

A Generalized Forward-Backward Splitting*

Hugo Raguet[†], Jalal Fadili[‡], and Gabriel Peyré[†]

Abstract. This paper introduces a generalized forward-backward splitting algorithm for finding a zero of a sum of maximal monotone operators $B + \sum_{i=1}^n A_i$, where B is cocoercive. It involves the computation of B in an explicit (forward) step and the parallel computation of the resolvents of the A_i 's in a subsequent implicit (backward) step. We prove the algorithm's convergence in infinite dimension and its robustness to summable errors on the computed operators in the explicit and implicit steps. In particular, this allows efficient minimization of the sum of convex functions $f + \sum_{i=1}^n g_i$, where f has a Lipschitz-continuous gradient and each g_i is simple in the sense that its proximity operator is easy to compute. The resulting method makes use of the regularity of f in the forward step, and the proximity operators of the g_i 's are applied in parallel in the backward step. While the forward-backward algorithm cannot deal with more than $n = 1$ nonsmooth function, we generalize it to the case of arbitrary n . Examples on inverse problems in imaging demonstrate the advantage of the proposed methods in comparison to other splitting algorithms.

Key words. forward-backward algorithm, monotone operator splitting, nonsmooth convex optimization, proximal splitting, image processing, sparsity

AMS subject classifications. 47N10, 90C25, 94A08

DOI. 10.1137/120872802

1. Introduction. Throughout this paper, \mathcal{H} denotes a real Hilbert space endowed with scalar product $\langle \cdot | \cdot \rangle$ and associated norm $\| \cdot \|$, Id is the identity operator on \mathcal{H} , and n is a positive integer.

1.1. Structured monotone inclusion and minimization problems. We consider the following monotone inclusion problem:

$$(1.1) \quad \text{Find } x \in \left\{ \text{zer} \left(B + \sum_{i=1}^n A_i x \right) \stackrel{\text{def}}{=} \left\{ x \in \mathcal{H} \mid 0 \in Bx + \sum_{i=1}^n A_i x \right\} \right\},$$

where $B : \mathcal{H} \mapsto \mathcal{H}$ is cocoercive and, for all i , $A_i : \mathcal{H} \mapsto 2^{\mathcal{H}}$ is a maximal monotone set-valued map. While such inclusion problems arise in various fields, our main motivation is to solve convex minimization problems. Indeed, it is well known that the subdifferential ∂g_i of a function $g_i \in \Gamma_0(\mathcal{H})$ is a maximal monotone map, $\Gamma_0(\mathcal{H})$ being the class of lower semicontinuous, proper, convex functions from \mathcal{H} to $]-\infty, +\infty]$. If, moreover, $f \in \Gamma_0(\mathcal{H})$ is differentiable with a Lipschitz-continuous gradient, then the Baillon–Haddad theorem [5] asserts that ∇f is cocoercive. Defining $F \stackrel{\text{def}}{=} f + \sum_{i=1}^n g_i$, the set of minimizers of F verifies

$$\text{argmin } F = \text{zer} \left(\nabla f + \sum_{i=1}^n \partial g_i \right),$$

*Received by the editors April 9, 2012; accepted for publication (in revised form) March 29, 2013; published electronically July 9, 2013.

<http://www.siam.org/journals/siims/6-3/87280.html>

[†]CEREMADE, CNRS-Université Paris-Dauphine, 75775 Paris Cedex 16, France (raguet@ceremade.dauphine.fr, gabriel.peyre@ceremade.dauphine.fr).

[‡]GREYC, CNRS-ENSICAEN-Université de Caen 6, 14050 Caen Cedex, France (Jalal.Fadili@greyc.ensicaen.fr).

provided that the following conditions hold:

$$(H1) \operatorname{argmin} F \neq \emptyset,$$

$$(H2) (0, \dots, 0) \in \operatorname{sri}\{(x - y_1, \dots, x - y_n) \mid x \in \mathcal{H} \text{ and } \forall i, y_i \in \operatorname{dom} g_i\},$$

where $\operatorname{dom} g \stackrel{\text{def}}{=} \{x \in \mathcal{H} \mid g(x) < +\infty\}$ denotes the domain of a function g and $\operatorname{sri} C$ denotes the strong relative interior of a nonempty convex subset C of \mathcal{H} [6]. Therefore, identifying B with ∇f and the A_i 's with the ∂g_i 's, solving (1.1) allows us to solve

$$(1.2) \quad \min_{x \in \mathcal{H}} \left\{ F(x) \stackrel{\text{def}}{=} f(x) + \sum_{i=1}^n g_i(x) \right\} .$$

The structured monotone inclusion problem (1.1) is fairly general, and a wide range of iterative algorithms for solving it take advantage of the specific properties of the operators involved in the summand. As we will see, one crucial property is the possibility of computing the *resolvent* of a maximal monotone operator A , denoted J_A . It is defined as (see section 4.1 for details)

$$J_A x = y \quad \stackrel{\text{def}}{\Leftrightarrow} \quad x \in y + Ay .$$

For a given $x \in \mathcal{H}$, computing $J_A x$ is in itself a monotone inclusion problem, but it turns out that it can be solved explicitly for many operators; e.g., the action of the resolvent can be easily computed in closed form. Our interest is in *splitting* methods to solve (1.1): iterative algorithms that evaluate individually the operator B (cocoercive) and the resolvents J_{A_i} , at various points of \mathcal{H} , but not the resolvents of sums.

The next section recalls several important previous works on splitting algorithms, focusing on their application to convex optimization.

1.2. Splitting methods for minimization problems. If g is a function in $\Gamma_0(\mathcal{H})$, the resolvent of its subdifferential, $J_{\partial g}$, can be shown (see section 4.1) to be equal to the Moreau proximity operator of g [60], defined for all $x \in \mathcal{H}$ as

$$\operatorname{prox}_g(x) \stackrel{\text{def}}{=} \operatorname{argmin}_{y \in \mathcal{H}} \frac{1}{2} \|x - y\|^2 + g(y) .$$

Again, this can be solved explicitly for many functions; such functions are dubbed “simple.”

Another important property of some part of a functional to be minimized is differentiability. Recalling (1.2), the forward-backward algorithm applies if f is differentiable with a Lipschitz-continuous gradient, and $n \equiv 1$ with g_1 simple. This scheme consists in performing alternatively a gradient-descent (corresponding to an explicit step on the function f) followed by a proximal step (corresponding to an implicit step on the function g_1). Such a scheme can be understood as a generalization of the projected gradient method. This algorithm, which finds its roots in numerical analysis for PDEs, has been well studied for solving monotone inclusion and convex optimization problems [9, 19, 27, 45, 57, 62, 73, 75]. Accelerated multistep versions for convex optimization have been proposed [8, 61], which enjoy a faster convergence rate of $O(1/k^2)$ on the objective F in the general case, where k is the iteration counter.

Other splitting methods do not require any smoothness on any part of the composite functional F . The Douglas–Rachford (DR) scheme was originally developed to find the zeros of the sum of two linear operators [31], and then two nonlinear operators in [53] or two maximal

monotone operators in [55]; see also [20, 37]. This scheme applies to minimizing $g_1 + g_2$, provided that g_1 and g_2 are simple. The backward-backward algorithm [1, 7, 20, 54, 62] can be used to minimize $F = g_1 + g_2$ when the functions involved are the indicator functions of nonempty closed convex sets or involve Moreau envelopes. Interestingly, if one of the functions g_1 or g_2 is a Moreau envelope and the other is simple, the backward-backward algorithm amounts to a forward-backward scheme.

If L is a bounded injective linear operator, it is possible to minimize $F = g_1 \circ L + g_2$ by applying these splitting schemes on the Fenchel–Rockafellar dual problem. It was shown that applying the DR scheme leads to the alternating direction method of multipliers (ADMM) [37, 44, 45, 46, 47]. For nonnecessarily injective L and g_2 strongly convex with a Lipschitz-continuous gradient, the forward-backward algorithm can be applied to the Fenchel–Rockafellar dual [22, 40]. Dealing with an arbitrary bounded linear operator L can be achieved using primal-dual methods motivated by the classical Kuhn–Tucker theory. Starting from methods to solve saddle function problems such as the Arrow–Hurwicz method [3] and its modification [66], the extragradient method [49], this problem has received a lot of attention more recently [11, 14, 18, 59, 70, 74].

It is also possible to extend the DR algorithm to an arbitrary number $n > 2$ of simple functions. Inspired by the method of partial inverses [71, section 5], most methods rely either explicitly or implicitly on introducing auxiliary variables and bringing back the original problem to the case $n \equiv 2$ in the product space \mathcal{H}^n . Doing so yields iterative schemes in which one performs independent parallel proximal steps on each of the simple functions and then computes the next iterate by essentially averaging the results. Variants have been proposed in [24], and in [38], which describes a general projective framework that does not reduce the problem to the case $n \equiv 2$. These extensions, however, do not apply to the forward-backward scheme that can handle only $n \equiv 1$. It is at the heart of this paper to present such an extension.

Recently proposed methods extend existing splitting schemes to handle the sum of any number of $n \geq 2$ composite functions of the form $g_i = h_i \circ L_i$, where the h_i 's are simple and the L_i 's are bounded linear operators. Let us denote by L_i^* the adjoint operator of L_i . If L_i satisfies $L_i L_i^* = \nu \text{Id}$ for any $\nu > 0$ (it is a so-called *tight frame*), $h_i \circ L_i$ is simple as soon as h_i is simple and L_i^* is easy to compute [23]. This case thus reduces to the previously reviewed ones. If L_i is not a tight frame but $(\text{Id} + L_i^* L_i)$ or $(\text{Id} + L_i L_i^*)$ is easily invertible, it is again possible to reduce the problem to the previous cases by introducing, for each L_i , an auxiliary variable belonging to its range. Note, however, that, if solved with the DR algorithm on the product space, the auxiliary variables are also duplicated, which would increase significantly the dimensionality of the problem. Some dedicated parallel implementations were specifically designed for the case where $(\sum_i L_i^* L_i)$ or $(\sum_i L_i L_i^*)$ is (easily) invertible; see, for instance, [36, 63]. If the L_i 's satisfy none of the above properties, it is still possible to call on primal-dual methods, either by writing $F = \sum_{i=1}^n h_i \circ L_i = g \circ L$ with $L(x) = (L_i(x))_i$ and $g((x_i)_i) = \sum_i h_i(x_i)$ (see, for instance, [33]) or by solving on the product space $F((x_i)_i) = \sum_i h_i(L_i x_i) + \iota_{\mathcal{S}}((x_i)_i)$ [11], where $\iota_{\mathcal{S}}$ is the indicator function of the closed convex set \mathcal{S} defined in section 4.2.

In spite of the wide range of already existing proximal splitting methods, none seems to satisfactorily address explicitly the case where $n > 1$ and f is smooth but not necessarily

simple. A workaround that has been proposed previously uses nested algorithms to compute the proximity operator of $\sum_i g_i$ within subiterations (see, for instance, [17, 32]); this leads to practical as well as theoretical difficulties in selecting the number of subiterations. More recently, [59] proposed an algorithm for minimizing $F = f + g$ under linear constraints. We show in section 2.3 how this can be adapted to address the general problem (1.2) while achieving full splitting of the proximity operators of the g_i 's and using the gradient of f . In preparing a first draft of this paper, we became aware that other authors [26, 28, 76] have independently and concurrently developed primal-dual algorithms to solve problems that encompass the one we consider here. These approaches and algorithms are, however, different from ours in many important ways. This will be discussed in detail in section 2.3 especially in relation to [26]. We also report some numerical experiments in section 3 suggesting that our primal algorithm is more adapted for imaging problems of the form (1.2).

1.3. Applications in image processing. Many imaging applications require solving ill-posed inverse problems to recover high quality images from low-dimensional and noisy observations. These challenging problems necessitate the use of regularization through prior models to capture the geometry of natural signals, images, or videos. Numerical solution of inverse problems can be achieved through minimization of objective functionals, with respect to a high-dimensional variable, that takes into account both a term for fidelity to the observations and regularization terms reflecting the priors. Clearly, such functionals are composite by construction, hence fitting in the framework of (1.2). Section 3 details several examples of such inverse problems.

In many situations, this leads to the optimization of a convex functional that can be split into the sum of convex smooth and nonsmooth terms. The smooth part of the objective is in some cases a data fidelity term and reflects some specific knowledge about the forward model, i.e., the noise and the measurement/degradation operator. This is, for instance, the case if the operator is linear and the noise is additive Gaussian, in which case the data fidelity is a quadratic function. The most successful regularizations that have been advocated are nonsmooth, which typically allows one to preserve sharp and intricate structures in the recovered data. Among such priors, sparsity-promoting ones have become popular, e.g., the ℓ_1 -norm of coefficients in a wisely chosen dictionary [56], or a total variation (TV) prior [69]. To better model the data, composite priors can be constructed by summing several suitable regularizations; see, for instance, the morphological diversity framework [72]. The proximity operator of the ℓ_1 -norm penalization is a simple soft-thresholding [30], whereas the use of complex or mixed regularization priors justifies the splitting of nonsmooth terms in several simpler functions (see section 3 for concrete examples).

The composite structure of convex optimization problems arising when solving inverse problems in the form of a sum of simple and/or smooth functions involving linear operators explains the popularity of proximal splitting schemes in imaging science. Depending on the structure of the objective functional as detailed in the previous section, one can resort to the appropriate splitting algorithm. For instance, the forward-backward algorithm and its modifications have become popular for sparse regularization with a smooth data fidelity; see, for instance, [8, 10, 16, 27, 29, 41, 43]. The DR and its parallelized extensions were also used in a variety of inverse problems implying only nonsmooth functions; see, for instance,

[12, 17, 23, 24, 34, 32, 35, 67]. The ADMM (which is nothing but DR on the dual) was also applied to some linear inverse problems in [2, 42]. Primal-dual schemes [14, 33] are among the most flexible schemes for handling more complicated priors. The interested reader may refer to [72, Chapter 7] and [25] for extensive reviews.

1.4. Contributions and paper organization. This paper introduces a novel generalized forward-backward (GFB) algorithm to solve the monotone inclusion (1.1). The algorithm achieves full splitting where all operators are used separately: an explicit step for B (single-valued) and a parallelized implicit step through the resolvent of the A_i 's. We prove convergence of the algorithm even when summable errors may contaminate the iterations. To the best of our knowledge, it is among the first algorithms to tackle the case where $n > 1$ (see section 2.3 for relation to other works). Although our numerical results are reported only on imaging applications, the algorithm may prove useful for many other applications such as machine learning, statistical estimation, or optimal control.

Section 2 presents the algorithm and states our main theoretical result, before commenting on some relevant aspects and on alternatives in the literature. Numerical examples are reported in section 3 to show the usefulness of this approach for imaging problems. The convergence proof is deferred to section 4, after recalling some preliminary results on monotone operator theory.

