



HAL
open science

Theoretical and numerical comparison of first-order algorithms for cocoercive equations and smooth convex optimization

Luis M Briceño-Arias, Nelly Pustelnik

► **To cite this version:**

Luis M Briceño-Arias, Nelly Pustelnik. Theoretical and numerical comparison of first-order algorithms for cocoercive equations and smooth convex optimization. *Signal Processing*, 2023, 206 (C). hal-03405740

HAL Id: hal-03405740

<https://hal.science/hal-03405740>

Submitted on 27 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proximal or gradient steps for cocoercive operators

Luis M. Briceño-Arias*, Nelly Pustelnik †

January 18, 2021

Abstract

This paper provides a theoretical and numerical comparison of classical first order splitting methods for solving smooth convex optimization problems and cocoercive equations. In a theoretical point of view, we compare convergence rates of gradient descent, forward-backward, Peaceman-Rachford, and Douglas-Rachford algorithms for minimizing the sum of two smooth convex functions when one of them is strongly convex. A similar comparison is given in the more general cocoercive setting under the presence of strong monotonicity and we observe that the convergence rates in optimization are strictly better than the corresponding rates for cocoercive equations for some algorithms. We obtain improved rates with respect to the literature in several instances exploiting the structure of our problems. In a numerical point of view, we verify our theoretical results by implementing and comparing previous algorithms in well established signal and image inverse problems involving sparsity. We replace the widely used ℓ_1 norm by the Huber loss and we observe that fully proximal-based strategies have numerical and theoretical advantages with respect to methods using gradient steps. In particular, Peaceman-Rachford is the more performant algorithm in our examples.¹

Keywords: Proximal algorithms, convergence rates, cocoercive equations, smooth convex optimization, Huber loss, sparse inverse problems.

1 Introduction

The resolution of many signal processing problems relies on the minimization of a sum of a data-fidelity term and a penalisation. This formulation can be encountered either in standard variational strategies [1], mainly used in the past 20 years, or into more recent deep-learning framework [2].

Formally, the associated optimization problem writes

$$\underset{x \in \mathcal{H}}{\text{minimize}} f(x) + g(x), \quad (1)$$

where \mathcal{H} denotes a real Hilbert space, and $f: \mathcal{H} \rightarrow]-\infty, +\infty]$ and $g: \mathcal{H} \rightarrow]-\infty, +\infty]$ are very often considered as proper lower semicontinuous convex functions.

Since almost twenty years, a large panel of efficient first-order algorithms have been derived in order to solve (1) under different assumptions on functions f and g (see [3, 4, 5] for an exhaustive list). From stronger to weaker assumptions, the gradient method [6, 7] is implementable if f and g are smooth, forward-backward splitting (FBS) [8, 9] can be applied when either f or g is smooth, while Peaceman-Rachford splitting (PRS) [10] and Douglas-Rachford splitting (DRS) [10, 11, 12] are applicable without any smoothness assumption. When a function is not smooth, FBS, PRS, DRS use proximal (implicit) steps for the function, which amounts to solve a non-linear equation. Since solving a non-linear equation at each iteration can be computationally costly, a common practice is to choose gradient steps when the function is smooth. However, nowadays there exists a wide class of functions whose proximal steps are explicit or easy to compute² and activating f and g via proximal steps can be advantageous numerically [13]. In this context, it becomes important to provide a theoretical comparison of algorithmic schemes involving gradient and/or proximal steps for solving (1). We focus our analysis on first order methods when f and g is smooth and proximal steps of both functions are easy to compute. The theoretical analysis of several first-order methods in this context provides interesting insights of the structural properties of first-order algorithms to be considered in more general frameworks.

From the signal processing user point of view, the choice of the most efficient algorithm for a specific data processing problem with the form of (1) is a complicated task. In order to tackle this problem, the convergence

*L. M. Briceño-Arias, Department of Mathematics, Universidad Técnica Federico Santa María, Santiago, Chile.

†N. Pustelnik is with Univ Lyon, Ens de Lyon, Univ Lyon 1, CNRS, Laboratoire de Physique, Lyon, 69342, France.

¹This work is supported by Agencia Nacional de Investigación y Desarrollo (ANID) from Chile, under grant FONDECYT 1190871 and also by the ANR (Agence Nationale de la Recherche) from France ANR-19-CE48-0009 Multisc'In. The authors acknowledge the support from LIA-MSD, CNRS-France.

²See, e.g., <http://proximity-operator.net/>

rate is a useful tool in order to provide a theoretical comparison among algorithms. However, the theoretical behaviour of an algorithmic scheme may differ considerably from its numerical efficiency, which enlightens the importance of obtaining sharp convergence rates exploiting the properties of f and g . In this context, sharp linear convergence rates can be obtained for several splitting algorithms under strong convexity of f and/or g [14, 15, 16, 17, 18], which can be extended when the strong convexity is satisfied on particular manifolds in the case of partly smooth functions [19, 20]. Moreover, sub-linear convergence rates of some first order methods depending on the KL-exponent are obtained in [21] when $f + g$ is a KL-function (see [22]). Since KL-exponents are usually difficult to compute [23], we focus on global strong convexity assumptions when we aim at finding linear convergence rates.

All previous discussion also holds in the context of monotone operators, which appear naturally from primal-dual first order optimality conditions of optimization problems involving linear operators (see, e.g., [24, 25, 26, 27]). We generalize our study of splitting algorithms involving implicit and/or explicit steps in the context of cocoercive equations. In the presence of strong monotonicity we compare linear convergence rates of the methods in this context.

Contributions – In the absence of strong convexity, we compare the nonexpansive properties of the operators governing the gradient method, FBS, PRS, and DRS for solving (1) when f and g are smooth. We obtain a new averaged nonexpansive constant for FBS, generalizing [3, Proposition 26.1(iv)(d)]. The averaged nonexpansive constants of all methods are preserved in the case of cocoercive equations. We also prove the averaged nonexpansive property of the operator defining PRS in the fully cocoercive setting, which is new as far as we know. This allows us to guarantee the weak convergence of PRS in the fully cocoercive setting, complementing [10, Corollary 1].

In the case when f is strongly convex, we compare the Lipschitz-continuous constants of the operators governing the gradient method, FBS, PRS, and DRS, which leads to a comparison of their linear convergence rates. This gives a theoretical support to the results obtained in [13] for the strongly convex case. In the context of strongly monotone cocoercive equations, we provide the linear convergence rates of the four algorithms under study, which are larger than the rates in the optimization context. We also provide an improved convergence rate for DRS inspired in [28, 15], which exploits the fully smooth context, which is replicated in the cocoercive setting. In addition, we derive an improved convergence rate for gradient method in the strongly monotone and cocoercive setting inspired in [28]. In this framework, we improve several convergence rates in the literature including [9, 29, 30, 10, 18].

A third contribution is to provide several experiments comparing the theoretical rates and the numerical behaviour of the four methods under study in the presence of high and low strong convexity parameters. We obtain that proximal-based schemes PRS and DRS are more efficient than EA and FBS in the context of piecewise constant denoising and image restoration.

Outline – In Section 2 we provide the results and concepts needed throughout the paper and the state-of-the-art on convergence properties of the algorithms under study. In Section 3 we provide and compare the averaged nonexpansive and the Lipschitz continuous constants of the operators governing the methods under study in the cocoercive setting and our results are refined in Section 4 for the particular smooth convex optimization context. We finish with numerical experiments in Section 5.

2 Preliminaries, problem, and state of the art

Throughout this paper, \mathcal{H} is a real Hilbert space endowed with the inner product $\langle \cdot | \cdot \rangle$. In this section we provide our notation, concepts, and results needed on this paper split in fixed point theory, monotone operator theory, convex analysis, and convergence of several algorithms.

2.1 Fixed point theory

An operator $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ is ω -Lipschitz continuous for some $\omega \in [0, +\infty[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|\Phi x - \Phi y\| \leq \omega \|x - y\|, \quad (2)$$

and Φ is nonexpansive if it is 1-Lipschitz continuous. The following convergence result, derived from [3, Theorem 1.50], is known as the Banach-Picard theorem and asserts the strong and linear convergence of iterations generated by repeatedly applying a ω -Lipschitz continuous operator when $\omega \in [0, 1[$.

Proposition 1. *Let $\omega \in [0, 1[$, let $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ be a ω -Lipschitz continuous operator, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = \Phi x_k. \quad (3)$$

Then, $\text{Fix } \Phi = \{\hat{x}\}$ for some $\hat{x} \in \mathcal{H}$ and we have

$$(\forall k \in \mathbb{N}) \quad \|x_k - \hat{x}\| \leq \omega^k \|x_0 - \hat{x}\|. \quad (4)$$

Moreover, $(x_k)_{k \in \mathbb{N}}$ converges strongly to \hat{x} with linear convergence rate ω .

In the case when $\omega = 1$, Proposition 1 is no longer valid, which is confirmed by the case when $\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a rotation.

An operator $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ is μ -averaged nonexpansive for some $\mu \in]0, 1]$ if, for every $x \in \mathcal{H}$ and $y \in \mathcal{H}$,

$$\|\Phi x - \Phi y\|^2 \leq \|x - y\|^2 - \left(\frac{1-\mu}{\mu}\right) \|(\text{Id} - \Phi)x - (\text{Id} - \Phi)y\|^2, \quad (5)$$

and Φ is firmly nonexpansive if it is $1/2$ -averaged. Observe that Φ is nonexpansive if and only if Φ is 1 -averaged. The following result derived from [3, Proposition 5.16] guarantees the weak convergence of the sequence generated by repeatedly applying a μ -averaged nonexpansive operator for some $\mu \in]0, 1[$.

Proposition 2. *Let $\mu \in]0, 1[$, let $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ be a μ -averaged nonexpansive operator such that $\text{Fix } \Phi \neq \emptyset$, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = \Phi x_k. \quad (6)$$

Then $(x_k)_{k \in \mathbb{N}}$ converges weakly to a point in $\text{Fix } \Phi$.

2.2 Monotone operator theory

Let $\mathcal{M}: \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a set-valued operator. The graph of \mathcal{M} is $\text{gra}(\mathcal{M}) = \{(x, u) \in \mathcal{H} \times \mathcal{H} \mid u \in \mathcal{M}x\}$, \mathcal{M} is monotone if it satisfies, for every (x, u) and (y, v) in $\text{gra}(\mathcal{M})$,

$$\langle u - v \mid x - y \rangle \geq 0, \quad (7)$$

and it is maximally monotone if its graph is maximal (in the sense of inclusions) among the graphs of monotone operators in $\mathcal{H} \times \mathcal{H}$. For every monotone operator $\mathcal{M}: \mathcal{H} \rightarrow 2^{\mathcal{H}}$, $J_{\mathcal{M}} = (\text{Id} + \mathcal{M})^{-1}$ is the resolvent of \mathcal{M} , which is single-valued. In addition, if \mathcal{M} is maximally monotone, then $J_{\mathcal{M}}$ is everywhere defined and firmly nonexpansive [3, Proposition 23.8].

For every $\eta \in [0, +\infty[$, we define the class \mathcal{C}_{η} of η -cocoercive operators $\mathcal{M}: \mathcal{H} \rightarrow \mathcal{H}$ satisfying, for every x and y in \mathcal{H} ,

$$\langle \mathcal{M}x - \mathcal{M}y \mid x - y \rangle \geq \eta \|\mathcal{M}x - \mathcal{M}y\|^2. \quad (8)$$

In particular, \mathcal{C}_0 is the class of single-valued monotone operators. Note that, if $\mathcal{M} \in \mathcal{C}_{\eta}$ for some $\eta > 0$, then \mathcal{M} is η^{-1} -Lipschitz continuous, by applying Cauchy-Schwartz inequality in (8), and maximally monotone in view of [3, Corollary 20.28].

An operator $\mathcal{M}: \mathcal{H} \rightarrow \mathcal{H}$ is ρ -strongly monotone for some $\rho \in]0, +\infty[$ if, for every x and y in \mathcal{H} ,

$$\langle \mathcal{M}x - \mathcal{M}y \mid x - y \rangle \geq \rho \|x - y\|^2. \quad (9)$$

2.3 Convex analysis

We denote by $\Gamma_0(\mathcal{H})$ the class of functions $h: \mathcal{H} \rightarrow]-\infty, +\infty]$ which are proper, lower semicontinuous, and convex. For every $h \in \Gamma_0(\mathcal{H})$, the maximally monotone operator

$$\partial h: x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H}) h(x) + \langle y - x \mid u \rangle \leq h(y)\} \quad (10)$$

is the subdifferential of h and $\text{Argmin}_{x \in \mathcal{H}} h(x)$ is the set of solutions to the problem of minimizing h over \mathcal{H} . For every $h \in \Gamma_0(\mathcal{H})$, it follows from [3, Proposition 17.4] that $\hat{x} \in \text{Argmin}_{x \in \mathcal{H}} h(x)$ if and only if $0 \in \partial h(\hat{x})$ and the proximity operator of h is defined by

$$\text{prox}_h: x \mapsto \arg \min_{y \in \mathcal{H}} \left(h(y) + \frac{1}{2} \|y - x\|^2 \right), \quad (11)$$

which is well defined because the objective function in (11) is strongly convex. We have $\text{prox}_h = J_{\partial h}$ and it reduces to P_C , the projection operator onto a closed convex set C , when $h = \iota_C$ is the indicator function of C which takes the value 0 in C and $+\infty$ outside.

