ is $\widetilde{\eta}$-cocoercive in $\mathcal{Z}_P$. The result then follows from Theorem 3.7, which remains valid in this context. □

Remark 4.3: When there is no mismatch, $M = \sqrt{2}(H^*H + \kappa\,\mathrm{Id}_{\mathcal{H}})^{1/2}$ and we recover the conditions derived in [40, Theorem 3.1], [26, Theorem 3.1] for the convergence of sequences $(x_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ generated by (IV.4).

Similarly to Theorem 3.9, we can provide an estimate of the distance between a Kuhn-Tucker pair $(\widehat{x}, \widehat{u})$ of the original problem and a fixed point $(\widetilde{x}, \widetilde{u})$ of $\overline{T}$.

Proposition 4.4: Assume that $0 \in \mathrm{sri}\,(\mathrm{Ran}(D) - \mathrm{dom}(g))$ and L, given in (IV.6), is a cocoercive operator.

Let $\rho \in]0, +\infty[$. If $\kappa \geq \rho$ or $g \circ D$ is strongly convex with modulus ρ, then there exists a unique vector $\widetilde{x}$ in $\overline{\mathcal{F}}_1$, defined in Theorem 4.1(ii), and a unique solution $\widehat{x}$ to the primal problem (II.1). Moreover, inequality (III.36) holds.

In addition, if g is β-Lipschitz differentiable with $\beta \in [0, +\infty[$, there exists a unique $(\widetilde{x}, \widetilde{u}) \in \overline{\mathcal{F}}$, defined in (IV.7), and a unique solution $\widehat{u}$ to the dual problem. Finally, inequality (III.37) is satisfied.

V. The mismatched Combettes - Pesquet algorithm

A. CP algorithm

In this section, we explore the benefits of the Combettes - Pesquet algorithm (CP), which relies on Tseng's splitting (II.6) to solve (II.1)-(II.4). This algorithm was introduced in [22, Theorem 4.2] and constitutes a generalization of the one in [9]. It reads

CP-iterations for (II.1) **and** (II.4)**:**

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} v_{1,n} = x_n - \gamma\,(H^*(Hx_n - y) + D^*u_n) \\ p_{1,n} = \mathrm{prox}_{\gamma f}(v_{1,n}) \\ v_{2,n} = u_n + \gamma D x_n \\ p_{2,n} = \mathrm{prox}_{\gamma g^*}(v_{2,n}) \\ q_{2,n} = p_{2,n} + \gamma D p_{1,n} \\ q_{1,n} = p_{1,n} - \gamma\,(H^*(Hp_{1,n} - y) + D^*p_{2,n}) \\ x_{n+1} = x_n - v_{1,n} + q_{1,n} \\ u_{n+1} = u_n - v_{2,n} + q_{2,n}, \end{array} \right. \tag{V.1}$$

where $\gamma > 0$ and $(x_0, u_0) \in \mathcal{Z}$. By setting, for every $n \in \mathbb{N}$,

$$z_n = (x_n, u_n)\,,\ v_n = (v_{1,n}, v_{2,n})\,,\ p_n = (p_{1,n}, p_{2,n})\,,\ \text{ and }\ q_n = (q_{1,n}, q_{2,n})\,, \tag{V.2}$$

iterations (V.1) can be rewritten as

Tseng iterations:

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} v_n = z_n - \gamma Q(z_n) \\ p_n = J_{\gamma\mathscr{M}}(v_n) \\ q_n = p_n - \gamma Q(p_n) \\ z_{n+1} = z_n - v_n + q_n, \end{array} \right. \tag{V.3}$$

with

$$\mathscr{M}\colon \mathcal{Z} \to 2^{\mathcal{Z}}\colon (x,u) \mapsto (\partial f(x), \partial g^*(u)) \tag{V.4}$$

$$Q\colon \mathcal{Z} \to \mathcal{Z}\colon (x,u) \mapsto (H^*Hx + D^*u - H^*y, -Dx)\,. \tag{V.5}$$

Since (V.3) is an instance of Tseng's algorithm, according to [9, Theorem 2.5 (ii)], if $\operatorname{zer}(\mathscr{M}+Q) \neq \varnothing$ and $\gamma \in]0, 1/\vartheta[$, sequences $(z_n)_{n\in\mathbb{N}} = \big((x_n,u_n)\big)_{n\in\mathbb{N}}$ and $(p_n)_{n\in\mathbb{N}}$ converge weakly to some $\widehat{z} = (\widehat{x},\widehat{u}) \in \operatorname{zer}(\mathscr{M}+Q)$. Therefore, sequences $(x_n)_{n\in\mathbb{N}}$ (resp. $(u_n)_{n\in\mathbb{N}}$) generated by (V.1) converge weakly to $\widehat{x}$ (resp. $\widehat{u}$), which is a solution to (II.1) (resp. (II.4)).

B. Convergence properties of the twice mismatched CP algorithm

For algorithm (V.1), we consider a mismatched form obtained by substituting a fixed operator $\overline{K} \in \mathcal{B}(\mathcal{G},\mathcal{H})$ for H^* as well as an operator $V^* \in \mathcal{B}(\mathcal{L},\mathcal{H})$ for D^*. This leads to the following iterations:

Mismatched CP-iterations:

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} v_{1,n} = x_n - \gamma\left(\overline{K}(Hx_n - y) + V^*u_n\right) \\ p_{1,n} = \operatorname{prox}_{\gamma f}(v_{1,n}) \\ v_{2,n} = u_n + \gamma(Dx_n + \varepsilon u_n) \\ p_{2,n} = \operatorname{prox}_{\gamma g^*}(v_{2,n}) \\ q_{2,n} = p_{2,n} + \gamma(Dp_{1,n} + \varepsilon p_{2,n}) \\ q_{1,n} = p_{1,n} - \gamma\left(\overline{K}(Hp_{1,n} - y) + V^*p_{2,n}\right) \\ x_{n+1} = x_n - v_{1,n} + q_{1,n} \\ u_{n+1} = u_n - v_{2,n} + q_{2,n}, \end{array} \right. \tag{V.6}$$

where $\varepsilon > 0$ is an additional parameter. In addition to the presence of mismatched adjoints, we see that there are two algorithmic modifications in the updates rules of the variables $v_{2,n}$ and $q_{2,n}$.
By making the change of variables (V.2), algorithm (V.6) can be rewritten, in the product space $\mathcal{Z}$, as

Tseng form of (V.6)**:**

$$\text{for } n = 0, 1, \ldots \quad \left\lfloor \begin{array}{l} v_n = z_n - \gamma\widetilde{Q}(z_n) \\ p_n = J_{\gamma\mathscr{M}}(v_n) \\ q_n = p_n - \gamma\widetilde{Q}(p_n) \\ z_{n+1} = z_n - v_n + q_n, \end{array} \right. \tag{V.7}$$

where

$$\widetilde{Q}\colon \mathcal{Z} \to \mathcal{Z}\colon (x,u) \mapsto \big(Lx + V^*u - \overline{K}y, -Dx + \varepsilon u\big) \tag{V.8}$$

and

$$L = \overline{K}H, \tag{V.9}$$

as in section III.
We first provide some preliminary results about operator $\widetilde{Q}$.

Proposition 5.1: Let $\lambda_{\min}$ defined in Theorem 3.3 and

$$^{\varepsilon}\lambda_{\min} = \lambda_{\min} - \frac{1}{4\varepsilon}\|V - D\|^2_{\mathcal{H},\mathcal{L}}, \tag{V.10}$$

$$^{\varepsilon}\vartheta_1 = \max\{\|L\|_{\mathcal{H},\mathcal{H}}, \varepsilon\} + \max\{\|D\|_{\mathcal{H},\mathcal{L}}, \|V\|_{\mathcal{H},\mathcal{L}}\}, \tag{V.11}$$

$$^{\varepsilon}\vartheta_2 = \sqrt{\|L\|^2_{\mathcal{H},\mathcal{H}} + \|V\|^2_{\mathcal{H},\mathcal{L}} + \|D\|^2_{\mathcal{H},\mathcal{L}} + \varepsilon^2}, \tag{V.12}$$

and $^{\varepsilon}\vartheta = \min\{{}^{\varepsilon}\vartheta_1, {}^{\varepsilon}\vartheta_2\}$. We have the following properties:

(i) $\widetilde{Q}$ is Lipschitz continuous with constant $^{\varepsilon}\vartheta$.
(ii) If $^{\varepsilon}\lambda_{\min} \geq 0$,
then $\widetilde{Q}$ is monotone.
(iii) If $^{\varepsilon}\lambda_{\min} > 0$, then $\widetilde{Q}$ is strongly monotone and cocoercive.