2. Generalized forward-backward splitting.

2.1. The algorithmic scheme. We consider problem (1.1), where all operators are maximal monotone, B is β -cocoercive with $\beta \in]0, +\infty[$; i.e.,

$$\forall x, y \in \mathcal{H}, \quad \beta \|Bx - By\|^2 \leq \langle Bx - By | x - y \rangle ,$$

and for all i and all $\gamma > 0$, $J_{\gamma A_i}$ (the resolvent of γA_i) is easy to compute. Our proposed GFB algorithm is detailed in Algorithm 1.

Algorithm 1. A generalized forward-backward algorithm for solving (1.1). $\beta \in]0, +\infty[$ is a cocoercivity constant of B .

Require $(z_i)_{i \in \llbracket 1, n \rrbracket} \in \mathcal{H}^n$, $(w_i)_{i \in \llbracket 1, n \rrbracket} \in]0, 1]^n$ such that $\sum_{i=1}^n w_i = 1$,
 $\gamma \in]0, 2\beta[$, $\lambda_k \in]0, \min(\frac{3}{2}, \frac{1}{2} + \frac{\beta}{\gamma}) [\forall k \in \mathbb{N}$.

Initialization

$x \leftarrow \sum_i w_i z_i$;
 $k \leftarrow 0$.

Main iteration

repeat

for $i \in \llbracket 1, n \rrbracket$ **do**
(2.1) $z_i \leftarrow z_i + \lambda_k (J_{\frac{\gamma}{w_i} A_i} (2x - z_i - \gamma Bx) - x)$;
 $x \leftarrow \sum_i w_i z_i$;
 $k \leftarrow k + 1$.

until convergence;

Return x .

To state our main theorem that ensures the convergence of the algorithm and its robustness to summable errors, for each i let $\varepsilon_{1,k,i}$ be the error at iteration k when computing $J_{\frac{\gamma}{w_i}A_i}$, and let $\varepsilon_{2,k}$ be the error at iteration k when computing B . An inexact GFB algorithm generates sequences $(z_{i,k})_{k \in \mathbb{N}}$, $i \in \llbracket 1, n \rrbracket$, and $(x_k)_{k \in \mathbb{N}}$, such that for all $i \in \llbracket 1, n \rrbracket$ and $k \in \mathbb{N}$,

$$(2.2) \quad z_{i,k+1} = z_{i,k} + \lambda_k \left(J_{\frac{\gamma}{w_i}A_i} (2x_k - z_{i,k} - \gamma_k (Bx_k + \varepsilon_{2,k})) + \varepsilon_{1,k,i} - x_k \right) .$$

Theorem 2.1. *Suppose that $\text{zer} (B + \sum_{i=1}^n A_i) \neq \emptyset$. Suppose that the following assumptions are satisfied:*

- (i) $0 < \inf_{k \in \mathbb{N}} \lambda_k \leq \sup_{k \in \mathbb{N}} \lambda_k < \min \left(\frac{3}{2}, \frac{1}{2} + \frac{\beta}{\gamma} \right)$, and
- (ii) $\sum_{k=0}^{+\infty} \|\varepsilon_{2,k}\| < +\infty$, and for all i , $\sum_{k=0}^{+\infty} \|\varepsilon_{1,k,i}\| < +\infty$.

Then the sequence $(x_k)_{k \in \mathbb{N}}$ defined in (2.2) converges weakly toward a solution of (1.1). Moreover, if $\forall k \in \mathbb{N}$, $\lambda_k \leq 1$, then the convergence is strong if either B is uniformly monotone, or $\times_{i=1}^n w_i^{-1} A_i$ is uniformly monotone. The latter is true, for instance, if $\forall i \in \llbracket 1, n \rrbracket$, A_i is uniformly monotone with its modulus φ being also subadditive or convex.

The definition of uniform monotonicity and the function φ is provided in section 4.1.

The following corollary specializes Theorem 2.1 to the case of convex optimization problems of the form (1.2).

Corollary 2.2. *Suppose that ∇f is Lipschitz continuous with constant $1/\beta$ and that (H1)–(H2) are verified. In Algorithm 1 and in (2.2), substitute B with ∇f and A_i with ∂g_i (i.e., $J_{\frac{\gamma}{w_i}A_i}$ with $\text{prox}_{\frac{\gamma}{w_i}g_i}$). Then, under assumptions (i)–(ii) of Theorem 2.1, the sequence $(x_k)_{k \in \mathbb{N}}$ converges weakly towards a minimizer of (1.2). Moreover, if $\forall k \in \mathbb{N}$, $\lambda_k \leq 1$, then $(x_k)_{k \in \mathbb{N}}$ converges strongly to the unique minimizer of (1.2) if either f is uniformly convex, or $\times_{i=1}^n w_i^{-1} \partial g_i$ is uniformly monotone. The latter is true, for instance, if $\forall i \in \llbracket 1, n \rrbracket$, g_i is uniformly convex with its modulus φ being also subadditive or convex.*

The proofs are detailed in section 4.

Remark 2.1. Recall that a function $f \in \Gamma_0(\mathcal{H})$ is uniformly convex of modulus $\varphi : [0, +\infty[\rightarrow [0, +\infty]$ if φ is a nondecreasing function that vanishes only at 0, such that for all x and y in $\text{dom } f$, the following holds:

$$\forall \rho \in]0, 1[, \quad f(\rho x + (1 - \rho)y) + \rho(1 - \rho)\varphi(\|x - y\|) \leq \rho f(x) + (1 - \rho)f(y) .$$

The formulation of Algorithm 1 is general, but it can be simplified for practical purposes. In particular, the auxiliary variables z_i can all be initialized to 0, the weights w_i can be set equally to $1/n$, and for simplicity the relaxation parameters λ_k can be set to 1, constant along iterations. This is what has been done in the numerical experiments described in section 3.

2.2. Special instances. Our GFB algorithm can be viewed as a hybrid splitting algorithm whose special instances turn out to be classical splitting methods, namely, the forward-backward and DR algorithms.

Relaxed forward-backward. For $n \equiv 1$, the core update operator (2.1) of Algorithm 1 specializes to

$$x \leftarrow x + \lambda_k (J_{\gamma A} (x - \gamma Bx) - x) ,$$

so that x_k given by (2.2) follows the iterations of the relaxed forward-backward algorithm [20, section 6]. In this case, convergence can be ensured with step-size γ varying along iterations; see the discussion in Remark 4.3. For convex minimization problems, known results

on convergence rate analysis (on the objective in general) and accelerated versions of the forward-backward algorithm [8, 9, 61] might inspire one to study those of our GFB (for the case where $\beta > 0$).

Relaxed Douglas–Rachford. If we set $B \equiv 0$, the update of the auxiliary variables in (2.1) becomes

$$z_i \leftarrow z_i + \lambda_k \left(J_{\frac{\gamma}{w_i} A_i} (2x - z_i) - x \right) ,$$

so that $(z_{i,k})_i$ given by (2.2) follows the iterations of the relaxed DR algorithm on the product space \mathcal{H}^n for solving $0 \in \sum_i A_i x$ [24, 71]. The convergence statements of Theorem 4.17(a)–(c) hold by replacing the conditions on the relaxation parameters by $\forall k \in \mathbb{N}, \lambda_k \in]0, 2[$ and $\sum_{k \in \mathbb{N}} \lambda_k (2 - \lambda_k) = +\infty$; this extends Remark 4.1 to $\alpha = \frac{1}{2}$ by Proposition 4.12 (see section 4.5).

Resolvents of the sum of monotone operators. Our GFB scheme provides yet another way to compute the resolvent of the sum of maximal monotone operators $(A_i)_i$. Given a point $y \in \text{ran}(\text{Id} + \sum_i A_i)$, set in (1.1) $B : x \mapsto x - y$ and $\beta \equiv 1$. It would be interesting to compare this algorithm with the DR and Dykstra-based variants [21]. This is left to a future work.

2.3. Relation to other works.

Relation to [59]. The authors in [59, section 5.3, problem (51)] describe an instance of the “block-decomposition” hybrid proximal extragradient (HPE) for minimizing $F = f + g$ under linear constraints. Equation (1.2) can be cast in an equivalent linearly constrained convex programming,

$$(2.3) \quad \min_{z=(z_i)_i \in \mathcal{H}^n} f(\sum_i w_i z_i) + \sum_i g_i(z_i) \quad \text{such that} \quad P_{\mathcal{S}^\perp}(z) = \mathbf{0} ,$$

where $P_{\mathcal{S}^\perp}$ is the orthogonal projector on $\mathcal{S}^\perp \stackrel{\text{def}}{=} \{z = (z_i)_i \in \mathcal{H}^n \mid \sum_i w_i z_i = 0\}$. As $P_{\mathcal{S}^\perp}$ is self-adjoint, z is an optimal solution if and only if there exists $v = (v_i)_i \in \mathcal{H}^n$ such that

$$\mathbf{0} \in (\nabla f(\sum_j w_j z_j))_i + (\partial g_i(z_i)/w_i)_i + P_{\mathcal{S}^\perp}(v) \quad \text{and} \quad P_{\mathcal{S}^\perp}(z) = \mathbf{0} ,$$

and the minimizer of F is given by $x = \sum_i w_i z_i$.

Let $\varsigma \in]0, 1]$ and $\gamma = \varsigma \frac{2\varsigma\beta}{1 + \sqrt{1 + 4\varsigma^2\beta^2}}$. Transposed to our setting, their iterations are presented in Algorithm 2.

Algorithm 2. Iterations of block-decomposition HPE [59].

```

repeat
  for  $i \in \llbracket 1, n \rrbracket$  do
     $z_i \leftarrow \text{prox}_{\frac{\gamma}{w_i} g_i}(\gamma^2 x + (1 - \gamma^2)z_i - \gamma \nabla f(x) + \gamma(v_i - u))$ ;
  for  $i \in \llbracket 1, n \rrbracket$  do
     $v_i \leftarrow v_i - \gamma z_i + \gamma x$ ;
   $x \leftarrow \sum_i w_i z_i$ ;
   $u \leftarrow \sum_i w_i v_i$ .
until convergence;

```

The update of the z_i 's in this iteration bears similarities with the one in Algorithm 1, where the γ 's play analogous roles. Nonetheless, the two algorithms are different. For instance, our algorithm solves the primal problem, while Algorithm 2 solves both the primal and dual problems. In addition, the objective in [59] is to study complexity, hence the different set of assumptions.

In preparing a revised draft of this paper, it came to our attention that another adaptation of the block-decomposition HPE, exploiting the specific properties of the linear constraints $P_{\mathcal{S}^\perp}(\mathbf{z}) = 0$ and changing the metric, leads to the iterations (2.2) with $\forall k \in \mathbb{N}$, $\lambda_k = 1$, i.e., no under- or overrelaxation. This could be another framework for studying convergence properties of GFB.

Relation to [26]. Combettes and Pesquet independently developed another algorithm for solving a general class of problems that covers (1.1). They rely on the classical Kuhn–Tucker theory and propose a primal-dual splitting algorithm for solving monotone inclusions involving a mixture of sums, linear compositions, and parallel sums (inf-convolution in convex optimization) of set-valued and Lipschitz operators. More precisely, the authors exploit the fact that the primal and dual problems have a similar structure and cast the problem as finding a zero of the sum of a Lipschitz-continuous monotone map with a maximal monotone operator whose resolvent is easily computable. They solve the corresponding monotone inclusion using an inexact version of Tseng's forward-backward-forward splitting algorithm [75].

Removing the parallel sum, taking the linear operators as the identity in [26, problem (1.1)], and assuming that the Lipschitz part is also cocoercive, one recovers problem (1.1). For the sake of simplicity and space saving we do not fully reproduce Combettes and Pesquet's algorithm. However, adapted to the optimization problem $\min_{x \in \mathcal{H}} f(x) + \sum_i g_i(L_i x)$, where each L_i is a bounded linear operator, their scheme is presented in Algorithm 3 (g_i^* is the Legendre–Fenchel conjugate of g_i).

Algorithm 3. Iterations of primal-dual algorithm of [26].

Choose a sequence $(\gamma_k)_{k \in \mathbb{N}}$ in $[\epsilon, (1 - \epsilon)/\zeta]$, where $\zeta \stackrel{\text{def}}{=} \beta^{-1} + \sqrt{\sum_i \|L_i\|^2}$ and $\epsilon \in]0, 1/(1 + \zeta)[$.
 $k \leftarrow 0$.

repeat

$y \leftarrow x - \gamma_k (\nabla f(x) + \sum_{i=1}^n L_i^*(v_i))$

for $i \in \llbracket 1, n \rrbracket$ **do**

$z_i \leftarrow v_i + \gamma_k L_i(x)$;

$v_i \leftarrow v_i - z_i + \text{prox}_{\gamma_k g_i^*}(z_i) + \gamma_k L_i(y)$;

$x \leftarrow x - \gamma_k (\nabla f(y) + \sum_{i=1}^n L_i^*(\text{prox}_{\gamma_k g_i^*}(z_i)))$;

$k \leftarrow k + 1$

until convergence;

Recall that the proximity operator of g_i^* can be easily deduced from that of g_i using Moreau's identity. Taking $L_i = \text{Id}$ in Algorithm 3 solves (1.2). While we solve the primal problem, their algorithm solves both the primal and dual problems. Note, however, that it requires two calls to the gradient of f per iteration.

3. Numerical experiments. This section exemplifies the applicability of our GFB splitting algorithm on image processing problems by solving some regularized inverse problems. The problems are selected so that other splitting algorithms can be applied as well and compared fairly. The parameters involved were manually selected for each compared algorithm to achieve its best performance, for instance, in terms of energy decay. In the following, Id denotes the identity operator on the appropriate space to be understood from the context, N is a positive integer, and $\mathcal{I} \equiv \mathbb{R}^{N \times N}$ is the set of images of size $N \times N$ pixels.

3.1. Variational image restoration. We consider a class of inverse problem regularizations, where one wants to recover an (unknown) high resolution image $y_0 \in \mathcal{I}$ from noisy low resolution observations $y = \Phi y_0 + w \in \mathcal{I}$. We report results using several ill-posed linear operators $\Phi : \mathcal{I} \rightarrow \mathcal{I}$, and focus our attention on convolution and a masking operator and a combination of these operators. In the numerical experiments, the noise vector $w \in \mathcal{I}$ is a realization of an additive white Gaussian noise of variance σ_w^2 .

The restored image $\hat{y}_0 = W\hat{x}$ is obtained by optimizing the coefficients $\hat{x} \in \mathcal{H}$ in a redundant wavelet frame [56], where $W : \mathcal{H} \rightarrow \mathcal{I}$ is the wavelet synthesis operator. The wavelet atoms are normalized so that W is a Parseval tight frame, i.e., it satisfies $WW^* = \text{Id}$. In this setting, the coefficients are vectors $x \in \mathcal{H} \equiv \mathcal{I}^J$, where the redundancy $J = 3J_0 + 1$ depends on the number of resolution levels J_0 of the wavelet transform.