For every $L \geq 0$, we consider the class $\mathcal{C}_L^{1,1}(\mathcal{H})$ of functions $h: \mathcal{H} \rightarrow \mathbb{R}$ satisfying:

- h is Gâteaux differentiable in \mathcal{H} , i.e., for every $x \in \mathcal{H}$ there exists a linear bounded operator $Dh(x): \mathcal{H} \rightarrow \mathbb{R}$ such that, for every $d \in \mathcal{H}$,

$$Dh(x)d = \lim_{t \downarrow 0} \frac{h(x + td) - h(x)}{t} = \langle \nabla h(x) \mid d \rangle, \quad (12)$$

where we denote by $\nabla h(x) \in \mathcal{H}$ the Riesz-Fréchet representant, and

- $\nabla h: \mathcal{H} \rightarrow \mathcal{H}$ is L -Lipschitz continuous.

Observe that, in view of [3, Corollary 17.42], every function in $\mathcal{C}_L^{1,1}(\mathcal{H})$ is Fréchet differentiable. The following proposition is a direct consequence of [3, Proposition 18.15] and asserts that every convex function $h \in \mathcal{C}_L^{1,1}(\mathcal{H})$ satisfies that ∇h is L^{-1} -cocoercive and viceversa. This result provides a subclass of $\mathcal{C}_{1/L}$ composed by gradients of convex functions in $\mathcal{C}_L^{1,1}(\mathcal{H})$.

Proposition 3. *Let $L \geq 0$ and let $h: \mathcal{H} \rightarrow \mathbb{R}$ be a convex function. Then the following are equivalent:*

1. $h \in \mathcal{C}_L^{1,1}(\mathcal{H})$.
2. h is Fréchet differentiable and, for every $x \in \mathcal{H}$ and $y \in \mathcal{H}$,

$$\langle x - y \mid \nabla h(x) - \nabla h(y) \rangle \leq L \|x - y\|^2. \quad (13)$$

3. h is Fréchet differentiable and $\nabla h \in \mathcal{C}_{1/L}$.

A function $h \in \mathcal{C}_L^{1,1}(\mathcal{H})$ is ρ -strongly convex, for some $\rho \in]0, +\infty[$, if $h - \frac{\rho}{2} \|\cdot\|_2^2$ is convex or, equivalently, if ∇h is ρ -strongly monotone.

For further details and properties of monotone operators and convex functions in Hilbert spaces, the reader is referred to [3].

2.4 Problem and algorithms

In this paper we study several splitting algorithms in the context of the monotone inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \mathcal{A}x + \mathcal{B}x, \quad (14)$$

where $\mathcal{A}: \mathcal{H} \rightarrow 2^{\mathcal{H}}$ and $\mathcal{B}: \mathcal{H} \rightarrow 2^{\mathcal{H}}$ are maximally monotone operators. The problem in (14) models several problems in game theory [31], and optimization problems as considered in signal and image processing [4, 32, 33, 1, 2], among other areas. In the particular case when $\mathcal{A} = \partial f$ and $\mathcal{B} = \partial g$ for some functions f and g in $\Gamma_0(\mathcal{H})$, the convex optimization problem (under standard qualification conditions)

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + g(x), \quad (15)$$

is an important particular instance of the problem in (14) in view of [3, Proposition 17.4].

In order to solve the problem in (14), the algorithms we consider generate recursive sequences via Banach-Picard iterations of the form

$$x_{k+1} = \Phi x_k, \quad (16)$$

where $x_0 \in \mathcal{H}$ and $\Phi: \mathcal{H} \rightarrow \mathcal{H}$ is a suitable nonexpansive operator which incorporates resolvents and/or explicit computations of \mathcal{A} and \mathcal{B} and such that we can recover a solution in $(\mathcal{A} + \mathcal{B})^{-1}(\{0\})$ from its fixed points. More precisely, in this paper we study the following algorithms for solving the problem in (14).

Explicit algorithm (EA) – It corresponds to apply (16) with the explicit operator

$$\Phi = G_{\tau(\mathcal{A} + \mathcal{B})} := \text{Id} - \tau(\mathcal{A} + \mathcal{B}), \quad (17)$$

for some $\tau > 0$. EA can be seen as an explicit Euler discretization of the dynamical system governed by $\mathcal{A} + \mathcal{B}$ in the single-valued case [34, Section 2.4]. In the particular case when $\mathcal{A} = \nabla f$ and $\mathcal{B} = \nabla g$ for smooth convex functions f and g , EA corresponds to gradient descent [6, 7]. It is clear that

$$(\forall \tau > 0) \quad (\mathcal{A} + \mathcal{B})^{-1}(0) = \text{Fix } G_{\tau(\mathcal{A} + \mathcal{B})}. \quad (18)$$

Proximal Point Algorithm (PPA) – It is proposed in [35] for a variational inequality problem and by [36] in the maximally monotone context. This algorithm corresponds to the iteration in (16) governed by the resolvent

$$\Phi = J_{\tau(\mathcal{A} + \mathcal{B})} = (\text{Id} + \tau(\mathcal{A} + \mathcal{B}))^{-1}. \quad (19)$$

for some $\tau > 0$. PPA can be seen as an implicit discretization of the dynamical system governed by $\mathcal{A} + \mathcal{B}$ [34, Section 2.3]. In the particular case when $\mathcal{A} = \partial g$ and $\mathcal{B} = \partial f$ for some convex functions f and g satisfying standard qualification conditions, $J_{\tau(\mathcal{A} + \mathcal{B})} = \text{prox}_{\tau(f+g)}$ is the proximity operator defined in (11) and motivates the name to the algorithm. Each (implicit) step of PPA includes the resolution of a non-linear equation, but, in a large class of operators, this equation has an explicit solution or it is easy to solve. It is clear that

$$(\forall \tau > 0) \quad (\mathcal{A} + \mathcal{B})^{-1}(0) = \text{Fix } J_{\tau(\mathcal{A} + \mathcal{B})}. \quad (20)$$

Forward-Backward splitting (FBS) – It follows from (16) with the Forward-Backward operator

$$\Phi = T_{\tau\mathcal{B},\tau\mathcal{A}} = J_{\tau\mathcal{B}} \circ G_{\tau\mathcal{A}} = (\text{Id} + \tau\mathcal{B})^{-1}(\text{Id} - \tau\mathcal{A}), \quad (21)$$

for some $\tau > 0$, which alternates explicit and implicit steps. In the case when $\mathcal{A} = \nabla f$ and $\mathcal{B} = \partial g$, for some f and g in $\Gamma_0(\mathcal{H})$, $J_{\tau\mathcal{B}} = \text{prox}_{\tau g}$ for every $\tau > 0$ and FBS is the proximal gradient algorithm (see, e.g., [37]). This method finds its roots in the projected gradient method [38] (case $g = \iota_C$ for some closed convex set C). In the context of variational inequalities appearing in some PDE's, a generalization of the projected gradient method is proposed in [39, 40, 41].

It follows from [3, Proposition 26.1(iv)(a)] that

$$(\forall \tau > 0) \quad (\mathcal{A} + \mathcal{B})^{-1}(0) = \text{Fix } T_{\tau\mathcal{B},\tau\mathcal{A}}. \quad (22)$$

Peaceman-Rachford splitting (PRS) – This scheme follows from (16) with the Peaceman-Rachford operator

$$\Phi = R_{\tau\mathcal{B},\tau\mathcal{A}} = (2J_{\tau\mathcal{B}} - \text{Id}) \circ (2J_{\tau\mathcal{A}} - \text{Id}), \quad (23)$$

for some $\tau > 0$. PRS is first proposed in [42] for solving some linear systems derived from discretizations of PDE's and it is studied in the non-linear monotone case in [10]. It follows from [3, Proposition 26.1(iii)(b)] that

$$(\forall \tau > 0) \quad (\mathcal{A} + \mathcal{B})^{-1}(0) = J_{\tau\mathcal{A}}(\text{Fix } R_{\tau\mathcal{B},\tau\mathcal{A}}). \quad (24)$$

As before, we recover PRS in the optimization context by using the identity $J_{\partial h} = \text{prox}_h$ for $h \in \Gamma_0(\mathcal{H})$.

Douglas-Rachford splitting (DRS) – This scheme follows from (16) with Douglas-Rachford operator

$$\Phi = S_{\tau\mathcal{B},\tau\mathcal{A}} = \frac{\text{Id} + R_{\tau\mathcal{B},\tau\mathcal{A}}}{2} = J_{\tau\mathcal{B}}(2J_{\tau\mathcal{A}} - \text{Id}) + \text{Id} - J_{\tau\mathcal{A}}, \quad (25)$$

for some $\tau > 0$, which is the average between Id and $R_{\tau\mathcal{B},\tau\mathcal{A}}$. The algorithm is first proposed for solving some linear systems derived from discretizations of PDE's [43] and it is studied in the non-linear monotone case in [10]. It follows from [3, Proposition 26.1(iii)(b)] that

$$(\forall \tau > 0) \quad (\mathcal{A} + \mathcal{B})^{-1}(0) = J_{\tau\mathcal{A}}(\text{Fix } S_{\tau\mathcal{B},\tau\mathcal{A}}). \quad (26)$$

As before, we recover DRS in the optimization context by using the identity $J_{\partial h} = \text{prox}_h$ for $h \in \Gamma_0(\mathcal{H})$.

2.5 State-of-the-art on convergence of algorithms

It is well known that, for every η -cocoercive operator \mathcal{M} and every $\tau \in]0, 2\eta[$, $G_{\tau\mathcal{M}}$ is averaged nonexpansive [3, Proposition 4.39] and, therefore, EA converges weakly to a point in $\mathcal{M}^{-1}(\{0\})$ in view of Proposition 2. Therefore, if \mathcal{A} and \mathcal{B} are cocoercive, $\mathcal{M} = \mathcal{A} + \mathcal{B}$ is cocoercive and EA achieves weak convergence to a solution to (14). If $\mathcal{M} = \mathcal{A} + \mathcal{B}$ is additionally strongly monotone and $\tau \in]0, 2\eta[$, $G_{\tau\mathcal{M}}$ is Lipschitz continuous with constant in $]0, 1[$ [17, Fact 7] and EA achieves linear convergence in view of Proposition 1.

On the other hand, for every $\tau > 0$ and any maximally monotone operator \mathcal{M} , $J_{\tau\mathcal{M}}$ is firmly nonexpansive [3, Proposition 23.8], which allows us to prove the weak convergence of PPA to a point in $\mathcal{M}^{-1}(\{0\})$. If we additionally assume strong monotonicity of \mathcal{M} , we obtain that $J_{\tau\mathcal{M}}$ is Lipschitz continuous with constant in $]0, 1[$ and PPA achieves linear convergence [3, Proposition 23.13]. However, when $\mathcal{M} = \mathcal{A} + \mathcal{B}$, the computation of $J_{\tau\mathcal{M}}$ can be difficult, and other splitting methods as EA, FBS, PRS, and DRS can be considered in order to reduce the computational time by iteration.

In the case of FBS, the weak convergence of the iterations generated by (16) with $\Phi = T_{\tau\mathcal{B},\tau\mathcal{A}}$ is guaranteed if \mathcal{A} is α -cocoercive and $\tau \in]0, 2\alpha[$ [3, Theorem 26.14]. This is a consequence of the averaged nonexpansiveness of $T_{\tau\mathcal{B},\tau\mathcal{A}}$ in this context [3, Proposition 26.1(iv)(d)]. If additionally we assume the strong monotonicity of \mathcal{A} or \mathcal{B} , the linear convergence of FBS is guaranteed [3, Theorem 26.16], which follows from the Lipschitz continuity of $T_{\tau\mathcal{B},\tau\mathcal{A}}$ with Lipschitz constant in $]0, 1[$. In [29] the authors provide a detailed analysis of the convergence rates of FBS in the strongly monotone context.

