Convergence Rate Comparison of Proximal Algorithms for Non-Smooth Convex Optimization With an Application to Texture Segmentation

L. M. Briceño-Arias and N. Pustelnik, *Member, IEEE*

***Abstract*—In this paper we provide a theoretical and numerical comparison of convergence rates of forward-backward, Douglas-Rachford, and Peaceman-Rachford algorithms for minimizing the sum of a convex proper lower semicontinuous function and a strongly convex differentiable function with Lipschitz continuous gradient. Our results extend the comparison made in [1], when both functions are smooth, to the context where only one is assumed differentiable. Optimal step-sizes and rates of the three algorithms are compared theoretically and numerically in the context of texture segmentation problem, obtaining very sharp estimations and illustrating the high efficiency of Peaceman-Rachford splitting.**

***Index Terms*—Convex optimization, convergence rates, proximal algorithms, texture segmentation.**

I. Introduction

DURING the last twenty years, a huge number of algorithms have been developed in order to perform signal/image analysis involving sparsity constraints. Most of these algorithms belong to the class of proximal algorithms including forward-backward iterations, Douglas-Rachford splitting, Alternating Direction Method of Multipliers (ADMM), primal-dual proximal schemes (see [2]–[4] and references therein).

The selection of an appropriate algorithm to minimize a specific objective function is a tedious task and its choice mainly depends on the properties of the involved functions (differentiability, Lipschitz gradient, closed form expression of the proximity operator) but also depends on the user expertise for a class of specific algorithmic schemes. However, a common practical rule is to consider a gradient activation step when the function is differentiable with a Lipschitz gradient and a proximal step when the function is not. Such a choice has no theoretical roots as very few works propose an exhaustive numerical comparison between proximal versus gradient step activation in the context of the inverse problems (see a contrario [1], [5]).

A fair comparison between numerical schemes relies on convergence rate analysis requiring to study strongly convex objective functions, as encountered in specific tasks of data processing such as signal/image denoising or texture segmentation [6], [7]. A tight convergence rate integrates the properties of the involved functions (strong convexity and Lipschitz constants) and the step-size parameter. From this general rate expression, we can derive the optimal step-size and the associated optimal convergence rate. The benefit of such a procedure is twofold as it allows us to identify the best step-size parameter, which is often a tedious task when done empirically, and it allows us to compare algorithm rates without requiring to implement them.

In [1], a deep analysis of gradient descent, forward-backward (FBS), Douglas-Rachford (DRS), and Peaceman-Rachford (PRS) schemes was provided in the context of a sum of two differentiable functions having Lipschitz gradient. In this context, [1] provides a parallel comparison of convergence rates of the method mentioned above recovering and improving some results in [8]–[15]. It provided numerical experiments allowing us to compare numerical and theoretical rates, highlighting the good behavior of fully proximal activation schemes, especially of Peaceman-Rachford scheme. However, the assumption of differentiability on both functions is quite restrictive in applications, which inspire us to extend our analysis to non-smooth objective functions. In this context the question of how larger (worse) are the convergence rates by removing the smoothness assumption on one function naturally appears.

Contribution: The present work is an extension of our previous contribution in [1] when we relax a differentiability assumption on one of the two functions. The convergence rate and optimal step-size are provided in this context and we observe that the convergence rates of FBS and PRS remain unchanged while that of DRS is larger in general. The theoretical bounds are compared to experimental ones in a context of scale-free texture segmentation involving a strong convex data-fidelity term.

Notation: Throughout this paper, $\mathcal{H}$ is a real Hilbert space endowed with the inner product $\langle \cdot \mid \cdot \rangle$ and associated norm $\|\cdot\| = \sqrt{\langle \cdot \mid \cdot \rangle}$. We denote by $\Gamma_0(\mathcal{H})$ the class of functions $h : \mathcal{H} \to]-\infty, +\infty]$ which are proper, lower semicontinuous (l.s.c), and convex. For every $L > 0$, we consider the class

Manuscript received April 19, 2022; accepted May 17, 2022. Date of publication June 8, 2022; date of current version June 17, 2022. The work of L. M. Briceño-Arias was supported in part by Centro de Modelamiento Matemático (CMM) under Grants ACE210010 and FB210005, BASAL funds for centers of excellence from ANID-Chile, in part by Grant Redes 180032, and in part by Grant FONDECYT 1190871 from ANID-Chile. The work of N. Pustelnik was supported by the Agence Nationale de la Recherche (ANR) from France under Grant ANR-19-CE48-0009 Multisc'In. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Shiwen He. *(Corresponding author: N. Pustelnik.)*

L. M. Briceño-Arias is with the Department of Mathematics, Universidad Técnica Federico Santa María, Santiago, RM 8940897, Chile (e-mail: luis.briceno@usm.cl).