Proof: Since $\widetilde{Q}$ is an affine operator, its monotonicity, Lipschitz continuity, and cocoercivity properties are the same as those of the linear operator

$$\overline{Q} = \widetilde{Q} + (\overline{K}y, 0). \tag{V.13}$$

(i) On the one hand

$$\begin{aligned} \|\overline{Q}\|_{\mathcal{Z},\mathcal{Z}} &\leq \left\|\begin{bmatrix} L & 0 \\ 0 & \varepsilon\operatorname{Id}_{\mathcal{L}} \end{bmatrix}\right\|_{\mathcal{Z},\mathcal{Z}} + \left\|\begin{bmatrix} 0 & V^* \\ -D & 0 \end{bmatrix}\right\|_{\mathcal{Z},\mathcal{Z}} \\ &\leq \max\{\|L\|_{\mathcal{H},\mathcal{H}}, \varepsilon\} + \max\{\|D\|_{\mathcal{H},\mathcal{L}}, \|V\|_{\mathcal{H},\mathcal{L}}\} = {}^{\varepsilon}\vartheta_1. \end{aligned} \tag{V.14}$$

On the other hand, for every $z = (x, u) \in \mathcal{Z}$,

$$\begin{aligned}\|\overline{Q}z\|_{\mathcal{Z}}^2 &= \|Lx + V^*u\|_{\mathcal{H}}^2 + \| - Dx + \varepsilon u\|_{\mathcal{L}}^2 \\ &\leq \|LL^* + V^*V\|_{\mathcal{H},\mathcal{H}}\|z\|_{\mathcal{Z}}^2 + \|DD^* + \varepsilon^2 \operatorname{Id}_{\mathcal{L}}\|_{\mathcal{L},\mathcal{L}}\|z\|_{\mathcal{Z}}^2 \\ &\leq (\|L\|_{\mathcal{H},\mathcal{H}}^2 + \|V\|_{\mathcal{H},\mathcal{L}}^2 + \|D\|_{\mathcal{H},\mathcal{L}}^2 + \varepsilon^2)\|z\|_{\mathcal{Z}}^2, \end{aligned} \tag{V.15}$$

which implies that

$$\|\overline{Q}\|_{\mathcal{Z},\mathcal{Z}} \leq {}^{\varepsilon}\vartheta_2. \tag{V.16}$$

In summary, $\overline{Q}$, and thus $\widetilde{Q}$, are Lipschitz continuous with constant $\|\overline{Q}\|_{\mathcal{Z},\mathcal{Z}} \leq {}^{\varepsilon}\vartheta$.

(ii) By using Cauchy-Schwarz inequality, for every $z = (x, u) \in \mathcal{Z}$,

$$\begin{aligned}\left\langle \overline{Q}z, z\right\rangle_{\mathcal{Z}} &= \langle Lx, x\rangle_{\mathcal{H}} + \varepsilon\|u\|_{\mathcal{L}}^2 + \langle u, (V - D)x\rangle_{\mathcal{L}} \\ &\geq \lambda_{\min}\|x\|_{\mathcal{H}}^2 + \varepsilon\|u\|_{\mathcal{L}}^2 - \|u\|_{\mathcal{L}}\|(V - D)x\|_{\mathcal{L}} \\ &\geq \lambda_{\min}\|x\|_{\mathcal{H}}^2 + \varepsilon\|u\|_{\mathcal{L}}^2 - \|V - D\|_{\mathcal{H},\mathcal{L}}\|u\|_{\mathcal{L}}\|x\|_{\mathcal{H}} \\ &= \begin{bmatrix}\|x\|_{\mathcal{H}} & \|u\|_{\mathcal{L}}\end{bmatrix} C \begin{bmatrix}\|x\|_{\mathcal{H}} \\ \|u\|_{\mathcal{L}}\end{bmatrix}, \end{aligned} \tag{V.17}$$

where

$$C = \begin{bmatrix}\lambda_{\min} & -\frac{1}{2}\|V - D\|_{\mathcal{H},\mathcal{L}} \\ -\frac{1}{2}\|V - D\|_{\mathcal{H},\mathcal{L}} & \varepsilon\end{bmatrix}. \tag{V.18}$$

C is positive semidefinite if and only if

$$\begin{cases}\operatorname{tr}(C) = \lambda_{\min} + \varepsilon \geq 0 \\ \det(C) = \lambda_{\min}\varepsilon - \frac{1}{4}\|V - D\|_{\mathcal{H},\mathcal{L}}^2 \geq 0,\end{cases} \tag{V.19}$$

that is ${}^{\varepsilon}\lambda_{\min} \geq 0$. we deduce from (V.17) that, subject to this condition, $\overline{Q}$ and thus $\widetilde{Q}$ are monotone.

(iii) Assume now that ${}^{\varepsilon}\lambda_{\min} > 0$. Then C is positive definite and its smallest eigenvalue is

$${}^{\varepsilon}\upsilon = \frac{\lambda_{\min} + \varepsilon - \sqrt{(\lambda_{\min} + \varepsilon)^2 - 4\,{}^{\varepsilon}\lambda_{\min}\varepsilon}}{2} > 0. \tag{V.20}$$

It follows from (V.17) that $\overline{Q}$ is strongly monotone with constant ${}^{\varepsilon}\upsilon$.

We have then, for every $z \in \mathcal{Z}$,

$$\left\langle \overline{Q}z, z\right\rangle_{\mathcal{Z}} \geq {}^{\varepsilon}\upsilon\,\|z\|_{\mathcal{Z}}^2 \geq {}^{\varepsilon}\upsilon\,\frac{\|\overline{Q}z\|_{\mathcal{Z}}^2}{\|\overline{Q}\|_{\mathcal{Z},\mathcal{Z}}^2} \geq \frac{{}^{\varepsilon}\upsilon}{({}^{\varepsilon}\vartheta)^2}\,\|\overline{Q}z\|_{\mathcal{Z}}^2.$$

This shows that $\overline{Q}$ (and thus $\widetilde{Q}$) is cocoercive with constant

$${}^{\varepsilon}\eta = \frac{{}^{\varepsilon}\upsilon}{({}^{\varepsilon}\vartheta)^2}. \tag{V.21}$$

□

Conditions of convergence for the mismatched CP algorithm are deduced from this result.

Proposition 5.2: Let ${}^{\varepsilon}\lambda_{\min}$ and ${}^{\varepsilon}\vartheta$ be defined as in Theorem 5.1. Let $\gamma \in]0, ({}^{\varepsilon}\vartheta)^{-1}[$. Assume that $\operatorname{zer}(\mathscr{M} + \widetilde{Q}) \neq \varnothing$ and ${}^{\varepsilon}\lambda_{\min} \geq 0$. Then the sequences $((x_n, u_n))_{n\in\mathbb{N}}$ and $((p_{1,n}, p_{2,n}))_{n\in\mathbb{N}}$ generated by Algorithm (V.6) converge weakly to $(\widetilde{x}, \widetilde{u}) \in \operatorname{zer}(\mathscr{M} + \widetilde{Q})$.

Proof: Under the considered assumptions, $\widetilde{Q}$ is monotone and ${}^{\varepsilon}\vartheta$-Lipschitzian. Thus, the result follows from standard conditions for the convergence of Tseng's algorithm applied to (V.7). □ We now characterize the set of limit points.

Proposition 5.3: Let ${}^{\varepsilon}g$ be the Moreau envelope of g of parameter $\varepsilon > 0$ defined as

$$(\forall v \in \mathcal{L}) \quad {}^{\varepsilon}g(v) = \inf_{w\in\mathcal{L}} g(w) + \frac{1}{2\varepsilon}\|w - v\|_{\mathcal{L}}^2. \tag{V.22}$$

Let $(\widetilde{x}, \widetilde{u}) \in \mathcal{Z}$. Then $(\widetilde{x}, \widetilde{u}) \in \operatorname{zer}(\mathscr{M} + \widetilde{Q})$ if and only if $(\widetilde{x}, \widetilde{u})$ belongs to

$${}^{\varepsilon}\overline{\mathcal{F}} = \left\{(x, u) \in \mathcal{Z} \mid \overline{K}y \in \partial f(x) + Lx + V^*u,\ u = \nabla\,{}^{\varepsilon}g(Dx) = \frac{Dx - \operatorname{prox}_{\varepsilon g}(Dx)}{\varepsilon}\right\}, \tag{V.23}$$

which is nonempty if $L + \partial f + V^* \circ \nabla\,{}^{\varepsilon}g \circ D$ is surjective.

(i) If ${}^{\varepsilon}\lambda_{\min}$ defined by (V.10) is nonnegative, then ${}^{\varepsilon}\overline{\mathcal{F}}$ is closed and convex.

(ii) Let ${}^{\varepsilon}\overline{\mathcal{F}}_1 = \{x \in \mathcal{H} \mid (x, \nabla {}^{\varepsilon}g(Dx)) \in {}^{\varepsilon}\overline{\mathcal{F}}\}$ and let

$$ {}^{\varepsilon}\lambda_{1,\min} = \lambda_{\min} - \frac{1}{\varepsilon}\|V - D\|_{\mathcal{H},\mathcal{L}}\|D\|_{\mathcal{H},\mathcal{L}} \geq 0. \tag{V.24}$$

${}^{\varepsilon}\overline{\mathcal{F}}_1$ has at most one element if one of the following conditions holds:

(a) $L + \partial f + V^* \circ \nabla {}^{\varepsilon}g \circ D$ is strictly monotone.