If \mathcal{A} is not cocoercive the convergence of FBS is not guaranteed and, if it is not single-valued, it is not applicable. In these contexts PRS and DRS can be used if $J_{\mathcal{A}}$ is not difficult to compute. In the case when \mathcal{A} and \mathcal{B} are merely maximally monotone, reflections $2J_{\mathcal{A}} - \text{Id}$ and $2J_{\mathcal{B}} - \text{Id}$ are merely nonexpansive, and the convergence of PRS is not guaranteed. This motivates the average with Id in (25), which allows to obtain an averaged nonexpansive operator for DRS with weak convergence to a solution. Under the cocoercivity assumption on \mathcal{A} , the weak convergence of PRS is guaranteed in [10, Corollary 1& Remark 2(2)]. If in addition we suppose the strong monotonicity of \mathcal{A} , the reflection $2J_{\mathcal{A}} - \text{Id}$ is Lipschitz continuous with constant in $]0, 1[$ [28] and, therefore, PRS converges linearly and strongly to a solution. This property also holds for DRS, but with a larger convergence rate. Of course, previous properties are inherited by the algorithms in

the particular optimization context, sometimes with better convergence rates by exploiting the variational formulation [17, 16, 14, 15].

In summary, without any cocoercivity on problem (14) the only available convergent method is DRS, if resolvents are easy to compute. However, in the fully cocoercive setting all the methods under study are convergent and can be implemented, and there is no theoretical/numerical comparison of these methods in the literature in this context. In this paper, as stated in Section 1, we restrict ourselves to classes \mathcal{C}_η for some $\eta \geq 0$ ($\mathcal{C}_{1/\eta}^{1,1}(\mathcal{H})$ in the optimization setting), in order to provide a simple context suitable to theoretical and numerical comparisons of the algorithms described above. We start by studying cocoercive equations.

3 Cocoercive equations

In this section we study properties of different numerical schemes for solving the following cocoercive equation.

Problem 1. *Let $(\alpha, \beta) \in]0, +\infty[^2$ and let $\mathcal{A} \in \mathcal{C}_\alpha$ and $\mathcal{B} \in \mathcal{C}_\beta$. The problem is to*

$$\text{find } x \in \mathcal{H} \text{ such that } \mathcal{A}x + \mathcal{B}x = 0, \quad (27)$$

under the assumption that solutions exist.

We split our theoretical study in the general cocoercive setting and the strongly monotone case.

3.1 General cocoercive case

In the general cocoercive case, the following proposition provides the averaging constants of the operators defining EA, FBS, PRS, and DRS, which implies their weak convergence in view of Proposition 2.

Proposition 4. *Let $\tau > 0$. In the context of Problem 1, the following hold:*

1. *Suppose that $\tau \in]0, 2\beta\alpha/(\beta + \alpha)[$. Then $G_{\tau(\mathcal{A}+\mathcal{B})}$ is $\mu_G(\tau)$ -averaged, where*

$$\mu_G(\tau) := \frac{\tau(\beta + \alpha)}{2\beta\alpha} \in]0, 1[. \quad (28)$$

2. *Suppose that $\tau \in]0, 2\alpha[$. Then $T_{\tau\mathcal{B}, \tau\mathcal{A}}$ is $\mu_T(\tau)$ -averaged, where*

$$\mu_T(\tau) := \frac{2\tau(\beta + \alpha)}{4\beta\alpha + \tau(4\alpha - \tau)} \in]0, 1[. \quad (29)$$

3. *$R_{\tau\mathcal{B}, \tau\mathcal{A}}$ is $\mu_R(\tau)$ -averaged, where*

$$\mu_R(\tau) := \frac{\tau}{\frac{\alpha\beta}{\alpha+\beta} + \tau} \in]0, 1[. \quad (30)$$

4. *$S_{\tau\mathcal{B}, \tau\mathcal{A}}$ is $\mu_S(\tau) = \frac{\mu_R(\tau)}{2}$ -averaged.*

The proof is provided in Appendix 6. Note that, in the absence of cocoercivity for \mathcal{B} ($\beta = 0$) it follows from (29) that $\mu_T(\tau) = 2\alpha/(4\alpha - \tau)$ which coincides with the constant in [3, Proposition 26.1(iv)(d)] and our constant is smaller in general.

The averaged nonexpansivity of the reflection $2J_{\mathcal{A}} - \text{Id}$ when \mathcal{A} is cocoercive is studied in [28]. As far as we know, the averaged nonexpansivity of $R_{\tau\mathcal{B}, \tau\mathcal{A}}$ in the context of Problem 1 is a new result.

In the case when $\mathcal{B} = 0$, Problem 1 reduces to

$$\text{find } x \in \mathcal{H} \text{ such that } \mathcal{A}x = 0, \quad (31)$$

and, for every $\tau > 0$, $G_{\tau(\mathcal{A}+\mathcal{B})} = T_{\tau\mathcal{B}, \tau\mathcal{A}} = G_{\tau\mathcal{A}}$ and $S_{\tau\mathcal{A}, \tau\mathcal{B}} = S_{\tau\mathcal{B}, \tau\mathcal{A}} = J_{\tau\mathcal{A}}$. Therefore, since (8) yields $\mathcal{B} = 0 \in \mathcal{C}_\beta$ for every $\beta > 0$, by considering $\beta \rightarrow +\infty$ in Proposition 4, we obtain the following known results (see, e.g., [3, Proposition 4.39] and [28, Proposition 5.2]).

Proposition 5. *Let $\tau \in]0, +\infty[$, $\alpha \in]0, +\infty[$, and suppose that $\mathcal{A} \in \mathcal{C}_\alpha$. Then, the following hold.*

1. *If $\tau \in]0, 2\alpha[$, then $G_{\tau\mathcal{A}}$ is $\tau/(2\alpha)$ -averaged nonexpansive and $G_{2\alpha\mathcal{A}}$ is nonexpansive.*
2. *$J_{\tau\mathcal{A}}$ is $\tau/(2(\tau + \alpha))$ -averaged nonexpansive.*

Now, suppose that $\mathcal{A} = \mathcal{L}^* \circ \mathcal{M} \circ \mathcal{L}$, where $\mathcal{M} \in \mathcal{C}_\eta$, for some $\eta \in]0, +\infty[$. An advantage of using $G_{\mathcal{A}}$ instead of $J_{\mathcal{A}}$ is that the former is explicit while the latter rarely have a closed form expression. In the following result we recall some specific frameworks in which $J_{\mathcal{A}}$ can be computed explicitly in terms of $J_{\mathcal{M}}$.

Proposition 6. Let \mathcal{H} and \mathcal{G} be real Hilbert spaces, let $\mathcal{L}: \mathcal{H} \rightarrow \mathcal{G}$ be a real bounded operator such that $\mathcal{L} \circ \mathcal{L}^*$ is invertible, let $\mathcal{M} \in \mathcal{C}_\eta$, for some $\eta \in]0, +\infty[$, and set $\mathcal{A} = \mathcal{L}^* \circ \mathcal{M} \circ \mathcal{L}$. Then \mathcal{A} is α -cocoercive with $\alpha = \eta/\|\mathcal{L}\|^2$ and the following holds.

1. For every $\tau > 0$, $J_{\tau\mathcal{A}} = \text{Id} - \mathcal{L}^* \circ (\mathcal{L}\mathcal{L}^*)^{-1} \circ (\text{Id} - J_{\tau(\mathcal{L}\mathcal{L}^*), \mathcal{M}}) \circ \mathcal{L}$.
2. Suppose that $\mathcal{L} \circ \mathcal{L}^* = \mu \text{Id}$, for some $\mu > 0$. Then, for every $\tau > 0$, $J_{\tau\mathcal{A}} = \text{Id} - \mu^{-1}\mathcal{L}^* \circ (\text{Id} - J_{\tau\mu\mathcal{M}}) \circ \mathcal{L}$.
3. Suppose that $\mathcal{H} = \mathcal{H}_1 \oplus \dots \oplus \mathcal{H}_n$ and that $\mathcal{L} \circ \mathcal{L}^* = \mathcal{D}$, where $\mathcal{D}: (x_1, \dots, x_n) \mapsto (\chi_i x_i)_{1 \leq i \leq n}$ and $\chi_i > 0$ for every $i \in \{1, \dots, n\}$. Then, for every $\tau > 0$, $J_{\tau\mathcal{A}} = \text{Id} - \mathcal{L}^* \circ \mathcal{D}^{-1} \circ (\text{Id} - J_{\tau\mathcal{D}\mathcal{M}}) \circ \mathcal{L}$.

Proof. The cocoercivity follows from [3, Proposition 4.12] and the formulas are derived from [3, Proposition 23.25] (see also [44] for 3 in finite dimensions). \square

3.2 Cocoercive and strongly monotone case

In the context of Problem 1, suppose in addition that \mathcal{A} is ρ -strongly monotone, for some $\rho \in]0, \alpha^{-1}[$. Under this additional assumption, there exists a unique solution $\hat{x} \in \mathcal{A}^{-1}(0)$ and the operators $G_{\tau(\mathcal{A}+\mathcal{B})}$, $T_{\tau\mathcal{B}, \tau\mathcal{A}}$, $R_{\tau\mathcal{B}, \tau\mathcal{A}}$ and $S_{\tau\mathcal{B}, \tau\mathcal{A}}$ defined in (17)–(25) are $\omega(\tau)$ -Lipschitz continuous for some $\omega(\tau) \in]0, 1[$, under suitable conditions on τ . The Lipschitz continuous constant of each algorithm corresponds to its linear convergence rate in view of Proposition 1, which allows the user to compare not only numerically but also theoretically the convergence behaviour of each method. In the next proposition, we summarize the convergence rates for the schemes governed by the operators defined in (17)–(25) aiming to solve Problem 1.

Proposition 7. In the context of Problem 1, let $\tau > 0$ and let $\rho \in]0, \alpha^{-1}[$. Suppose that \mathcal{A} is ρ -strongly monotone. The following hold:

1. Suppose that $\tau \in]0, 2\beta\alpha/(\beta + \alpha)[$. Then $G_{\tau(\mathcal{A}+\mathcal{B})}$ is $\omega_G(\tau)$ -Lipschitz continuous, where

$$\omega_G(\tau) := \sqrt{1 - \frac{2\tau\rho}{\alpha(2\beta - \tau)}(2\beta\alpha - \tau(\beta + \alpha))} \in]0, 1[. \quad (32)$$

In particular, the minimum in (32) is achieved at

$$\tau^* = \frac{2\beta\alpha}{\sqrt{\beta + \alpha}(\sqrt{\beta + \alpha} + \sqrt{\beta})} \quad (33)$$

and

$$\omega_G(\tau^*) = \sqrt{1 - \frac{4\rho\beta\alpha}{(\sqrt{\beta + \alpha} + \sqrt{\beta})^2}}. \quad (34)$$

2. Suppose that $\tau \in]0, 2\alpha[$. Then $T_{\tau\mathcal{B}, \tau\mathcal{A}}$ is $\omega_{T_1}(\tau)$ -Lipschitz continuous, where

$$\omega_{T_1}(\tau) := \sqrt{1 - \frac{\tau\rho}{\alpha}(2\alpha - \tau)} \in]0, 1[. \quad (35)$$

In particular, the minimum in (35) is achieved at

$$\tau^* = \alpha \quad \text{and} \quad \omega_{T_1}(\tau^*) = \sqrt{1 - \alpha\rho}. \quad (36)$$

3. Suppose that $\tau \in]0, 2\beta[$. Then $T_{\tau\mathcal{A}, \tau\mathcal{B}}$ is $\omega_{T_2}(\tau)$ -Lipschitz continuous, where

$$\omega_{T_2}(\tau) := \frac{1}{1 + \tau\rho} \in]0, 1[. \quad (37)$$

In particular, the minimum in (37) is achieved at

$$\tau^* = 2\beta \quad \text{and} \quad \omega_{T_2}(\tau^*) = \frac{1}{1 + 2\beta\rho}. \quad (38)$$

4. $R_{\tau\mathcal{B}, \tau\mathcal{A}}$ and $R_{\tau\mathcal{A}, \tau\mathcal{B}}$ are $\omega_R(\tau)$ -Lipschitz continuous, where

$$\omega_R(\tau) = \sqrt{\frac{\alpha - 2\tau\rho\alpha + \tau^2\rho}{\alpha + 2\tau\rho\alpha + \tau^2\rho}} \in]0, 1[. \quad (39)$$

In particular, the minimum in (39) is achieved at

$$\tau^* = \sqrt{\frac{\alpha}{\rho}} \quad \text{and} \quad \omega_R(\tau^*) = \sqrt{\frac{1 - \sqrt{\alpha\rho}}{1 + \sqrt{\alpha\rho}}}. \quad (40)$$

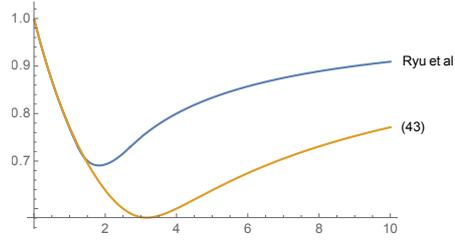


Figure 1: Comparison between the Lipschitz constants in [18] and (41) for DRS when $\beta = 1$, $\rho = 0.3$, and $\alpha = 3$.

