

Solving inverse problems with overcomplete transforms and convex optimization techniques

Chaâri L.^a, Pustelnik N.^a, Chaux C.^a, and Pesquet J.-C.^a

^a Université Paris-Est, Laboratoire d'Informatique Gaspard Monge and UMR-CNRS 8049, Champs-sur-Marne, 77454 Marne-la-Vallée, France

ABSTRACT

Many algorithms have been proposed during the last decade in order to deal with inverse problems. Of particular interest are convex optimization approaches that consist of minimizing a criteria generally composed of two terms: a data fidelity (linked to noise) term and a prior (regularization) term. As image properties are often easier to extract in a transform domain, frame representations may be fruitful. Potential functions are then chosen as priors to fit as well as possible empirical coefficient distributions. As a consequence, the minimization problem can be considered from two viewpoints : a minimization along the coefficients or along the image pixels directly. Some recently proposed iterative optimization algorithms can be easily implemented when the frame representation reduces to an orthonormal basis. Furthermore, it can be noticed that in this particular case, it is equivalent to minimize the criterion in the transform domain or in the image domain. However, much attention should be paid when an overcomplete representation is considered. In that case, there is no longer equivalence between coefficient and image domain minimization. This point will be developed throughout this paper. Moreover, we will discuss how the choice of the transform may influence parameters and operators necessary to implement algorithms.

Keywords: inverse problems, wavelets, redundant transforms, convex optimization, proximal operator, restoration

1. INTRODUCTION

Because of sensor imperfection and acquisition mode, observed data are often noisy and degraded by a linear operator. This operator and noise properties may actually depend on the considered application. For instance, in optical satellite imaging, the linear operator models a blur and the noise can be assumed Gaussian. In the biomedical area, the ill-conditioned linear operator in parallel Magnetic Resonance Imaging (pMRI) represents a sensitivity matrix and the noise is Gaussian, whereas in Positron Emission Tomography (PET), the acquisition process is modelled by a large-size projection matrix and the noise is Poisson distributed.

To solve such inverse problems, some methods, like the Wiener filter, directly process the image. However, since the 90's, wavelet transforms¹ have been often used in restoration methods.²⁻⁴ Nevertheless, the classical (critically decimated) discrete wavelet decomposition suffers from some drawbacks like a lack of shift invariance and poor orientation properties. To circumvent these difficulties, various overcomplete frame representations have been introduced. The undecimated wavelet transform has been proposed having the desired shift-invariance properties but at a price of a high redundancy. More recently, geometrical transforms have been proposed such as curvelets,⁵ dual-trees⁶ or grouplets,⁷ which are often able to achieve the shift-invariance property and to perform directional analyses.

Among all the existing methods which have been developed to solve inverse problems, we are particularly interested in convex optimization techniques which have gained much popularity during the last past years. More precisely, these approaches consist of minimizing a criterion, often constituted of two terms, that can be interpreted from a Bayesian perspective : a term (named the fidelity term) is linked to the noise nature and the second one (named regularization term) is linked to some a priori one can have concerning the target image. A large choice of a priori functions can be envisaged. However, it can be convenient to define these

E-mail: {lotfi.chaari,nelly.pustelnik,caroline.chaux,jean-christophe.pesquet}@univ-paris-est.fr

functions in the wavelet transform domain and then, to choose them so as to model the wavelet transform coefficient distribution. The a priori term can be chosen equal to an ℓ_1 norm (promoting sparsity) for example or a generalized Gaussian prior can be employed.

At this point, two strategies can be envisaged: the criterion can be minimized with respect to the image (Analysis Approach) or to the wavelet coefficients (Synthesis Approach). Some works comparing the two approaches have been recently published. Elad et al.⁸ considered a general inverse problem involving Gaussian noise and showed that the two approaches are non longer equivalent when overcomplete transforms are used. More recently, a study was conducted in⁹ when the criterion to minimize is differentiable. The proposed restoration method is based on Nesterov¹⁰ algorithm.

