Proximal Gradient Algorithm in the Presence of Adjoint Mismatch

Marion Savanier, Emilie Chouzenoux, Jean-Christophe Pesquet, Cyril Riddell, Yves Trousset

Proximal Gradient Algorithm in the Presence of Adjoint Mismatch

Marion Savanier[1,2], Emilie Chouzenoux[1], Jean-Christophe Pesquet[1], Cyril Riddell[2] and Yves Trousset[2]
(1) Univ. Paris-Saclay, CentraleSupélec, CVN, Inria, Gif-sur-Yvette, France.
(2) GE Healthcare, Buc, France.

***Abstract*—The proximal gradient algorithm is a popular iterative algorithm to deal with penalized least-squares minimization problems. Its simplicity and versatility allow one to embed non-smooth penalties efficiently. In the context of inverse problems arising in signal and image processing, a major concern lies in the computational burden when implementing minimization algorithms. For instance, in tomographic image reconstruction, a bottleneck is the cost for applying the forward linear operator and its adjoint [1], [2]. Consequently, it often happens that these operators are approximated numerically, so that the adjoint property is no longer fulfilled. In this paper, we focus on the proximal gradient algorithm stability properties when such an adjoint mismatch arises. By making use of tools from convex analysis and fixed point theory, we establish conditions under which the algorithm can still converge to a fixed point. We provide bounds on the error between this point and the solution to the minimization problem. We illustrate the applicability of our theoretical results through numerical examples in the context of computed tomography.**

***Index Terms*—Proximal gradient algorithm, adjoint mismatch, convergence analysis, fixed point methods, image reconstruction, computed tomography.**

I. Introduction

Inverse problems appear in various fields of science and engineering. Linear inverse problems stem from observation models of the form:

$$\mathbf{y} = \mathbf{H}\overline{\mathbf{x}} + \mathbf{b}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the vector of observed data, $\overline{\mathbf{x}} \in \mathbb{R}^N$ is the unknown signal, $\mathbf{H} \in \mathbb{R}^{M\times N}$ is a linear operator, and $\mathbf{b} \in \mathbb{R}^M$ is a noise term. The aim is then to recover an estimate of $\overline{\mathbf{x}}$ given $\mathbf{H}$ and $\mathbf{y}$. The fact that $\mathbf{H}$ is usually poorly conditioned makes the inverse problem solution very sensitive to noise. An efficient strategy of resolution is to recast the problem as the minimization of a sum of two terms: a data fidelity term and a penalization term. When the data fidelity term is quadratic, the problem falls into the class of penalized least-squares minimizations. A very simple way for minimizing such a composite cost function is the proximal gradient algorithm (PGA) [3], which is an instance of the forward-backward algorithm [4]. The basic idea behind this scheme is to alternate an explicit step of gradient descent on the data fidelity term with an implicit proximity step on the non necessarily smooth penalization term. The observation matrix $\mathbf{H}$ and its adjoint $\mathbf{H}^\top$ are involved in the computation of the gradient step. In practical implementations of PGA, it may however happen that the adjoint of the forward operator $\mathbf{H}$ is purposefully replaced by an approximation of it, denoted hereafter by $\mathbf{K} \in \mathbb{R}^{N\times M}$.

For instance, approximation of the adjoint operator occurs in large scale tomographic imaging [5], as practiced in industrial non-destructive testing and diagnostic medical imaging [6], [7]. The acquisition pipeline consists in rotating an X-ray source and a detector around an object of interest. The forward operation is named the projector and its adjoint, the backprojector. These operators are acting in a continuous domain, and their discretization is not straightforward because the sampling rate on a Cartesian grid is not constant by rotation. Accuracy is increased when the discretization is performed at a high sampling rate, at the price of a high computation burden. Symmetric projection/backprojection pairs are available but they rely on advanced interpolation schemes and are not GPU-friendly [8], [9]. Thus, in practice, while oversampling in the image domain is performed and leads to an accurate projection operator, the discretization of the backprojector is often simplified, resulting in an adjoint mismatch [10], [11]. Hence, instabilities are expected in the convergence of the reconstruction algorithm, as errors may accumulate over iterations [12].