5. $S_{\tau\mathcal{B},\tau\mathcal{A}}$ and $S_{\tau\mathcal{A},\tau\mathcal{B}}$ are $\omega_S(\tau)$ -Lipschitz continuous, where

$$\omega_S(\tau) = \min \left\{ \frac{1 + \omega_R(\tau)}{2}, \frac{\beta + \tau^2 \rho}{\beta + \tau \beta \rho + \tau^2 \rho} \right\} \in]0, 1[. \quad (41)$$

In particular, the minimum in (41) is achieved at

$$\tau^* = \begin{cases} \sqrt{\frac{\alpha}{\rho}}, & \text{if } \beta \leq \frac{4\alpha}{(1 + \sqrt{1 - \alpha\rho})^2}; \\ \sqrt{\frac{\beta}{\rho}}, & \text{otherwise,} \end{cases} \quad (42)$$

and

$$\omega_S(\tau^*) = \begin{cases} \frac{1 + \sqrt{1 - \alpha\rho}}{1 + \sqrt{1 - \alpha\rho} + \sqrt{\alpha\rho}}, & \text{if } \beta \leq \frac{4\alpha}{(1 + \sqrt{1 - \alpha\rho})^2}; \\ \frac{2}{2 + \sqrt{\beta\rho}}, & \text{otherwise.} \end{cases} \quad (43)$$

The proof is provided in Appendix 7. Observe that Proposition 7(1) is a new result, in which the Lipschitz constant of the explicit operator is improved with respect to considering a single operator when a splitting is possible (see Remark 1). Proposition 7(2) provides a smaller Lipschitz-constant for operator $T_{\tau\mathcal{B},\tau\mathcal{A}}$ than in [9, Remarque 3.1(2)], [29, Theorem 2.4], [30, Proposition 1(d)], and [3, Proposition 26.16(ii)], by exploiting the cocoercivity of \mathcal{A} . On the other hand, in Proposition 7(3) we obtain a better Lipschitz constant for $T_{\tau\mathcal{A},\tau\mathcal{B}}$ than in [30, Proposition 1(d)] and [29, Theorem 2.4], and we recover the Lipschitz constant in [3, Proposition 26.16(i)], but we obtain a smaller Lipschitz constant by allowing $\tau = 2\beta$. The Lipschitz constant of $R_{\tau\mathcal{A},\tau\mathcal{B}}$ and $R_{\tau\mathcal{B},\tau\mathcal{A}}$ in (39) is obtained in [28, Theorem 7.4], and it is smaller than Lipschitz constants in [28, Theorem 6.5 & Theorem 5.6] which are also valid in our context. The constant in (41) is provided in [28, Theorem 7.4] and it is tighter than the constant obtained in [10, Proposition 4], which does not take advantage of the full cocoercivity of the problem. The Lipschitz constant of $S_{\tau\mathcal{A},\tau\mathcal{B}}$ and $S_{\tau\mathcal{B},\tau\mathcal{A}}$ in (41) is obtained from [28, Theorem 5.6 & Theorem 7.4] by exploiting the cocoercivity of \mathcal{A} and \mathcal{B} . When α is large with respect to β , our constant is sharper than the constant in [18, Corollary 4.2] (see Figure 1), which is obtained via computer-assisted analysis. This is because the cocoercivity of \mathcal{A} is not considered in [18].

In the case when $\mathcal{B} = 0$, by taking $\beta \rightarrow +\infty$ in parts 1 (or 2) and 3 of Proposition 7 we obtain as a direct consequence the following result for EA and PPA in the strongly monotone case. The Lipschitz continuous constant of EA obtained in [17, Fact 7] with a geometric proof is complemented with analytic arguments in the proof of Proposition 7. The constant of PPA is proved in [3, Proposition 23.13].

Proposition 8. *Suppose that $\mathcal{A} \in \mathcal{C}_\alpha$ is ρ -strongly monotone, for some $\alpha \in]0, +\infty[$ and $\rho \in]0, \alpha^{-1}[$. Then the following hold.*

1. For every $\tau \in]0, 2\alpha[$, $G_{\tau\mathcal{A}}$ is $\omega_{G_0}(\tau)$ -Lipschitz continuous, where

$$\omega_{G_0} := \sqrt{1 - \frac{\tau\rho}{\alpha}(2\alpha - \tau)} \in]0, 1[. \quad (44)$$

2. For every $\tau > 0$, $J_{\tau\mathcal{A}}$ is $\omega_J(\tau)$ -Lipschitz continuous, where

$$\omega_J(\tau) := \frac{1}{1 + \tau\rho} \in]0, 1[. \quad (45)$$

Remark 1. *Observe that $\mathcal{A} + \mathcal{B}$ is $\beta\alpha/(\beta + \alpha)$ -cocoercive [3, Proposition 4.12] and ρ -strongly monotone. Moreover, for every $\tau \in]0, 2\beta\alpha/(\beta + \alpha)[$ we have*

$$\frac{\tau\rho}{\beta\alpha}(2\beta\alpha - \tau(\beta + \alpha)) < \frac{2\tau\rho}{\alpha(2\beta - \tau)}(2\beta\alpha - \tau(\beta + \alpha)). \quad (46)$$

Therefore ω_G defined in (32) is strictly lower than ω_{G_0} in (44). Moreover, in the case when $\mathcal{B} = 0$ ($\beta \rightarrow \infty$), both functions coincide. This new result implies that the gradient operator takes advantage of the splitting when a part of the monotone inclusion is strongly monotone.

4 Smooth convex optimization

In this section we restrict our attention to the following particular instance of Problem 1.

Problem 2. Let $f \in \mathcal{C}_{1/\alpha}^{1,1}(\mathcal{H})$ and $g \in \mathcal{C}_{1/\beta}^{1,1}(\mathcal{H})$, for some $\alpha \in]0, +\infty[$ and $\beta \in]0, +\infty[$. The problem is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + g(x), \quad (47)$$

under the assumption that solutions exist.

Proposition 4 provides the averaged nonexpansive constants of the operators $G_{\tau(\nabla g + \nabla f)}$, $T_{\tau \nabla g, \tau \nabla f}$, $R_{\tau \nabla g, \tau \nabla f}$, and $S_{\tau \nabla g, \tau \nabla f}$ and, from Proposition 5, those of operators $G_{\nabla f}$ and prox_f in the case when $f = 0$. As before, this guarantees the weak convergence of all methods to a solution to Problem 2.

4.1 Strongly convex case

In the context of Problem 2, suppose in addition that

$$(\exists \rho \in]0, 1/\alpha]) \quad f \text{ is } \rho\text{-strongly convex.} \quad (48)$$

Under this context, there exists a unique solution to Problem 2, which is denoted by \hat{x} . Since $\mathcal{A} = \nabla f$ is cocoercive and strongly monotone, Proposition 7 provides Lipschitz constants of the operators governing the numerical schemes under study. The following result is a refinement of Proposition 7, in which the Lipschitz constants are improved by using the convex optimization structure of the problem.

Proposition 9. In the context of Problem 2, suppose that f is ρ -strongly convex, for some $\rho \in]0, \alpha^{-1}[$, and let $\tau > 0$. Then, the following hold:

1. Suppose that $\tau \in]0, 2\beta\alpha/(\beta + \alpha)[$. Then, $G_{\tau(\nabla g + \nabla f)}$ is $r_G(\tau)$ -Lipschitz continuous, where

$$r_G(\tau) := \max \{ |1 - \tau\rho|, |1 - \tau(\beta^{-1} + \alpha^{-1})| \} \in]0, 1[. \quad (49)$$

In particular, the minimum in (49) is achieved at

$$\tau^* = \frac{2}{\rho + \alpha^{-1} + \beta^{-1}} \quad (50)$$

and

$$r_G(\tau^*) = \frac{\alpha^{-1} + \beta^{-1} - \rho}{\alpha^{-1} + \beta^{-1} + \rho}. \quad (51)$$

2. Suppose that $\tau \in]0, 2\alpha[$. Then $T_{\tau \nabla g, \tau \nabla f}$ is $r_{T_1}(\tau)$ -Lipschitz continuous, where

$$r_{T_1}(\tau) := \max \{ |1 - \tau\rho|, |1 - \tau\alpha^{-1}| \} \in]0, 1[. \quad (52)$$

In particular, the minimum in (52) is achieved at

$$\tau^* = \frac{2}{\rho + \alpha^{-1}} \quad \text{and} \quad r_{T_1}(\tau^*) = \frac{\alpha^{-1} - \rho}{\alpha^{-1} + \rho}. \quad (53)$$

3. Suppose that $\tau \in]0, 2\beta[$. Then $T_{\tau \nabla f, \tau \nabla g}$ is $r_{T_2}(\tau)$ -Lipschitz continuous, where

$$r_{T_2}(\tau) := \frac{1}{1 + \tau\rho} \in]0, 1[. \quad (54)$$

In particular, the minimum in (54) is achieved at

$$\tau^* = 2\beta \quad \text{and} \quad r_{T_2}(\tau^*) = \frac{1}{1 + 2\beta\rho}. \quad (55)$$

4. $R_{\tau \nabla g, \tau \nabla f}$ and $R_{\tau \nabla f, \tau \nabla g}$ are $r_R(\tau)$ -Lipschitz continuous, where

$$r_R(\tau) = \max \left\{ \frac{1 - \tau\rho}{1 + \tau\rho}, \frac{\tau\alpha^{-1} - 1}{\tau\alpha^{-1} + 1} \right\} \in]0, 1[. \quad (56)$$

In particular, the minimum in (56) is achieved at

$$\tau^* = \sqrt{\frac{\alpha}{\rho}} \quad \text{and} \quad r_R(\tau^*) = \frac{1 - \sqrt{\alpha\rho}}{1 + \sqrt{\alpha\rho}}. \quad (57)$$

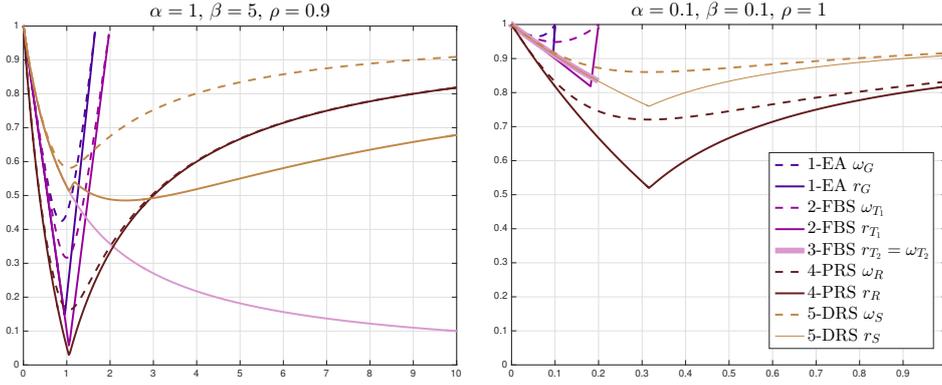


Figure 2: Comparison of the convergence rates of EA, FBS, PRS, DRS obtained in Proposition 9 (continuous lines) and Proposition 7 (dashed lines) for two choices of α , β , and ρ . Note that optimization rates are better than cocoercive rates in general.

5. $S_{\tau\nabla g, \tau\nabla f}$ and $S_{\tau\nabla f, \tau\nabla g}$ are $r_S(\tau)$ -Lipschitz continuous, where

$$r_S(\tau) = \min \left\{ \frac{1 + r_R(\tau)}{2}, \frac{\beta + \tau^2 \rho}{\beta + \tau \beta \rho + \tau^2 \rho} \right\} \in]0, 1[\quad (58)$$

and r_R is defined in (56). In particular, the optimal step-size and the minimum in (58) are

$$(\tau^*, r_S(\tau^*)) = \begin{cases} \left(\sqrt{\frac{\alpha}{\rho}}, \frac{1}{1 + \sqrt{\alpha\rho}} \right), & \text{if } \beta \leq 4\alpha; \\ \left(\sqrt{\frac{\beta}{\rho}}, \frac{2}{2 + \sqrt{\beta\rho}} \right), & \text{otherwise.} \end{cases} \quad (59)$$

The Lipschitz constant of the operators $G_{\nabla g + \nabla f}$ and $T_{\nabla g, \nabla f}$ are consequence of [14, Theorem 3.1] (see also [17, Fact 3] for a geometric interpretation). We provide an alternative shorter and more direct proof of Proposition 9(1)-(2) in Appendix 8, in which we use some techniques from [45, Section 2.1.3]. The Lipschitz constant of $T_{\nabla f, \nabla g}$ is a direct consequence of Proposition 7(3) and (56) is obtained in [15, Theorem 2], which improves several constants in the literature. The Lipschitz constant in (58) is obtained by combining [15, Theorem 2] and [28, Theorem 5.6].