(b) ${}^{\varepsilon}\lambda_{1,\min} \geq 0$ and ${}^{\varepsilon}g \circ D + f$ is strictly convex.

$\overline{\mathcal{F}}_1$ is a singleton if one of the following conditions holds:

(c) $L + \partial f + V^* \circ \nabla {}^{\varepsilon}g \circ D$ is strongly monotone.

(d) ${}^{\varepsilon}\lambda_{1,\min} > 0$

(e) ${}^{\varepsilon}\lambda_{1,\min} \geq 0$, and f is strongly convex or [g^* is Lipschitz-differentiable and D^*D is strongly positive].

Proof: The proof follows the same lines as in the proof of Theorem 3.5. In the following, we point out the main differences. By using (III.9) and (V.8), we have

$$\begin{aligned} &(\widetilde{x}, \widetilde{u}) \in \operatorname{zer}(\mathscr{M} + \widetilde{Q}) \\ \Leftrightarrow &\begin{cases} 0 \in \partial f(\widetilde{x}) + L\widetilde{x} + V^*\widetilde{u} - \overline{K}y \\ 0 \in \partial g^*(\widetilde{u}) - D\widetilde{x} + \varepsilon\widetilde{u}. \end{cases} \end{aligned} \tag{V.25}$$

We know that $({}^{\varepsilon}g)^* = g^* + \frac{\varepsilon}{2}\|\cdot\|^2 \Rightarrow \partial({}^{\varepsilon}g)^*(\widetilde{u}) = \partial g^*(\widetilde{u}) + \varepsilon\widetilde{u}$ [4, Proposition 14.1] and ${}^{\varepsilon}g$ is differentiable with gradient $\nabla {}^{\varepsilon}g = \varepsilon^{-1}(\mathrm{Id}_{\mathcal{L}} - \operatorname{prox}_{\varepsilon g})$ [4, Proposition 12.30]. We thus deduce that

$$\begin{aligned} &(\widetilde{x}, \widetilde{u}) \in \operatorname{zer}(\mathscr{M} + \widetilde{Q}) \\ \Leftrightarrow &\begin{cases} \overline{K}y \in \partial f(\widetilde{x}) + L\widetilde{x} + V^*\widetilde{u} \\ D\widetilde{x} \in \partial({}^{\varepsilon}g)^*(\widetilde{u}) \end{cases} \\ \Leftrightarrow &\begin{cases} \overline{K}y \in \partial f(\widetilde{x}) + L\widetilde{x} + V^*\widetilde{u} \\ \widetilde{u} = \nabla {}^{\varepsilon}g(D\widetilde{x}). \end{cases} \end{aligned} \tag{V.26}$$

This shows that $(\widetilde{x}, \widetilde{u}) \in {}^{\varepsilon}\overline{\mathcal{F}}$.

(i) According to Theorem 5.1(i)-(ii), if ${}^{\varepsilon}\lambda_{\min} \geq 0$, $\widetilde{Q}$ is monotone and continuous. Since $\mathscr{M}$ is maximally monotone, $\mathscr{M} + \widetilde{Q}$ is also maximally monotone and $\operatorname{zer}(\mathscr{M} + \widetilde{Q})$ is closed and convex.

(ii) We can perform the decomposition

$$L + \partial f + V^* \circ \nabla {}^{\varepsilon}g \circ D = \partial f + D^* \circ \nabla {}^{\varepsilon}g \circ D + L + (V - D)^* \circ \nabla {}^{\varepsilon}g \circ D. \tag{V.27}$$

Subdifferential $\partial f + D^* \circ \nabla {}^{\varepsilon}g \circ D = \partial(f + {}^{\varepsilon}g \circ D)$ is maximally monotone. Using the Cauchy-Schwarz inequality, for every $(x, x') \in \mathcal{H}^2$,

$$\begin{aligned} &\langle L(x - x'), x - x'\rangle_{\mathcal{H}} + \langle (V - D)^*\nabla {}^{\varepsilon}g(Dx) - (V - D)^*\nabla {}^{\varepsilon}g(Dx'), x - x'\rangle_{\mathcal{H}} \\ &\geq \lambda_{\min}\|x - x'\|_{\mathcal{H}}^2 - \big\|(V - D)^*\big(\nabla {}^{\varepsilon}g(Dx) - \nabla {}^{\varepsilon}g(Dx')\big)\big\|_{\mathcal{H}}\|x - x'\|_{\mathcal{H}} \\ &\geq \lambda_{\min}\|x - x'\|_{\mathcal{H}}^2 - \|V - D\|_{\mathcal{H},\mathcal{L}}\|\nabla {}^{\varepsilon}g(Dx) - \nabla {}^{\varepsilon}g(Dx')\|_{\mathcal{L}}\|x - x'\|_{\mathcal{H}} \\ &\geq \lambda_{\min}\|x - x'\|_{\mathcal{H}}^2 - \frac{1}{\varepsilon}\|V - D\|_{\mathcal{H},\mathcal{L}}\|D(x - x')\|_{\mathcal{L}}\|x - x'\|_{\mathcal{H}} \\ &\geq {}^{\varepsilon}\lambda_{1,\min}\|x - x'\|_{\mathcal{H}}^2, \end{aligned} \tag{V.28}$$

where we have used the ε^{-1}-Lipschitz continuity of $\nabla {}^{\varepsilon}g$. This shows that $L + (V - D)^* \circ \nabla {}^{\varepsilon}g \circ D$ is monotone, if ${}^{\varepsilon}\lambda_{1,\min} \geq 0$. In addition, it is strongly monotone (hence, strictly monotone) if ${}^{\varepsilon}\lambda_{1,\min} > 0$. Since this operator is also continuous, it is maximally monotone in both cases. The rest of the proof is similar to that of Theorem 3.5, by noticing that ${}^{\varepsilon}g$ is strongly convex $\Leftrightarrow$ $({}^{\varepsilon}g)^*$ is Lipschitz-differentiable $\Leftrightarrow$ g^* is Lipschitz-differentiable.

□

We now provide a bound on the distance between an optimal pair of solutions $(\widehat{x}, \widehat{u})$ to problem (II.1)-(II.4) and $(\widetilde{x}, \widetilde{u}) \in {}^{\varepsilon}\overline{\mathcal{F}}$.

Proposition 5.4: Let ${}^{\varepsilon}\lambda_{\min}$ be defined by (V.10). Assume that ${}^{\varepsilon}\lambda_{\min} > 0$, f is strongly convex with modulus $\rho > 0$, and g is Lipschitz-differentiable with constant $\beta > 0$. Then, there exists a unique pair $\widetilde{z} = (\widetilde{x}, \widetilde{u}) \in {}^{\varepsilon}\overline{\mathcal{F}}$ and a unique solution $(\widehat{x}, \widehat{u})$ to the primal-dual problem. In addition,

$$\sqrt{\|\widetilde{x} - \widehat{x}\|_{\mathcal{H}}^2 + \|\widetilde{u} - \widehat{u}\|_{\mathcal{L}}^2} \leq \frac{1}{\mu}\left(\|(\overline{K} - H^*)(H\widehat{x} - y)\|_{\mathcal{H}} + \sqrt{\|V - D\|_{\mathcal{H},\mathcal{L}}^2 + \varepsilon^2}\,\|\widehat{u}\|_{\mathcal{L}}\right), \tag{V.29}$$

where $\mu = \min\{\rho, 1/\beta\}$.

Proof: f is ρ-strongly convex and g is β-Lispchitz differentiable (i.e., g^* is β^{-1}-strongly convex), ∂f and ∂g^* are strongly monotone with constants ρ and $1/\beta$, respectively. $\mathscr{M}$ is thus strongly monotone with constant μ. Since Q is continuous and monotone, $\mathscr{M}+Q$ is maximally monotone and strongly monotone. The existence of a unique zero $\widehat{z}$ to $\mathscr{M}+Q$ is thus guaranteed by [4, Corollary 23.37]. Similarly, it follows from Theorem 5.1(i) and Theorem 5.1(iii) that $\widetilde{Q}$ is continuous and strongly monotone. Hence $\mathscr{M}+\widetilde{Q}$ is maximally monotone and strongly monotone and $\mathrm{zer}(\mathscr{M}+\widetilde{Q})$ is a singleton $\{\widetilde{z}\}$.
For every $\gamma>0$,

$$\widehat{z}\in\mathrm{zer}(\mathscr{M}+Q) \quad\Leftrightarrow\quad \widehat{z}=J_{\gamma\mathscr{M}}(\widehat{z}-\gamma Q\widehat{z}) \tag{V.30}$$

$$\widetilde{z}\in\mathrm{zer}(\mathscr{M}+\widetilde{Q}) \quad\Leftrightarrow\quad \widetilde{z}=J_{\gamma\mathscr{M}}(\widetilde{z}-\gamma\widetilde{Q}\widetilde{z}) \tag{V.31}$$