The general variational problem for the recovery reads as

$$(3.1) \quad \min_{x \in \mathcal{H}} \{F(x) \equiv \frac{1}{2} \|y - \Phi Wx\|^2 + \mu \|x\|_{1,2}^B + \nu \|Wx\|_{\text{TV}}\} .$$

The first term in the summand is the *data fidelity* term, which is taken to be a squared ℓ_2 -norm to reflect the additive white Gaussianity of the noise. The second and third terms are *regularizations*, enforcing priors assumed to be satisfied by the original image. The first regularization is a ℓ_1/ℓ_2 -norm by blocks, inducing structured sparsity on the solution. The second regularization is a discrete *total variation seminorm*, inducing sparsity on the gradient of the restored image. The scalars μ and ν are weights—so-called regularization parameters—to balance between each terms of the energy F . We now detail the properties of each of these three terms.

3.1.1. Data fidelity $\frac{1}{2} \|y - \Phi Wx\|^2$. For the inpainting inverse problem, one considers a masking operator

$$(My)_p \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } p \in \Omega , \\ y_p & \text{otherwise ,} \end{cases}$$

where Ω is a set of pixels, taking into account missing or defective sensors; we will denote $\rho = |\Omega|/N^2$ the ratio of missing pixels. For the deblurring inverse problem, we consider a convolution with a discrete Gaussian filter of width σ_K , $K : y \mapsto G_{\sigma_K} * y$, normalized to a unit mass. In the following, Φ will be either M , K , or the composition of both, MK .

Denoting $L \stackrel{\text{def}}{=} \Phi W$, the fidelity term thus reads as $f(x) = \frac{1}{2} \|y - Lx\|^2$. The function f corresponds to the smooth term in (1.2). Its gradient $\nabla f : x \mapsto L^*(Lx - y)$ is Lipschitz continuous with constant $\beta^{-1} \leq \|\Phi W\|^2 = 1$.

For any $\gamma > 0$, the proximity operator of f reads as

$$(3.2) \quad \text{prox}_{\gamma f}(x) = (\text{Id} + \gamma L^* L)^{-1} (x + \gamma L^* y) .$$

The vector L^*y can be precomputed, but inverting $\text{Id} + \gamma L^*L$ may be in general computationally demanding. For inpainting or deblurring alone, as W is associated to a Parseval tight frame, the Sherman–Morrison–Woodbury formula gives

$$(3.3) \quad \begin{aligned} (\text{Id} + \gamma L^*L)^{-1} &= \text{Id} - L^*(\text{Id} + \gamma LL^*)^{-1}L \\ &= \text{Id} - W^*\Phi^*(\text{Id} + \gamma\Phi\Phi^*)^{-1}\Phi W . \end{aligned}$$

Since M (resp., K) is a diagonal operator in the pixel domain (resp., Fourier domain), (3.3) can be computed in $O(N^2)$ (resp., $O(N^2 \log N)$) operations. However, the composite case $L \equiv MKW$ is more involved. A possible workaround is to introduce an auxiliary variable, replacing $f : \mathcal{H} \rightarrow \mathbb{R}$ by $\tilde{f} : \mathcal{H} \times \mathcal{I} \rightarrow]-\infty, +\infty]$ defined by

$$(3.4) \quad \tilde{f}(x, u) = \frac{1}{2}\|y - Mu\|^2 + \iota_{C_{KW}}(x, u) = g_1(u) + g_2(x, u) ,$$

where $C_{KW} \stackrel{\text{def}}{=} \{(x, u) \in \mathcal{H} \times \mathcal{I} \mid u = KWx\}$ and ι_C is the indicator function of the closed convex set C ; i.e., $\iota_C(v) = 0$ if $v \in C$, and $+\infty$ otherwise. Only then, $\text{prox}_{\gamma g_1}$ can be computed from (3.2), and $\text{prox}_{\gamma g_2}$ is the orthogonal projection on $\ker([\text{Id}, -KW])$ [12, 33], which involves an inversion similar to that in (3.3).

3.1.2. Regularization $\mu\|x\|_{1,2}^{\mathcal{B}}$. Sparsity-promoting regularizations with a synthesis-type prior over wavelet (or other transformed) coefficients are popular for solving a wide range of inverse problems [56]. Figure 1(a) shows an example of orthogonal wavelet coefficients of a natural image where most of the coefficients have small amplitude. A way to enforce this “sparsity” is to include in (3.1) the ℓ_1 -norm of the coefficients $\|x\|_1 = \sum_p |x_p|$.

The presence of edges or textures creates structured local dependencies in the wavelet coefficients of natural images. A way to take into account those dependencies is to replace the absolute value of the coefficients in the ℓ_1 -norm by the ℓ_2 -norm of groups (or *blocks*) of coefficients [64, 77, 50, 15, 39]. This is known as the mixed ℓ_1/ℓ_2 -norm, defined here as

$$(3.5) \quad \|x\|_{1,2}^{\mathcal{B}} = \sum_{\mathbf{b} \in \mathcal{B}} \mu_{\mathbf{b}} \|x_{\mathbf{b}}\| = \sum_{\mathbf{b} \in \mathcal{B}} \mu_{\mathbf{b}} \sqrt{\sum_{p \in \mathbf{b}} x_p^2} ,$$

where p indexes the coefficients, the blocks \mathbf{b} are sets of indices, the *block-structure* \mathcal{B} is a collection of blocks, and $x_{\mathbf{b}} \stackrel{\text{def}}{=} (x_p)_{p \in \mathbf{b}}$ is a subvector of x indexed by \mathbf{b} . The positive scalars $\mu_{\mathbf{b}}$ are weights tuning the influence of each block. Equation (3.5) defines a norm on \mathcal{H} as soon as \mathcal{B} covers the whole space; i.e., $\forall p \in \llbracket 1, N \rrbracket^2 \times \llbracket 1, J \rrbracket, \exists \mathbf{b} \in \mathcal{B} : p \in \mathbf{b}$ and $\mu_{\mathbf{b}} > 0$. Note that for $\mathcal{B} \equiv \bigcup_p \{p\}$ and $\mu_{\{p\}} \equiv 1$ for all p , $\|x\|_{1,2}^{\mathcal{B}}$ reduces to the ℓ_1 -norm.

We mentioned in the introduction that the proximal operator of the ℓ_1 -norm is the coefficientwise soft-thresholding. Similarly, it is easy to show that whenever \mathcal{B} is a disjoint partition where the blocks are *nonoverlapping*, i.e., $\forall \mathbf{b}, \mathbf{b}' \in \mathcal{B}, \mathbf{b} \cap \mathbf{b}' = \emptyset$, the proximity operator of $\|\cdot\|_{1,2}^{\mathcal{B}}$ is the blockwise soft-thresholding

$$\text{prox}_{\mu\|\cdot\|_{1,2}^{\mathcal{B}}}((x_{\mathbf{b}})_{\mathbf{b}}) = (\Theta_{\mu_{\mathbf{b}} \cdot \mu}(x_{\mathbf{b}}))_{\mathbf{b}} ,$$

with

$$\Theta_{\tau}(x_{\mathbf{b}}) = \begin{cases} 0 & \text{if } \|x_{\mathbf{b}}\| < \tau , \\ \left(1 - \frac{\tau}{\|x_{\mathbf{b}}\|}\right) x_{\mathbf{b}} & \text{otherwise ,} \end{cases}$$

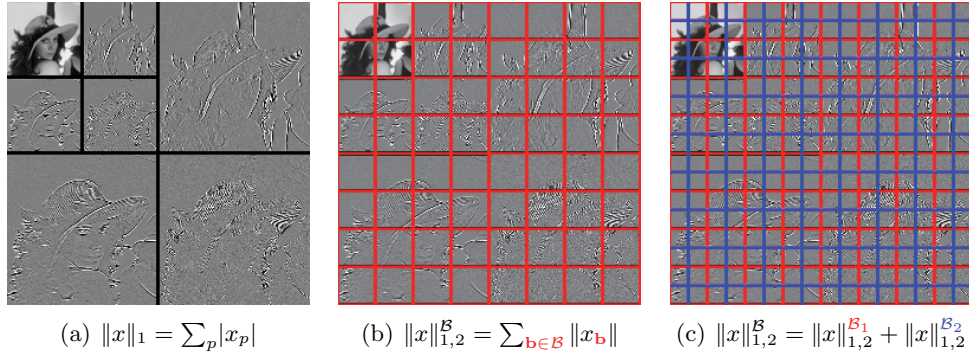


Figure 1. Illustration of the block ℓ_1/ℓ_2 -norm. (a) sparsity of the image in an orthogonal wavelet decomposition (gray pixels corresponds to low coefficients); (b) a nonoverlapping block structure; (c) splitting of a more complex overlapping block structure into two nonoverlapping layers.

and the coefficients x_p not covered by \mathcal{B} are left unaltered.

Nonoverlapping block structures break the *translation invariance* that is underlying most traditional image models. To restore this invariance, one can consider overlapping blocks, as illustrated in Figure 1(c). Computing $\text{prox}_{\|\cdot\|_{1,2}^{\mathcal{B}}}$ in this case is not as simple as for the nonoverlapping case, because the blocks cannot be treated independently. For tree-structured blocks (i.e., $\mathbf{b} \cap \mathbf{b}' \neq \emptyset \Rightarrow \mathbf{b} \subset \mathbf{b}'$ or $\mathbf{b}' \subset \mathbf{b}$), [48] proposes a method involving the computation of a min-cost flow. This could be computationally expensive and does not address the general case anyway. Instead, it is always possible to decompose the block structure as a finite union of nonoverlapping substructures $\mathcal{B} = \bigcup_i \mathcal{B}_i$. The resulting term can finally be split into $\|x\|_{1,2}^{\mathcal{B}} = \sum_{\mathbf{b} \in \mathcal{B}} \|x_{\mathbf{b}}\| = \sum_i \sum_{\mathbf{b} \in \mathcal{B}_i} \|x_{\mathbf{b}}\| = \sum_i \|x\|_{1,2}^{\mathcal{B}_i}$, where each $\|\cdot\|_{1,2}^{\mathcal{B}_i}$ is simple.

In our numerical experiments where $\mathcal{H} \equiv \mathcal{I}^J$, coefficients within each resolution level (from 1 to J_0) and each subband are grouped according to all possible square spatial blocks of size $S \times S$, which can be decomposed into S^2 nonoverlapping block structures.

3.1.3. Regularization $\nu\|Wx\|_{\text{TV}}$. The second regularization favors piecewise-smooth images, by inducing sparsity on its gradient [69]. The TV seminorm can be viewed as a specific instance of the ℓ_1/ℓ_2 -norm, $\|y\|_{\text{TV}} = \| \text{grad } y \|_{1,2}^{\mathcal{B}_{\text{TV}}}$, with

$$\text{grad} : \begin{cases} \mathcal{I} & \longrightarrow \mathcal{I}^2, \\ y & \longmapsto (V * y, H * y) \end{cases} \quad \text{and} \quad \|(v, h)\|_{1,2}^{\mathcal{B}_{\text{TV}}} = \sum_{p \in [1, N]^2} \sqrt{v_p^2 + h_p^2},$$

where the image gradient is computed by finite differences through convolution with a vertical filter V and a horizontal filter H , and \mathcal{B}_{TV} is clearly nonoverlapping. For some special gradient filters, the modified TV seminorm can be split into simple functions; see, for instance, [24, 67]. However, we consider more conventional filters

$$V = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$$

centered in the upper-left corner. Introducing an auxiliary variable as advocated in (3.4), the main difficulty remains to invert the operator $(\text{Id} + \gamma \text{grad} \circ \text{grad}^*)$, where grad^* is the adjoint

of the gradient (i.e., the opposite of the divergence operator). Under appropriate boundary conditions, this can be done in the Fourier domain in $O(N^2 \log(N))$ operations.

3.2. Resolution with splitting methods.

3.2.1. Tested algorithms. We now give the details of the different splitting strategies required to apply the three tested algorithms to (3.1).

Generalized forward-backward (GFB). The problem is rewritten under the form (1.2) as

$$(3.6) \quad \min_{\substack{x \in \mathcal{H} \\ u \in \mathcal{I}^2}} \frac{1}{2} \|y - MKWx\|^2 + \mu \sum_{i=1}^{S^2} \|x\|_{1,2}^{\mathcal{B}_i} + \nu \|u\|_{1,2}^{\mathcal{B}_{\text{TV}}} + \iota_{C_{\text{grad} \circ W}}(x, u),$$

with $f(x) \equiv \frac{1}{2} \|y - MKWx\|^2$ and $n \equiv S^2 + 2$. The indicator function $\iota_{C_{\text{grad} \circ W}}$ is defined similarly as in (3.4). In Algorithm 1, we set equal weights $w_i \equiv 1/n$, a constant gradient step-size $\gamma \equiv 1.8\beta$, and a constant relaxation parameter $\lambda \equiv 1$.

Relaxed Douglas–Rachford (DR). Here the problem is split as

$$\min_{\substack{x \in \mathcal{H} \\ u_1 \in \mathcal{I} \\ u_2 \in \mathcal{I}^2}} \frac{1}{2} \|y - M u_1\|^2 + \iota_{C_{KW}}(x, u_1) + \mu \sum_{i=1}^{S^2} \|x\|_{1,2}^{\mathcal{B}_i} + \nu \|u_2\|_{1,2}^{\mathcal{B}_{\text{TV}}} + \iota_{C_{\text{grad} \circ W}}(x, u_2)$$

and solved with Algorithm 1, where $f \equiv 0$ and $n \equiv S^2 + 4$. As mentioned in section 2.2, this corresponds to a relaxed version of the DR algorithm. In our experiments, the best results were obtained for $\gamma \equiv 1/n$.

Primal-dual Chambolle–Pock (CHPO). A way to avoid operator inversions is to rewrite the original problem as

$$\min_{x \in \mathcal{H}} g(\Lambda x),$$

where

$$\Lambda : \begin{cases} \mathcal{H} & \longrightarrow & \mathcal{I} \times (\mathcal{H})^{S^2} \times \mathcal{I}^2, \\ x & \longmapsto & (MKWx, x, \dots, x, \text{grad} \circ Wx) \end{cases}$$

and

$$g : \begin{cases} \mathcal{I} \times (\mathcal{H})^{S^2} \times \mathcal{I}^2 & \longrightarrow & \mathbb{R}^+, \\ (u_1, x_1, \dots, x_{S^2}, g) & \longmapsto & \frac{1}{2} \|y - u_1\|^2 + \mu \sum_{i=1}^{S^2} \|x_i\|_{1,2}^{\mathcal{B}_i} + \nu \|g\|_{1,2}^{\mathcal{B}_{\text{TV}}}. \end{cases}$$

The operator Λ is a concatenation of linear operators, its adjoint is easy to compute, and g is simple, being a separable sum of simple functions. Note that this is not the only splitting possible. For instance, one can write the problem on a product space as $\min_{(x_i)_{i \in \mathcal{H}}} \iota_{\mathcal{S}}((x_i)_i) + \sum_i g_i(\Lambda_i x_i)$, where g_i is each of the functions in g above and Λ_i is each of the linear operators in Λ .