Remark 2. 1. When $\rho \approx 0$ the choice of (53) justifies the classical choice $\tau^* \approx 2\alpha$. This case arises naturally in several inverse problems and, in particular, in sparse image restoration which is studied in detail in Section 5.3.

2. Note that the Lipschitz continuous constants obtained in Proposition 9(1) and 9(2) are strictly lower than the constants obtained in Proposition 7(1) and 7(2) in the cocoercive case, as it can be verified in Figure 2.

3. Figures 3 and 2 illustrate the Lipschitz constants in Proposition 9. From Figure 3 (first row), we can observe that for α and β fixed, the larger is the strong monotony constant ρ , the better is the convergence rate. Additionally, for α and β fixed, Peaceman-Rachford iterations $R_{\tau\nabla g, \tau\nabla f}$ and the forward-backward iterations $T_{\tau\nabla g, \tau\nabla f}$ and $T_{\tau\nabla f, \tau\nabla g}$ are the algorithms achieving the best convergence rates.

From the second row of Figure 3, we conclude that the smaller is the Lipschitz constant of the strongly convex function (larger is α), the better is the convergence rate at exception of $T_{\tau\nabla f, \tau\nabla g}$. In the last case, the Lipschitz constant depends only on the strongly convex parameter ρ . Once again, for ρ and β fixed, we observe that the Peaceman-Rachford iterations $R_{\tau\nabla g, \tau\nabla f}$ and the forward-backward iterations $T_{\tau\nabla g, \tau\nabla f}$ achieve the best convergence rate.

From the third row of Figure 3, where ρ and α are fixed, we observe that a smaller Lipschitz constant of ∇g (larger β) affects positively to $G_{\tau\nabla g, \tau\nabla f}$, $T_{\tau\nabla f, \tau\nabla g}$, and $S_{\tau\nabla g, \tau\nabla f}$. Last convergence rate takes advantage of the fully smooth context of Problem 2 and it is a new result.

From Figure 2, we observe the benefit of the refinement of convergence rates in the optimization framework (dashed line) with respect to the cocoercive case (solid line) in all methods at exception of $T_{\tau\nabla f, \tau\nabla g}$, whose rate is the same. We also observe that in general Peaceman-Rachford iterations $R_{\tau\nabla g, \tau\nabla f}$ has the better convergence rate for several configurations of (α, β, ρ) .

In the case when $g = 0 \in \mathcal{C}_0^{1,1}(\mathcal{H})$, Problem 2 reduces to minimize f over \mathcal{H} and $G_{\nabla g + \nabla f} = T_{\nabla g, \nabla f} = G_{\nabla f}$ and $T_{\nabla f, \nabla g} = S_{\nabla f, \nabla g} = S_{\nabla g, \nabla f} = \text{prox}_f$. Therefore, by taking $\beta \rightarrow +\infty$ in Proposition 9, we recover the following known results (see also [28, Proposition 5.2] and [3, Proposition 4.39]).

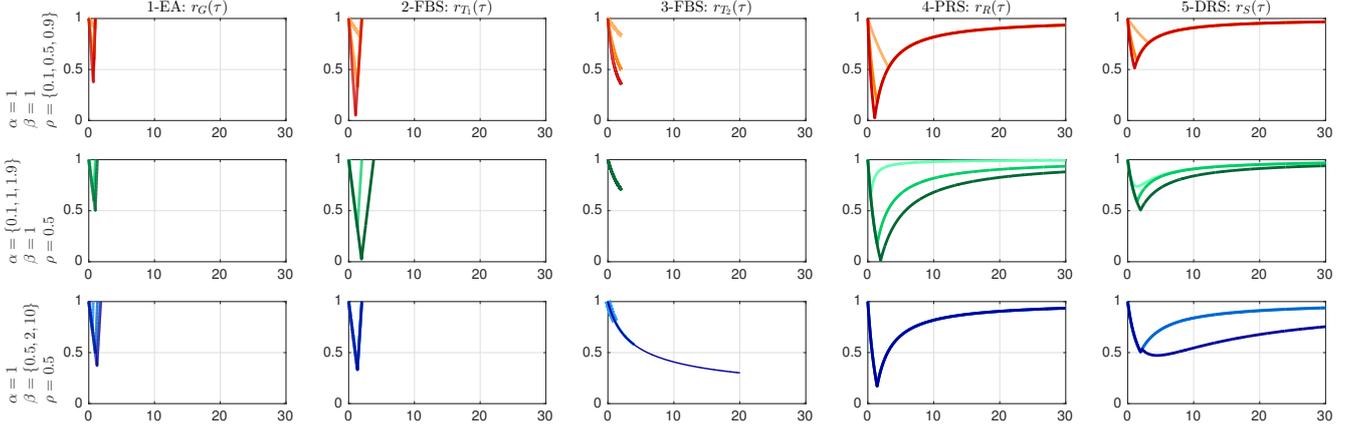


Figure 3: Convergence rate (Lipschitz continuous constant) provided in Proposition 9 w.r.t. the step-size parameter τ for different choices of α , β , and ρ .

Proposition 10. *Let $\tau \in]0, +\infty[$, $\alpha \in]0, +\infty[$, $\rho \in]0, +\infty[$, and suppose that $f \in \mathcal{C}_{1/\alpha}^{1,1}(\mathcal{H})$ and that f is ρ -strongly convex. Then, the following hold.*

1. *Suppose that $\tau \in]0, 2\alpha[$. Then $G_{\tau \nabla f}$ is $r_{G_0}(\tau)$ -Lipschitz continuous, where*

$$r_{G_0}(\tau) := \max \{ |1 - \tau\rho|, |1 - \tau\alpha^{-1}| \} \in]0, 1[. \quad (60)$$

2. *$\text{prox}_{\tau f}$ is $r_J(\tau)$ -Lipschitz continuous, where*

$$r_J(\tau) := \frac{1}{1 + \tau\rho} \in]0, 1[. \quad (61)$$

Remark 3. *Note that ω_{G_0} obtained in Proposition 8 in the cocoercive operator context achieve its optimal value in $\tau^* = \alpha$, in which case $\omega_{G_0}(\alpha) = \sqrt{1 - \rho\alpha}$. In the convex optimization context, r_{G_0} is strictly lower than ω_{G_0} , $\tau^* = 2/(\rho + \alpha^{-1})$ and $r_{G_0}(\tau^*) = (\rho\alpha - 1)/(\rho\alpha + 1)$. On the other hand, from Proposition 8 and Proposition 10 we have $r_J = \omega_J$ and, since $\lim_{\tau \rightarrow +\infty} r_J(\tau) = 0$, the larger the choice of τ , the better the convergence rate of resolvent iterations. In the Section 5.1 we compare Lipschitz constants r_{G_0} , ω_{G_0} , and $r_J = \omega_J$ in the particular case of ordinary least squares in order to illustrate the pros and cons of each approach.*

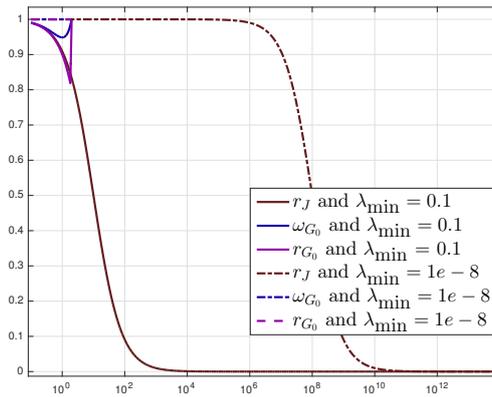


Figure 4: Comparison of strict contraction constants (i.e. linear convergence rate) ω_{G_0} , r_{G_0} , and r_J (prox) w.r.t τ for different values of $\kappa = \frac{\lambda_{\min}}{\lambda_{\max}}$ when $\lambda_{\max} = 1$ and $\lambda_{\min} = \{0.1, 1e-8\}$. We can observe that the smaller is λ_{\min} the closer to 1 is the convergence rate of the gradient descent. However, for sufficiently large step-size parameter τ , the proximal point algorithm achieves very good rate (close to 0)

5 Numerical experiments

The theoretical results provided in the previous sections are now illustrated on standard data processing examples with different level of complexity: Ordinary least squares, piecewise-constant denoising, and image restoration.

5.1 Ordinary least squares

Set $\mathcal{H} = \mathbb{R}^N$ and suppose that $f = \frac{1}{2}\|A \cdot -a\|^2$, where A is a $M \times N$ real matrix and $a \in \mathbb{R}^M$. We denote by λ_{\min} (resp. λ_{\max}) the smallest (resp. largest) eigenvalue of $A^\top A$ and we suppose that $\lambda_{\max} > \lambda_{\min} \geq 0$. Then, since $\nabla f: x \mapsto A^\top(Ax - a)$ is λ_{\max} -Lipschitz continuous and λ_{\min} -strongly monotone, we have that $f \in \mathcal{C}_{\lambda_{\max}}^{1,1}(\mathbb{R}^N)$ and that f is λ_{\min} -strongly convex (if $\lambda_{\min} = 0$ it is just convex). Hence, the problem of minimizing f over \mathcal{H} is equivalent to solve the classical least-squares problem

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \frac{1}{2}\|Ax - a\|^2. \quad (62)$$

In this context, for every $\tau > 0$, we have for every $\tau \in]0, 2/\lambda_{\max}[$, $G_{\tau \nabla f}: x \mapsto (\text{Id} - \tau A^\top A)x + \tau A^\top a$ and $\text{prox}_{\tau f}: x \mapsto (\text{Id} + \tau A^\top A)^{-1}(x + \tau A^\top a)$. Therefore, it is clear from their affine linear structure that $G_{\tau \nabla f}$ and $\text{prox}_{\tau f}$ are Lipschitz continuous with constants, for every $\tau \in]0, 2/\lambda_{\max}[$,

$$\|\text{Id} - \tau A^\top A\| = \max\{|1 - \tau \lambda_{\min}|, |1 - \tau \lambda_{\max}|\} \in]0, 1], \quad (63)$$

and, for every $\tau > 0$,

$$\|(\text{Id} + \tau A^\top A)^{-1}\| = (1 + \tau \lambda_{\min})^{-1} \in]0, 1] \quad (64)$$

respectively, and the constants are strictly less than 1 in the case when $\ker A = \{0\}$ ($\lambda_{\min} > 0$). Note that, constants in (63) and (64) coincide with the theoretical constants $r_{G_0}(\tau)$ and $r_J(\tau)$ in Proposition 10. Moreover, the minimum value of r_{G_0} is attained at $\tau^* = 2/(\lambda_{\max} + \lambda_{\min})$, in which case $r_{G_0}(\tau^*) = (\kappa - 1)/(\kappa + 1)$, where $\kappa = \lambda_{\min}/\lambda_{\max}$ is the condition number of $A^\top A$. This constant is strictly lower than $\min \omega_{G_0} = \sqrt{1 - \kappa}$, where ω_{G_0} is defined in Proposition 8 as discussed in Remark 3. A comparison of the behaviour of the constants is illustrated in Figure 4. An important advantage of resolvent iterations is that $\inf r_J = \lim_{\tau \rightarrow +\infty} r_J(\tau) = 0$ and, therefore, the strict contraction constant of $J_{\tau A}$ can be arbitrarily small as τ increases while $\min r_{G_0} = (\kappa - 1)/(\kappa + 1) > 0$ can be close to 1 in the case of bad conditioned problems.

Next, we compare in Figure 5 the theoretical bounds described in Proposition 10 with the numerical behaviour of Banach-Picard iterations governed by $G_{\tau \nabla f}$ and $\text{prox}_{\tau f}$ for some $\tau > 0$, in the context of (62) when A models the matrix associated with a standard 2D uniform convolution periodic filter of size 3×3 . In this case, we have a bad conditioned matrix $A^\top A$, but $\lambda_{\min} > 0$ and $\kappa = 5.87 \cdot 10^{-6}$ (resp. $1.97 \cdot 10^{-7}$) when $N = 400$ (resp. $N = 2500$).

From this experimental results, we confirm that both numerically and theoretically, the largest is the proximal step, the fastest is the convergence. Moreover, the benefit in terms of iterates of the proximal step compared to gradient descent step is clearly illustrated through these experiments.

According to the theoretical and numerical results provided in this section, proximal point algorithm clearly appears to achieve better convergence rate even for badly conditioned matrices with full rank. However, an efficient practical implementation does not only rely on the convergence rate but on the cost of each iterations.