N. Pustelnik is with the Univ Lyon, Ens de Lyon, Univ Lyon 1, CNRS, Laboratoire de Physique, 69342 Lyon, France , and with ISPGroup, INMA, ICTEAM Institute, UCLouvain, 1348 Louvain-la-Neuve, Belgium (e-mail: nelly.pustelnik@lens-lyon.fr).

Digital Object Identifier 10.1109/LSP.2022.3179169

$\mathscr{C}_L^{1,1}(\mathcal{H})$ of functions $h : \mathcal{H} \to \mathbb{R}$ satisfying h is differentiable in $\mathcal{H}$ and $\nabla h : \mathcal{H} \to \mathcal{H}$ is L-Lipschitz continuous, i.e., for every $(x, y) \in \mathcal{H} \times \mathcal{H}$, $\|\nabla h(x) - \nabla h(y)\| \leq L\|x - y\|$. Given $\rho > 0$, $h \in \mathscr{C}_L^{1,1}(\mathcal{H})$ is $\rho-$strongly convex if $h - \frac{\rho}{2}\|\cdot\|_2^2$ is convex. The proximity operator of h is defined by $\operatorname{prox}_h : x \mapsto \arg\min_{y\in\mathcal{H}}(h(y) + \frac{1}{2}\|y - x\|^2)$, which is well defined since, for every $x \in \mathcal{H}$, $h(\cdot) + \|\cdot - x\|^2/2$ is strongly convex and, hence, it admits a unique minimizer. Id and $\circ$ denote the the identity operator and the composition operation, respectively.

II. Optimization Problem and Proximal Schemes

Through this work we consider the following convex optimization problem.

Problem 1: Let $f \in \Gamma_0(\mathcal{H})$ and let $g \in \mathscr{C}_L^{1,1}(\mathcal{H})$ be $\rho-$strongly convex for some $L \in]0, +\infty[$ and $\rho \in]0, L[$. The problem considered is to

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

under the assumption that solutions exist.

The algorithms we consider generate recursive sequences of the form

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

where $\Phi : \mathcal{H} \to \mathcal{H}$ is a suitable operator which incorporates the proximity operator of f and proximity or gradient steps on g and such that we can recover a solution to Problem 1 from points in $\operatorname{Fix}\Phi = \{x \in \mathcal{H} \big| x = \Phi x\}$. In this work we study three algorithms for solving Problem 1: forward-backward, Douglas-Rachford, and Peaceman-Rachford.

• **Forward-backward splitting (FBS)** – The Forward-backward operator reads

$$\Phi = T_{\tau f,\tau g} = \operatorname{prox}_{\tau f} \circ (\mathrm{Id} - \tau\nabla g), \tag{2}$$

for some $\tau > 0$. This scheme alternates an explicit gradient step and an implicit subgradient step. It follows from [2, Proposition 26.1(iv)(a)] that

$$(\forall \tau > 0) \quad \arg\min(f + g) = \operatorname{Fix} T_{\tau f,\tau g}. \tag{3}$$

This scheme is also known as the proximal gradient algorithm (see, e.g., [16]).

• **Peaceman-Rachford splitting (PRS)** – The operator associated to this scheme is

$$\Phi = R_{\tau f,\tau g} = (2\operatorname{prox}_{\tau f} - \mathrm{Id}) \circ (2\operatorname{prox}_{\tau g} - \mathrm{Id}), \tag{4}$$

for some $\tau > 0$. It follows from [2, Prop. 26.1(iii)(b)] that

$$(\forall \tau > 0) \quad \arg\min(f + g) = \operatorname{prox}_{\tau g}(\operatorname{Fix} R_{\tau f,\tau g}). \tag{5}$$

• **Douglas-Rachford splitting (DRS)** – This scheme is formulated as:

$$\Phi = S_{\tau f,\tau g} = \operatorname{prox}_{\tau f}(2\operatorname{prox}_{\tau g} - \mathrm{Id}) + \mathrm{Id} - \operatorname{prox}_{\tau g}, \tag{6}$$

for some $\tau > 0$, which is the average between Id and $R_{\tau f,\tau g}$. It follows from [2, Prop. 26.1(iii)(b)] that

$$(\forall \tau > 0) \quad \arg\min(f + g) = \operatorname{prox}_{\tau g}(\operatorname{Fix} S_{\tau f,\tau g}). \tag{7}$$

There exists at least two ways to prove the convergence of the sequence of $(x_k)_{k\in\mathbb{N}}$ to $\arg\min(f + g)$ of such proximal schemes.