The two approaches are not disconnected. Depending on the considered problem, it can be easier to express the minimization in the coefficient space or in the pixel space. Duality properties relate the two minimization problems (see¹¹ where a denoising example involving Gaussian noise is investigated). A large panel of approaches are available regarding the Synthesis Approach. Of particular interest are the proximal algorithms like the forward-backward algorithm¹² or the Parallel PROximal Algorithm (PPRA).¹³ While the first one requires to compute a gradient and a proximity operator,¹⁴ the second one requires to compute the proximity operators of J (possibly greater than 2) convex functions so to minimize their sum.

In such proximal algorithms, some parameters play a prominent role in the restoration process. For example, the step-size of the forward-backward algorithm must be carefully chosen in order to guarantee a good convergence behaviour. This may require the computation of the upper frame constant of the overcomplete transform, which is not always known. Furthermore, for the considered algorithms the computation of some proximity operators are needed. However, when frames are involved, the computation of the proximity operator of the composition of the frame operator and some convex function has to be performed, which is not always obvious. The aim of the paper is to describe general convex optimization techniques to solve inverse problem whatever the analysis or synthesis approach is chosen.

The organization of the paper is as follows: Section 2 presents the context of the study. The differences between the analysis and synthesis approaches are emphasized. In Section 3, algorithmic issues are discussed, which are illustrated in Section 4. A conclusion is given in Section 5

2. PROBLEM FORMULATION

Generally speaking, an inverse problem consists of recovering some unknown signal/image $\bar{y} \in \mathbb{R}^L$ from its degraded observation $z \in \mathbb{R}^M$ through a linear operator $H : \mathbb{R}^L \rightarrow \mathbb{R}^M$. When an additive noise $n \in \mathbb{R}^M$ is considered, a general formulation of the observed model is the following:

$$z = H\bar{y} + n. \quad (1)$$

An example of real world inverse problems with zero-mean Gaussian additive noise is image reconstruction in parallel Magnetic Resonance Imaging (pMRI). However, other non-additive noise models may be encountered in applications like Poisson noise in Positron Emission Tomography (PET) or confocal microscopy, Gamma distributed noise in Synthesis Aperture Radar (SAR),... Note that for denoising problems, the operator H reduces to the identity one.

The inverse problem in (1) can be solved using conventional estimators by minimizing some distance d between the solution and the observation. In this case, the estimation procedure relies on the optimization of the following optimality criterion:

$$\min_y d(Hy, z). \quad (2)$$

When the considered distance is quadratic, which means that the observation noise is Gaussian, the Least Squares (LS) estimator is often used in the literature.

When $M \geq L$, the problem is overdetermined and well-posed, unless H is ill-conditioned. However, when $M < L$, the problem is underdetermined with an infinity of solutions that may match the observation model, and hence it is ill-posed. When the problem is ill-posed, one usually resorts to regularization¹⁵ techniques to stabilize the solution. In this context, a regularization function r is introduced bringing some prior

information about the target solution. This regularization function may be designed in order to emphasize some features of the signal under investigation like sparsity or some local smoothness properties. Moreover, when designing the regularization function, it has been found in many application fields that using *frame* representations let one have better results. The next section will give a brief overview of the frame concept.

2.1 The frame concept

Let K be an integer greater or equal to L . A family of vectors $(e_k)_{1 \leq k \leq K}$ in the finite-dimensional space \mathbb{R}^L is a frame when there exists two constants $\underline{\mu}$ and $\bar{\mu}$ in $]0, +\infty[$ such that

$$(\forall y \in \mathbb{R}^L), \quad \underline{\mu} \|y\|^2 \leq \sum_{k=1}^K |\langle y | e_k \rangle|^2 \leq \bar{\mu} \|y\|^2. \quad (3)$$