Since $\mathscr{M}$ is strongly monotone with constant μ, $J_{\gamma\mathscr{M}}$ is Lipschitz continuous with constant $1/(1+\gamma\mu)$ [4, Proposition 23.13]. From (V.30) and (V.31), we deduce that

$$\begin{aligned}\|\widetilde{z}-\widehat{z}\|_{\mathcal{Z}} &\leq \frac{1}{1+\gamma\mu}\|\widetilde{z}-\widehat{z}-\gamma(\widetilde{Q}\widetilde{z}-Q\widehat{z})\|_{\mathcal{Z}}\\ &\leq \frac{1}{1+\gamma\mu}\|(\mathrm{Id}_{\mathcal{Z}}-\gamma\widetilde{Q})(\widetilde{z}-\widehat{z})-\gamma(\widetilde{Q}-Q)\widehat{z}\|_{\mathcal{Z}}\\ &\leq \frac{1}{1+\gamma\mu}\left(\|\mathrm{Id}_{\mathcal{Z}}-\gamma\widetilde{Q}\|_{\mathcal{Z},\mathcal{Z}}\|\widetilde{z}-\widehat{z}\|_{\mathcal{Z}}+\gamma\|(\widetilde{Q}-Q)\widehat{z}\|_{\mathcal{Z}}\right).\end{aligned} \tag{V.32}$$

According to Theorem 5.1(iii), $\widetilde{Q}$ is cocoercive with constant ${}^{\varepsilon}\eta$ given by (V.21). Therefore, by assuming that $\gamma\in]0,{}^{\varepsilon}\eta]$, we have $\|\mathrm{Id}_{\mathcal{Z}}-\gamma\widetilde{Q}\|_{\mathcal{Z},\mathcal{Z}}\leq 1$. We deduce from (V.32) that

$$\begin{aligned}\|\widetilde{z}-\widehat{z}\|_{\mathcal{Z}} &\leq \frac{1}{\mu}\|(\widetilde{Q}-Q)\widehat{z}\|_{\mathcal{Z}}\\ &= \frac{1}{\mu}\left\|\left((\overline{K}-H^*)(H\widehat{x}-y)+(V-D)^*\widehat{u},\varepsilon\widehat{u}\right)\right\|_{\mathcal{Z}}\\ &\leq \frac{1}{\mu}\left(\left\|\left((\overline{K}-H^*)(H\widehat{x}-y),0\right)\right\|_{\mathcal{Z}}+\left\|\left((V-D)^*\widehat{u},\varepsilon\widehat{u}\right)\right\|_{\mathcal{Z}}\right)\\ &\leq \frac{1}{\mu}\left(\|(\overline{K}-H^*)(H\widehat{x}-y)\|_{\mathcal{H}}+\sqrt{\|V-D\|^2_{\mathcal{H},\mathcal{L}}+\varepsilon^2}\,\|\widehat{u}\|_{\mathcal{L}}\right).\end{aligned} \tag{V.33}$$

□

Remark 5.5:

(i) In the absence of mismatch on D^* (i.e., $V^*=D^*$), one can choose $\varepsilon=0$ in (V.6) and (V.8). $\mathrm{zer}(\mathscr{M}+\widetilde{Q})$ is then equal to the set $\overline{\mathcal{F}}$ characterized in Theorem 3.5. If $\overline{\mathcal{F}}\neq\varnothing$, $\lambda_{\min}\geq 0$, and $\gamma\in]0,({}^{0}\vartheta)^{-1}[$, then the sequences $((x_n,u_n))_{n\in\mathbb{N}}$ and $((p_{1,n},p_{2,n}))_{n\in\mathbb{N}}$ generated by Algorithm (V.6) converge weakly to $(\widetilde{x},\widetilde{u})\in\overline{\mathcal{F}}$. Theorem 3.9 applies to evaluate the error incurred by the mismatch.

(ii) When $H=\overline{K}=0$, it follows from [39, Theorem 1.2] that, if f is strongly convex with modulus $\rho>0$, $(\widetilde{x},\widetilde{u})\in{}^{\varepsilon}\overline{\mathcal{F}}$ and $\widehat{x}$ is the solution to the primal problem, then

$$\|\widetilde{x}-\widehat{x}\|_{\mathcal{H}}\leq\frac{1}{\rho}\|V-D\|_{\mathcal{H},\mathcal{L}}\|\widetilde{u}\|_{\mathcal{L}}. \tag{V.34}$$

VI. Numerical experiments

This section illustrates the applicability of our theoretical results to the resolution of 2D image reconstruction problems arising in Computed Tomography (CT). All the simulations presented in this section are performed using the ASTRA Toolbox [54], [53] implemented in Matlab.

A. Example 1: reconstruction from few CT views

We aim at recovering an image $\overline{x}$ with N pixels, reshaped as a vector in the Euclidean space $\mathcal{H}=\mathbb{R}^N$. A set of noisy tomographic projections $p\in\mathcal{G}=\mathbb{R}^S$ of the original image $\overline{x}$ is available, according to the following observation model:

$$p=R\overline{x}+b \tag{VI.1}$$

where $R\in\mathbb{R}^{S\times N}$ is the forward operator called projector, and b is an additive i.i.d. zero-mean Gaussian noise. The projector is a discrete implementation of the Radon transform, chosen here as the line-length ray-driven projector [56]. An approximate adjoint of R, denoted by $B\in\mathbb{R}^{N\times S}$, is derived from an alternative discretization model. In our simulations, B is the pixel-driven backprojector which is particularly suited for a GPU implementation compared to the adjoint of R [32]. For (u,v) i.i.d. uniformly sampled on $([0,1]^N)^2$, the average over 20 realizations of the ratio $\langle Ru\mid v\rangle/\langle u\mid Bv\rangle$ is 1.005.
Our goal is to retrieve an estimate of $\overline{x}$ given p, the projector R, and its surrogate adjoint B.

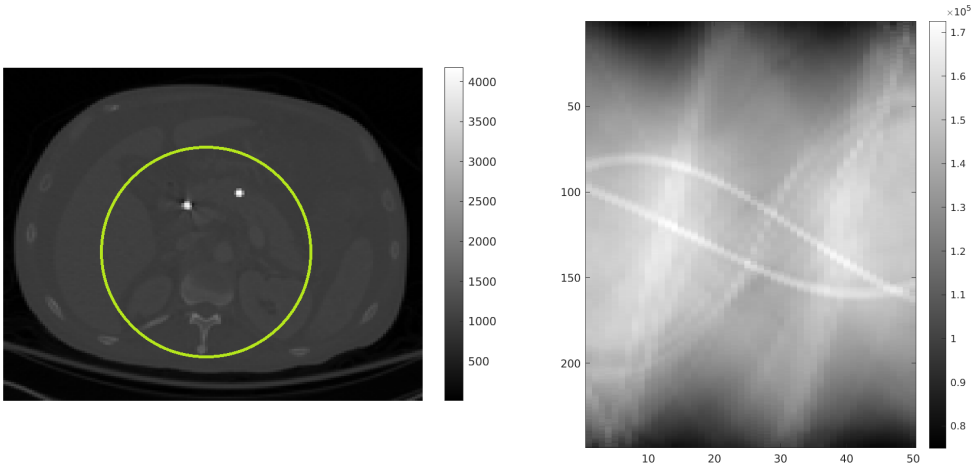

Fig. 1: Original image $\overline{x}$ with highlighted FOV (left) and observed projections p (right).

1) Data: Let us specify the data and settings used for this experiment. In model (VI.1), $\overline{x}$ is an axial slice of an abdomen of size 41 cm whose values range from 1000 to 3000, containing intense inserts with pixel intensity ranges in $[3500, 4200]$. These values correspond to positive Hounsfield Unit (HU) (air is 0 HU and water is 1000 HU). The source-to-object distance is 800 mm, and the source-to-image distance is 1200 mm, which corresponds to a magnification factor of 1.5, typically encountered in clinical scanners. The projector R describes a fan-beam geometry over 180° using 50 regularly spaced angular steps. The detector has a length of 40 cm. The bin grid is twice upsampled with respect to the pixel grid: the detector has 250 bins of size 1.6 mm, so that $S = 250 \times 250$. The image is reconstructed on a grid of $N = 160 \times 160$ pixels, with size $2 \times 1.6/1.5 = 2.13$ mm. Data p is obtained after adding 1% relative Gaussian noise on $R\overline{x}$. The inverse problem (VI.1) is highly ill-posed because of the small detector field of view (FOV) and the limited angular density. This problem corresponds to a common setup in CT when the detector is not large enough to measure the projections of large body parts such as the abdomen, as it happens in image guidance for interventional radiology. The set of pixels in the image whose projections belong to the detector FOV defines the so-called FOV image. Because of truncation, the exterior of the FOV has to be estimated for the reconstruction of the FOV to be accurate. The size of the reconstruction grid is thus slightly larger than the support of the FOV. Figure 1 shows the original image $\overline{x}$ with its FOV and the observed data p. Due to the use of a short detector, the projections suffer from axial truncation, as shown by non-zero values at the border of some vertical lines of the sinogram.