A first approach is to consider operators $\Phi : \mathcal{H} \to \mathcal{H}$ which are $r-$Lipschitz continuous for some $r \in [0, 1[$, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|\Phi x - \Phi y\| \leq r\|x - y\|, \tag{8}$$

and then we prove the convergence of the generated sequence following the Banach-Picard theorem stated below.

Theorem 1: [2, Theorem 1.50] Let $r \in [0, 1[$, let $\Phi : \mathcal{H} \to \mathcal{H}$ be a $r-$Lipschitz continuous operator, and let $x_0 \in \mathcal{H}$. Set

$$(\forall k \in \mathbb{N}) \quad x_{k+1} = \Phi x_k. \tag{9}$$

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

$$(\forall k \in \mathbb{N}) \quad \|x_k - \hat{x}\| \leq r^k\|x_0 - \hat{x}\|. \tag{10}$$

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

The second way to prove convergence relies on the theory of averaged nonexpansive operators [2, Theorem 5.23]. This context of study benefit of less restrictive assumptions of f and g to prove convergence, but at the price of weaker convergence results. Indeed, the convergence of the sequence $(x_k)_{k\in\mathbb{N}}$ to an element in $\operatorname{Argmin}(f + g)$ for FBS, DRS, and PRS is only weak and convergence rates are sub-linear in general.

III. Convergence Rates

In the next result we provide the linear convergence rates of FBS, DRS, and PRS in the context of Problem 1.

Proposition 1: Consider the context of Problem 1 and let $\tau > 0$. Then, the following statements hold:

1) Suppose that $\tau \in \left]0, \frac{2}{L}\right[$. Then $T_{\tau f,\tau g}$ is $r_T(\tau)-$Lipschitz continuous, where

$$r_T(\tau) := \max\{|1 - \tau\rho|, |1 - \tau L|\} \in]0, 1[. \tag{11}$$

In particular, the minimum of the function r_T defined in (11) is achieved at

$$\tau^* = \frac{2}{\rho + L} \quad \text{and} \quad r_T(\tau^*) = \frac{L - \rho}{L + \rho}. \tag{12}$$

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

$$r_R(\tau) = \max\left\{\frac{1 - \tau\rho}{1 + \tau\rho}, \frac{\tau L - 1}{\tau L + 1}\right\} \in]0, 1[. \tag{13}$$

In particular, the minimum of the function r_R defined in (13) is achieved at

$$\tau^* = \sqrt{\frac{1}{L\rho}} \quad \text{and} \quad r_R(\tau^*) = \frac{1 - \sqrt{\frac{\rho}{L}}}{1 + \sqrt{\frac{\rho}{L}}}. \tag{14}$$

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

$$r_S(\tau) = \frac{1 + r_R(\tau)}{2} \in]0, 1[\tag{15}$$

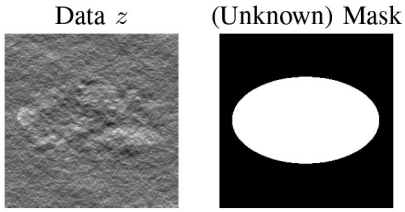

Fig. 1. Piecewise homogeneous fractal texture generated according to [7, Section 4] using mask (right).

and r_R is defined in (13). In particular, the minimum of the function r_S defined in (15) is achieved at

$$\tau^* = \sqrt{\frac{1}{L\rho}} \quad \text{and} \quad r_S(\tau^*) = \frac{1 + r_R(\tau^*)}{2}. \tag{16}$$

Proof: We give the proof in the Appendix. ■

These results allows us to identify the optimal rates for each scheme and the optimal step-size, which is often a tedious task in practice. A step-size $\tau \approx 2/L$ is commonly used for FBS while large values of τ are often used for DRS in order to obtain a faster convergence. This is theoretically justified with the optimal step-sizes obtained in (12) and (16) when ρ is small, i.e., when the problem is close to loose the strong convexity.