If $\underline{\mu} = \bar{\mu}$, $(e_k)_{1 \leq k \leq K}$ is called a *tight* frame. The bounded linear frame analysis operator F and its adjoint synthesis frame operator F^* are defined as

$$F: \mathbb{R}^L \rightarrow \mathbb{R}^K \quad (4)$$

$$y \mapsto (\langle y | e_k \rangle)_{1 \leq k \leq K},$$

$$F^*: \mathbb{R}^K \rightarrow \mathbb{R}^L \quad (5)$$

$$(\xi_k)_{1 \leq k \leq K} \mapsto \sum_{k=1}^K \xi_k e_k.$$

When $F^{-1} = F^*$, $(e_k)_{k \in \mathbb{K}}$ is an orthonormal basis. A simple example of a redundant frame is the union of $\mu \in \mathbb{N}^*$ orthonormal bases. In this case, the frame is tight with $\mu = \underline{\mu} = \bar{\mu}$ and thus, we have $F^* \circ F = \mu \text{Id}$ where Id is the identity operator.

2.2 Regularization

Under a Bayesian framework, the design of the regularization function is performed through the choice of a prior distribution that models the signal under investigation in a given space. Recovering the unknown signal is then performed based on its posterior distribution.

To go through this Bayesian formulation in more details, we will first assume the following:

- the observed signal z is a realization of a random variable Z ,
- the signal to recover \bar{y} is a realization of a random variable Y .

Using a frame representation, and combining the likelihood and the appropriate prior distribution, the posterior probability distribution may be easily derived. To do so, two competing approaches are possible to deal with: Analysis Approach (AA) and Synthesis Approach (SA).

2.2.1 The Analysis Approach (AA)

If one adopts AA, the prior distribution will be designed based on a direct transformation of the signal to recover. In this case, the posterior probability distribution of the unknown signal is given by:

$$p(y|z) \propto p(z|y)p(y). \quad (6)$$

Assuming appropriate exponential shape for the likelihood and the prior probability distribution, the latter may be expressed as:

$$p(z|y) \propto \exp(-d(Hy, z)) \quad (7)$$

and

$$p(y) \propto \exp(-r(Fy)). \quad (8)$$

The prior defines a proper probability density if, for example,

$$\forall x \in \mathbb{R}^L, \quad r(x) \geq \rho \|x\|^p \quad (9)$$

where $\rho > 0$ and $p > 0$. We have indeed

$$\int_{\mathbb{R}^L} e^{-r(Fy)} dy \leq \int_{\mathbb{R}^L} e^{-\rho \|Fy\|^p} dy \leq \int_{\mathbb{R}^L} e^{-\rho \mu^{p/2} \|y\|^p} dy < \infty. \quad (10)$$

The posterior probability in (6) can then be reexpressed as:

$$p(y|z) \propto \exp(-d(Hy, z) - r(Fy)). \quad (11)$$

2.2.2 The Synthesis Approach (SA)

When dealing with SA, the prior distribution will be designed based on the representation of the signal in a given dictionary such that $\bar{y} = F^* \bar{x}$. In this case, an estimate \hat{x} of \bar{x} will be first found before getting $\hat{y} = F^* \hat{x}$. The posterior probability of the unknown signal is given by:

$$p(x|z) \propto p(z|F^*x)p(x). \quad (12)$$

Assuming the same form for the likelihood and the following shape of the prior

$$p(x) \propto \exp(-r(x)) \quad (13)$$

the posterior probability distribution in this case is given by:

$$p(x|z) \propto \exp(-d(HF^*x, z) - r(x)). \quad (14)$$

2.3 Analysis/Synthesis estimation procedure

Based on the posterior probability distribution in each case, the Posterior Mean (PM) or the Maximum A Posteriori (MAP) estimators may be used to recover an estimate \hat{y} of y . The PM can be used only when easy integration of the posterior distribution can be performed. However, an approximation of it may be obtained by Monte Carlo simulation methods. This approach will not be detailed in this paper, but more details can be found in¹⁶ and references therein. Our attention will be focused on the MAP estimator since it is somehow more familiar than the PM one.