Recently, some authors have investigated conditions of convergence of specific forms of PGA in the presence of adjoint mismatch and the impact over its asymptotic solution. Among them [13] gave conditions of convergence in the finite dimensional case without regularization, generalizing the earlier work of [5]. They later proposed in [14] a modification of the iterative scheme to ensure convergence despite the mismatch. However, up to our knowledge, those works only focus on the case when the cost function is a least-squares term without any regularization, PGA being reduced to a simple iterative linear scheme. The approach followed in the analysis of [13] could be extended in a direct manner to the case when the regularization term is a quadratic function. However, in the presence of a more generic prior, inducing a nonlinear proximal operator, the convergence of PGA becomes much more challenging to study. It is worth mentioning the work in [15], which also dealt with a least-squares cost function reformulated as a set of constraints onto convex sets that,

This work was supported by the European Research Council Starting Grant MAJORIS ERC-2019-STG-850925 and the ANRT CIFRE Convention 2018/1587.

therefore, does not rely on a proximal gradient-based scheme.

In this paper, we provide the first analysis of PGA in the presence of adjoint mismatch, when solving a penalized least-squares problem. In Section II, we give necessary conditions to preserve the convergence of PGA with an adjoint mismatch. Furthermore, we study how this mismatch converts into a discrepancy between the fixed point of the algorithm and the "true" minimizer of the original objective function. Finally, a numerical example arising from computed tomography is provided in Section IV. Section V concludes this work.

II. Optimization background

A. Notation and definitions

In this paper, the underlying signal space is the N-dimensional Euclidean space $\mathbb{R}^N$ endowed with the standard scalar product $\langle\cdot,\cdot\rangle$ and the norm $\|\cdot\|$. Moreover, $|||\mathbf{L}|||$ denotes the spectral norm of squared matrix $\mathbf{L}$ and $\operatorname{Ker}\mathbf{L}$ designates its nullspace. The class of functions which are proper, convex, lower-semicontinuous on $\mathbb{R}^N$ and take values in $\mathbb{R}\cup\{+\infty\}$ is denoted by $\Gamma_0(\mathbb{R}^N)$. For every $g\in\Gamma_0(\mathbb{R}^N)$, $\operatorname{dom} g$ is the domain of g and ∂g is the subdifferential of g. If $\mathbf{x}\in\mathbb{R}^N$, the proximity operator of g at $\mathbf{x}$ is defined as [16]

$$\operatorname{prox}_g(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^N} \left(g(\mathbf{z}) + \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|^2 \right). \tag{2}$$

We say that $f:\mathbb{R}^N\mapsto\mathbb{R}^N$ is coercive if

$$\lim_{\|\mathbf{x}\|\to+\infty} f(\mathbf{x}) = +\infty \tag{3}$$

A key property to ensure the convergence of PGA is the cocoercivity of the involved gradient operator or of its approximation. Let us recall that operator $A\colon\mathbb{R}^N\to\mathbb{R}^N$ is η-cocoercive with $\eta\in[0,+\infty[$ if, for every $(\mathbf{x},\mathbf{y})\in(\mathbb{R}^N)^2$,

$$\eta\|A\mathbf{x}-A\mathbf{y}\|^2 \leqslant \langle \mathbf{x}-\mathbf{y}, A\mathbf{x}-A\mathbf{y}\rangle. \tag{4}$$

B. Proximal gradient algorithm for the penalized least-squares criterion

To find an estimate of $\overline{\mathbf{x}}$, defined in (1), we focus on solving the following penalized least squares criterion:

$$\operatorname*{minimize}_{\mathbf{x}\in\mathbb{R}^N} \ \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{x}\|^2 + g(\mathbf{x}) + \frac{\kappa}{2}\|\mathbf{x}\|^2, \tag{5}$$

where $g\in\Gamma_0(\mathbb{R}^N)$ is a suitable possibly non-smooth regularization function and $\kappa\in[0,+\infty[$. When $\kappa>0$, the objective function in (5) is strongly convex and we deal with an elastic net-like penalization [17]. For optimization problem (5), PGA reads, for every $n\in\mathbb{N}$,