The optimal rates in (12), (14), and (16) permit the identification of the faster strategy by using the strong convexity and the smoothness of the problem. In the next section, we explore numerically the tightness of these optimal rates and we compare them in the context of texture segmentation problem.

IV. Application to Texture Segmentation

Texture descriptors – In this section we focus on segmentation of autosimilar (or fractal) textures characterized by their local variance and local regularity. This model appears in a wide class of imagery problems as motivated in [7]. We denote $\mathrm{z} \in \mathbb{R}^{\mathrm{N}}$ the texture under study and $\mathcal{L}_{j,n}$ the associated multi-resolution coefficient at scale 2^j and location n [18]. For instance, $\mathcal{L}_{j,n}$ can refer to the absolute value of a wavelet coefficient or a wavelet leader coefficient. Following [18], [19] the local variance (which is proportional to some $\eta_n \in \mathbb{R}$) and local regularity, modeled with the local Hölder exponent h_n, are related with $\mathcal{L}_{j,n}$ when it goes to fine scales as follows

$$\mathcal{L}_{j,n} \simeq \eta_n 2^{j h_n}. \tag{17}$$

Define $\boldsymbol{h} = (h_n)_{n\in\{1,\dots,N\}}$ and $\boldsymbol{v} = (\log_2 \eta_n)_{n\in\{1,\dots,N\}}$.

Texture segmentation – In this work, we focus on a minimization problem inspired by [7], where we essentially replace the discrete horizontal/vertical finite difference filtering with a wavelet transform. This modification leads to an optimization problem satisfying the hypotheses of Problem 1, which keeps the essence of the model in [7]. The problem is to

$$\underset{(\boldsymbol{v},\boldsymbol{h})\in\mathbb{R}^N\times\mathbb{R}^N}{\text{minimize}} \frac{1}{2}\sum_{j=j_1}^{j_2} \|\boldsymbol{v} + j\boldsymbol{h} - \log_2 \boldsymbol{\mathcal{L}}_j\|_{\text{Fro}}^2 + \chi_h\|W\boldsymbol{h}\|_1 + \chi_v\|W\boldsymbol{v}\|_1, \tag{18}$$

where $\|\cdot\|_{\text{Fro}}$ denotes the Frobenius norm, $1 \le j_1 < j_2$ are integers, $W \in \mathbb{R}^{N\times N}$ models an orthonormal Haar-wavelet transform, and $\chi_h, \chi_v > 0$ denote the regularization parameters. In order to fit (18) in the context of Problem 1, let us set

$$\begin{cases} g : (\boldsymbol{v},\boldsymbol{h}) \mapsto \frac{1}{2}\sum_{j=j_1}^{j_2} \|\boldsymbol{v} + j\boldsymbol{h} - \log_2 \boldsymbol{\mathcal{L}}_j\|_{\text{Fro}}^2 \\ f : (\boldsymbol{v},\boldsymbol{h}) \mapsto \chi_h\|W\boldsymbol{h}\|_1 + \chi_v\|W\boldsymbol{v}\|_1 \end{cases}$$

and note that $f \in \Gamma_0(\mathbb{R}^N \times \mathbb{R}^N)$. Moreover, set $M_m = \sum_{j=j_1}^{j_2} j^m$ for $m \in \{0,1,2\}$, $\mathbf{1} = (1,\dots,1)^\top \in \mathbb{R}^{j_2-j_1+1}$, $\boldsymbol{j} = (j_1,\dots,j_2)^\top \in \mathbb{R}^{j_2-j_1+1}$, $A = \begin{bmatrix}\mathbf{1} & \boldsymbol{j}\end{bmatrix}$,

$$L = \lambda_{\max}(A^*A) = \frac{(M_0+M_2) + \sqrt{(M_0-M_2)^2 + 4M_1^2}}{2},$$

and

$$\rho = \lambda_{\min}(A^*A) = \frac{(M_0+M_2) - \sqrt{(M_0-M_2)^2 + 4M_1^2}}{2}.$$