If one wants to deal with the MAP estimator by maximizing the posterior distribution established with AA and SA, the inherent optimization problems will be the following:

- AA:

$$\hat{y} = \arg \min_y d(Hy, z) + r(Fy), \quad (15)$$

- SA:

$$\hat{y} = F^* \arg \min_x [d(HF^*x, z) + r(x)]. \quad (16)$$

We can notice that in both cases, the penalized criterion is based on two terms: data fidelity $d(\cdot, \cdot)$ and penalty $r(\cdot)$. The regularization function r generally involves a regularization parameter which balances the solution between the two terms. Philosophically, AA and SA are quite different. However, when the used frame is an orthonormal basis, the two approaches are strictly equivalent. For more general frames, the equivalence is not always ensured unless for special cases of frames or regularization function. In fact, Elad et al.⁸ proved the equivalence between AA and SA when a redundant frame is used for the case of Gaussian noise and quadratic regularization functions. However, for other noise classes or for more useful penalty terms like an ℓ_1 norm (Laplace prior), the two problems become quite different and may give different results. Moreover, it has been reported in Elad et al. work⁸ that the AA may be more robust to estimation errors in each solution component apart. Good results in image deconvolution, for instance, have been obtained using AA in Weiss et al.¹⁷ and SA in Chaux et al.¹². For both approaches, efficient algorithms have been proposed like the Nesterov algorithm^{10,17}, the forward-backward algorithm^{12,18}, the Douglas-Rachford algorithm,¹⁹ the PPRA¹³ or Dykstra-like algorithms.²⁰

3. OPTIMIZATION ISSUES

In this section we adopt the following notations:

- \mathcal{H} and \mathcal{G} denote separable Hilbert spaces with scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$ (for example $\mathcal{H} = \mathbb{R}^L$ or $\mathcal{H} = \mathbb{R}^K$).
- $\Gamma_0(\mathcal{H})$ (resp. $\Gamma_0(\mathcal{G})$) is the class of lower semicontinuous convex functions from \mathcal{H} (resp. \mathcal{G}) to $] - \infty, +\infty]$ which are not identically equal to $+\infty$.

By considering convex optimization algorithms^{12,13,19} to solve the minimization problems (15) and (16), some hurdles can be encountered. For example, the convergence rate of the forward-backward algorithm¹² depends on the upper frame constant $\bar{\mu}$. For other approaches such as the Douglas-Rachford algorithm¹⁹ or the PPRA¹³, the difficulty may stem from the computation of the proximity operator associated to $\varphi \circ T$ where $\varphi \in \Gamma_0(\mathcal{H})$ and $T : \mathcal{G} \rightarrow \mathcal{H}$ is a bounded linear operator.

In what follows, we will first recall the definition of the proximity operator, as well as some convex optimization schemes such as the forward-backward and PPRA. Computational solution of some problems when dealing with such algorithms will also be detailed.

3.1 Some convex optimization tools and algorithms

3.1.1 Proximity operator definition

For a function $\phi \in \Gamma_0(\mathcal{H})$, the proximity operator^{14,21} is such that:

$$(\forall \xi \in \mathcal{H}), \quad \text{prox}_\phi(\xi) = \arg \min_{\zeta \in \mathcal{H}} \phi(\zeta) + \frac{1}{2} \|\zeta - \xi\|^2. \quad (17)$$

In the recent literature^{12,22}, many useful properties of the proximity operator were given. This tool allows us to deal with a wide class of SA and AA minimization problems involving possibly non-differentiable criteria. Two of these algorithms are detailed in the next subsections.