2) Regularization: We provide an estimate of $\overline{x}$ by solving the following penalized least squares problem:

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \frac{1}{2}\|p - Rx\|^2 + f(x) + g(Dx) + \frac{\kappa}{2}\|x\|^2 \tag{VI.2}$$

with $\kappa \in [0, +\infty[$. Functions f and $g \circ D$ serve to enforce desired properties on the sought image. Here, we promote sparsity of the image vertical and horizontal gradients [13]. We additionally constraint the nonnegativity of the reconstructed pixel intensities. This leads us to set $f = \iota_{[0,+\infty[^N}$ where ι_C denotes the indicator function of a set C. Moreover, we set $g = \xi\|\cdot\|_{1,2}$ with $D = \nabla = \begin{bmatrix}\nabla^{\mathrm{h}} \\ \nabla^{\mathrm{v}}\end{bmatrix}$, where $\nabla^{\mathrm{h}} \in \mathbb{R}^{N\times N}$, $\nabla^{\mathrm{v}} \in \mathbb{R}^{N\times N}$ are, respectively, the horizontal and vertical discrete gradient operators (assuming zero-padding) and $\|\cdot\|_{1,2}$ is the $\ell_{1,2}$-norm of $\mathbb{R}^N$, so that $g \circ D$ is the discrete total variation penalty weighted by $\xi \in [0, +\infty[$ [47]. We set the regularization hyperparameter ξ to 800 through a grid search, minimizing the FOV image error.

3) Condat-Vũ algorithm: Problem (VI.2) can be rewritten as (II.1) with $H = \begin{bmatrix} R \\ \sqrt{\kappa}\mathrm{Id}_{\mathbb{R}^N}\end{bmatrix}$, $\kappa \in [0, +\infty[$, and $y = \begin{bmatrix} p \\ 0\end{bmatrix}$. The surrogate adjoint of H is $\overline{K} = \begin{bmatrix} B & \sqrt{\kappa}\mathrm{Id}_{\mathbb{R}^N}\end{bmatrix}$. For such a problem, we can apply the CV approach presented in section III. We run Algorithms (III.3) and (III.4) (i.e., CV algorithm without/with an adjoint mismatch, respectively) for $\kappa \in \{\kappa_1, \kappa_2\} \subset]0, +\infty[$ where $L = \overline{K}H = BR + \kappa\mathrm{Id}_{\mathbb{R}^N}$ is only cocoercive for $\kappa = \kappa_2$. In the latter case, the condition given in Theorem 3.5(d) holds, which proves the existence of a unique fixed point $\widetilde{x}$ of scheme (III.4) and its convergence is ensured according to Theorem 3.8. In contrast, nothing can be said about the convergence of the scheme in the case involving κ_1. We set $\kappa_1 = 1$ and κ_2 to $-\widetilde{\lambda}_{\min} + 0.01$ where $\widetilde{\lambda}_{\min} = -28.24$ is the minimum spectral value of $(BR + R^*B^*)/2$ estimated from the power iterative method. The cocoercivity constant η is computed using Theorem 3.4(ii).

The convergence parameter σ is set to 10^{-3}. For Algorithm (III.3), the step-size τ is set to $0.99/(8\sigma + 0.5/\theta)$ with $\theta = 1/\|H\|^2_{\mathcal{H},\mathcal{G}}$. To illustrate the instabilities incurred by the use of the mismatched adjoint $\overline{K}$ when using κ_1, the same step-size value τ is used as in the matched case. With κ_2, the convergence of Algorithm (III.4) is ensured, as stated by Theorem 3.8

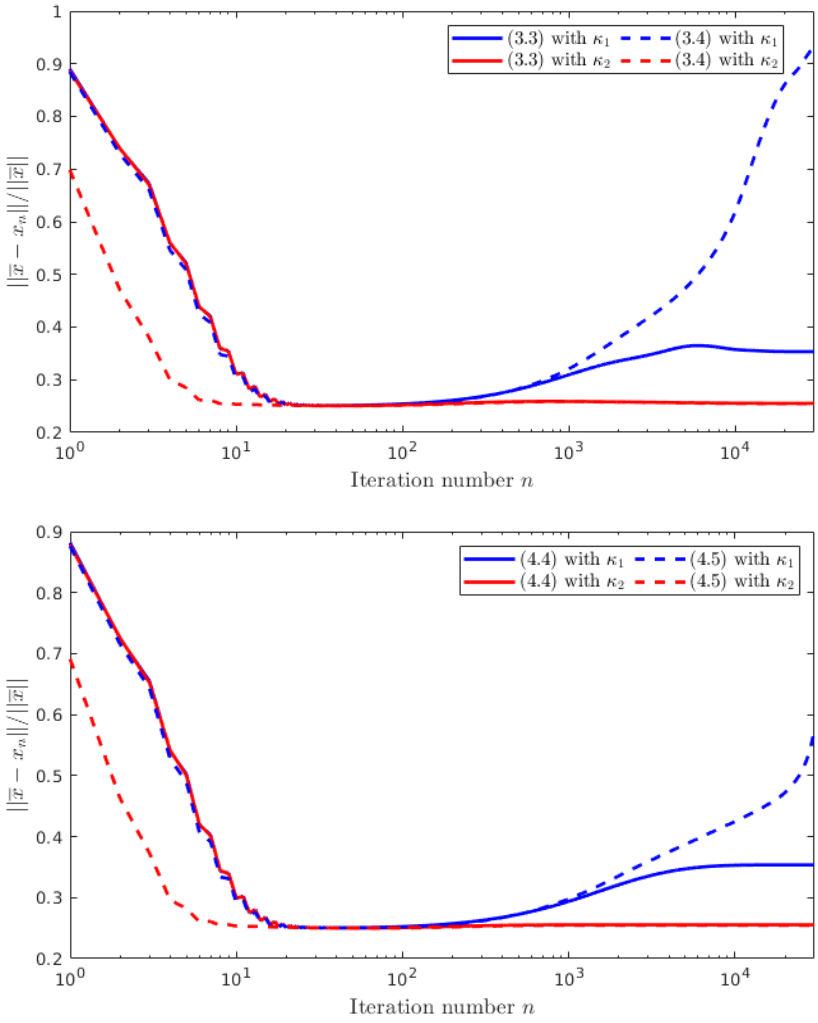

Fig. 2: Evolution of the error $(\|\overline{x}-x_n\|/\|\overline{x}\|)_n$ along iterations for Algorithms (III.3)-(III.4) (top) and Algorithms (IV.4)-(IV.5) (bottom), for two settings of parameter κ.

and Theorem 3.4, by setting $\tau = 0.99/(8\sigma + \|M\|^2_{\mathcal{H},\mathcal{H}}/4)$, where M is defined in (III.9). Both algorithms are run until a maximum number 3×10^4 of iterations is reached. Initial iterates x_0 and u_0 are set to zero.

4) Loris-Verhoeven algorithm: Problem (VI.2) can also be solved with the LV approach presented in section IV. More precisely, the cost function can be rewritten as in (II.1) by setting $H = R$, $y = p$, $D = \begin{bmatrix} \nabla \\ \mathrm{Id}_{\mathbb{R}^N} \end{bmatrix}$, $f = \frac{\kappa}{2}\|\cdot\|^2$, and

$$\big(\forall (z_1, z_2) \in (\mathbb{R}^N)^2\big) \quad g\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) = \|z_1\|_{1,2} + \iota_{[0,+\infty[^N}(z_2). \tag{VI.3}$$

Similarly to the CV case, $L = BR + \kappa \mathrm{Id}_{\mathbb{R}^N}$. Therefore, for the mismatched LV algorithm (IV.5), the existence and uniqueness of a fixed point $\widetilde{x}$ in $\overline{\mathcal{F}}_1$, defined in Theorem 4.1(ii)(c), is only guaranteed when $\kappa = \kappa_2$, but not when $\kappa = \kappa_1$. The convergence parameter σ is here set as $1.99/(9\tau)$. For Algorithm (IV.4) with both values of κ and Algorithm (IV.5) with κ_1, step-size τ is set to $1.99/(\|H\|^2_{\mathcal{H},\mathcal{G}} + \kappa)$. The convergence of Algorithm (IV.5) with κ_2 is ensured by setting τ to $3.99/\|M\|^2_{\mathcal{H},\mathcal{H}}$, where M is the same as in the CV case, in accordance with Theorem 4.2.