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \theta_n\big(\operatorname{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x}_n - \gamma\mathbf{H}^\top(\mathbf{H}\mathbf{x}_n-\mathbf{y})) - \mathbf{x}_n\big), \tag{6}$$

where $\mathbf{x}_0\in\mathbb{R}^N$ is the initial estimate, $(\theta_n)_{n\in\mathbb{N}}\in]0,1]$ are relaxation parameters and $\gamma\in]0,+\infty[$ is the gradient step size. An instance of this algorithm is the well-known iterative soft-thresholding algorithm (ISTA), obtained when $\kappa=0$ and g is the ℓ_1 norm used to promote the sparsity of the resulting solution [18]–[20] (see [21] for extensions to other regularization functions).

If $\theta_n\in[\epsilon,1]$ with $\epsilon\in]0,1[$ and $\gamma\in]0,2/(|||\mathbf{H}|||^2+\kappa)[$, the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ generated by Algorithm (6) converges to a solution to Problem (5) when such a solution exists [4], [19], [22]. Note that, although there exist variants of PGA which use an iteration dependent step size, we will assume for the sake of simplicity that the step size remains fixed.

III. Adjoint mismatch

A. Mismatched algorithm

As mentioned earlier, in some practical implementations, the adjoint $\mathbf{H}^\top$ is approximated by an operator $\mathbf{K}$. This leads to the following modified PGA iteration where, for every $n\in\mathbb{N}$,

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \theta_n\big(\operatorname{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x}_n - \gamma\mathbf{K}(\mathbf{H}\mathbf{x}_n-\mathbf{y})) - \mathbf{x}_n\big). \tag{7}$$

We propose to analyze the convergence properties of Algorithm (7), by relying on fixed point theory. To this end, we introduce operator

$$\begin{aligned} T_\gamma\colon \mathbb{R}^N &\to \mathbb{R}^N \\ \mathbf{x} &\mapsto \operatorname{prox}_{\gamma g}((1-\gamma\kappa)\mathbf{x} - \gamma\mathbf{K}(\mathbf{H}\mathbf{x}-\mathbf{y})) \end{aligned} \tag{8}$$

with $\gamma\in]0,+\infty[$. This operator will play a key role in the characterization of the limit points of (7). We will show that the convergence of Algorithm (7) is guaranteed under cocoercivity conditions on matrix $\mathbf{L}=\mathbf{K}\mathbf{H}+\kappa\mathbf{Id}$.

B. Regularity of the modified gradient descent operator

When $\mathbf{K}\neq\mathbf{H}^\top$, the main difficulty is that the operator $\kappa\mathbf{Id}+\mathbf{K}(\mathbf{H}\cdot-\mathbf{y})$ is no longer guaranteed to be the gradient of the smooth part of our objective function. We however give conditions under which it remains a cocoercive operator, thus allowing to preserve a stable behavior of the iterative scheme (7). The following result can be established:

Proposition III.1 *Let $\lambda_{\min}$ (resp. $\lambda_{\max}$) be the minimum (resp. maximum) eigenvalue of $(\mathbf{L}+\mathbf{L}^\top)/2$. Let $\lambda^+_{\min}$ be the minimum positive eigenvalue of $(\mathbf{L}+\mathbf{L}^\top)/2$ and let $\beta=|||\mathbf{L}-\mathbf{L}^\top|||/2$.*

(i) *If $\lambda_{\min}\geqslant 0$, then $\mathbf{L}$ is cocoercive with constant*

$$\eta = \begin{cases} 1/\left(\sqrt{\lambda_{\max}} + \frac{\beta}{\sqrt{\lambda^+_{\min}}}\right)^2 \\ \qquad \text{if } \dim(\operatorname{Ker}(\mathbf{L}+\mathbf{L}^\top)) = \dim(\operatorname{Ker}\mathbf{L}) \\ 1/\lambda_{\max} \quad \text{if } \beta = 0. \end{cases}$$