3.1.2 Forward-backward algorithm

Let us define the functions $f_A \in \Gamma_0(\mathbb{R}^L)$ and $f_S \in \Gamma_0(\mathbb{R}^K)$ from d (the data fidelity term) as follows:

$$(\forall y \in \mathbb{R}^L), \quad f_A(y) = d(Hy, z), \quad (18)$$

and

$$(\forall x \in \mathbb{R}^K), \quad f_S(y) = d(HF^*x, z). \quad (19)$$

In the following, the functions $d(\cdot, z)$ and the regularization term r are assumed to be in $\Gamma_0(\mathbb{R}^M)$ and $\Gamma_0(\mathbb{R}^K)$ respectively. When one of these functions is differentiable with a Lipschitz gradient, the forward-backward algorithm can be used. The forward-backward iterations are given by Algorithm 1 when $d(\cdot, z)$ is assumed to be differentiable with α -Lipschitz gradient.

The weak convergence^{12,22} of $(y_n)_{n \in \mathbb{N}}$ (resp. $(x_n)_{n \in \mathbb{N}}$) to a solution to Problem (15) (resp. Problem (16)) is secured when $\inf_{n \in \mathbb{N}} \lambda_n > 0$, $\sum_{n=0}^{+\infty} \|a_n\| < +\infty$ and $\sum_{n=0}^{+\infty} \|b_n\| < +\infty$. The sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ correspond to some error tolerances in the evaluation of the gradient and the proximity operator, respectively.

Algorithm 1 The forward-backward algorithm

AA	SA
1: Select $y_0 \in \mathbb{R}^L$	1: Select $x_0 \in \mathbb{R}^K$
2: for $n \in \mathbb{N}$ do	2: for $n \in \mathbb{N}$ do
3: Set $\gamma_n \in]0, 2/(\alpha \ H\ ^2)[$ and $\lambda_n \in]0, 1[$	3: Set $\gamma_n \in]0, 2/(\alpha \ HF^*\ ^2)[$ and $\lambda_n \in]0, 1[$
4: $y_{n+\frac{1}{2}} = y_n - \gamma_n (\nabla f_A(y_n) + b_n)$	4: $x_{n+\frac{1}{2}} = x_n - \gamma_n (\nabla f_S(x_n) + b_n)$
5: $p_n = \text{prox}_{\gamma_n r \circ F}(y_{n+\frac{1}{2}})$	5: $p_n = \text{prox}_{\gamma_n r}(x_{n+\frac{1}{2}})$
6: $y_{n+1} = y_n + \lambda_n (p_n + a_n - y_n)$	6: $x_{n+1} = x_n + \lambda_n (p_n + a_n - x_n)$
7: end for	7: end for
8: After convergence, take $\hat{y} = y_{n+1}$.	8: After convergence, take $\hat{y} = F^* x_{n+1}$.

3.1.3 Parallel PROximal Algorithm (PPRA)

The PPRA detailed in Algorithm 2 aims at minimizing a sum of $J \in \mathbb{N}$ functions $(f_j)_{1 \leq j \leq J} \in \Gamma_0(\mathcal{H})^J$ non-necessarily differentiable, for which the proximity operator can be computed. The problem is thus the following:

$$\min_{\xi \in \mathcal{H}} \sum_{j=1}^J f_j(\xi). \quad (20)$$

Algorithm 2 Parallel PROximal Algorithm (PPRA)

- 1: Let $\gamma \in]0, +\infty[$
 - 2: $\forall j \in \{1, \dots, J\}$, set $(\omega_j)_{1 \leq j \leq J} \in]0, 1]^J$ such that $\sum_{j=1}^J \omega_j = 1$.
 - 3: $\forall j \in \{1, \dots, J\}$, set $(u_{j,0})_{1 \leq j \leq J} \in \mathcal{H}^J$ and $\xi_0 = \sum_{j=1}^J \omega_j u_{j,0}$.
 - 4: **for** $n \in \mathbb{N}$ **do**
 - 5: **for** $j = 1, \dots, J$ **do**
 - 6: $p_{j,n} = \text{prox}_{\gamma/\omega_j f_j} u_{j,n} + a_{j,n}$
 - 7: **end for**
 - 8: $p_n = \sum_{j=1}^J \omega_j p_{j,n}$
 - 9: $\lambda_n \in]0, 2[$
 - 10: **for** $j = 1, \dots, J$ **do**
 - 11: $u_{j,n+1} = u_{j,n} + \lambda_n (2 p_n - \xi_n - p_{j,n})$
 - 12: **end for**
 - 13: $\xi_{n+1} = \xi_n + \lambda_n (p_n - \xi_n)$
 - 14: **end for**
-