After straightforward computations, we deduce that g is ρ-strongly convex and differentiable with a L-Lipschitz gradient. Note that Cauchy–Schwarz inequality yields $M_1 = \mathbf{1}^\top\boldsymbol{j} < \|\mathbf{1}\|\|\boldsymbol{j}\| = \sqrt{M_0 M_2}$, which implies that $\rho > 0$. Moreover, the proximity operator of g (provided in [7]) is, for every $(\boldsymbol{v},\boldsymbol{h}) \in \mathbb{R}^N \times \mathbb{R}^N$, $\text{prox}_g(\boldsymbol{v},\boldsymbol{h}) = (\boldsymbol{p},\boldsymbol{q})$, where

$$\begin{cases} \boldsymbol{p} = \frac{(1+M_2)\left(\sum_j \log_2 \mathcal{L}_j + \boldsymbol{v}\right) - M_1\left(\sum_j j\log_2 \mathcal{L}_j + \boldsymbol{h}\right)}{(1+M_0)(1+M_2) - M_1^2}, \\ \boldsymbol{q} = \frac{(1+M_0)\left(\sum_j j\log_2 \mathcal{L}_j + \boldsymbol{h}\right) - M_1\left(\sum_j \log_2 \mathcal{L}_j + \boldsymbol{v}\right)}{(1+M_0)(1+M_2) - M_1^2}, \end{cases}$$

and the proximity operator of f has the closed form expression:

$$\text{prox}_f : (\boldsymbol{v},\boldsymbol{h}) \mapsto \begin{bmatrix} W^{-1}\text{prox}_{\chi_h\|\cdot\|_1} W\boldsymbol{v} \\ W^{-1}\text{prox}_{\chi_v\|\cdot\|_1} W\boldsymbol{h} \end{bmatrix}.$$

TABLE I
Theoretical Optimal Rates for FBS, PRS, and DRS

FBS	$j_1 = 1$	$j_1 = 2$	$j_1 = 3$
$j_2 = 3$	0.958	0.991	ND
$j_2 = 4$	0.965	0.988	0.997
$j_2 = 5$	0.972	0.988	0.996
PRS	$j_1 = 1$	$j_1 = 2$	$j_1 = 3$
$j_2 = 3$	**0.743**	**0.874**	ND
$j_2 = 4$	**0.764**	**0.857**	**0.928**
$j_2 = 5$	**0.786**	**0.856**	**0.911**
DRS	$j_1 = 1$	$j_1 = 2$	$j_1 = 3$
$j_2 = 3$	0.872	0.937	ND
$j_2 = 4$	0.882	0.929	0.964
$j_2 = 5$	0.893	0.928	0.956

The highest rate for each (j_1, j_2) is displayed in bold.

Theoretical comparisons – We apply the theoretical analysis to the texture segmentation optimization problem defined in (18). The optimal rates only depend on the regression scales j_1 and j_2 and their values are summarized in Table I. By a simple evaluation of the rate, we can identify that PRS is the scheme with smallest convergence rate.

V. Experimental Validation

Generation of synthetic data – A piecewise homogeneous fractal texture is generated according to [7, Section 4] using the mask displayed in Fig. 1(b), with $N = 256 \times 256$. The mask is composed with two regions: i) a background (in black), with

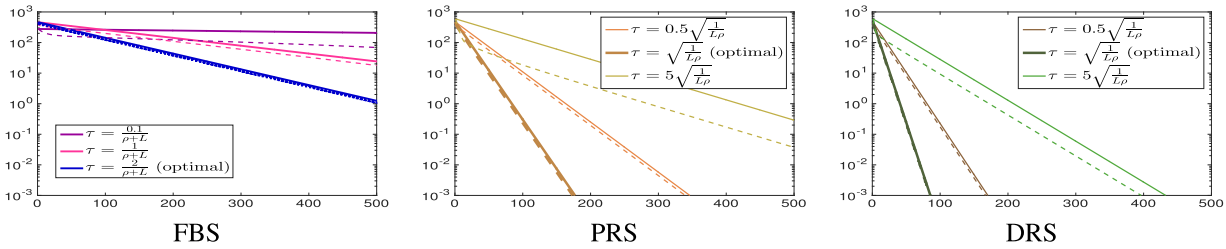

Fig. 2. We compare the theoretical upper bound (i.e., $r_\Phi(\tau)^k\|x_0 - x_\infty\|_2$) versus the numerical error (i.e., $\|x_k - x_\infty\|_2$) w.r.t. the number of iterations, where $r_\Phi(\tau)$ stands for each of the theoretical linear rates defined in Proposition 1 when $\Phi \in \{T, R, S\}$. We consider y-log scale for FBS, DRS and PRS schemes for solving the segmentation problem over the range $j \in \{2, 3, 4\}$, $\chi_v = 0.1$, and $\chi_h = 200$.