To solve this, here we use the primal-dual relaxed Arrow–Hurwicz algorithm described in [14]. According to the notation in that paper, we set the parameters $\sigma \equiv 1$, $\tau \equiv \frac{0.9}{\sigma(1+S^2+8)}$, and $\theta \equiv 1$.

Block-decomposition hybrid proximal extragradient (HPE). We split the problem written in (3.6) according to (2.3) and set equal weights $w_i \equiv 1/n$. According to section 2.3, we set the parameter $\varsigma \equiv 0.9$.

Primal-dual Combettes–Pesquet (CoPE). Finally, the problem takes its simplest form

$$(3.7) \quad \min_{x \in \mathcal{H}} \frac{1}{2} \|y - MKWx\|^2 + \mu \sum_{i=1}^{S^2} \|x\|_{1,2}^{\beta_i} + \nu \|\text{grad} \circ Wx\|_{1,2}^{\beta_{\text{TV}}} .$$

As long as $\nu \equiv 0$ (no TV-regularization), this is exactly (3.6); we apply Algorithm 3 where $L_i \equiv \text{Id}$ for all i and $\gamma \equiv 0.9/(1 + S)$. However, with TV-regularization, we avoid the introduction of the auxiliary variable u with $L_{S^2+1} \equiv \text{grad} \circ W$ and $\gamma \equiv 0.9/(1 + \sqrt{S^2 + 8})$.

3.2.2. Results. All experiments were performed on discrete images of width $N \equiv 256$, with values in the range $[0, 1]$. The additive white Gaussian noise has standard deviation $\sigma_w \equiv 2.5 \cdot 10^{-2}$. The reconstruction operator W uses separable bidimensional Daubechies wavelets with two vanishing moments. It is implemented such that each atom has norm 2^{-j} , with $j \in \llbracket 1, J_0 \rrbracket$ and where J_0 is the coarsest resolution level. Accordingly, we set the weights $\mu_{\mathbf{b}}$ in the ℓ_1/ℓ_2 -norm to 2^{-j} at the resolution level j of the coefficients in block \mathbf{b} . We use $J_0 \equiv 4$, resulting in a dictionary with redundancy $J = 3J_0 + 1 = 13$. All algorithms are implemented in MATLAB.¹

Results are presented in Figures 2–5. Only one image is shown here, but we obtained very similar results on other natural images (see <http://www.ceremade.dauphine.fr/~raguet/gfb/>). For each problem, the five algorithms were run 1000 iterations (initialized at zero), while we monitored their objective functional values F along iterations. F_{\min} is fixed as the minimum value reached over the five algorithms (in our experiments, this was always that of GFB), and evolution of the objectives compared to F_{\min} is displayed for the first 100 iterations. Because the computational complexity of an iteration may vary between algorithms, computation times for 100 iterations (no parallel implementation) are given beside the curves. Below the objective decay graph, one can find from left to right the original image, the degraded image, and the restored image after 100 iterations of GFB. Degraded and restored images quality are given in terms of signal-to-noise ratio (SNR).

Comparison to algorithms that do not use the (gradient) explicit step (CHPo, DR). For the first three experiments, there is no TV regularization. In the deblurring task (Figure 2), blocks of size 2×2 are used. GFB is slightly faster than the others, while the iteration cost of CHPo is much higher for this problem. When increasing the block size (inpainting, Figure 3, size 4×4) computation times tend to be similar, but the decay of the objective provided by GFB is clearly faster than that of other algorithms. The advantage of using the gradient information becomes even more salient in the composite case (i.e., $\Phi \equiv MK$): in Figure 4, DR performs hardly better than CHPo. Indeed, in contrast to the previous cases (see section 3.1.1), f is not simple any longer, and the introduction of the auxiliary variable decreases the efficiency of each iteration of DR. This phenomenon is further illustrated in the last case, where the TV is added, introducing another auxiliary variable.

Comparison to algorithms that use the (gradient) explicit step (HPE, CoPE). In the first experiment where n is small, the iterations of the suggested block-decomposition HPE and CoPE are almost as efficient as those of GFB but take more time to compute, especially for

¹The codes for reproducing the experiments, as well as results on other images, are available at <http://www.ceremade.dauphine.fr/~raguet/gfb/>.

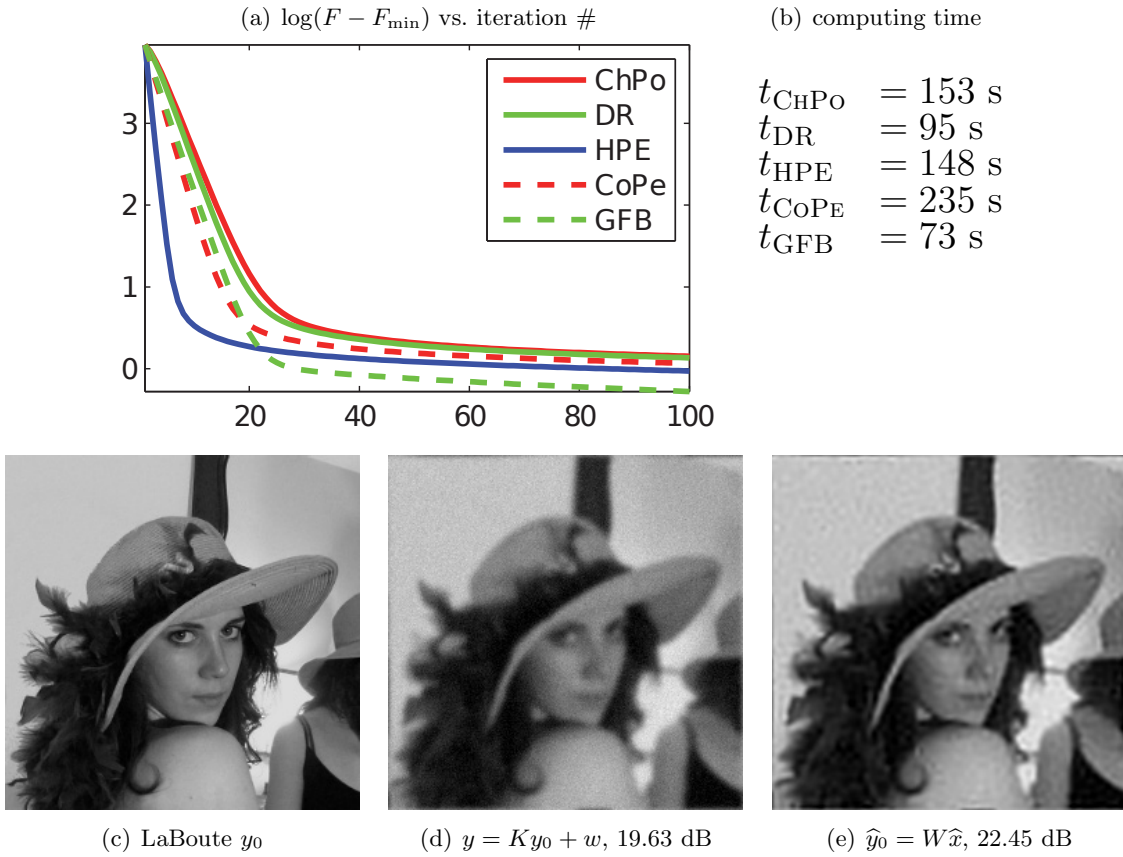


Figure 2. Deblurring: $\sigma = 2$; $\mu = 1.3 \cdot 10^{-3}$; $S = 2$; $\nu = 0$.

CoPE, which needs two calls to ∇f . Recall, however, that HPE and CoPE solve both the primal and dual problems. In the second setting, HPE and CoPE are hardly better than DR. They perform better in the composite setting (i.e., $\Phi \equiv MK$) but require more computational time than GFB. In the last setting, iterations of CoPE are still not as efficient as those of GFB, despite the higher computational load due to the composition by the linear operator $\text{grad} \circ W$.

Finally, let us note that in the composite case, the SNR of the restored image is greater when using both regularizations rather than one or the other separately. Moreover, we observed that mixed regularizations yield restorations more robust to variations of the parameters μ and ν . These arguments seem to be in favor of mixed regularizations.

4. Convergence proofs. This section is dedicated to the proof of convergence of the GFB. We first recall some essential definitions and properties of monotone operator theory that are necessary to our exposition. The interested reader may refer to [6, 65] for a comprehensive treatment. As we will deal with maximal monotone operator splitting, we then introduce specific notation on the product space \mathcal{H}^n . Finally, the proof itself is detailed in two steps. First, we derive an equivalent *fixed point equation* satisfied by any solution of (1.1). From this, we draw an algorithmic scheme (equivalent to GFB) and establish its convergence properties

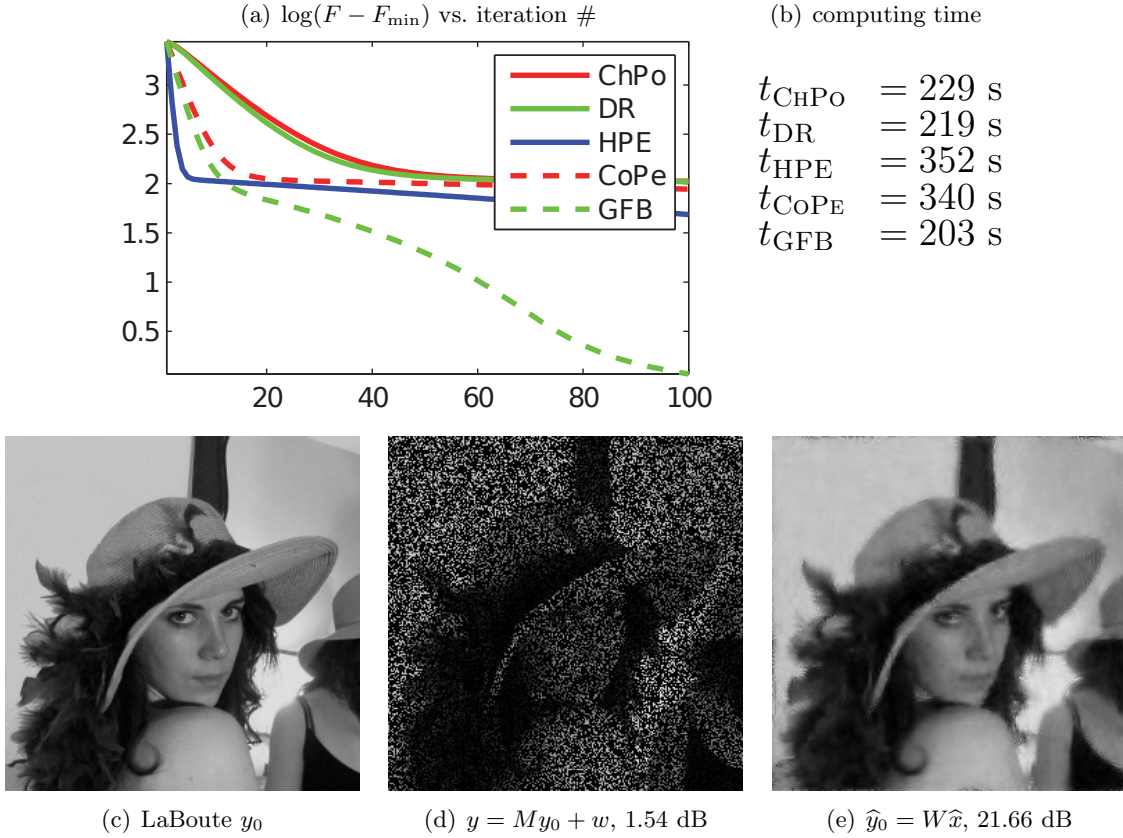


Figure 3. Inpainting: $\rho = 0.7$; $\mu = 2.6 \cdot 10^{-3}$; $S = 4$; $\nu = 0$.

and its robustness to summable errors.

4.1. Definitions and properties. In the following, $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ is a set-valued operator, and $T : \text{dom } T = \mathcal{H} \rightarrow \mathcal{H}$ is a full-domain (see below), single-valued operator. Id denotes the identity operator on \mathcal{H} .

Definition 4.1 (graph, inverse, domain, range, and zeros). The graph of A is the set $\text{gra } A \stackrel{\text{def}}{=} \{(x, y) \in \mathcal{H}^2 \mid y \in Ax\}$. The inverse of A is the operator whose graph is $\text{gra } A^{-1} \stackrel{\text{def}}{=} \{(x, y) \in \mathcal{H}^2 \mid (y, x) \in \text{gra } A\}$. The domain of A is $\text{dom } A \stackrel{\text{def}}{=} \{x \in \mathcal{H} \mid Ax \neq \emptyset\}$. The range of A is $\text{ran } A \stackrel{\text{def}}{=} \{y \in \mathcal{H} \mid \exists x \in \mathcal{H} : y \in Ax\}$, and its zeros set is $\text{zer } A \stackrel{\text{def}}{=} \{x \in \mathcal{H} \mid 0 \in Ax\} = A^{-1}(0)$.

Definition 4.2 (resolvent and reflection operators). The resolvent of A is the operator $J_A \stackrel{\text{def}}{=} (\text{Id} + A)^{-1}$. The reflection operator associated to J_A is the operator $R_A \stackrel{\text{def}}{=} 2J_A - \text{Id}$.

Definition 4.3 (maximal monotone operator). A is monotone if

$$\forall x, y \in \mathcal{H}, \quad (u \in Ax \text{ and } v \in Ay) \Rightarrow \langle u - v \mid x - y \rangle \geq 0.$$

It is, moreover, maximal monotone if its graph is not strictly contained in the graph of any other monotone operator.