(ii) *If $\lambda_{\min}>0$, then $\mathbf{L}$ is cocoercive with constant*

$$\eta = \frac{2}{|||(\mathbf{Id} + (\mathbf{L}-\mathbf{L}^\top)(\mathbf{L}+\mathbf{L}^\top)^{-1})(\mathbf{L}+\mathbf{L}^\top)^{1/2}|||^2}.$$

Remark III.2 A simple condition for ensuring that $\lambda_{\min}$ is positive is to choose $\kappa>-\widetilde{\lambda}_{\min}$ where $\widetilde{\lambda}_{\min}$ is the minimum eigenvalue of $(\mathbf{K}\mathbf{H}+\mathbf{H}^\top\mathbf{K}^\top)/2$. In this case, $\lambda^+_{\min}=\lambda_{\min}>0$, while $\operatorname{Ker}(\mathbf{L}+\mathbf{L}^\top)$ and $\operatorname{Ker}\mathbf{L}$ reduce to the null space.

C. Fixed points

Although our mismatched PGA may no longer minimize an objective function, it still converges to a fixed point under suitable conditions. It is easy to check that the fixed points of iteration (7) are those of operator T_γ. We now provide conditions ensuring that such a set of fixed points $\operatorname{Fix} T_\gamma$ is non empty and reduces to a singleton.

Proposition III.3

(i) *Let $\gamma \in]0, +\infty[$ and let $\widetilde{\mathbf{x}} \in \mathbb{R}^N$. We have $\widetilde{\mathbf{x}} \in \operatorname{Fix} T_\gamma$ if and only if $\widetilde{\mathbf{x}}$ belongs to*

$$\mathcal{F} = \left\{ \mathbf{x} \in \mathbb{R}^N \;\middle|\; 0 \in \mathbf{K}(\mathbf{H}\mathbf{x} - y) + \partial g(\mathbf{x}) + \kappa \mathbf{x} \right\}. \quad (9)$$

(ii) *If $\lambda_{\min} \geqslant 0$, then $\mathcal{F}$ is a closed and convex set.*

(iii) *Assume that $\mathbf{L}$ is cocoercive. $\mathcal{F}$ is nonempty if one of the following condition holds:*

a) $\operatorname{dom} \partial g = \mathbb{R}^N$ *and*

$$\mathbf{x} \mapsto \frac{1}{2}\langle \mathbf{x} \mid \mathbf{L}\mathbf{x} \rangle + g(\mathbf{x}) \quad (10)$$

is coercive;

b) $\operatorname{dom} g$ *is bounded.*

(iv) *In addition, $\mathcal{F}$ is a singleton if $\lambda_{\min} \geqslant 0$ and one of the following condition holds:*

d) $\lambda_{\min} \neq 0$*;*

e) g *is strongly convex.*

Eq. (9) highlights the fact that any fixed point of T_γ is a solution to an equilibrium rather than satisfies some optimality condition.

Furthermore, when the conditions in Remark III.2 are met, the above result guarantees the existence of a unique fixed point $\widetilde{\mathbf{x}}$ for T_γ, which can be viewed as an approximation to the minimizer of Problem (5). We now quantify the distance between these vectors.

Proposition III.4 *Let $\widehat{\mathbf{x}}$ be the minimizer of Problem* (5)*. The following upper bound on the error incurred by the mismatch holds:*

$$\|\widetilde{\mathbf{x}} - \widehat{\mathbf{x}}\| \leqslant \chi \, |||\mathbf{H}^\top - \mathbf{K}||| \, \|\mathbf{H}\widehat{\mathbf{x}} - \mathbf{y}\| \quad (11)$$

where $\chi = (\kappa + 2\lambda_{\min})^{-1}$.