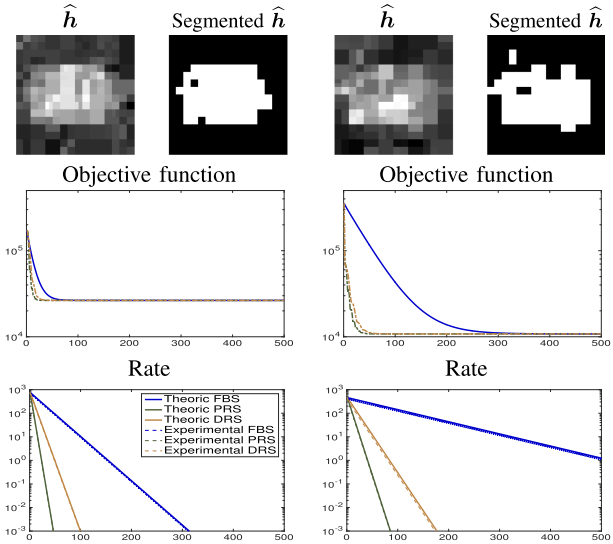

Fig. 3. Segmentation results and associated convergence behavior with FBS, DRS, and PRS schemes considering y-log scale. (right) Range $j \in \{1, 2, 3\}$, $\chi_v = 0.1$, and $\chi_h = 40$. (left) Range $j \in \{2, 3, 4\}$, $\chi_v = 0.1$, and $\chi_h = 200$. (1st row) Results obtained from the minimization of (18). (2nd row) Objective function (18) w.r.t the number of iterations. (3 rd row) Theoretical upper bound (i.e., $r_\Phi(\tau)^k\|x_0 - x_\infty\|_2$) versus numerical error (i.e., $\|x_k - x_\infty\|_2$) w.r.t number of iterations, where $r_\Phi(\tau)$ denote the theoretical linear convergence rates defined in Proposition 1 with $\Phi \in \{T, R, S\}$.

variance and local regularity (0.6, 0.5), a central ellipse (in white) of parameters (1.1,0.9).

Algorithmic rate comparisons for different regression ranges – The first set of experiments displayed in Fig. 3 provides a comparison between the theoretical and numerical convergence rates for two different ranges $(j_1, j_2) \in \{(1, 3), (2, 4)\}$. The choice of (j_1, j_2) affects the strong convexity constant and, hence, the theoretical convergence rate as observed in Table I. As expected, we observe that the scale of regression affects the segmentation results. Additionally, we note that the numerical error fits quite well the theoretical upper bound for the three algorithms under study, placing PRS as the most efficient algorithm in this context. For instance, when $(j_1, j_2) = (1, 3)$, in order to achieve an accuracy of 1e-1, PRS requires around 50 iterations, DRS around 120 iterations, and FBS around 700 iterations.

Optimal versus not optimal rate – In Fig. 2, we display the convergence behavior for each scheme FBS, PRS, and DRS for optimal τ^* and non-optimal step-sizes. We can observe that the rate is very tight for the optimal step-size and looses the tightness as the step-size deviates from the optimal value. The optimal step-size according to the theory always leads to the fastest numerical scheme.

VI. Conclusion

This work provides accurate convergence rate for FBS, DRS, and PRS for minimizing the sum of a proper l.s.c. convex function and a strongly convex differentiable function with Lipschitz gradient. The proposed results allows us to evaluate the optimal step-size for each scheme based on the constant of the strong convexity and Lipschitz constant of the gradient. We validate the theoretical results on a texture segmentation problem relying on fractal analysis. Peaceman-Rachford appears to be the faster scheme for this specific study case of texture segmentation.