The inversion involved in the proximity operator in the context of OLS is always possible, but it can be computationally costly and needs to be performed efficiently in order to have similar benefit in time. When the inversion in proximity operator step is performed by using Matlab inversion is very efficient at small size, else it is equivalent to Backslash strategy. When dimensionality start to be high, an efficient inversion should exploit the circulant form of the operator A in order to invert it considering Fourier diagonalization. Considering, periodic boundary effects, this type of inversion is always possible for time/space-invariant filters.

5.2 Piecewise constant denoising

Piecewise constant denoising (also referred as change-point detection) is a very well documented problem of signal processing literature and it is of interest for numerous signal processing application going from genomics [46] to geophysics studies [47].

The standard formulation is dedicated to piecewise constant signal $\bar{x} \in \mathbb{R}^N$ degraded with a Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$, whose degraded version is denoted $z = \bar{x} + \varepsilon$. An illustration of \bar{x} (resp. z) is provided in solid black line (resp. gray) in Figure 6 (top).

The estimation of a piecewise constant signal \hat{x} from degraded data z has been addressed by several strategies going from Cusum procedures [48], hierarchical Bayesian inference frameworks [49], or functional optimization

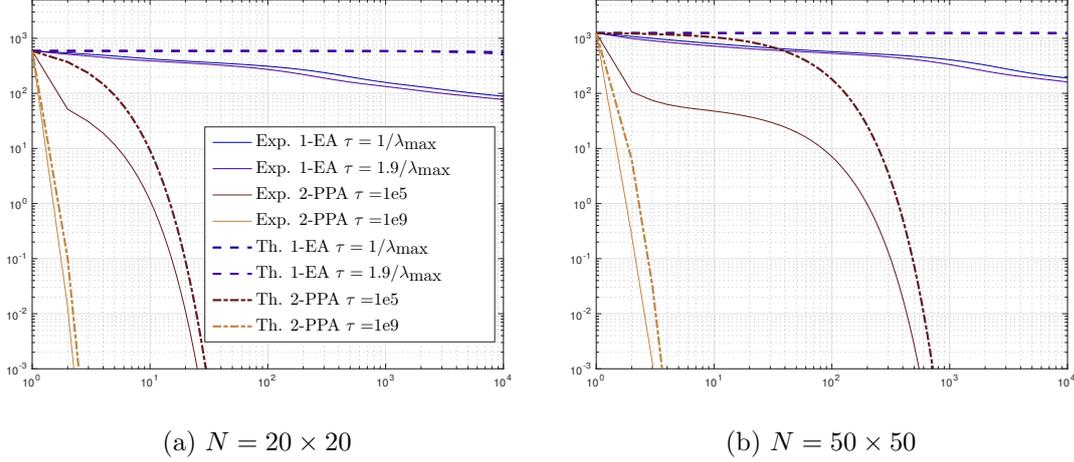


Figure 5: Numerical and theoretical comparisons of PPA and EA for several choices of step-size parameter τ and different sizes of images when A models a 2D periodic filtering associated with a uniform blur of size 7×7 leading to $\lambda_{\min} = \{5.87 \cdot 10^{-6}, 1.97 \cdot 10^{-7}\}$ and $\lambda_{\max} = 1$.

formulations involving ℓ_1 -norm or the ℓ_0 -pseudo-norm of the first differences of the signal (see e.g. [50] and references therein). In the latter context, we consider the minimization problem:

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \frac{1}{2} \|x - z\|_2^2 + \chi h(Lx), \quad (65)$$

where $L \in \mathbb{R}^{N-1 \times N}$ denotes the first order discrete difference operator

$$(\forall n \in \{1, \dots, N-1\}) \quad (Lx)_n = \frac{1}{2}(x_n - x_{n-1})$$

and h denotes the Huber loss, the smooth approximation of the ℓ_1 -norm parametrized by $\mu > 0$, defined by (see, e.g., [13, Example 2.5])

$$h : \mathbb{R}^{N-1} \rightarrow \mathbb{R} : (\zeta_i)_{1 \leq i \leq m} \mapsto \sum_{i=1}^{N-1} h_i(\zeta_i) \quad (66)$$

and

$$h_i : \zeta \mapsto \begin{cases} |\zeta| - \frac{\mu}{2}, & \text{if } |\zeta| > \mu; \\ \frac{|\zeta|^2}{2\mu}, & \text{if } |\zeta| \leq \mu. \end{cases} \quad (67)$$

Note that, since

$$h'_i : \zeta \mapsto \begin{cases} \frac{\zeta}{|\zeta|}, & \text{if } |\zeta| > \mu; \\ \frac{\zeta}{\mu}, & \text{if } |\zeta| \leq \mu, \end{cases}$$

we have $h \in \mathcal{C}_{1/\mu}^{1,1}(\mathbb{R}^{N-1})$. By setting $f = \frac{1}{2} \|\cdot - z\|_2^2$ and $g = \chi h(L\cdot)$, (65) is a particular instance of Problem 2, where f is $\rho = 1$ strongly convex, $\alpha = 1$, and $\beta = \frac{\mu}{\chi \|L\|^2}$ and it can be solved by the following two explicit schemes:

- 1- **EA**: Use $G_{\tau(\nabla g + \nabla f)}$ with the step-size τ^* in (50).
- 2- **FBS**: Use $T_{\tau \nabla f, \tau \nabla g}$ with the step-size τ^* in (55).

Moreover, the proximity operator of h can be computed explicitly via

$$\text{prox}_{\tau h} : (\zeta_i)_{1 \leq i \leq m} \mapsto (\text{prox}_{\tau \phi} \zeta_i)_{1 \leq i \leq m} \quad (68)$$

for some $\tau > 0$, where

$$\text{prox}_{\tau \phi} : \zeta \mapsto \begin{cases} \zeta - \frac{\tau \zeta}{|\zeta|}, & \text{if } |\zeta| > \tau + \mu; \\ \frac{\mu \zeta}{\tau + \mu}, & \text{if } |\zeta| \leq \tau + \mu, \end{cases} \quad (69)$$

but the proximity operator of $h \circ L$ is not explicit because of the influence of operator L . By exploiting the separable structure of h , we obtain the following equivalent formulation of (65):

$$\min_{x \in \mathcal{H}} \frac{1}{2} \|x - z\|_2^2 + \chi h_{\mathbb{T}_1}(L_{\mathbb{T}_1} x) + \chi h_{\mathbb{T}_2}(L_{\mathbb{T}_2} x), \quad (70)$$

where $\mathbb{I}_1 = \{1, 3, \dots\}$ and $\mathbb{I}_2 = \{2, 4, \dots\}$ are the sets of odd and even indices and, for $k \in \{1, 2\}$, $h_{\mathbb{I}_k}(y_{\mathbb{I}_k}) = \sum_{i \in \mathbb{I}_k} h_i(y_i)$, and $L_{\mathbb{I}_k} \in \mathbb{R}^{|\mathbb{I}_k| \times N}$ denotes the sub-matrix of L associated with the \mathbb{I}_k rows. Since $L_{\mathbb{I}_1} L_{\mathbb{I}_1}^\top = \text{Id}/2$ and $L_{\mathbb{I}_2} L_{\mathbb{I}_2}^\top = \text{Id}/2$ the split formulation (70) allows for the following closed form expressions of the proximity operator of $h_{\mathbb{I}_k} \circ L_{\mathbb{I}_k}$, for $k \in \{1, 2\}$ (see Proposition 6(2)), for every $\tau > 0$,

$$\text{prox}_{\tau h_{\mathbb{I}_k} \circ L_{\mathbb{I}_k}} : z \mapsto z - 2L_{\mathbb{I}_k}^\top (\text{Id} - J_{\frac{\tau}{2} \nabla h_{\mathbb{I}_k}})(L_{\mathbb{I}_k} z),$$

where $J_{\frac{\tau}{2} \nabla h_{\mathbb{I}_k}} : (\zeta_i)_{i \in \mathbb{I}_k} \mapsto (\text{prox}_{\frac{\tau}{2} \phi} \zeta_i)_{i \in \mathbb{I}_k}$. By setting $\tilde{f} = \frac{1}{2} \|\cdot - z\|_2^2 + \chi h_{\mathbb{I}_2}(L_{\mathbb{I}_2} \cdot)$ and $\tilde{g} = \chi h_{\mathbb{I}_1}(L_{\mathbb{I}_1} \cdot)$, we write (70) as Problem 2, where \tilde{f} is $\rho = 1$ strongly convex, $\alpha = \frac{\mu}{\mu + \chi \|L_{\mathbb{I}_2}\|^2}$, and $\beta = \frac{\mu}{\chi \|L_{\mathbb{I}_1}\|^2}$. This approach gives raise to 4 alternative methods for solving (70).

3- **FBS 2:** Use $T_{\tau \nabla \tilde{g}, \tau \nabla \tilde{f}}$ with the step-size τ^* in (53).

4- **FBS 3:** Use $T_{\tau \nabla \tilde{f}, \tau \nabla \tilde{g}}$ with the step-size τ^* in (55).

5- **PRS:** Use $R_{\tau \nabla \tilde{f}, \tau \nabla \tilde{g}}$ with the step-size τ^* in (57).

6- **DRS:** Use $S_{\tau \nabla \tilde{f}, \tau \nabla \tilde{g}}$ with the step-size τ^* in (59).

We consider an approximation of the unique solution \hat{x} to (65), by applying PRS with a large number of iterations. In view of Section 2.4, 1-EA, 2-FBS, 3-FBS2, and 4-FBS3 are initialized with $x_0 = z$, while using

$$z = \text{prox}_{\gamma f}(x_n) \Leftrightarrow (\text{Id} + \gamma \nabla f) y_n = x_n$$

proximal based procedures 5-PRS and 6-DRS are initialized by $x_0 = z + \tau \nabla f(z)$, in order to provide similar initializations.

The numerical and theoretical convergence rate are displayed in Figures 6 for different settings of μ and χ leading to sharper or smoother estimates depending of the configuration. When $\mu = 10^{-4}$ the performance are similar than what is expected for ℓ_1 -minimization.

From Figure 6 (bottom), we can observe that PRS iterations provides the best theoretical and experimental rates when the optimal step-size is selected. DRS iterations also provides a good behaviour, while EA and FBS strategies relying on the splitting $f = \frac{1}{2} \|\cdot - z\|_2^2$ and $g = \chi h(L \cdot)$ appears less efficient than the one involving the splitting $\tilde{f} = \frac{1}{2} \|\cdot - z\|_2^2 + \chi h_{\mathbb{I}_2}(L_{\mathbb{I}_2} \cdot)$ and $\tilde{g} = \chi h_{\mathbb{I}_1}(L_{\mathbb{I}_1} \cdot)$. Similar conclusion can be observed from Figure 6 (top), where the optimal solution is reached after 100 iterations for DRS (light brown) and PRS (dark brown) while gradient based procedures require much more iterations. This is especially true when μ is small, leading to a large Lipschitz constant.

5.3 Image restoration

Another classical image processing problem is image restoration that consists in recovering an image $\bar{x} \in \mathbb{R}^N$ with N pixels from degraded observations $z = A\bar{x} + \varepsilon$, degraded by a linear degradation $A \in \mathbb{R}^{M \times N}$ and a white Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$. When data are assumed to be sparse, as commonly encountered in astrophysics or microscopy, the restoration can be achieved by solving

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad \frac{1}{2} \|Ax - z\|_2^2 + \chi h(x), \quad (71)$$

where $\chi > 0$ denotes the regularization parameter and h the Huber penalization as defined in (66).

We evaluate the theoretical and the experimental rates for several algorithmic scheme when $f = \frac{1}{2} \|A \cdot - z\|_2^2$ and $g = \chi h$, leading to $\rho = \lambda_{\min}$ strongly convex, $\alpha = \lambda_{\max}^{-1}$, and $\beta = \frac{\mu}{\chi}$

1- **EA:** Use $G_{\tau(\nabla g + \nabla f)}$ with the step-size τ^* in (50).

2- **FBS:** Use $T_{\tau \nabla g, \tau \nabla f}$ with the step-size τ^* in (53).

3- **FBS:** Use $T_{\tau \nabla f, \tau \nabla g}$ with the step-size τ^* in (55).

4- **PRS:** Use $R_{\tau \nabla f, \tau \nabla g}$ with the step-size τ^* in (57).

5- **DRS:** Use $S_{\tau \nabla f, \tau \nabla g}$ with the step-size τ^* in (59).