The resulting upper bound on the error is thus proportional to the product of the norm of the mismatch on the adjoint and the norm of an approximation of the noise in the observation model. When κ is large, the distance to the "true" minimizer is smaller but the obtained solution becomes more biased with respect to the ground truth.

D. Convergence result

The convergence of the mismatched PGA follows from standard results concerning the forward-backward algorithm for solving monotone inclusion problems [16]:

Proposition III.5 *Assume that $\mathbf{L}$ is η-cocoercive. Let $\gamma \in]0, 2\eta[$ and $\delta = 2 - \gamma/(2\eta)$. Let $(\theta_n)_{n\in\mathbb{N}}$ be a sequence in $[0, \delta]$ such that $\sum_{n\in\mathbb{N}} \theta_n(\delta - \theta_n) = +\infty$. Suppose that $\mathcal{F} \neq \varnothing$. Then the sequence $(\mathbf{x}_n)_{n\in\mathbb{N}}$ generated by Algorithm* (7) *converges to a point $\widetilde{\mathbf{x}} \in \mathcal{F}$.*

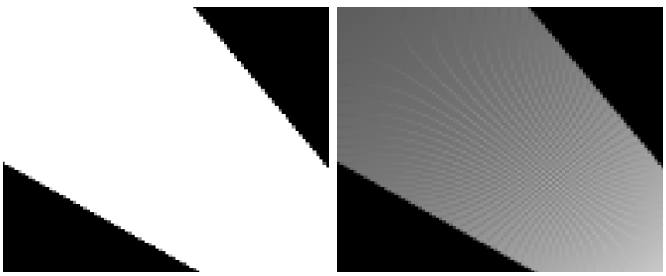
If $\mathbf{L}$ is self-adjoint positive semidefinite (i.e. $\beta = 0$ and $\lambda_{\min} \geqslant 0$), then $\lambda_{\max} = |||\mathbf{L}|||$ and Propositions III.1 and III.5 provide $2/|||\mathbf{L}|||$ as a strict upper bound on step size γ in order to guarantee the convergence of the algorithm. This recovers the classical upper bound on the step size value for the algorithm (7) in the special case when $\mathbf{K} = \mathbf{H}^\top$.

IV. Application to tomographic reconstruction

A. Problem statement

We now illustrate our theoretical results through an example of tomographic image reconstruction in fan-beam geometry with a truncated field of view. This acquisition model arises for instance in medical application for image-guidance in interventional radiology and surgery, where the detector has a limited size and truncation is unavoidable when scanning large body parts such as the abdomen. For such an under-determined inverse problem, iterative reconstruction methods have proven their superiority over filtered backprojection [23]. In observation model (1), $\mathbf{H}$ is the projection matrix, $\mathbf{y}$ represents the tomographic measurements, $\overline{\mathbf{x}}$ is a phantom, and $\mathbf{b}$ an additive i.i.d. Gaussian noise drawn from $\mathcal{N}(0, 0.2)$. To quantify the error introduced by $\mathbf{K}$, we compute δ as the average over 20 realizations of the ratio $\langle \mathbf{H}\mathbf{u}, \mathbf{v} \rangle / \langle \mathbf{u}, \mathbf{K}\mathbf{v} \rangle$ with $(\mathbf{u}, \mathbf{v})$ uniformly sampled in $([0, 1]^N)^2$. The farther δ from 1, the farther $\mathbf{K}$ from $\mathbf{H}^\top$.