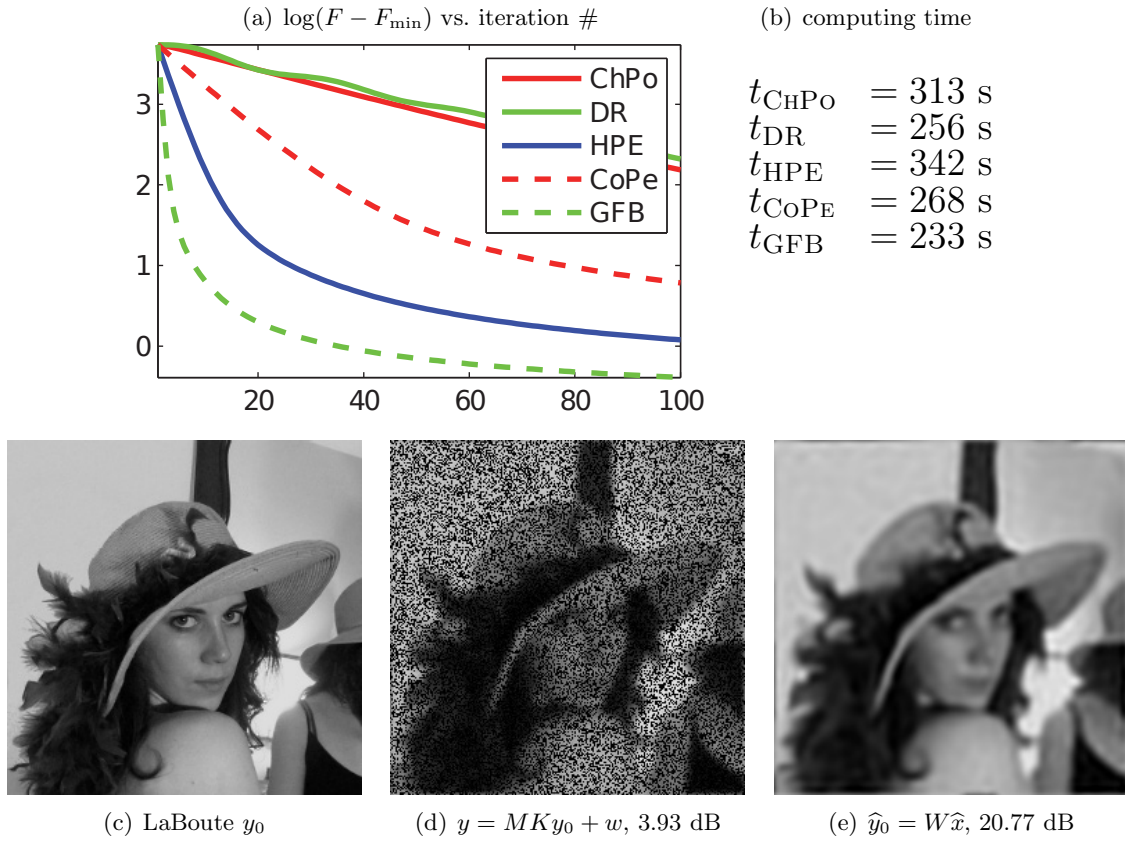


Figure 4. Composite: $\sigma = 2$; $\rho = 0.4$; $\mu = 1.0 \cdot 10^{-3}$; $S = 4$; $\nu = 0$.

Definition 4.4 (uniformly monotone operator). A is uniformly monotone of modulus $\varphi : [0, +\infty[\rightarrow [0, +\infty[$ if φ is a nondecreasing function that vanishes only at 0 such that

$$\forall x, y \in \mathcal{H}, \quad (u \in Ax \text{ and } v \in Ay) \Rightarrow \langle u - v | x - y \rangle \geq \varphi(\|x - y\|).$$

Definition 4.5 (nonexpansive and α -averaged operators). T is nonexpansive if

$$\forall x, y \in \mathcal{H}, \quad \|Tx - Ty\| \leq \|x - y\|.$$

For $\alpha \in]0, 1[$, T is α -averaged if there exists R nonexpansive such that $T = (1 - \alpha)\text{Id} + \alpha R$. We denote by $\mathcal{A}(\alpha)$ the class of α -averaged operators on \mathcal{H} . In particular, $\mathcal{A}(\frac{1}{2})$ is the class of firmly nonexpansive operators.

Definition 4.6 (cocoercive operator). For $\beta \in]0, +\infty[$, T is β -cocoercive if $\beta T \in \mathcal{A}(\frac{1}{2})$.

The following lemma gives some useful characterizations of firmly nonexpansive operators.

Lemma 4.7. The following statements are equivalent:

- (i) T is firmly nonexpansive;
- (ii) $2T - \text{Id}$ is nonexpansive;
- (iii) $\forall x, y \in \mathcal{H}, \|Tx - Ty\|^2 \leq \langle Tx - Ty | x - y \rangle$;
- (iv) T is the resolvent of a maximal monotone operator A , i.e., $T = J_A$.

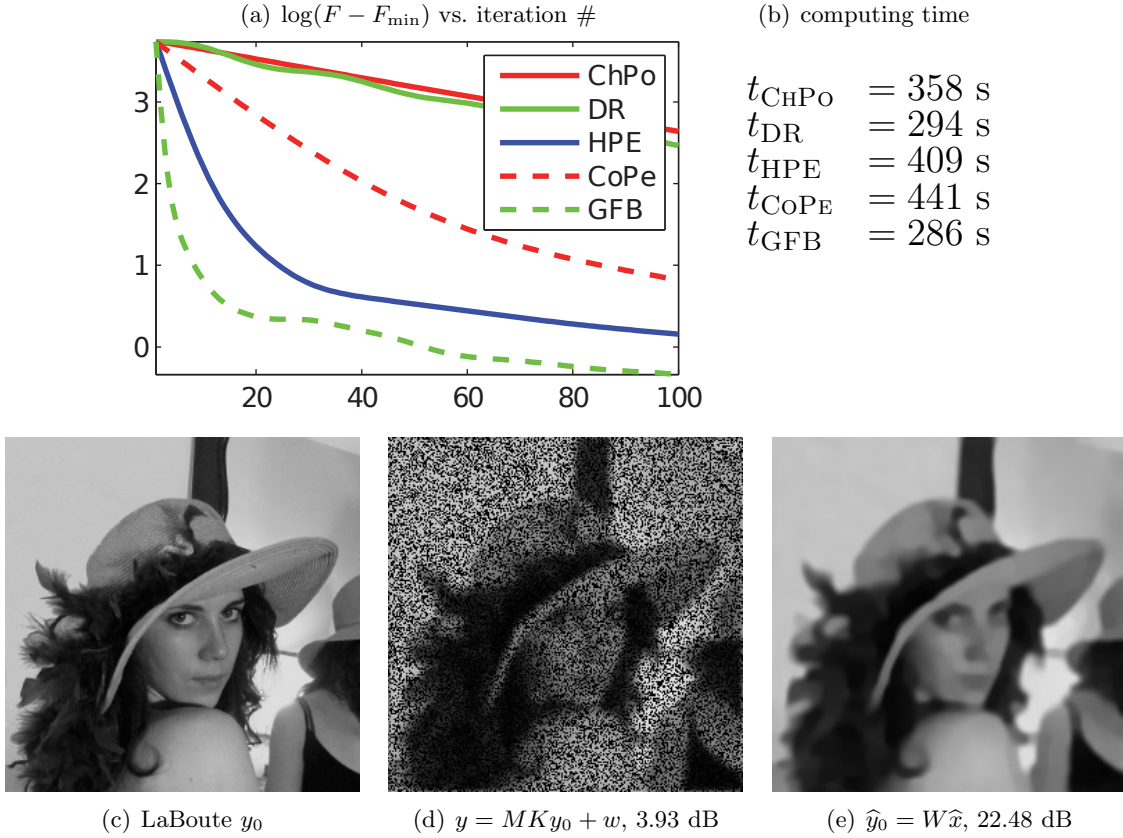


Figure 5. Composite: $\sigma = 2$; $\rho = 0.4$; $\mu = 5.0 \cdot 10^{-4}$; $S = 4$; $\nu = 5.0 \cdot 10^{-3}$.

Proof. (i) \Leftrightarrow (ii). $T \in \mathcal{A}(\frac{1}{2}) \Leftrightarrow T = \frac{\text{Id}+R}{2}$ for some R nonexpansive.

(i) \Leftrightarrow (iii). For the proof, see [78].

(i) \Leftrightarrow (iv). For the proof, see [58]. ■

Note that with (iii), one retrieves the characterization of the cocoercivity given in section 2.1. It follows by the Cauchy–Schwarz inequality that β -cocoercivity implies $1/\beta$ -Lipschitz continuity, but the converse is not true in general. It turns out, however, to be the case for gradients of convex functionals. We summarize here some properties of the subdifferential.

Lemma 4.8. *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ be a convex differentiable function, with $1/\beta$ -Lipschitz-continuous gradient, $\beta \in]0, +\infty[$, and let $g \in \Gamma_0(\mathcal{H})$. Then,*

(i) $\beta \nabla f \in \mathcal{A}(\frac{1}{2})$, i.e., is firmly nonexpansive;

(ii) ∂g is maximal monotone;

(iii) the resolvent of ∂g is the proximity operator of g , i.e., $\text{prox}_g = J_{\partial g}$.

Proof. (i) This is the Baillon–Haddad theorem [5].

(ii) For the proof, see [68].

(iii) For the proof, see [60]. ■

4.2. Product space. Let $(w_i)_{i \in \llbracket 1, n \rrbracket} \in]0, 1]^n$ such that $\sum_{i=1}^n w_i = 1$. We consider $\mathcal{H} \stackrel{\text{def}}{=} \mathcal{H}^n$ endowed with the scalar product $\langle \cdot | \cdot \rangle$, defined as

$$\forall \mathbf{x} = (x_i)_i, \mathbf{y} = (y_i)_i \in \mathcal{H}, \quad \langle \mathbf{x} | \mathbf{y} \rangle = \sum_{i=1}^n w_i \langle x_i | y_i \rangle ,$$

and with the corresponding norm $\| \cdot \|$. $\mathcal{S} \subset \mathcal{H}$ denotes the nonempty closed convex set defined by $\mathcal{S} \stackrel{\text{def}}{=} \{ \mathbf{x} = (x_i)_i \in \mathcal{H} \mid x_1 = x_2 = \dots = x_n \}$, whose orthogonal complement is the closed linear subspace \mathcal{S}^\perp . We denote by Id the identity operator on \mathcal{H} , and we define the canonical isometry

$$C : \mathcal{H} \rightarrow \mathcal{S}, x \mapsto (x, \dots, x) .$$

$\iota_{\mathcal{S}} : \mathcal{H} \rightarrow]-\infty, +\infty]$ and $N_{\mathcal{S}} : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ are, respectively, the *indicator function* and the *normal cone* of \mathcal{S} , that is

$$\iota_{\mathcal{S}}(\mathbf{x}) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } \mathbf{x} \in \mathcal{S} , \\ +\infty & \text{otherwise} \end{cases} \quad \text{and} \quad N_{\mathcal{S}}(\mathbf{x}) \stackrel{\text{def}}{=} \begin{cases} \mathcal{S}^\perp & \text{if } \mathbf{x} \in \mathcal{S} , \\ \emptyset & \text{otherwise} . \end{cases}$$

Since \mathcal{S} is nonempty closed and convex, it is straightforward to see that $N_{\mathcal{S}}$ is maximal monotone.

We also introduce the following concatenated operators. Fix B and the A_i 's in problem (1.1). Given $\gamma = (\gamma_i)_{i \in \llbracket 1, n \rrbracket} \in]0, +\infty[^n$, we define

$$\gamma \bullet \mathbf{A} : \mathcal{H} \rightarrow 2^{\mathcal{H}}, \mathbf{x} = (x_i)_i \mapsto \bigtimes_{i=1}^n \gamma_i A_i(x_i) ;$$

i.e., its graph is

$$\begin{aligned} \text{gra } \gamma \bullet \mathbf{A} &\stackrel{\text{def}}{=} \bigtimes_{i=1}^n \text{gra } \gamma_i A_i \\ &= \{ (\mathbf{x}, \mathbf{y}) \in \mathcal{H}^2 \mid \mathbf{x} = (x_i)_i, \mathbf{y} = (y_i)_i, \text{ and } \forall i, y_i \in \gamma_i A_i x_i \} , \end{aligned}$$

and $\mathbf{B} : \mathcal{H} \rightarrow \mathcal{H}, \mathbf{x} = (x_i)_i \mapsto (Bx_i)_i$. Using the maximal monotonicity of the A_i 's and the β -cocoercivity of B , it is an easy exercise to establish that $\gamma \bullet \mathbf{A}$ is maximal monotone and \mathbf{B} is β -cocoercive on \mathcal{H} .

4.3. Fixed point equation. Now that we have all necessary material, let us characterize solutions of (1.1).

Proposition 4.9. *Let $(w_i)_{i \in \llbracket 1, n \rrbracket} \in]0, 1]^n$. For any $\gamma > 0$, $x \in \mathcal{H}$ is a solution of (1.1) if and only if there exists $(z_i)_{i \in \llbracket 1, n \rrbracket} \in \mathcal{H}^n$ such that*

$$(4.1) \quad \begin{cases} \forall i, z_i = R_{\frac{\gamma}{w_i} A_i}(2x - z_i - \gamma Bx) - \gamma Bx , \\ x = \sum_i w_i z_i . \end{cases}$$

Proof. Setting $\gamma > 0$, we have the equivalence

$$0 \in Bx + \sum_i A_i x \Leftrightarrow \exists (z_i)_i \in \mathcal{H}^n : \begin{cases} \forall i, w_i(x - z_i - \gamma Bx) \in \gamma A_i x, \\ x = \sum_i w_i z_i. \end{cases}$$

Now,

$$\begin{aligned} w_i(x - z_i - \gamma Bx) \in \gamma A_i x &\Leftrightarrow (2x - z_i - \gamma Bx) - x \in \frac{\gamma}{w_i} A_i x \\ &\text{(by Lemma 4.7(iv)) } \Leftrightarrow x = J_{\frac{\gamma}{w_i} A_i}(2x - z_i - \gamma Bx) \\ &\Leftrightarrow 2x - (2x - z_i) = 2J_{\frac{\gamma}{w_i} A_i}(2x - z_i - \gamma Bx) \\ &\quad - (2x - z_i - \gamma Bx) - \gamma Bx \\ &\Leftrightarrow z_i = R_{\frac{\gamma}{w_i} A_i}(2x - z_i - \gamma Bx) - \gamma Bx. \quad \blacksquare \end{aligned}$$

From now on, to lighten the notation, we denote $P_{\mathcal{S}} \stackrel{\text{def}}{=} J_{N_{\mathcal{S}}}$ and $R_{\mathcal{S}} \stackrel{\text{def}}{=} R_{N_{\mathcal{S}}}$. Before formulating our fixed point equation, we need the following preparatory lemma.

Lemma 4.10. For all $\mathbf{z} = (z_i)_i \in \mathcal{H}$, $\mathbf{b} = (b_i)_i \in \mathcal{S}$, and $\boldsymbol{\gamma} = (\gamma_i)_i \in]0, +\infty[^n$,

- (i) $P_{\mathcal{S}}$ is the orthogonal projector on \mathcal{S} , and $P_{\mathcal{S}}\mathbf{z} = \mathbf{C}(\sum_i w_i z_i)$;
- (ii) $R_{\mathcal{S}}(\mathbf{z} - \mathbf{b}) = R_{\mathcal{S}}\mathbf{z} - \mathbf{b}$;
- (iii) $R_{\boldsymbol{\gamma}\bullet\mathbf{A}}\mathbf{z} = (R_{\gamma_i A_i}(z_i))_i$.