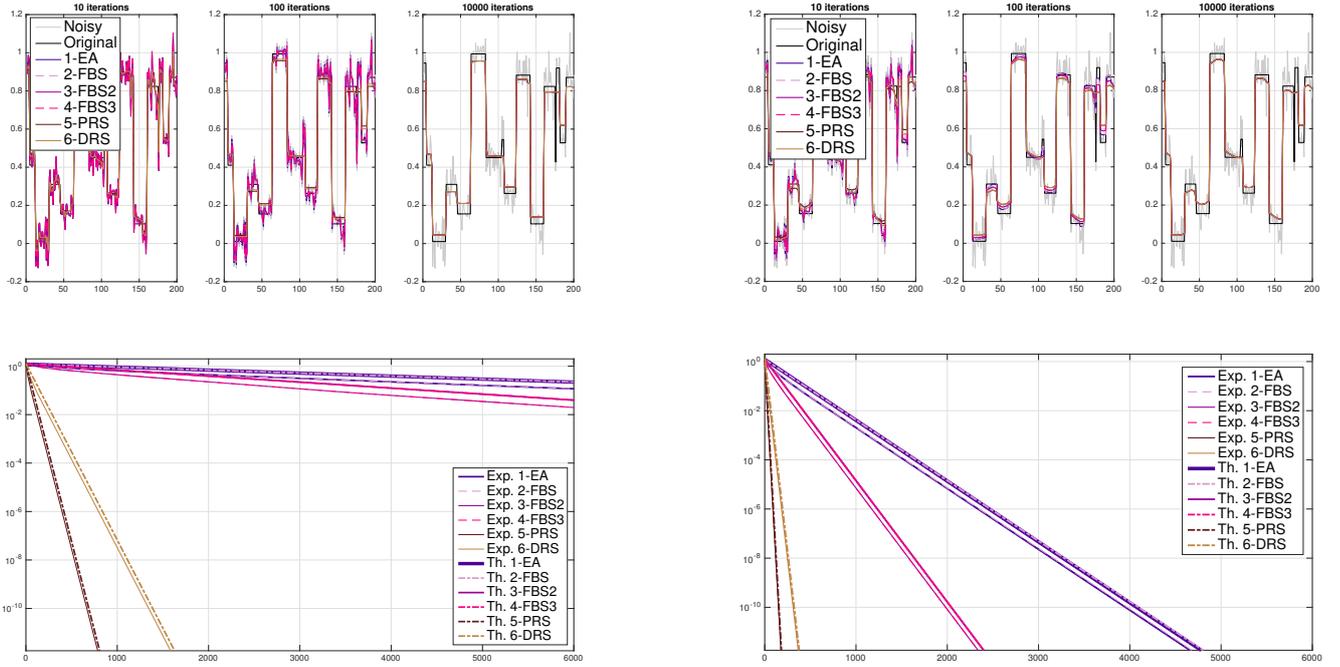


Figure 6: Piecewise constant denoising estimates after 10, 100, and 10000 iterations with $\chi = 0.7$ and $\mu = 0.0001$ (top, left) and $\chi = 0.7$ and $\mu = 0.002$ (top, right). We can observe that the piecewise constant estimate is obtained after 100 iterations for DRS or PRS while EA or FBS requires much more iterations (bottom). We also exhibit the experimental and theoretical rates associated with each implemented methods for optimal step-size τ . The behaviour is in accordance to the results observed on the first row.

The results are displayed in Figure 7 on an image with $N = 3600$ pixels when A models the matrix associated with a standard 2D uniform convolution periodic filter of size 7×7 . In this case, we have a bad conditioned matrix $A^T A$, but $\lambda_{\min} = 8.5 \cdot 10^{-9} > 0$ and $\lambda_{\max} = 1$. The results are obtained considering $\chi = 0.005$ and $\mu = 0.01$ and the images are displayed in Figure 7(top) leading to a sparse solution despite the fact that an approximation of the ℓ_1 -norm is considered. The initialization has been set as in Section 5.2 for all algorithmic schemes.

From Figure 7 (bottom), we observe that the strong convexity constant is so small that the theoretical rate cannot be usable (all close to 1) in practice to compare methods, which illustrates the limit of strong convexity on standard inverse problems. For this reason appropriate choice of τ is selected manually for each method. We can observe that full proximal strategy (4-PRS, 5-PRS) stay the more efficient.

References

- [1] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet, “Wavelet-based image deconvolution and reconstruction,” *Wiley Encyclopedia of EEE*, 2016.
- [2] J.A. Fessler, “Optimization methods for MR image reconstruction,” *IEEE Sig. Proc. Mag.*, vol. 37, no. 1, pp. 33–40, 2020.
- [3] H.H. Bauschke and P.L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2017.
- [4] P.L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, (H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Editors), Springer, pp. 185–212, 2011.
- [5] N. Parikh and S. Boyd, “Proximal algorithms,” *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [6] A. Cauchy, “Méthode générale pour la résolution des systèmes d’équations simultanées,” *C. R. Acad. Sci. Paris*, vol. 22, pp. 536–538, 1847.

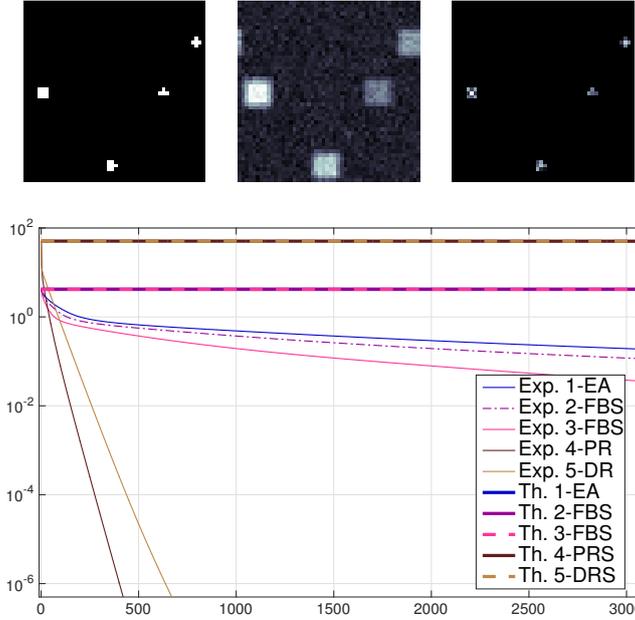


Figure 7: (top) From left to right: Original image \bar{x} , degraded image z , restored image \hat{x} when $\chi = 0.005$ and $\mu = 0.01$. (bottom) Experimental and theoretical rates associated with each implemented methods for optimal step-size τ when considering gradient, FB2 and FB3 while $\tau = 115$ for PR and DR.

- [7] H. B. Curry, “The method of steepest descent for non-linear minimization problems,” *Quart. Appl. Math.*, vol. 2, pp. 258–261, 1944.
- [8] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.
- [9] B. Mercier, *Inéquations variationnelles de la mécanique*, vol. 1 of *Publications Mathématiques d’Orsay 80 [Mathematical Publications of Orsay 80]*, Université de Paris-Sud, Département de Mathématique, Orsay, 1980.
- [10] P.-L. Lions and B. Mercier, “Splitting algorithms for the sum of two nonlinear operators,” *SIAM J. Numer. Anal.*, vol. 16, no. 6, pp. 964–979, 1979.
- [11] J. Eckstein and D. P. Bertsekas, “On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators,” *Math. Program.*, vol. 55, pp. 293–318, 1992.
- [12] P. L. Combettes and J.-C. Pesquet, “A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 564–574, 2007.
- [13] P. L. Combettes and L. E. Glaudin, “Proximal activation of smooth functions in splitting algorithms for convex image recovery,” *SIAM J. Imaging Sci.*, vol. 12, no. 4, pp. 1905–1935, 2019.
- [14] A. B. Taylor, J. M. Hendrickx, and F. Glineur, “Exact worst-case convergence rates of the proximal gradient method for composite convex minimization,” *J. Optim. Theory Appl.*, vol. 178, no. 2, pp. 455–476, 2018.
- [15] P. Giselsson and S. Boyd, “Linear convergence and metric selection for Douglas-Rachford splitting and ADMM,” *IEEE Trans. Automat. Control*, vol. 62, no. 2, pp. 532–544, 2017.
- [16] D. Davis and W. Yin, “Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions,” *Math. Oper. Res.*, vol. 42, no. 3, pp. 783–805, 2017.
- [17] E. K. Ryu, R. Hannah, and W. Yin, “Scaled relative graph: Nonexpansive operators via 2D euclidean geometry,” <https://arxiv.org/pdf/1902.09788>, 2019.
- [18] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson, “Operator splitting performance estimation: Tight contraction factors and optimal parameter selection,” *SIAM J. Optim.*, vol. 30, no. 3, pp. 2251–2271, 2020.
- [19] A. S. Lewis, “Active sets, nonsmoothness, and sensitivity,” *SIAM J. Optim.*, vol. 13, no. 3, pp. 702–725 (2003), 2002.

- [20] J. Liang, J. Fadili, and G. Peyré, “Activity identification and local linear convergence of forward-backward-type methods,” *SIAM J. Optim.*, vol. 27, no. 1, pp. 408–437, 2017.
- [21] H. Attouch, J. Bolte, and B. F. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods,” *Math. Program.*, vol. 137, no. 1-2, Ser. A, pp. 91–129, 2013.
- [22] J. Bolte, A. Daniilidis, and A. Lewis, “The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems,” *SIAM J. Optim.*, vol. 17, no. 4, pp. 1205–1223, 2006.
- [23] J. Bolte, Trong P. Nguyen, J. Peypouquet, and B. W. Suter, “From error bounds to the complexity of first-order descent methods for convex functions,” *Math. Program.*, vol. 165, no. 2, Ser. A, pp. 471–507, 2017.
- [24] L. M. Briceño-Arias and P. L. Combettes, “A monotone + skew splitting model for composite monotone inclusions in duality,” *SIAM J. Optim.*, vol. 21, no. 4, pp. 1230–1250, 2011.
- [25] B. He and X. Yuan, “Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective,” *SIAM J. Imaging Sci.*, vol. 5, no. 1, pp. 119–149, 2012.
- [26] L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.
- [27] B. C. Vũ, “A splitting algorithm for dual monotone inclusions involving cocoercive operators,” *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, 2013.
- [28] P. Giselsson, “Tight global linear convergence rate bounds for Douglas–Rachford splitting,” *Journal of Fixed Point Theory and Applications*, vol. 19, no. 4, pp. 2241–2270, 2017.
- [29] G. H.-G. Chen and R. T. Rockafellar, “Convergence rates in forward-backward splitting,” *SIAM J. Optim.*, vol. 7, no. 2, pp. 421–444, 1997.
- [30] P. Tseng, “Applications of a splitting algorithm to decomposition in convex programming and variational inequalities,” *SIAM J. Control Optim.*, vol. 29, no. 1, pp. 119–138, 1991.
- [31] L. M. Briceño Arias and P. L. Combettes, “Monotone operator methods for Nash equilibria in non-potential games,” in *Computational and analytical mathematics*, vol. 50 of *Springer Proc. Math. Stat.*, pp. 143–159. Springer, New York, 2013.
- [32] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, “Proximal algorithms for multi-component image processing,” *J. Math. Imaging Vision*, vol. 41, no. 1, pp. 3–22, 2011.
- [33] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, “Image restoration: Total variation, wavelet frames, and beyond,” *Journal of the American Mathematical Society*, vol. 25, pp. 1033–1089, 2012.
- [34] J. Peypouquet and S. Sorin, “Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time,” *J. Convex Anal.*, vol. 17, no. 3-4, pp. 1113–1163, 2010.
- [35] B. Martinet, “Régularisation d’inéquations variationnelles par approximations successives,” *Rev. Française Informat. Recherche Opérationnelle*, vol. 4, no. Sér. R-3, pp. 154–158, 1970.
- [36] R. T. Rockafellar, “Monotone operators and the proximal point algorithm,” *SIAM J. Control Optim.*, vol. 14, no. 5, pp. 877–898, 1976.
- [37] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [38] E. S. Levitin and B. T. Polyak, “Constrained minimization methods,” *USSR Computational mathematics and mathematical physics*, vol. 6, no. 5, pp. 1–50, 1966.
- [39] H. Brezis and M. Sibony, “Méthodes d’approximation et d’itération pour les opérateurs monotones,” *Arch. Rational MEch. Anal.*, vol. 28, pp. 59–82, 1967/1968.
- [40] B. Mercier, *Lectures on topics in finite element solution of elliptic problems*, vol. 63 of *Tata Institute of Fundamental Research Lectures on Mathematics and Physics*, Tata Institute of Fundamental Research, Bombay, 1979.
- [41] M. Sibony, “Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone,” *Calcolo*, vol. 7, pp. 65–183, 1970.