B. Implementation



(a) $\mathbf{K}$ (b) $\mathbf{H}^\top$

Fig. 1: Zoom on the backprojection of a uniform view

We consider a grid of $N = 128 \times 128$ pixels of size 1 mm sampling a geometrical abdomen phantom (values between 0.036 and 1.5). The source-to-object distance and the source-to-image distance were respectively set to 180 mm and 270 mm. The detector has 128 bins of size 0.53 mm. We simulated 90 projections at uniformly spaced angular positions within interval $[0°, 180°]$, so that $M = 90 \times 128$. This simulation was performed with the ASTRA Toolbox [24] on Matlab, which allows the explicit computation of matrices $\mathbf{H}$ and $\mathbf{K}$. ASTRA GPU implementation of the projector and backprojector is such that the operators are not matched: the backprojector is pixel-driven [25] as it is always the case in

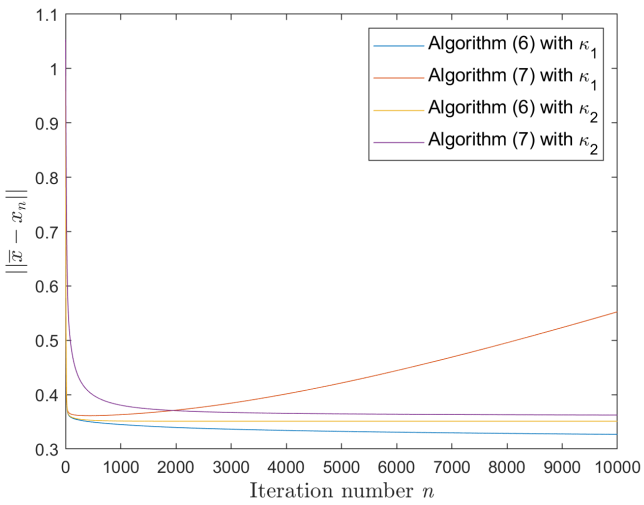


Fig. 2: Decay of the error along iterations for Algorithms (6) and (7) and two choices of κ parameter.

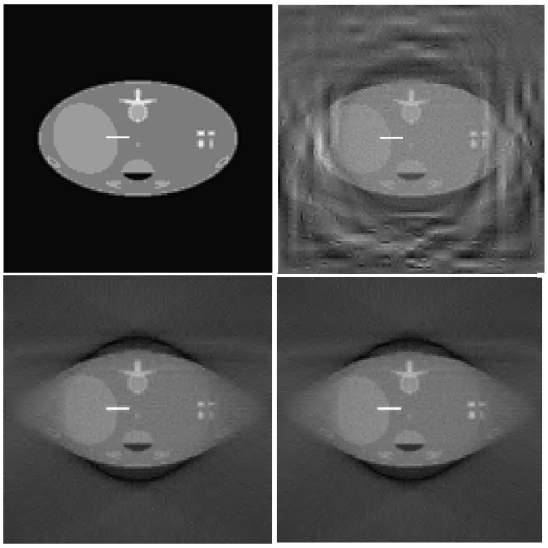


Fig. 3: Original phantom (top left), and reconstruction results using Algorithm (7) with κ_1 (top right), Algorithm (6) with κ_2 (bottom left) and Algorithm (7) with κ_2 (bottom right).

analytical reconstruction approaches, whereas the projector is ray-driven [26]. Matrix $\mathbf{H}$ was computed using ASTRA line-length ray driven projector [27] and $\mathbf{K}$ was generated using a pixel-driven backprojector, which led to $\delta = 1.0076$. In our settings, $\mathbf{H}^\top$ contains 1.0778% nonzero elements whereas this percentage decreases to 0.89% for $\mathbf{K}$. Figure 1 shows the backprojection of constant measurements at a single angle using either $\mathbf{H}^\top$ or $\mathbf{K}$.

An estimate of $\overline{\mathbf{x}}$ is obtained by adopting a compressed sensing approach [28]–[30]. More precisely, we solve the penalized least squares problem (5) with $g = \lambda\|\mathbf{W} \cdot \|_1$, $\mathbf{W}$ being the Symlet 2 wavelet transform, and $\lambda > 0$ the regularization parameter. We ran algorithms (6) and (7), for two choices of κ. With value $\kappa = \kappa_1$, $\mathbf{L}$ is not cocoercive, while with value $\kappa = \kappa_2$, $\mathbf{L}$ becomes cocoercive, and thus the convergence of scheme (7) is ensured. In practice, we computed the smallest negative eigenvalue $\widetilde{\lambda}_{\min}$ of the symmetric matrix $(\mathbf{KH} + \mathbf{H}^\top\mathbf{K}^\top)/2$. Then κ_2 was chosen slightly greater than $|\widetilde{\lambda}_{\min}|$ so that $\lambda_{\min} > 0$, while κ_1 is taken as a small value not satisfying this requirement. Altogether, we have $\kappa_1 = 0.01$, $\kappa_2 = 6.5$, $\lambda = 0.45$, $\theta_n \equiv 1$, and $\gamma = 1.9/(|||\mathbf{H}|||^2 + \kappa)$.