Proof. (i) From Lemma 4.8(iii), we have for $\mathbf{z} \in \mathcal{H}$

$$P_{\mathcal{S}}(\mathbf{z}) = \operatorname{argmin}_{\mathbf{y} \in \mathcal{S}} \|\mathbf{z} - \mathbf{y}\| \stackrel{\text{def}}{=} \operatorname{proj}_{\mathcal{S}}(\mathbf{z}).$$

Now, $\operatorname{argmin}_{\mathbf{y} \in \mathcal{S}} \|\mathbf{z} - \mathbf{y}\|^2 = \mathbf{C}(\operatorname{argmin}_{y \in \mathcal{H}} \sum_i w_i \|z_i - y\|^2)$, where the unique minimizer of $\sum_i w_i \|z_i - y\|^2$ is the barycenter of $(z_i)_i$, i.e., $\sum_i w_i z_i$.

(ii) $P_{\mathcal{S}}$ is obviously linear, and so is $R_{\mathcal{S}}$. Since $\mathbf{b} \in \mathcal{S}$, $R_{\mathcal{S}}\mathbf{b} = \mathbf{b}$, and the result follows.

(iii) This is a consequence of the separability of $\boldsymbol{\gamma}\bullet\mathbf{A}$ in terms of the components of \mathbf{z} implying that $J_{\boldsymbol{\gamma}\bullet\mathbf{A}}\mathbf{z} = (J_{\gamma_i A_i} z_i)_i$. The result follows from the definition of $R_{\boldsymbol{\gamma}\bullet\mathbf{A}}$. \blacksquare

In what follows, we denote the set of *fixed points* of an operator $\mathbf{T} : \mathcal{H} \rightarrow \mathcal{H}$ by $\operatorname{fix} \mathbf{T} \stackrel{\text{def}}{=} \{\mathbf{z} \in \mathcal{H} \mid \mathbf{T}\mathbf{z} = \mathbf{z}\}$.

Proposition 4.11. $(z_i)_{i \in [1, n]} \in \mathcal{H}^n$ satisfies (4.1) if and only if $\mathbf{z} = (z_i)_i$ is a fixed point of the following operator:

$$(4.2) \quad \begin{aligned} \mathcal{H} &\longrightarrow \mathcal{H}, \\ \mathbf{z} &\longmapsto \frac{1}{2} [R_{\boldsymbol{\gamma}\bullet\mathbf{A}} R_{\mathcal{S}} + \mathbf{Id}] [\mathbf{Id} - \gamma \mathbf{B} P_{\mathcal{S}}](\mathbf{z}), \end{aligned}$$

with $\boldsymbol{\gamma} = (\frac{\gamma}{w_i})_i$.

Proof. Using Lemma 4.10 in (4.1), we have $\mathbf{C}(x) = P_{\mathcal{S}}\mathbf{z}$, $\mathbf{C}(Bx) = \mathbf{B} P_{\mathcal{S}}(\mathbf{z})$, and $R_{\mathcal{S}} - \gamma \mathbf{B} P_{\mathcal{S}} = R_{\mathcal{S}}[\mathbf{Id} - \gamma \mathbf{B} P_{\mathcal{S}}]$. Altogether, this yields

$$\begin{aligned} \mathbf{z} \text{ satisfies (4.1)} &\Leftrightarrow \mathbf{z} = R_{\boldsymbol{\gamma}\bullet\mathbf{A}} R_{\mathcal{S}} [\mathbf{Id} - \gamma \mathbf{B} P_{\mathcal{S}}] \mathbf{z} - \gamma \mathbf{B} P_{\mathcal{S}} \mathbf{z} \\ &\Leftrightarrow 2\mathbf{z} = R_{\boldsymbol{\gamma}\bullet\mathbf{A}} R_{\mathcal{S}} [\mathbf{Id} - \gamma \mathbf{B} P_{\mathcal{S}}] \mathbf{z} + [\mathbf{Id} - \gamma \mathbf{B} P_{\mathcal{S}}] \mathbf{z} \\ &\Leftrightarrow \mathbf{z} = \frac{1}{2} [R_{\boldsymbol{\gamma}\bullet\mathbf{A}} R_{\mathcal{S}} + \mathbf{Id}] [\mathbf{Id} - \gamma \mathbf{B} P_{\mathcal{S}}] \mathbf{z}. \quad \blacksquare \end{aligned}$$

4.4. Properties of the fixed point operator. Expression (4.2) gives us the operator on which is based our GFB scheme. We first study the properties of this operator that will play a crucial role in the convergence proof.

Proposition 4.12. *For any $\gamma \in]0, +\infty[^m$, define*

$$(4.3) \quad \mathbf{T}_{1,\gamma} : \begin{array}{l} \mathcal{H} \longrightarrow \mathcal{H} , \\ z \longmapsto \frac{1}{2} [R_{\gamma\mathbf{e}\mathbf{A}} R_{\mathcal{S}} + \text{Id}] z . \end{array}$$

Then, $\mathbf{T}_{1,\gamma}$ is firmly nonexpansive; i.e., $\mathbf{T}_{1,\gamma} \in \mathcal{A}(\frac{1}{2})$.

Proof. From Lemma 4.7(i) \Leftrightarrow (ii), $R_{\gamma\mathbf{e}\mathbf{A}_i}$ and $R_{\mathcal{S}}$ are nonexpansive, and so is $R_{\gamma\mathbf{e}\mathbf{A}}$ in view of Lemma 4.10(iii). Finally, as a composition of nonexpansive operators, $R_{\gamma\mathbf{e}\mathbf{A}} R_{\mathcal{S}}$ is also nonexpansive, and the proof is complete by definition of $\mathcal{A}(\frac{1}{2})$. ■

Proposition 4.13. *For any $\gamma \in]0, 2\beta[$, define*

$$(4.4) \quad \mathbf{T}_{2,\gamma} : \begin{array}{l} \mathcal{H} \longrightarrow \mathcal{H} , \\ z \longmapsto [\text{Id} - \gamma \mathbf{B}P_{\mathcal{S}}] z . \end{array}$$

Then, $\mathbf{T}_{2,\gamma} \in \mathcal{A}(\frac{\gamma}{2\beta})$.

Proof. By hypothesis, $\beta B \in \mathcal{A}(\frac{1}{2})$, and so is $\beta \mathbf{B}$. Then, from Lemma 4.7(iii), we have for any $\mathbf{x}, \mathbf{y} \in \mathcal{H}$

$$(4.5) \quad \begin{aligned} \|\beta \mathbf{B}P_{\mathcal{S}}\mathbf{x} - \beta \mathbf{B}P_{\mathcal{S}}\mathbf{y}\|^2 &\leq \langle \beta \mathbf{B}P_{\mathcal{S}}\mathbf{x} - \beta \mathbf{B}P_{\mathcal{S}}\mathbf{y} | P_{\mathcal{S}}\mathbf{x} - P_{\mathcal{S}}\mathbf{y} \rangle \\ &= \langle \beta P_{\mathcal{S}}\mathbf{B}P_{\mathcal{S}}\mathbf{x} - \beta P_{\mathcal{S}}\mathbf{B}P_{\mathcal{S}}\mathbf{y} | \mathbf{x} - \mathbf{y} \rangle \\ &= \langle \beta \mathbf{B}P_{\mathcal{S}}\mathbf{x} - \beta \mathbf{B}P_{\mathcal{S}}\mathbf{y} | \mathbf{x} - \mathbf{y} \rangle , \end{aligned}$$

where we derive the first equality from the fact that $P_{\mathcal{S}}$ is self-adjoint (Lemma 4.10(i)), and the second equality using that for all $\mathbf{x} \in \mathcal{H}$, $\mathbf{B}P_{\mathcal{S}}\mathbf{x} \in \mathcal{S}$. From Lemma 4.7(iii) \Leftrightarrow (i), we establish that $\beta \mathbf{B}P_{\mathcal{S}} \in \mathcal{A}(\frac{1}{2})$. We conclude the proof using [20, Lemma 2.3]. ■

Proposition 4.14. *For all $\gamma \in]0, +\infty[^m$ and $\gamma \in]0, 2\beta[$, we have that $\mathbf{T}_{1,\gamma}\mathbf{T}_{2,\gamma} \in \mathcal{A}(\alpha)$, with $\alpha = \max(\frac{2}{3}, \frac{2}{1+2\beta/\gamma})$.*

Proof. As $\mathbf{T}_{1,\gamma}$ and $\mathbf{T}_{2,\gamma}$ are α -averaged operators by Propositions 4.12 and 4.13, it follows from [20, Lemma 2.2(iii)] that their composition is also α -averaged with the given value of α . ■

The following proposition defines a maximal monotone operator \mathbf{A}'_{γ} which will be useful for characterizing the operator $\mathbf{T}_{1,\gamma}$.

Proposition 4.15. *For all $\gamma \in]0, +\infty[^m$ there exists a maximal monotone operator \mathbf{A}'_{γ} such that $\mathbf{T}_{1,\gamma} = J_{\mathbf{A}'_{\gamma}}$. Moreover, for all $\gamma > 0$,*

$$(4.6) \quad \mathbf{y} = \mathbf{T}_{1,\gamma}\mathbf{T}_{2,\gamma}z \Leftrightarrow z - \mathbf{y} - \gamma \mathbf{B}P_{\mathcal{S}}z \in \mathbf{A}'_{\gamma}\mathbf{y} .$$

In particular,

$$\text{fix } \mathbf{T}_{1,\gamma}\mathbf{T}_{2,\gamma} = \text{zer}(\mathbf{A}'_{\gamma} + \gamma \mathbf{B}P_{\mathcal{S}}) .$$

Proof. The existence of \mathbf{A}'_{γ} is ensured by Proposition 4.12 and Lemma 4.7(iv). Then for $z \in \mathcal{H}$,

$$\begin{aligned} \mathbf{y} = \mathbf{T}_{1,\gamma}\mathbf{T}_{2,\gamma}z &\Leftrightarrow \mathbf{y} = (\text{Id} + \mathbf{A}'_{\gamma})^{-1}(\text{Id} - \gamma \mathbf{B}P_{\mathcal{S}})z \\ &\Leftrightarrow z - \gamma \mathbf{B}P_{\mathcal{S}}z - \mathbf{y} \in \mathbf{A}'_{\gamma}\mathbf{y} . \end{aligned}$$

Taking $\mathbf{y} = \mathbf{z}$ proves the second statement. \blacksquare

Now, let us examine the properties of \mathbf{A}'_γ .

Proposition 4.16. *For all $\gamma \in]0, +\infty[^n$ and $(\mathbf{u}, \mathbf{y}) \in \mathcal{H}^2$,*

$$(4.7) \quad \mathbf{u} \in \mathbf{A}'_\gamma \mathbf{y} \Leftrightarrow \mathbf{u}^{\mathcal{S}} - \mathbf{y}^\perp \in \gamma \bullet \mathbf{A} \left(\mathbf{y}^{\mathcal{S}} - \mathbf{u}^\perp \right) ,$$

where we denote for any $\mathbf{v} \in \mathcal{H}$, $\mathbf{v}^{\mathcal{S}} \stackrel{\text{def}}{=} P_{\mathcal{S}}(\mathbf{v})$ and $\mathbf{v}^\perp \stackrel{\text{def}}{=} P_{\mathcal{S}^\perp}(\mathbf{v})$.

Proof. First, by definition of $\mathbf{T}_{1,\gamma}$ we have

$$(4.8) \quad \begin{aligned} \mathbf{T}_{1,\gamma} &= \frac{1}{2} [(2J_{\gamma \bullet \mathbf{A}} - \mathbf{Id})(2P_{\mathcal{S}} - \mathbf{Id}) + \mathbf{Id}] \\ &= \frac{1}{2} [2J_{\gamma \bullet \mathbf{A}}(P_{\mathcal{S}} - P_{\mathcal{S}^\perp}) - (P_{\mathcal{S}} - P_{\mathcal{S}^\perp}) + P_{\mathcal{S}} + P_{\mathcal{S}^\perp}] \\ &= J_{\gamma \bullet \mathbf{A}}(P_{\mathcal{S}} - P_{\mathcal{S}^\perp}) + P_{\mathcal{S}^\perp} . \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{u} \in \mathbf{A}'_\gamma \mathbf{y} &\Leftrightarrow \mathbf{T}_{1,\gamma}(\mathbf{u} + \mathbf{y}) = \mathbf{y} \\ (\text{by (4.8)}) &\Leftrightarrow J_{\gamma \bullet \mathbf{A}} \left((\mathbf{u} + \mathbf{y})^{\mathcal{S}} - (\mathbf{u} + \mathbf{y})^\perp \right) = \mathbf{y} - (\mathbf{u} + \mathbf{y})^\perp = \mathbf{y}^{\mathcal{S}} - \mathbf{u}^\perp \\ &\Leftrightarrow (\mathbf{u} + \mathbf{y})^{\mathcal{S}} - (\mathbf{u} + \mathbf{y})^\perp - \mathbf{y}^{\mathcal{S}} + \mathbf{u}^\perp \in \gamma \bullet \mathbf{A} \left(\mathbf{y}^{\mathcal{S}} - \mathbf{u}^\perp \right) \\ &\Leftrightarrow \mathbf{u}^{\mathcal{S}} - \mathbf{y}^\perp \in \gamma \bullet \mathbf{A} \left(\mathbf{y}^{\mathcal{S}} - \mathbf{u}^\perp \right) . \quad \blacksquare \end{aligned}$$

4.5. Convergence. We are now ready to state the main convergence result of our relaxed and inexact GFB splitting algorithm (2.2) to solve (1.1).

Theorem 4.17. *Let $\gamma \in]0, 2\beta[$, and set $\boldsymbol{\gamma} = (\frac{\gamma}{w_i})_i \in]0, +\infty[^n$, let $(\lambda_k)_{k \in \mathbb{N}}$ be a sequence in $]0, \min(\frac{3}{2}, \frac{1}{2} + \frac{\beta}{\gamma})[$, set $\mathbf{z}_0 \in \mathcal{H}$, and for every $k \in \mathbb{N}$, set*

$$(4.9) \quad \mathbf{z}_{k+1} = \mathbf{z}_k + \lambda_k (\mathbf{T}_{1,\gamma}(\mathbf{T}_{2,\gamma} \mathbf{z}_k + \boldsymbol{\varepsilon}_{2,k}) + \boldsymbol{\varepsilon}_{1,k} - \mathbf{z}_k) ,$$

where $\mathbf{T}_{1,\gamma}$ (resp., $\mathbf{T}_{2,\gamma}$) is defined in (4.3) (resp., in (4.4)), and $\boldsymbol{\varepsilon}_{1,k}, \boldsymbol{\varepsilon}_{2,k} \in \mathcal{H}$. If

- (i) $\text{zer}(B + \sum_i A_i) \neq \emptyset$;
- (ii) $0 < \inf_{k \in \mathbb{N}} \lambda_k \leq \sup_{k \in \mathbb{N}} \lambda_k < \min(\frac{3}{2}, \frac{1}{2} + \frac{\beta}{\gamma})$;
- (iii) $\sum_{k=0}^{+\infty} \|\boldsymbol{\varepsilon}_{1,k}\| < +\infty$ and $\sum_{k=0}^{+\infty} \|\boldsymbol{\varepsilon}_{2,k}\| < +\infty$

are satisfied, then

- (a) $(\mathbf{T}_{1,\gamma} \mathbf{T}_{2,\gamma} \mathbf{z}_k - \mathbf{z}_k)_{k \in \mathbb{N}}$ converges strongly to 0;
- (b) $(\mathbf{z}_k)_{k \in \mathbb{N}}$ converges weakly to a point $\mathbf{z} \in \text{fix } \mathbf{T}_{1,\gamma} \mathbf{T}_{2,\gamma}$;
- (c) $(x_k \stackrel{\text{def}}{=} \sum_i w_i \mathbf{z}_{i,k})_{k \in \mathbb{N}}$ converges weakly to $x \stackrel{\text{def}}{=} \sum_i w_i \mathbf{z}_i \in \text{zer}(B + \sum_i A_i)$.
- (d) Moreover, if $\forall k \in \mathbb{N}, \lambda_k \leq 1$, $(x_k)_{k \in \mathbb{N}}$ converges strongly in each of the following cases:
 - (1) B is uniformly monotone.
 - (2) $\times_{i=1}^n w_i^{-1} A_i$ is uniformly monotone. For instance, this is true if $\forall i \in \llbracket 1, n \rrbracket$, A_i is uniformly monotone with the same modulus φ being also subadditive or convex.