- [42] D. W. Peaceman and H. H. Rachford, Jr., “The numerical solution of parabolic and elliptic differential equations,” *J. Soc. Indust. Appl. Math.*, vol. 3, pp. 28–41, 1955.
- [43] J. Douglas, Jr. and H. H. Rachford, Jr., “On the numerical solution of heat conduction problems in two and three space variables,” *Trans. Amer. Math. Soc.*, vol. 82, pp. 421–439, 1956.
- [44] N. Pustelnik, C. Chaux, and J.-C. Pesquet, “Parallel proXimal algorithm for image restoration using hybrid regularization,” *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2450–2462, Sep. 2011.
- [45] Y. Nesterov, *Introductory lectures on convex optimization*, vol. 87 of *Applied Optimization*, Kluwer Academic Publishers, Boston, MA, 2004.
- [46] J.-P. Vert and K. Bleakley, “Fast detection of multiple change-points shared by many signals using group LARS,” *Advances in Neural Information Processing Systems*, vol. 23, pp. 2343–2351, 2010.
- [47] B. Pascal, N. Pustelnik, P. Abry, J.-C. Géminard, and V. Vidal, “Parameter-free and fast nonlinear piecewise filtering. application to experimental physics,” *Annals of Telecommunications*, vol. 75, pp. 655–671, 2020.
- [48] M. Basseville and I.V. Nikiforov, “Detection of abrupt changes: Theory and application,” *Prentice-Hall, Inc., Upper Saddle River, NJ, USA*, 1993.
- [49] M. Lavielle and E. Lebarbier, “An application of MCMC methods for the multiple change-points problem,” *Signal Process.*, vol. 81, no. 1, pp. 39–53, 2001.
- [50] J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, and P. Abry, “Bayesian selection for the l2-potts model regularization parameter: 1D piecewise constant signal denoising,” *IEEE Trans. on Signal Processing*, vol. 65, no. 25, pp. 5215–5224, 2017.
- [51] P. L. Combettes and I. Yamada, “Compositions and convex combinations of averaged nonexpansive operators,” *J. Math. Anal. Appl.*, vol. 425, no. 1, pp. 55–70, 2015.

6 Proof of Proposition 4

1: It follows from [3, Proposition 4.12] that $\mathcal{A} + \mathcal{B}$ is $(\beta^{-1} + \alpha^{-1})^{-1} = \alpha\beta/(\alpha + \beta)$ -cocoercive. The result thus follows from [3, Proposition 4.39].

2: Since $\tau\mathcal{B}$ is β/τ -cocoercive, it follows from [28, Proposition 5.2] and [3, Proposition 4.39] that $J_{\tau\mathcal{B}}$ and $G_{\tau\mathcal{A}}$ are $\alpha_{\mathcal{B}} = \tau/(2(\tau + \beta)) \in]0, 1/2[$ and $\alpha_{\mathcal{A}} = \tau/(2\alpha)$ -averaged nonexpansive, respectively. Hence, we deduce from [51, Proposition 2.4] that $T_{\tau\mathcal{B}, \tau\mathcal{A}} = J_{\tau\mathcal{B}} \circ G_{\tau\mathcal{A}}$ is averaged with constant $(\alpha_{\mathcal{B}} + \alpha_{\mathcal{A}} - 2\alpha_{\mathcal{B}}\alpha_{\mathcal{A}})/(1 - \alpha_{\mathcal{B}}\alpha_{\mathcal{A}})$ which leads the result after simple computations.

3: Since $\tau\mathcal{A}$ and $\tau\mathcal{B}$ are α/τ - and β/τ -cocoercive, respectively, it follows from [28, Proposition 5.3] that $R_{\tau\mathcal{A}} = 2J_{\tau\mathcal{A}} - \text{Id}$ and $R_{\tau\mathcal{B}} = 2J_{\tau\mathcal{B}} - \text{Id}$ are $\tau/(\tau + \alpha)$ and $\tau/(\tau + \beta)$ averaged nonexpansive, respectively. Hence, since $R_{\tau\mathcal{B}, \tau\mathcal{A}} = R_{\tau\mathcal{B}} \circ R_{\tau\mathcal{A}}$, the averaging constant is obtained from [51, Proposition 2.4] as in 2.

4: Since $S_{\tau, \mathcal{B}, \mathcal{A}} = (\text{Id} + R_{\tau\mathcal{B}} \circ R_{\tau\mathcal{A}})/2$ we deduce the result from 3 and [3, Proposition 4.40].

7 Proof of Proposition 7

1: Set $\mathcal{M} = \mathcal{A} + \mathcal{B}$, fix $\tau \in]0, 2\beta\alpha/(\beta + \alpha)[\subset]0, 2\min\{\beta, \alpha\}[$, fix x and y in \mathcal{H} . From the ρ -strong monotonicity and α -cocoercivity of \mathcal{A} , we have, for every $\lambda \in]0, 1[$,

$$\begin{aligned} \langle \mathcal{M}x - \mathcal{M}y \mid x - y \rangle &= \langle \mathcal{B}x - \mathcal{B}y \mid x - y \rangle + \langle \mathcal{A}x - \mathcal{A}y \mid x - y \rangle \\ &\geq \beta \|\mathcal{B}x - \mathcal{B}y\|^2 + \lambda\alpha \|\mathcal{A}x - \mathcal{A}y\|^2 + (1 - \lambda)\rho \|x - y\|^2. \end{aligned}$$

Hence, noting that, for every $\varepsilon > 0$,

$$\|\mathcal{M}x - \mathcal{M}y\|^2 \leq (1 + \varepsilon) \|\mathcal{B}x - \mathcal{B}y\|^2 + (1 + \varepsilon^{-1}) \|\mathcal{A}x - \mathcal{A}y\|^2,$$

we deduce

$$\begin{aligned} \|G_{\tau\mathcal{M}}x - G_{\tau\mathcal{M}}y\|^2 &= \|x - y\|^2 - 2\tau \langle \mathcal{M}x - \mathcal{M}y \mid x - y \rangle + \tau^2 \|\mathcal{M}x - \mathcal{M}y\|^2 \\ &\leq \|x - y\|^2 - 2\tau\beta \|\mathcal{B}x - \mathcal{B}y\|^2 - 2\tau\lambda\alpha \|\mathcal{A}x - \mathcal{A}y\|^2 - 2\tau\rho(1 - \lambda) \|x - y\|^2 + \tau^2 \|\mathcal{M}x - \mathcal{M}y\|^2 \\ &\leq (1 - 2\tau\rho(1 - \lambda)) \|x - y\|^2 - \tau(2\beta - \tau(1 + \varepsilon)) \|\mathcal{B}x - \mathcal{B}y\|^2 - \tau(2\lambda\alpha - \tau(1 + \varepsilon^{-1})) \|\mathcal{A}x - \mathcal{A}y\|^2. \end{aligned}$$

Thus, the result follows by setting $\varepsilon = (2\beta - \tau)/\tau > 0$ and $\lambda = \frac{\tau\beta}{\alpha(2\beta - \tau)} \in]0, 1[$.

2: Fix $\tau \in]0, 2\alpha[$. It follows from 1 in the limit case when $\mathcal{B} = 0$ ($\beta \rightarrow +\infty$) that $G_{\tau\mathcal{A}}$ is $\omega_{T_1}(\tau)$ -Lipschitz continuous (see also [17, Fact 7]). Hence, the result follows from $T_{\tau\mathcal{B}, \tau\mathcal{A}} = J_{\tau\mathcal{B}} \circ G_{\tau\mathcal{A}}$ and the nonexpansivity of $J_{\tau\mathcal{B}}$.

3: Fix $\tau \in]0, 2\beta[$. It follows from [3, Proposition 23.13] that $J_{\tau\mathcal{A}}$ is $\omega_{T_2}(\tau)$ -Lipschitz continuous. The result follows from $T_{\tau\mathcal{A}, \tau\mathcal{B}} = J_{\tau\mathcal{A}} \circ G_{\tau\mathcal{B}}$ and the nonexpansivity of $G_{\tau\mathcal{B}}$, in view of Proposition 5(1).

4: First note that [28, Theorem 7.2] implies that $R_{\tau\mathcal{A}} = 2J_{\tau\mathcal{A}} - \text{Id}$ is $\omega_R(\tau)$ -Lipschitz continuous. Now, since $R_{\tau\mathcal{B}}$ is nonexpansive, we obtain that $R_{\tau\mathcal{B}}R_{\tau\mathcal{A}}$ and $R_{\tau\mathcal{A}}R_{\tau\mathcal{B}}$ are also $\omega_R(\tau)$ -Lipschitz continuous.

5: Since $S_{\tau\mathcal{B}, \tau\mathcal{A}} = (\text{Id} + R_{\tau\mathcal{B}, \tau\mathcal{A}})/2$ and $S_{\tau\mathcal{A}, \tau\mathcal{B}} = (\text{Id} + R_{\tau\mathcal{A}, \tau\mathcal{B}})/2$, this result is a consequence of [28, Lemma 3.3 & Theorem 5.6], and 4.

In all the cases, the minima are obtained via simple computations.

8 Proof of Proposition 9

1: Set $h = f + g$. Since g is convex and Fréchet differentiable and $f \in \mathcal{C}_{1/\beta}^{1,1}(\mathcal{H})$ is ρ -strongly convex, we obtain that $\phi = h - \rho\|\cdot\|^2/2$ is convex and Fréchet differentiable. Moreover, since ∇f and ∇g are α^{-1} -Lipschitz continuous and β^{-1} -Lipschitz continuous, we have that $\nabla h = \nabla f + \nabla g$ is γ^{-1} -Lipschitz continuous, where $\gamma^{-1} = \alpha^{-1} + \beta^{-1}$, and thus $h \in \mathcal{C}_{1/\gamma}^{1,1}(\mathcal{H})$ and it is convex. Hence, since $\gamma^{-1} = \alpha^{-1} + \beta^{-1} > \rho + \beta^{-1} \geq \rho$, it follows from Proposition 3 that, for every x and y in \mathcal{H} ,

$$\langle x - y | \nabla\phi(x) - \nabla\phi(y) \rangle = \langle x - y | \nabla h(x) - \nabla h(y) \rangle - \rho\|x - y\|^2 \leq (\gamma^{-1} - \rho)\|x - y\|^2,$$

which yields $\phi \in \mathcal{C}_{\gamma^{-1} - \rho}^{1,1}(\mathcal{H})$ in view of Proposition 3. In addition, we have

$$G_{\tau\nabla h} = \text{Id} - \tau(\nabla\phi + \rho\text{Id}) = (1 - \tau\rho)\text{Id} - \tau\nabla\phi. \quad (72)$$

Now let $\tau \in]0, 2\beta\alpha/(\beta + \alpha)[=]0, 2\gamma[$ and denote $p = G_{\tau\nabla h}x$ and $q = G_{\tau\nabla h}y$. Since $\phi \in \mathcal{C}_{\gamma^{-1} - \rho}^{1,1}(\mathcal{H})$ and it is convex, it follows from (72), Proposition 3, and $\nabla\phi \in \mathcal{C}_{(\gamma^{-1} - \rho)^{-1}}$ that

$$\begin{aligned} \|p - q\|^2 &= (1 - \tau\rho)^2\|x - y\|^2 + \tau^2\|\nabla\phi(x) - \nabla\phi(y)\|^2 - 2\tau(1 - \tau\rho)\langle x - y | \nabla\phi(x) - \nabla\phi(y) \rangle \\ &\leq (1 - \tau\rho)^2\|x - y\|^2 + \tau(\tau(\gamma^{-1} + \rho) - 2)\langle x - y | \nabla\phi(x) - \nabla\phi(y) \rangle \\ &\leq (1 - \tau\rho)^2\|x - y\|^2 + \tau \max\{0, \tau(\gamma^{-1} + \rho) - 2\}(\gamma^{-1} - \rho)\|x - y\|^2 \\ &= \|x - y\|^2 \max\{(1 - \tau\rho)^2, (1 - \tau\gamma^{-1})^2\} \end{aligned}$$

and we obtain (49).

2: Let $\tau \in]0, 2\alpha[$. It follows from 1 that, in the case when $g = 0$ ($\beta^{-1} = 0$), $G_{\tau\nabla f}$ is $r_{T_1}(\tau)$ -Lipschitz continuous, where $r_{T_1}(\tau)$ is defined in (52). The result follows from $T_{\tau\nabla g, \tau\nabla f} = \text{prox}_{\tau g} \circ G_{\tau\nabla f}$ and the nonexpansivity of $\text{prox}_{\tau g}$.

3: Let $\tau \in]0, 2\beta[$. We deduce from Proposition 8(2) that $J_{\tau\nabla f} = \text{prox}_{\tau f}$ is $r_{T_2}(\tau)$ -Lipschitz continuous, where $r_{T_2}(\tau)$ is defined in (54). The result follows from $T_{\tau\nabla f, \tau\nabla g} = \text{prox}_{\tau f} \circ G_{\tau\nabla g}$ and the nonexpansivity of $G_{\tau\nabla g}$ guaranteed by Proposition 5(1).

4: See [15, Theorem 2].

5: It is a consequence of [15, Theorem 2] and [28, Theorem 5.6] in the particular case when $\mathcal{A} = \nabla f$ and $\mathcal{B} = \nabla g$.

In all the cases, the minimum is obtained via simple computations.