5) Results: Figure 2 displays the normalized root mean square error (NMSE) defined as $(\|\overline{x}-x_n\|/\|\overline{x}\|)_n$, computed along the iterations when applying CV Algorithms (III.3)-(III.4) and LV Algorithms (IV.4)-(IV.5). We recall that Algorithms (III.3) and (IV.4) require the use of the exact adjoint of H. The plots confirm that, with value κ_1, both CV and LV algorithms converge when this exact adjoint H^* is used, but diverge when H^* is replaced by $\overline{K}$, as was expected from our theoretical analysis. In the latter case, CV and LV algorithms show an initial convergence trend before diverging. We notice that on this example, the mismatched CV (III.4) diverges faster than the mismatched LV (IV.5). When reaching the maximal number of iterations, the NMSE associated to (III.4) is 0.93 whereas the NMSE associated to (IV.5) is 0.57.
When using κ_2, all algorithms converge to close fixed points, as expected by our theory. The corresponding NMSE values are

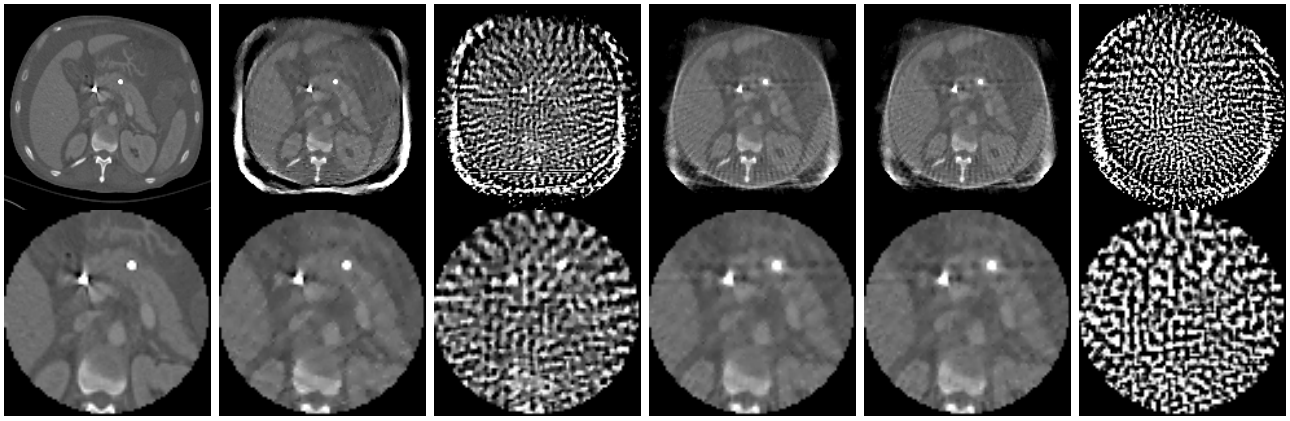

Fig. 3: Reconstructed images (top) and zoomed FOVs (bottom). From left to right: $\overline{x}$, reconstructions obtained using (III.3) with κ_1, (III.4) with κ_1, (III.3) with κ_2, (III.4) with κ_2, (IV.5) with κ_1.

0.251 for (III.3), 0.242 for (III.4), 0.253 for (IV.4) and 0.252 for (IV.5). Remarkably, our mismatched algorithms (III.4)-(IV.5) with κ_2 lead to lower reconstruction error in the first iterations of the algorithms.
Reconstructed images and their FOVs are displayed in Figure 3 using the same windowing. Note that the reconstructions obtained using (IV.4) look the same as those obtained with (III.3). Likewise, the same reconstruction is obtained with (III.4) and (IV.5), when κ_2 is used. For all the reconstructed images, we also provide the NMSE and the maximum absolute error (MAE) computed in the FOV image, defined as $\max_{i\in\{1,\dots,N\}} |[\mathrm{mask}_{\mathrm{FOV}}(\overline{x}-x)]_i|$ where $[\mathrm{mask}_{\mathrm{FOV}}(x_n)]_i = [x_n]_i$ if pixel i of x_n is in the FOV, and $[\mathrm{mask}_{\mathrm{FOV}}(x_n)]_i = 0$ otherwise. When parameter κ_2 is used, the reconstructed image obtained by CV/LV with the mismatched adjoint $\overline{K}$ (MAE=461, NMSE=0.040) is very similar to the image obtained without mismatch (MAE=495, NMSE=0.041). In contrast, combining the setting κ_1 with the mismatched adjoint yields reconstructions that are highly deteriorated by high-frequency patterns leading to a higher error (MAE=2735, NMSE=0.637 for CV and MAE=1249, NMSE=0.249 for LV) compared to the solution obtained when using the exact adjoint (MAE=215, NMSE=0.026 for both CV and LV).

B. Example 2: reconstruction from Poisson data

In this second example, we focus on an another acquisition scenario, namely we consider the case when the projection data p contain photon count views which follow a Poisson distribution:

$$p = \mathcal{P}(R\overline{x}), \tag{VI.4}$$

that each component p_s of $p \in \mathbb{R}^S$, $s \in \{1,\dots,S\}$, is drawn independently from a Poisson distribution with mean $[R\overline{x}]_s$. The goal is again to restore an estimate of $\overline{x}$ given p, R, and its mismatched adjoint B.

1) Data: This second test problem uses a fan-beam geometry with 200 views. In model (VI.4), $\overline{x}$ is an axial slice of an abdomen. Contrary to Example 1, it is now made only of an anatomical background (see Figure 4). We set the source-to-object distance, the source-to-image distance, the detector length, the number of bins, and the bin size as in Example 1. Hence, $S = 250 \times 200$. Projections p are then simulated by using model (VI.4). The image is reconstructed on a discrete grid of $N = 220 \times 220$ pixels, with size 2.13 mm. The mismatched backprojector B is derived from the same discretization scheme as B in Example 1.

2) Regularization: Due to the presence of Poisson noise, the data discrepancy term in the cost function differs from the one in Example 1. Namely, we make use of the negative log-likelihood of the image given the data [37], to define $\ell \circ R$ with

$$(\forall z = (z_s)_{1\le s\le S} \in \mathbb{R}^S) \qquad \ell(z) = \sum_{s=1}^{S} \mathcal{KL}(z_s, p_s) \tag{VI.5}$$

and

$$(\forall (u,v) \in \mathbb{R}^2) \qquad \mathcal{KL}(u,v) = \begin{cases} -v\log u + u & \text{if } u>0, v>0 \\ u & \text{if } u\ge 0, v=0 \\ +\infty & \text{otherwise.} \end{cases} \tag{VI.6}$$

Furthermore, we introduce a nonnegativity constraint on the components of the solution and a Tikhonov-based penalty. Altogether, the resulting minimization problem reads

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \chi\,\ell(Rx) + \frac{1}{2}\|Hx\|^2 + \iota_{[0,+\infty[^N}(x) \tag{VI.7}$$

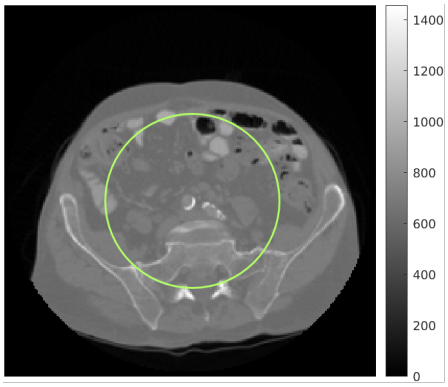

Fig. 4: Phantom $\overline{x}$ with highlighted FOV

where $\chi \in [0, +\infty[$ weights the Poisson data fidelity term and we define the linear operator $H = \begin{bmatrix} \Delta \\ \kappa \mathrm{Id}_{\mathbb{R}^N} \end{bmatrix}$. Moreover, $\Delta \in \mathbb{R}^{N \times N}$ refers to the 2D discrete Laplacian operator (here, implemented in the 2D Fourier domain), and $\kappa \in [0, +\infty[$. Problem (VI.7) can be rewritten in the form (II.1) with $y = 0_{\mathbb{R}^{2N}}$, $D = R$, $g = \chi\, \ell$, and $f = \iota_{[0,+\infty[^N}$.

3) Combettes-Pesquet algorithm: Problem (VI.7) is solved with CP algorithm (V.1)-(V.6). In our configuration, an adjoint mismatch only arises on D (i.e., the projector R), and again denoting by B the mismatched adjoint. We hence have $V = B^*$ and $\overline{K} = H^*$ in Algorithm (V.6).

To guarantee the convergence of Algorithm (V.6) to a unique fixed point, we must choose (κ, ε) to satisfy the conditions of Theorem 5.1 and Theorem 5.3. We first set κ and then choose $4\varepsilon = \|V - D\|^2_{\mathcal{H},\mathcal{L}}/(\kappa^2 - 0.02)$ so that ${}^{\varepsilon}\lambda_{\min} = 0.02 > 0$ in (V.10). In particular, we consider the setting $(\kappa_2, \varepsilon_2) = (3.3, 6.6)$ for which convergence is guaranteed and the setting $(\kappa_1, \varepsilon_1) = (0.6, 0)$ for which it is not guaranteed.