Proof. (a) Denoting $\mathbf{T} \stackrel{\text{def}}{=} \mathbf{T}_{1,\gamma} \mathbf{T}_{2,\gamma}$, we have, for all $k \in \mathbb{N}$,

$$(4.10) \quad \mathbf{z}_{k+1} = \mathbf{z}_k + \lambda_k (\mathbf{T} \mathbf{z}_k + \boldsymbol{\varepsilon}_k - \mathbf{z}_k) ,$$

with $\boldsymbol{\varepsilon}_k \stackrel{\text{def}}{=} \mathbf{T}_{1,\gamma} (\mathbf{T}_{2,\gamma} \mathbf{z}_k + \boldsymbol{\varepsilon}_{2,k}) - \mathbf{T}_{1,\gamma} (\mathbf{T}_{2,\gamma} \mathbf{z}_k) + \boldsymbol{\varepsilon}_{1,k}$. Proposition 4.12 shows that $\mathbf{T}_{1,\gamma} \in \mathcal{A}(\frac{1}{2})$ is in particular nonexpansive, so that $\|\boldsymbol{\varepsilon}_k\| \leq \|\boldsymbol{\varepsilon}_{2,k}\| + \|\boldsymbol{\varepsilon}_{1,k}\|$, and we deduce from (iii) that $\sum_{k=0}^{+\infty} \|\boldsymbol{\varepsilon}_k\| < +\infty$. Moreover, by Proposition 4.14, $\mathbf{T} \in \mathcal{A}(\alpha)$ with $\alpha = \max(\frac{2}{3}, \frac{2}{1+2\beta/\gamma})$. In particular, \mathbf{T} is nonexpansive, and thus $\text{fix } \mathbf{T}$ is closed and convex. Now, for $k \in \mathbb{N}$, setting $\mathbf{T}_k \stackrel{\text{def}}{=} \mathbf{Id} + \lambda_k (\mathbf{T} - \mathbf{Id})$, the iterations (4.10) can be rewritten as

$$(4.11) \quad \mathbf{z}_{k+1} = \mathbf{T}_k \mathbf{z}_k + \lambda_k \boldsymbol{\varepsilon}_k .$$

Since (ii) provides for all $k \in \mathbb{N}$, $\alpha_k \stackrel{\text{def}}{=} \lambda_k \alpha < 1$, [20, Lemma 2.2(i)] shows that $\mathbf{T}_k \in \mathcal{A}(\alpha_k)$, and (4.11) is thus a particular instance of [20, Algorithm 4.1]. Also, it is clear that for all $k \in \mathbb{N}$, $\text{fix } \mathbf{T}_k = \text{fix } \mathbf{T}$. Thus with Propositions 4.9 and 4.11, (i) provides $\bigcap_{k \in \mathbb{N}} \text{fix } \mathbf{T}_k = \text{fix } \mathbf{T} \neq \emptyset$. According to (ii), $\inf_{k \in \mathbb{N}} \lambda_k > 0$ and $\sup_{k \in \mathbb{N}} \alpha_k < 1$, so we deduce from [20, Theorem 3.1 and Remark 3.4] that

$$(4.12) \quad \sum_{k \in \mathbb{N}} \left\| \mathbf{T}_k \mathbf{z}_k - \mathbf{z}_k \right\|^2 < +\infty$$

and that $(\mathbf{z}_k)_{k \in \mathbb{N}}$ is *quasi-Fejér monotone* with respect to $\text{fix } \mathbf{T}$. By definition of \mathbf{T}_k , (4.12) gives $\sum_{k \in \mathbb{N}} \lambda_k^2 \|\mathbf{T} \mathbf{z}_k - \mathbf{z}_k\|^2 < +\infty$, which in turn implies $\mathbf{T} \mathbf{z}_k - \mathbf{z}_k \rightarrow 0$ since $\inf_{k \in \mathbb{N}} \lambda_k > 0$.

(b) \mathbf{T} being nonexpansive, it follows from the semiclosed principle [13], [6, Corollary 4.18] that any weak cluster point of $(\mathbf{z}_k)_{k \in \mathbb{N}}$ belongs to $\text{fix } \mathbf{T}$, so that [6, Theorem 5.5] provides weak convergence toward $\mathbf{z} \in \text{fix } \mathbf{T}$.

(c) For any $y \in \mathcal{H}$, $\langle y | x_k - x \rangle = \langle y | \sum_i w_i (\mathbf{z}_{i,k} - \mathbf{z}_i) \rangle = \sum_i w_i \langle y | \mathbf{z}_{i,k} - \mathbf{z}_i \rangle = \langle \mathbf{C}(y) | \mathbf{z}_k - \mathbf{z} \rangle$. So, (b) provides weak convergence of $(x_k)_{k \in \mathbb{N}}$ toward x , which is a zero of $B + \sum_i A_i$ by Proposition 4.9.

(d) If, moreover, $\forall k \in \mathbb{N}, \lambda_k \leq 1$, in view of Propositions 4.12 and 4.13, (4.9) is immediately a particular instance of [20, Algorithm 4.1]. In particular, [20, Theorem 3.1 and Remark 3.4] provides

$$(4.13) \quad \sum_{k \in \mathbb{N}} \left\| (\mathbf{Id} - \mathbf{T}_{2,\gamma}) \mathbf{z}_k - (\mathbf{Id} - \mathbf{T}_{2,\gamma}) \mathbf{z} \right\|^2 = \sum_{k \in \mathbb{N}} \gamma^2 \left\| \mathbf{B} \mathbf{P}_{\mathcal{S}} \mathbf{z}_k - \mathbf{B} \mathbf{P}_{\mathcal{S}} \mathbf{z} \right\|^2 < +\infty .$$

(d)(1). Now, if B is uniformly monotone, then we have, for all $k \in \mathbb{N}$,

$$\begin{aligned} \langle \mathbf{B} \mathbf{P}_{\mathcal{S}} \mathbf{z}_k - \mathbf{B} \mathbf{P}_{\mathcal{S}} \mathbf{z} | \mathbf{z}_k - \mathbf{z} \rangle &= \sum_i w_i \langle B(\sum_i w_i \mathbf{z}_{i,k}) - B(\sum_i w_i \mathbf{z}_i) | \mathbf{z}_{i,k} - \mathbf{z}_i \rangle \\ &= \langle B(\sum_i w_i \mathbf{z}_{i,k}) - B(\sum_i w_i \mathbf{z}_i) | \sum_i w_i (\mathbf{z}_{i,k} - \mathbf{z}_i) \rangle \\ &\geq \varphi(\|\mathbf{z}_k - \mathbf{z}\|) . \end{aligned}$$

From (b) and (4.13), we deduce that the right-hand side of the last inequality converges to 0. In view of the properties of φ , we obtain strong convergence of $(x_k)_{k \in \mathbb{N}}$ toward x .

(d)(2). Let $\mathbf{u} = -\gamma \mathbf{B} \mathbf{P}_{\mathcal{S}} \mathbf{z}$ and $\forall k \in \mathbb{N}$,

$$\mathbf{y}_k = \mathbf{T}_{1,\gamma} \mathbf{T}_{2,\gamma} \mathbf{z}_k \quad \text{and} \quad \mathbf{u}_k = (\mathbf{z}_k - \mathbf{y}_k) - \gamma \mathbf{B} \mathbf{P}_{\mathcal{S}} \mathbf{z}_k .$$

We then have

$$\|u_k - u\| \leq \|y_k - z_k\| + \gamma \|BP_S z_k - BP_S z\| .$$

It then follows from (a) and (4.13) that u_k converges strongly to u . On the other hand, by Proposition 4.15, we have

$$u \in A'_\gamma z \quad \text{and} \quad u_k \in A'_\gamma y_k .$$

Therefore, applying Proposition 4.16 to the pairs (z, u) and (y_k, u_k) , and using the fact that $\times_{i=1}^n w_i^{-1} A_i$ is uniformly monotone, we obtain

$$\langle (u^S - z^\perp) - (u_k^S - y_k^\perp) \mid (z^S - u^\perp) - (y_k^S - u_k^\perp) \rangle \geq \varphi \left(\left\| (z^S - u^\perp) - (y_k^S - u_k^\perp) \right\| \right)$$

for some nondecreasing function $\varphi : [0, +\infty[\rightarrow [0, +\infty]$ that vanishes only at 0. Now,

$$\begin{aligned} & \langle (u^S - z^\perp) - (u_k^S - y_k^\perp) \mid (z^S - u^\perp) - (y_k^S - u_k^\perp) \rangle \\ &= \langle (u^S - u_k^S) - (z^\perp - y_k^\perp) \mid (z^S - y_k^S) - (u^\perp - u_k^\perp) \rangle \\ &= \langle u^S - u_k^S \mid z^S - y_k^S \rangle + \langle z^\perp - y_k^\perp \mid u^\perp - u_k^\perp \rangle = \langle u - u_k \mid z - y_k \rangle . \end{aligned}$$

Moreover,

$$\begin{aligned} \varphi \left(\left\| (z^S - u^\perp) - (y_k^S - u_k^\perp) \right\| \right) &= \varphi \left(\left\| (z^S - y_k^S) - (u^\perp - u_k^\perp) \right\| \right) \\ &= \varphi \left(\sqrt{\|z^S - y_k^S\|^2 + \|u^\perp - u_k^\perp\|^2} \right) \\ &\geq \varphi \left(\|z^S - y_k^S\| \right) \end{aligned}$$

since $\varphi \circ \sqrt{\cdot} : [0, +\infty[\rightarrow [0, +\infty]$ is nondecreasing. Altogether, we arrive at

$$\varphi \left(\|z^S - y_k^S\| \right) \leq \langle u - u_k \mid z - y_k \rangle .$$

By (a) and (b), y_k converges weakly to z , and we have shown that u_k converges strongly to u . This proves that $\langle u - u_k \mid z - y_k \rangle \rightarrow 0$, and therefore y_k^S converges strongly to $z^S = x$ in view of the properties of φ . The latter in conjunction with (a) implies that $z_k^S = x_k$ converges strongly to x .

It remains to show the special cases implying uniform monotonicity in (d)(2). Indeed, if $\forall i \in \llbracket 1, n \rrbracket$, A_i is uniformly monotone with the same modulus φ which is also convex, then

for $(\mathbf{x}, \mathbf{u}) \in \text{gra } \gamma \bullet \mathbf{A}$ and $(\mathbf{y}, \mathbf{v}) \in \text{gra } \gamma \bullet \mathbf{A}$,

$$\begin{aligned} \langle \mathbf{u} - \mathbf{v} \mid \mathbf{x} - \mathbf{y} \rangle &= \sum_i w_i \langle u_i - v_i \mid x_i - y_i \rangle \geq \sum_i w_i \gamma_i \varphi(\|x_i - y_i\|) \\ &\quad (0 < \inf_i \gamma_i < \infty) \geq \inf_i \gamma_i \sum_i w_i \varphi(\|x_i - y_i\|) \\ &\quad (\varphi \text{ is convex}) \geq \inf_i \gamma_i \varphi\left(\sum_i w_i \|x_i - y_i\|\right) \\ &\quad (\varphi \text{ is nondecreasing}) \geq \inf_i \gamma_i \varphi\left(\inf_i w_i \sqrt{\sum_i \|x_i - y_i\|^2}\right) \\ &\quad (\varphi \text{ is nondecreasing and } w_i \in]0, 1]) \geq \inf_i \gamma_i \varphi\left(\inf_i w_i \sqrt{\sum_i w_i \|x_i - y_i\|^2}\right) \\ &= \inf_i \gamma_i \varphi\left(\inf_i w_i \|\mathbf{x} - \mathbf{y}\|\right). \end{aligned}$$

The proof for the case where φ is subadditive follows the same lines using subadditivity instead of convexity in the inequalities, and replacing $\inf_i \gamma_i$ by γ (since by definition $\gamma_i w_i = \gamma$) and $\inf_i w_i$ by 1. ■

Remark 4.1. For statements (a)–(c), assumption (ii) can be weakened. More precisely, (ii) can be replaced by $\sum_{k \in \mathbb{N}} \lambda_k (1 - \alpha \lambda_k) = +\infty$, where $\alpha = \max\left(\frac{2}{3}, \frac{2}{1+2\beta/\gamma}\right)$, and (iii) by $\sum_{t \in \mathbb{N}} \lambda_k (\|\varepsilon_{1,k}\| + \|\varepsilon_{2,k}\|) < +\infty$. The proof would follow the same lines as in [20, Lemma 5.1].

Remark 4.2 (strong convergence). We have proved strong convergence of the sequence $(x_k)_{k \in \mathbb{N}}$, but we did not elaborate on strong convergence of $(z_k)_{k \in \mathbb{N}}$. It turns out that the sequence $(z_k)_{k \in \mathbb{N}}$ is indeed quasi-Fejér monotone with respect to $\text{fix } \mathbf{T}$. Thus, if $\text{int}(\text{fix } \mathbf{T}) \neq \emptyset$, [20, Lemma 2.8(iv)] provides strong convergence of $(z_k)_{k \in \mathbb{N}}$, and therefore of $(x_k)_{k \in \mathbb{N}}$. An alternative sufficient condition is that \mathbf{A}'_γ is semiregular; see [4, Definition 2.3] and also [27, Condition 3.2] in the case of convex optimization. Semiregularity occurs, for instance, if the operator has a boundedly relatively compact domain (the intersection of its closure with any closed ball is compact); see [4, Proposition 2.4]. However, this condition is rather abstract, and it is not easy to translate it in terms of the properties of the individual A_i 's when $n > 1$.