Appendix

Proposition 1(1): We provide a simpler and shorter proof of the result in [17, Theorem 2.1], which is a contribution by itself. Since $g \in \mathscr{C}^{1,1}_L(\mathcal{H})$ is ρ-strongly convex, we have that $\phi = g - \rho\|\cdot\|^2/2$ is convex and differentiable. Hence, it follows from [2, Proposition 18.15] that, for every x and y in $\mathcal{H}$,

$$\begin{aligned}\langle x - y \mid \nabla\phi(x) - \nabla\phi(y)\rangle &= \langle x - y \mid \nabla g(x) - \nabla g(y)\rangle \\ &\quad -\rho\|x-y\|^2 \leq (L-\rho)\|x-y\|^2,\end{aligned}$$

which yields $\phi \in \mathscr{C}^{1,1}_{L-\rho}(\mathcal{H})$. In addition, by setting $G_{\tau g} = \mathrm{Id} - \tau\nabla g$, we have $T_{\tau f,\tau g} = \mathrm{prox}_{\tau f} \circ G_{\tau g}$ and

$$G_{\tau g} = \mathrm{Id} - \tau(\nabla\phi + \rho\,\mathrm{Id}) = (1 - \tau\rho)\,\mathrm{Id} - \tau\nabla\phi. \tag{19}$$

Now let $\tau \in]0, 2/L]$, let x and y in $\mathcal{H}$, and set $p = T_{\tau f,\tau g}x$ and $q = T_{\tau f,\tau g}y$. Since $\phi \in \mathscr{C}^{1,1}_{L-\rho}(\mathcal{H})$ is convex, it follows from [2, Proposition 12.28], (19), and [2, Proposition 18.15] that

$$\begin{aligned}&\|p - q\|^2 \leq \|G_{\tau g}x - G_{\tau g}y\|^2 \\ &= (1 - \tau\rho)^2\|x - y\|^2 + \tau^2\|\nabla\phi(x) - \nabla\phi(y)\|^2 \\ &\quad - 2\tau(1 - \tau\rho)\langle x - y \mid \nabla\phi(x) - \nabla\phi(y)\rangle \\ &\leq (1 - \tau\rho)^2\|x - y\|^2 \\ &\quad + \tau\left(\tau(L + \rho) - 2\right)\langle x - y \mid \nabla\phi(x) - \nabla\phi(y)\rangle \\ &\leq (1 - \tau\rho)^2\|x - y\|^2 = \|x - y\|^2 \max\{(1-\tau\rho)^2, (1-\tau L)^2\}\end{aligned}$$

and the result follows. *Proposition 1(2)&(3):* Both results are a consequence of [8, Theorem 2]. In all the cases, the minimum is obtained via simple computations.

REFERENCES

[1] L. M. Briceño-Arias and N. Pustelnik, "Proximal or gradient steps for cocoercive operators," 2020, *arXiv2101.06152*.

[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York, NY, USA: Springer, 2017.

[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*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., New York, NY, USA: Springer-Verlag, 2011, pp. 185–212.

[4] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations Trends Optim.*, vol. 1, no. 3, pp. 123–231, 2014.

[5] L. Patrick Combettes and L. E. Glaudin, "Proximal activation of smooth functions in splitting algorithms for convex image recovery," *SIAM J. Imag. Sci.*, vol. 12, no. 4, pp. 1905–1935, 2019.

[6] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1–2, pp. 89–97, Jan. 2004.

[7] B. Pascal, N. Pustelnik, and P. Abry, "Strongly convex optimization for joint fractal feature estimation and texture segmentation," *Appl. Comput. Harmon. Anal.*, vol. 54, pp. 303–322, 2021.

[8] P. Giselsson and S. Boyd, "Linear convergence and metric selection for Douglas-Rachford splitting and ADMM," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 532–544, Feb. 2017.

[9] B. Mercier, "Inéquations variationnelles de la mécanique,' vol. 1. Département de Mathématique, Université de Paris-Sud, Orsay, 1980.

[10] L. Demanet and X. Zhang, "Eventual linear convergence of the Douglas-Rachford iteration for basis pursuit," *Math. Comput.*, vol. 85, no. 297, pp. 209–238, 2016.

[11] P. Giselsson, "Tight global linear convergence rate bounds for Douglas-Rachford splitting," *J. Fixed Point Theory Appl.*, vol. 19, no. 4, pp. 2241–2270, 2017.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] S. Jaffard, "Wavelet techniques in multifractal analysis," in *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot*, vol. 72. M. Lapidus and M. van Frankenhuijsen Eds., 2004, pp. 91–152.

[19] H. Wendt, S. G. Roux, P. Abry, and S. Jaffard, "Wavelet leaders and bootstrap for multifractal analysis of images," *Signal Process.*, vol. 89, no. 6, pp. 1100–1114.