In Algorithm (V.1), parameter γ is set to $0.99/(4 + \kappa^2 + \|D\|_{\mathcal{H},\mathcal{L}})$. When Algorithm (V.6) is run with κ_1, γ is set as in the matched case. When Algorithm (V.6) is run with κ_2 and ε_2, we ensure its convergence using Theorem 5.2 by setting $\gamma = 0.99/{}^{\varepsilon}\vartheta$ with ${}^{\varepsilon}\vartheta = \min({}^{\varepsilon}\vartheta_1, {}^{\varepsilon}\vartheta_2) = {}^{\varepsilon}\vartheta_1$ where ${}^{\varepsilon}\vartheta_1$ and ${}^{\varepsilon}\vartheta_2$ are defined respectively by (V.11) and (V.12). We set the data fidelity parameter χ to 5000, and we perform 4000 iterations of CP algorithm. Similarly to Example 1, the initial iterates x_0 and u_0 are set to zero.

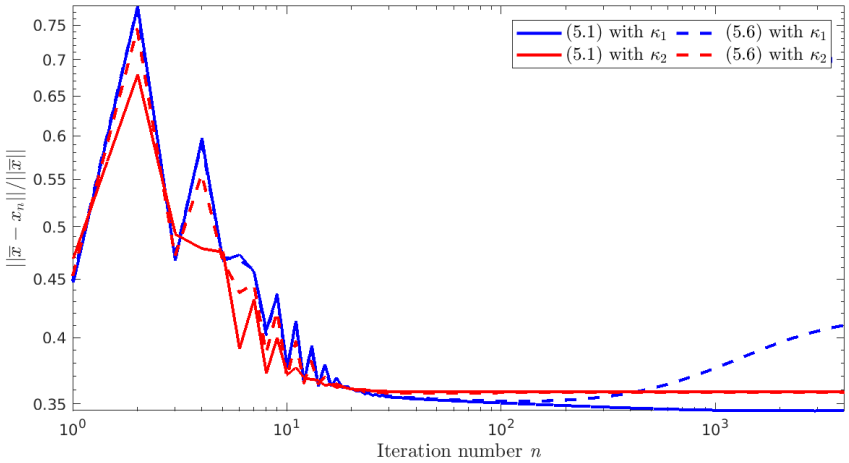

Fig. 5: Evolution of the error $\|\overline{x} - x_n\|/\|\overline{x}\|$ along iterations for CP Algorithms (V.1) and (V.6), for two settings of parameter κ.

4) Results: In Figure 5, we plot the relative error between the ground truth $\overline{x}$ and the estimate along the iterations. We observe the same behavior as in our previous example. Algorithm (V.6) with $(\kappa_1, \varepsilon_1)$ diverges quickly, while it converges to a fixed point with $(\kappa_2, \varepsilon_2)$. This fixed point is indistinguishable from the minimizer of (VI.7) with $\kappa = \kappa_2$, both in terms of NMSE/MAE and visual inspection (see Figure 6).

VII. Conclusion

We thoroughly analyzed a set of classical primal-dual proximal splitting algorithms when the adjoints of the involved linear operators are approximated. Our study strongly relies on fixed-point theory tools. We established necessary conditions to ensure

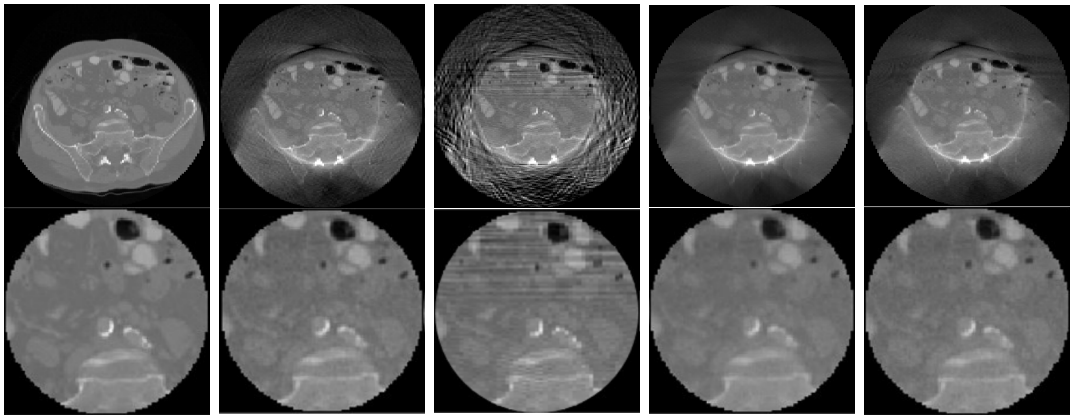

Fig. 6: Reconstructed images (top) and zoomed FOVs (bottom). From left to right: $\overline{x}$, reconstructed images using (V.1) with κ_1 (NMSE = 0.052, MAE = 201), (V.6) with κ_1 (NMSE = 0.079, MAE = 401), (V.1) with κ_2 (NMSE = 0.056, MAE = 232), (V.6) with κ_2 (NMSE = 0.055, MAE = 206).

the convergence of these unmatched schemes when applied to non-smooth convex penalized least-squares problems. We showed that unmatched algorithms solve more general monotone inclusions than those derived from the original minimization problem. We illustrated our results through two numerical examples of CT image reconstruction when the adjoint of the projector is inexact. Both a quadratic and a more sophisticated Poisson fidelity term have been considered in our experiments. In both cases, we showed that convergence can still be guaranteed for an unmatched projector-backprojector pair. It would be interesting to extend our analysis to algorithms for finding a zero of a sum of more than two maximally monotone operators [21]. Handling more general form of cost functions would also be of further interest in future research.

REFERENCES

[1] A. ALOTAIBI, P. L. COMBETTES, AND N. SHAHZAD, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn–Tucker set*, SIAM Journal on Optimization, 24 (2014), pp. 2076–2095, https://doi.org/10.1137/130950616.

[2] S. ARRIDGE, P. MAASS, O. ÖKTEM, AND C. SCHÖNLIEB, *Solving inverse problems using data-driven models*, Acta Numerica, 28 (2019), pp. 1 – 174.

[3] S. BANERT, J. RUDZUSIKA, O. ÖKTEM, AND J. ADLER, *Accelerated forward-backward optimization using deep learning*, 2021, https://doi.org/10.48550/arxiv.2105.05210.

[4] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis And Monotone Operator Theory In Hilbert Spaces*, Springer, New York, 2nd ed., 2017, https://hal.sorbonne-universite.fr/hal-01517477.

[5] S. BOYD, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning, 3 (2010), pp. 1–122, https://doi.org/10.1561/2200000016.

[6] R. I. BOŢ, E. R. CSETNEK, AND A. HEINRICH, *A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators*, SIAM Journal on Optimization, 23 (2013), pp. 2011–2036, https://doi.org/10.1137/12088255X.

[7] R. I. BOŢ AND C. HENDRICH, *Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization*, Journal of Mathematical Imaging and Vision, 49 (2014), pp. 551–568, https://doi.org/10.1007/s10851-013-0486-8.

[8] L. BRICEÑO-ARIAS AND S. LÓPEZ, *A projected primal–dual method for solving constrained monotone inclusions*, Journal of Optimization Theory and Applications, 180 (2019), https://doi.org/10.1007/s10957-018-1430-2.

[9] L. M. BRICEÑO-ARIAS AND P. L. COMBETTES, *A monotone+skew splitting model for composite monotone inclusions in duality*, SIAM Journal on Optimization, 21 (2011), pp. 1230–1250, https://doi.org/10.1137/10081602X.

[10] T. A. BUBBA, M. GALINIER, M. LASSAS, M. PRATO, L. RATTI, AND S. SILTANEN, *Deep neural networks for inverse problems with pseudodifferential operators: An application to limited-angle tomography*, SIAM Journal on Imaging Sciences, 14 (2021), pp. 470–505, https://doi.org/10.1137/20M1343075.

[11] M. BURGER, A. SAWATZKY, AND G. STEIDL, *First Order Algorithms in Variational Image Processing*, Springer International Publishing, Cham, 2016, pp. 345–407, https://doi.org/10.1007/978-3-319-41589-5_10.

[12] G. T. BUZZARD, S. H. CHAN, S. SREEHARI, AND C. A. BOUMAN, *Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium*, SIAM Journal on Imaging Sciences, 11 (2018), pp. 2001–2020, https://doi.org/10.1137/17M1122451.

[13] E. J. CANDES, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory, 52 (2006), pp. 489–509.

[14] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2010), pp. 120–145.

[15] C. CHAUX, P. L. COMBETTES, J.-C. PESQUET, AND V. R. WAJS, *A variational formulation for frame-based inverse problems*, Inverse Problems, 23 (2007), pp. 1495–1518, https://doi.org/10.1088/0266-5611/23/4/008.

[16] P. CHEN, J. C. HUANG, AND X. ZHANG, *A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration*, Inverse Problems, 29 (2013), p. 025011.