The stopping precision on the relative distance between two consecutive iterates is 10^{-7} and the maximum number of iterations is 10^4.

C. Results

Figure 2 displays the reconstruction error $\|\overline{x} - x_n\|$ with respect to the iterations when applying either Algorithm (6) or (7). The plots confirm that, with value κ_1, PGA only converges when the exact adjoint $\mathbf{H}^\top$ is used but diverges when using $\mathbf{K}$, as was expected from our theoretical analysis. More precisely, in the latter case, Algorithm (7) shows an initial convergence trend that reaches a minimum discrepancy point close to the minimizer obtained with $\mathbf{H}^\top$ before diverging. Nevertheless it would be difficult for the user to know when to stop the iterations so as to obtain this intermediary good solution. For value κ_2, both (6) and (7) converge to fixed points that are close to each other, again confirming our theoretical analysis. The corresponding reconstruction errors are 0.3512 and 0.3625 respectively. Note that PGA without mismatch requires less iterations to reach convergence than PGA using $\mathbf{K}$.

The reconstructed solutions are shown on Figure 3. We also computed the signal-to-noise ratio (SNR) over a central region of interest of 10×10 pixels. When value κ_2 is used, the reconstructed image obtained with $\mathbf{K}$ (SNR = 25.06 dB) looks almost the same as the image obtained without mismatch (SNR = 26.02 dB). In contrast, value κ_1 yields a reconstruction with $\mathbf{K}$ that is deteriorated by artifacts and a lower SNR (SNR = 22.32 dB) compared to the solution given by $\mathbf{H}^\top$ (SNR = 24.41 dB). In a nutshell, as soon as the convergence of PGA is ensured, an unmatched projector/backprojector pair gives a similar reconstruction quality than the matched pair but may lead to a slower convergence. Let us emphasize that, in a practical context, the decrease of the convergence rate would be compensated by a reduced computation cost for operator $\mathbf{K}$.

V. Conclusion

In this paper, we have presented necessary conditions to ensure the convergence of the proximal gradient algorithm when the adjoint of the linear operator involved in the differentiable part of the cost function is not fully computed but approximated. We have characterized the distance between the final solution and the solution to the original optimization problem. The technical conditions we obtained reconcile theory with practical implementations of PGA iteration in the context of X-ray tomographic imaging. A natural extension to this work will be to mitigate convergence instabilities when the adjoint operator is obtained by simplifying the physical model of the forward operator, as happens for instance in the context of SPECT imaging [5], [31].

REFERENCES

[1] E. Chouzenoux, F. Zolyniak, E. Gouillart, and H. Talbot, "A majorize-minimize memory gradient algorithm applied to X-ray tomography," in *Proceedings of the 20th IEEE International Conference on Image Processing (ICIP 2013)*, Melbourne, Australia, 15-18 Sep. 2013, pp. 1011–1015.

[2] L. E. Gueddari, E. Chouzenoux, A. Vignaud, J.-C. Pesquet, and P. Ciuciu, "Online MR image reconstruction for compressed sensing acquisition in T2* imaging," in *Proceedings of SPIE 11138, Wavelets and Sparsity XVIII*, no. 1113819, 9 Sep. 2019.

[3] 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.

[4] P. L. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, vol. 4 n°4, pp. 1164–1200, 2005.

[5] G. L. Zeng and G. T. Gullberg, "Unmatched projector/backprojector pairs in an iterative reconstruction algorithm," *IEEE Transactions on Medical Imaging*, 2000.