Remark 4.3 (nonstationary GFB). Convergence of the nonstationary version of our inexact GFB splitting algorithm, i.e., for a varying sequence $(\gamma_k)_{k \in \mathbb{N}}$, can also be established. More precisely, it can be shown that the statements of Theorem 4.17 hold under the additional assumption that $0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < 2\beta$ and $(\gamma_k - \gamma)_{k \in \mathbb{N}}$ is absolutely summable where $\gamma \in [\underline{\gamma}, \bar{\gamma}]$. The key idea underlying the proof consists in viewing the nonstationary method as a perturbed version of the stationary method with an additional error term (besides those previously considered in the implicit and explicit steps), and ensuring that this error is also summable; see the initial work of [51, 52] in this direction. This absolute summability assumption on $(\gamma_k - \gamma)_{k \in \mathbb{N}}$ can be dropped for $n \equiv 1$, in which case we recover the forward-backward algorithm.

Finally, let us explicit the relationship between Theorem 4.17 and the claims of section 2.1.

Proof of Theorem 2.1. It is straightforward to see that the vector whose coordinates are the sequences $(z_{i,k})_{k \in \mathbb{N}}$ defined in (2.2) follows iterations (4.9), with $\varepsilon_{1,k} = (\varepsilon_{1,k,i})_i$ and $\varepsilon_{2,k} = \mathbf{C}(-\gamma\varepsilon_{2,k})$, which are, of course, summable under the required assumptions. Applying Theorem 4.17, the claimed convergence properties follow. ■

Proof of Corollary 2.2. Under (H1)–(H2), [6, Theorems 16.2 and 16.37(i)] provides that $\operatorname{argmin}(f + \sum_i g_i) = \operatorname{zer}(\nabla f + \sum_i \partial g_i) \neq \emptyset$. Furthermore, Lemma 4.8(i) provides that ∇f is β -cocoercive, and Lemma 4.8(iii) shows that $J_{\frac{\gamma}{w_i}A_i}$ corresponds to $\operatorname{prox}_{\frac{\gamma}{w_i}g_i}$. Hence, weak convergence of $(x_k)_{k \in \mathbb{N}}$ toward a minimizer of (1.2) follows from Theorem 4.17(c). The proof of strong convergence is a consequence of Theorem 4.17(d) together with the fact that uniform convexity of a function in $\Gamma_0(\mathcal{H})$ implies uniform monotonicity of its subdifferential [6]. ■

5. Conclusion. We have introduced in this paper a novel splitting method for finding a zero of a sum of an arbitrary number of maximal monotone operators. It takes advantage of either the cocoercivity or the possibility of computing the resolvent of each operator separately. We provided theoretical guarantees on the convergence of the algorithm and its robustness to summable errors. We emphasized the corresponding novel primal proximal splitting method for minimizing convex functionals that are the sum of a smooth term and several simple functions. Our method generalizes some existing schemes and enlarges the class of problems that can be solved efficiently with proximal splitting methods. Numerical experiments on convex optimization for inverse problems show evidence of the advantages of our approach for large-scale imaging problems.

REFERENCES

- [1] F. ACKER AND M.-A. PRESTEL, *Convergence d'un schéma de minimisation alternée*, Ann. Fac. Sci. Toulouse Math. (5), 2 (1980), pp. 1–9.
- [2] M. V. AFONSO, J. M. BIOCAS-DIAS, AND M. A. T. FIGUEIREDO, *Fast image recovery using variable splitting and constrained optimization*, IEEE Trans. Image Process., 19 (2010), pp. 2345–2356.
- [3] K. J. ARROW, L. HURWICZ, AND H. UZAWA, *Studies in Linear and Non-linear Programming*, Stanford University Press, Stanford, CA, 1958.
- [4] H. ATTOUCH, L. M. BRICEÑO-ARIAS, AND P. L. COMBETTES, *A parallel splitting method for coupled monotone inclusions*, SIAM J. Control Optim., 48 (2010), pp. 3246–3270.
- [5] J.-B. BAILLON AND G. HADDAD, *Quelques propriétés des opérateurs angle-bornés et n -cycliquement monotones*, Israel J. Math., 26 (1977), pp. 137–150.
- [6] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer-Verlag, New York, 2011.
- [7] H. H. BAUSCHKE, P. L. COMBETTES, AND S. REICH, *The asymptotic behavior of the composition of two resolvents*, Nonlinear Anal., 60 (2005), pp. 283–301.
- [8] A. BECK AND M. TEOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
- [9] K. BREDIES AND D. A. LORENZ, *Linear convergence of iterative soft-thresholding*, J. Fourier Anal. Appl., 14 (2008), pp. 813–837.
- [10] L. M. BRICEÑO-ARIAS AND P. L. COMBETTES, *Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery*, Numer. Math. Theory Methods Appl., 2 (2009), pp. 485–508.
- [11] L. M. BRICEÑO-ARIAS AND P. L. COMBETTES, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250.
- [12] L. M. BRICEÑO-ARIAS, P. L. COMBETTES, J.-C. PESQUET, AND N. PUSTELNIK, *Proximal algorithms for multicomponent image recovery problems*, J. Math. Imaging Vision, 41 (2011), pp. 3–22.

- [13] F. E. BROWDER, *Convergence theorems for sequences of nonlinear operators in Banach spaces*, Math. Z., 100 (1967), pp. 201–225.
- [14] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
- [15] C. CHAUX, A. BENAZZA-BENYAHIA, AND J.-C. PESQUET, *A block-thresholding method for multispectral image denoising*, in Proceedings of the SPIE, Vol. 5914, San Diego, CA, 2005, pp. 1H-1–1H-13.
- [16] C. CHAUX, P. L. COMBETTES, J.-C. PESQUET, AND V. R. WAJS, *A variational formulation for frame-based inverse problems*, Inverse Problems, 23 (2007), pp. 1495–1518.
- [17] C. CHAUX, J.-C. PESQUET, AND N. PUSTELNIK, *Nested iterative algorithms for convex constrained image recovery problems*, SIAM J. Imaging Sci., 2 (2009), pp. 730–762.
- [18] G. CHEN AND M. TEBoulLE, *A proximal-based decomposition method for convex minimization problems*, Math. Programming, 64 (1994), pp. 81–101.
- [19] G. H.-G. CHEN AND R. T. ROCKAFELLAR, *Convergence rates in forward-backward splitting*, SIAM J. Optim., 7 (1997), pp. 421–444.
- [20] P. L. COMBETTES, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization, 53 (2004), pp. 475–504.
- [21] P. L. COMBETTES, *Iterative construction of the resolvent of a sum of maximal monotone operators*, J. Convex Anal., 16 (2009), pp. 727–748.
- [22] P. L. COMBETTES, D. DŨNG, AND B. C. VŨ, *Dualization of signal recovery problems*, Set-Valued Var. Anal., 18 (2010), pp. 373–404.
- [23] P. L. COMBETTES AND J.-C. PESQUET, *A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery*, IEEE J. Sel. Topics Signal Process., 1 (2007), pp. 564–574.
- [24] P. L. COMBETTES AND J.-C. PESQUET, *A proximal decomposition method for solving convex variational inverse problems*, Inverse Problems, 24 (2008), 065014.
- [25] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer-Verlag, New York, 2011, pp. 185–212.
- [26] P. L. COMBETTES AND J.-C. PESQUET, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum monotone operators*, Set-Valued Var. Anal., 20 (2012), pp. 307–330.
- [27] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
- [28] L. CONDAT, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., to appear.
- [29] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1541.
- [30] D. L. DONOHO, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.
- [31] J. DOUGLAS AND H. H. RACHFORD, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.
- [32] F.-X. DUPÉ, J. M. FADILI, AND J.-L. STARCK, *A proximal iteration for deconvolving Poisson noisy images using sparse representations*, IEEE Trans. Image Process., 18 (2009), pp. 310–321.
- [33] F.-X. DUPÉ, J. M. FADILI, AND J.-L. STARCK, *Linear inverse problems with various noise models and mixed regularizations*, in Proceedings of the 1st International ICST Workshop on New Computational Methods for Inverse Problems, Paris, 2011.
- [34] F.-X. DUPÉ, M. J. FADILI, AND J.-L. STARCK, *Inverse problems with Poisson noise: Primal and primal-dual splitting*, in Proceedings of the 18th IEEE International Conference on Image Processing (ICIP), Brussels, Belgium, 2011.
- [35] F.-X. DUPÉ, M. J. FADILI, AND J.-L. STARCK, *Deconvolution under Poisson noise using exact data fidelity and synthesis or analysis sparsity priors*, Stat. Methodol., 9 (2012), pp. 4–18.
- [36] J. ECKSTEIN, *Parallel alternating direction multiplier decomposition of convex programs*, J. Optim. Theory Appl., 80 (1994), pp. 39–62.
- [37] J. ECKSTEIN AND D. P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming, 55 (1992), pp. 293–318.
- [38] J. ECKSTEIN AND B. F. SVAITER, *General projective splitting methods for sums of maximal monotone operators*, SIAM J. Control Optim., 48 (2009), pp. 787–811.

- [39] E. ELHAMIFAR AND R. VIDAL, *Robust classification using structured sparse representation*, in Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1873–1879.
- [40] M. J. FADILI AND G. PEYRÉ, *Total variation projection with first order schemes*, IEEE Trans. Image Process., 20 (2010), pp. 657–669.
- [41] M. J. FADILI, J.-L. STARCK, AND F. MURTAGH, *Inpainting and zooming using sparse representations*, Comput. J., 52 (2009), pp. 64–79.
- [42] M. A. T. FIGUEIREDO AND J. M. BIOCAS-DIAS, *Restoration of Poissonian images using alternating direction optimization*, IEEE Trans. Image Process., 19 (2010), pp. 3133–3145.
- [43] M. A. T. FIGUEIREDO AND R. NOWAK, *An EM algorithm for wavelet-based image restoration*, IEEE Trans. Image Process., 12 (2003), pp. 906–916.
- [44] M. FORTIN AND R. GLOWINSKI, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, Elsevier Science Publishers, Amsterdam, 1983.
- [45] D. GABAY, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983, pp. 299–331.
- [46] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2 (1976), pp. 17–40.
- [47] R. GLOWINSKI AND P. LE TALLEC, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM, Philadelphia, 1989.
- [48] R. JENATTON, J. MAIRAL, G. OBOZINSKI, AND F. BACH, *Proximal Methods for Hierarchical Sparse Coding*, preprint, arXiv:1009.2139 [stat.ML], 2010.
- [49] G. M. KORPELEVICH, *An extragradient method for finding saddle points and for other problems*, Èkonom. i Mat. Metody, 12 (1976), pp. 747–756.
- [50] M. KOWALSKI AND B. TORRÉSANI, *Sparsity and persistence: Mixed norms provide simple signal models with dependent coefficients*, Signal, Image and Video Processing, 3 (2009), pp. 251–264.
- [51] B. LEMAIRE, *Stability of the iteration method for nonexpansive mappings*, Serdica Math. J., 22 (1996), pp. 331–340.
- [52] B. LEMAIRE, *Which fixed point does the iteration method select?*, in Recent Advances in Optimization: The 8th French-German Conference on Optimization, P. Gritzmann, R. Horst, E. Sachs, and R. Tichatschke, eds., Lecture Notes in Econom. and Math. Systems 452, Springer-Verlag, Berlin, 1997, pp. 154–167.
- [53] J. LIEUTAUD, *Approximation d'Opérateurs par des Méthodes de Décomposition*, Ph.D. thesis, Université de Paris, Paris, 1969.
- [54] P.-L. LIONS, *Une méthode itérative de résolution d'une inéquation variationnelle*, Israel J. Math., 31 (1978), pp. 204–208.
- [55] P.-L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.
- [56] S. MALLAT, *A Wavelet Tour of Signal Processing. The Sparse Way*, 3rd ed., Elsevier/Academic Press, Amsterdam, 2009.
- [57] B. MERCIER, *Lectures on Topics in Finite Element Solution of Elliptic Problems*, Tata Inst. Fund. Res. Lectures on Math. and Phys. 63, Tata Institute of Fundamental Research, Bombay, India, 1979.
- [58] O. J. MINTY, *Montone (nonlinear) operators in Hilbert space*, Duke Math. J., 29 (1962), pp. 341–346.
- [59] R. D. C. MONTEIRO AND B. F. SVAITER, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, SIAM J. Optim., 23 (2013), pp. 475–507.
- [60] J.-J. MOREAU, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
- [61] YU. NESTEROV, *Gradient methods for minimizing composite functions*, Math. Program., 140 (2013), pp. 125–161.
- [62] G. B. PASSTY, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Anal. Appl., 72 (1979), pp. 383–390.
- [63] J.-C. PESQUET AND N. PUSTELNIK, *A parallel inertial proximal optimization method*, Pac. J. Optim., 8 (2012), pp. 273–306.
- [64] G. PEYRÉ AND J. FADILI, *Group sparsity with overlapping partition functions*, in Proceedings of the 19th European Signal Processing Conference (EUSIPCO 2011), 2011, pp. 303–307.

- [65] R. R. PHELPS, *Convex Functions, Monotone Operators and Differentiability*, Lecture Notes in Math. 1364, 2nd ed., Springer-Verlag, Berlin, 1993.
- [66] L. D. POPOV, *A modification of the Arrow-Hurwitz method of search for saddle points*, Mat. Zametki, 28 (1980), pp. 777–784.
- [67] N. PUSTELNIK, C. CHAUX, AND J.-C. PESQUET, *Parallel proximal algorithm for image restoration using hybrid regularization*, IEEE Trans. Image Process., 20 (2011), pp. 2450–2462.
- [68] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [69] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [70] M. V. SOLODOV, *A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework*, Optim. Methods Softw., 19 (2004), pp. 557–575.
- [71] J. E. SPINGARN, *Partial inverse of a monotone operator*, Appl. Math. Optim., 10 (1983), pp. 247–265.
- [72] J.-L. STARCK, F. MURTAGH, AND M. J. FADILI, *Sparse Signal and Image Processing: Wavelets, Curvelets and Morphological Diversity*, Cambridge University Press, Cambridge, UK, 2010.
- [73] P. TSENG, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138.
- [74] P. TSENG, *Alternating projection-proximal methods for convex programming and variational inequalities*, SIAM J. Optim., 7 (1997), pp. 951–965.
- [75] P. TSENG, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446.
- [76] B. C. VŮ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681.
- [77] M. YUAN AND Y. LIN, *Model selection and estimation in regression with grouped variables*, J. R. Stat. Soc. Ser. B Stat. Methodol., 68 (2006), pp. 49–67.
- [78] E. H. ZARANTONELLO, *Projections on convex sets in Hilbert space and spectral theory. I. Projections on convex sets*, in Contributions to Nonlinear Functional Analysis, Academic Press, New York, 1971, pp. 237–341.