[17] E. CHOUZENOUX, A. JEZIERSKA, J. PESQUET, AND H. TALBOT, *A convex approach for image restoration with exact poisson–gaussian likelihood*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 2662–2682, https://doi.org/10.1137/15M1014395.

[18] E. CHOUZENOUX, J.-C. PESQUET, C. RIDDELL, M. SAVANIER, AND Y. TROUSSET, *Convergence of proximal gradient algorithm in the presence of adjoint mismatch*, Inverse Problems, 37 (2021), http://iopscience.iop.org/article/10.1088/1361-6420/abd85c.

[19] P. COMBETTES AND J.-C. PESQUET, *Proximal Splitting Methods in Signal Processing*, Springer-Verlag, NY, 2010, pp. 185–212.

[20] P. COMBETTES AND J.-C. PESQUET, *Deep neural network structures solving variational inequalities*, Set-Valued and Variational Analysis, 28 (2020), pp. 491–518, https://doi.org/10.1007/s11228-019-00526-z.

[21] P. L. COMBETTES AND J. PESQUET, *Fixed point strategies in data science*, IEEE Transactions on Signal Processing, 69 (2021), pp. 3878–3905.

[22] P. L. COMBETTES AND J.-C. PESQUET, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators*, Set-Valued and Variational Analysis, 20 (2011), pp. 307–330, https://doi.org/10.1007/s11228-011-0191-y.

[23] P. L. COMBETTES AND J.-C. PESQUET, *Proximal Splitting Methods in Signal Processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Bauschke, H. Burachik, R. Combettes, P. Elser, V. Luke, D. Wolkowicz, and H. (Eds.), eds., 2011, pp. 185–212.

[24] P. L. COMBETTES AND V. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal, 4 n°4 (2005), pp. 1164–1200.

[25] L. CONDAT, *A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013), pp. 460–479.

[26] L. CONDAT, D. KITAHARA, A. CONTRERAS, AND A. HIRABAYASHI, *Proximal splitting algorithms: A tour of recent advances, with new twists.*, arXiv: Optimization and Control, (2020).

[27] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1413–1457, https://doi.org/https://doi.org/10.1002/cpa.20042.

[28] Y. DONG, P. C. HANSEN, M. E. HOCHSTENBACH, AND N. A. BROGAARD RIIS, *Fixing nonconvergence of algebraic iterative reconstruction with an unmatched backprojector*, SIAM Journal on Scientific Computing, 41 (2019), pp. A1822–A1839.

[29] Y. DRORI, S. SABACH, AND M. TEBOULLE, *A simple algorithm for a class of nonsmooth convex-concave saddle-point problems*, Oper. Res. Lett., 43 (2015), pp. 209–214, https://doi.org/10.1016/j.orl.2015.02.001.

[30] T. ELFVING AND P. C. HANSEN, *Unmatched projector/backprojector pairs: Perturbation and convergence analysis*, SIAM Journal on Scientific Computing, 40 (2018), pp. A573–A591.

[31] E. ESSER, X. ZHANG, AND T. F. CHAN, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1015–1046, https://doi.org/10.1137/09076934X.

[32] F. XU AND K. MUELLER, *A comparative study of popular interpolation and integration methods for use in computed tomography*, in Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006., Apr 2006, pp. 1252–1255.

[33] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & Mathematics with Applications, 2 (1976), pp. 17–40, https://doi.org/10.1016/0898-1221(76)90003-1.

[34] J. HERTRICH, S. NEUMAYER, AND G. STEIDL, *Convolutional proximal neural networks and Plug-and-Play algorithms*, Linear Algebra and its Applications, 631 (2021), pp. 203–234, https://doi.org/10.1016/j.laa.2021.09.004.

[35] A. C. KAK AND M. SLANEY, *Principles of Computerized Tomographic Imaging*, Society of Industrial and Applied Mathematics, 2001.

[36] N. KOMODAKIS AND J.-C. PESQUET, *Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems*, IEEE Signal Processing Magazine, 32 (2015), pp. 31–54, https://doi.org/10.1109/MSP.2014.2377273.

[37] S. KULLBACK AND R. A. LEIBLER, *On information and sufficiency*, Ann. Math. Statist., 22 (1951), pp. 79–86.

[38] D. A. LORENZ, S. ROSE, AND F. SCHÖPFER, *The randomized Kaczmarz method with mismatched adjoint*, BIT Numerical Mathematics, 58 (2018), pp. 1079–1098.

[39] D. A. LORENZ AND F. SCHNEPPE, *Chambolle-pock's primal-dual method with mismatched adjoint*, 2022, https://doi.org/10.48550/arxiv.2201.04928.

[40] I. LORIS AND C. VERHOEVEN, *On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty*, Inverse Problems, 27 (2011), p. 125007.

[41] J. NUYTS, B. D. MAN, J. A. FESSLER, W. ZBIJEWSKI, AND F. J. BEEKMAN, *Modelling the physics in the iterative reconstruction for transmission computed tomography*, Physics in Medicine and Biology, 58 (2013), pp. R63–R96, https://doi.org/10.1088/0031-9155/58/12/r63.

[42] K. M. PER CHRISTIAN HANSEN, KEN HAYAMI, *Gmres methods for tomographic reconstruction with an unmatched back projector*, tech. report, 2022, https://arxiv.org/pdf/2110.01481.pdf.

[43] J.-C. PESQUET, A. REPETTI, M. TERRIS, AND Y. WIAUX, *Learning maximally monotone operators for image recovery*, SIAM Journal on Imaging Sciences, 14 (2021), pp. 1206–1237, https://doi.org/10.1137/20M1387961.

[44] Z. RAMZI, P. CIUCIU, AND J.-L. STARCK, *Benchmarking proximal methods acceleration enhancements for cs-acquired mr image analysis reconstruction*, in SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations Workshop, 2019, https://arxiv.org/abs/1807.04005.

[45] E. J. REID, L. F. DRUMMY, C. A. BOUMAN, AND G. T. BUZZARD, *Multi-resolution data fusion for super resolution imaging*, IEEE Transactions on Computational Imaging, 8 (2022), pp. 81–95, https://doi.org/10.1109/TCI.2022.3140551.

[46] C. RIDDELL, B. BENDRIEM, M. H. BOURGUIGNON, AND J. P. KERNEVEZ, *The approximate inverse and conjugate gradient: non-symmetrical algorithms for fast attenuation correction in SPECT*, Physics in Medicine and Biology, 40 (1995), pp. 269–281.

[47] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.

[48] H. SADEGHI, S. BANERT, AND P. GISELSSON, *Forward-backward splitting with deviations for monotone inclusions*, 2021, https://doi.org/10.48550/arxiv.2112.00776.

[49] M. SAVANIER, E. CHOUZENOUX, J.-C. PESQUET, AND C. RIDDELL, *Unmatched preconditioning of the proximal gradient algorithm*, tech. report, 2022, https://arxiv.org/abs/http://www.optimization-online.org/DB_FILE/2021/12/8741.pdf.

[50] E. Y. SIDKY, P. C. HANSEN, J. S. JØRGENSEN, AND X. PAN, *Iterative image reconstruction for CT with unmatched projection matrices using the generalized minimal residual algorithm*, tech. report, 2022.

[51] E. Y. SIDKY, J. H. JØRGENSEN, AND X. PAN, *Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm.*, Physics in medicine and biology, 57 10 (2012), pp. 3065–91.

[52] P. TSENG, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM Journal on Control and Optimization, 38 (2000), pp. 431–446, https://doi.org/10.1137/S0363012998338806.

[53] W. VAN AARLE, W. J. PALENSTIJN, J. CANT, E. JANSSENS, F. BLEICHRODT, A. DABRAVOLSKI, J. D. BEENHOUWER, K. J. BATENBURG, AND J. SIJBERS, *Fast and flexible x-ray tomography using the ASTRA toolbox*, Opt. Express, 24 (2016), pp. 25129–25147, https://doi.org/10.1364/OE.24.025129.

[54] W. VAN AARLE, W. J. PALENSTIJN, J. DE BEENHOUWER, T. ALTANTZIS, S. BALS, K. J. BATENBURG, AND J. SIJBERS, *The ASTRA toolbox: A platform for advanced algorithm development in electron tomography*, Ultramicroscopy, 157 (2015), pp. 35–47, https://doi.org/10.1016/j.ultramic.2015.05.002.

[55] B. C. VU, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics, 38 (2013), pp. 667–681, https://doi.org/10.1007/s10444-011-9254-8.

[56] G. ZENG AND G. GULLBERG, *A ray-driven backprojector for backprojection filtering and filtered backprojection algorithms*, in 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, 1993, pp. 1199–1201, https://doi.org/10.1109/NSSMIC.1993.701833.

[57] G. L. ZENG, *Counter examples for unmatched projector/backprojector in an iterative algorithm*, Chinese Journal of Academic Radiology, (2019).

[58] G. L. ZENG AND G. T. GULLBERG, *Unmatched projector/backprojector pairs in an iterative reconstruction algorithm*, IEEE Transactions on Medical Imaging, (2000).