[6] A. Kak and M. Slaney, *Principles of computerized tomographic imaging*, ser. Classics in Applied Mathematics. Philadelphia: SIAM, 1988.

[7] J.-Y. Buffiere, E. Maire, J. Adrien, J.-P. Masse, and E. Boller, "In situ experiments with X-ray tomography: an attractive tool for experimental mechanics," *Experimental Mechanics*, vol. 50, no. 3, pp. 289–305, Mar. 2010.

[8] C. Chapdelaine, N. Gac, A.-M. Djafari, and E. Parra-Denis, "New GPU implementation of Separable Footprint (SF) Projector and Backprojector : first results," in *Proceedings of the 5th International Conference on Image Formation in X-Ray Computed Tomography*, May 2018, pp. 314–317.

[9] R. Liu, L. Fu, B. De Man, and H. Yu, "GPU-based branchless distance-driven projection and backprojection," *IEEE Transactions on Computational Imaging*, pp. 1–1, Mar 2017.

[10] R. R. Galigekere, K. Wiesent, and D. W. Holdsworth, "Cone-beam reprojection using projection-matrices," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, Oct. 2003.

[11] G. Zeng, "Counter examples for unmatched projector/backprojector in an iterative algorithm," *Chinese Journal of Academic Radiology*, Apr 2019.

[12] F. Arcadu, M. Stampanoni, and F. Marone, "On the crucial impact of the coupling projector-backprojector in iterative tomographic reconstruction," *CoRR*, 2016.

[13] T. Elfving and P. C. Hansen, "Unmatched projector/backprojector pairs: Perturbation and convergence analysis," *SIAM Journal on Scientific Computing*, vol. 40, no. 1, pp. A573–A591, 2018.

[14] 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*, vol. 41, no. 3, pp. A1822–A1839, 2019.

[15] D. A. Lorenz, S. Rose, and F. Schöpfer, "The randomized kaczmarz method with mismatched adjoint," *BIT Numerical Mathematics*, vol. 58, pp. 1079–1098, 2018.

[16] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2017.

[17] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[18] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, Aug 2003.

[19] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1457, 2004.

[20] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A l1-unified variational framework for image restoration," in *Proceedings of the 8th European Conference on Computer Vision (ECCV 2004)*, ser. Lecture Notes in Computer Science, vol. 3024, May 2004, pp. 1–13.

[21] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame based inverse problems," *Inverse Problems*, vol. 23, no. 1, pp. 1495–1518, 2007.

[22] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases," *SIAM Journal on Optimization*, vol. 18, no. 4, pp. 1351–1376, 2007.

[23] C. Riddell, A. Savi, M. Gilardi, and F. Fazio, "Frequency weighted least squares reconstruction of truncated transmission SPECT data," *IEEE Transactions on Nuclear Science*, vol. 43, pp. 2292 – 2298, Sept 1996.

[24] 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," *Optics Express*, vol. 24, no. 22, pp. 25 129–25 147, Oct 2016.

[25] 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, pp. 1252–1255.

[26] P. M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Transactions on Medical Imaging*, vol. 1, no. 3, pp. 192–196, Nov 1982.

[27] G. L. Zeng and G. T. Gullberg, "A ray-driven backprojector for backprojection filtering and filtered backprojection algorithms," in *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, Oct 1993, pp. 1199–1201.

[28] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.

[29] M. Guay, W. Czaja, M. Aronova, and R. Leapman, "Compressed sensing electron tomography for determining biological structure," *Scientific Reports*, vol. 6, p. 27614, 06 2016.

[30] L. El Gueddari, E. Chouzenoux, A. Vignaud, and P. Ciuciu, "Calibration-less parallel imaging compressed sensing reconstruction based on OSCAR regularization," Tech. Rep., 2019, https://hal.inria.fr/hal-02292372/document.

[31] 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*, vol. 40, no. 2, pp. 269–281, feb 1995.