The sequence $(\xi_n)_{n \in \mathbb{N}}$ defined by Algorithm 2 converges weakly to a solution to Problem (20) under the following assumption.

Assumption 3.1

- (i) $\lim_{\|\xi\| \rightarrow +\infty} f_1(\xi) + \dots + f_J(\xi) = +\infty$.
- (ii) \mathcal{H} is finite-dimensional and $\cap_{j=1}^J \text{rint dom } f_j \neq \emptyset$.*
- (iii) $\sum_{n \in \mathbb{N}} \lambda_n (2 - \lambda_n) = +\infty$.
- (iv) $(\forall j \in \{1, \dots, J\}) \quad \sum_{n \in \mathbb{N}} \lambda_n \|a_{j,n}\| < +\infty$.

*The relative interior of a set S of \mathcal{H} is designated by $\text{rint } S$ and the domain of a function $f : \mathcal{H} \rightarrow]-\infty, +\infty]$ is $\text{dom } f = \{\xi \in \mathcal{H} | f(\xi) < +\infty\}$.

The sequence $(a_{j,n})_{n \in \mathbb{N}}$ in \mathcal{H} introduced in Algorithm 2 corresponds to possible errors (numerical errors for instance) in the computation of the proximity operators, which shows that convergence is ensured in spite of these errors.

The AA and SA criteria can be minimized by considering Algorithm 2. A special case when $J = 2$ is to take $f_1 = f_S$ and $f_2 = r$ for SA and $f_1 = f_A$ and $f_2 = r \circ F$ for AA.

3.2 Proximity operator and frame representation

In this section, we will focus on the computation of the proximity operator associated with $\varphi \circ T$ where $\varphi \in \Gamma_0(\mathcal{G})$ and $T: \mathcal{H} \rightarrow \mathcal{G}$ is a bounded linear operator. To do so, three methods are described: explicit form computation, splitting approach and iterative approach.

3.2.1 Explicit form

For some specific operators T , $\text{prox}_{\varphi \circ T}$ can be easily calculated. A helpful property introduced by Combettes and Pesquet¹⁹ states that:

Proposition 3.2 *Let \mathcal{G} be a real Hilbert space, let $\varphi \in \Gamma_0(\mathcal{G})$, and let $T: \mathcal{H} \rightarrow \mathcal{G}$ be a bounded linear operator. Suppose that the composition of T and T^* satisfies $T \circ T^* = \chi \text{Id}$, for some $\chi \in]0, +\infty[$. Then $\varphi \circ T \in \Gamma_0(\mathcal{H})$ and*

$$\text{prox}_{\varphi \circ T} = \text{Id} + \frac{T^*}{\chi} \circ (\text{prox}_{\chi\varphi} - \text{Id}) \circ T. \quad (21)$$

When a denoising problem ($H = \text{Id}$) is considered with SA using the PPRA, one of the proximity operators to compute is $\text{prox}_{d(\cdot, z) \circ F^*}$. The latter can be calculated in the restrictive framework of a tight frame ($F^* \circ F = \bar{\mu} \text{Id}$) by invoking Proposition (3.2) with $T = F^*$ and $\chi = \bar{\mu}$.

3.2.2 Splitting approach

In the case when the function $\varphi \in \Gamma_0(\mathbb{R}^N)$ is separable, which means that

$$(\forall \xi = (\xi^{(n)})_{1 \leq n \leq N}), \quad \varphi(\xi) = \sum_{n=1}^N \varphi_n(\xi^{(n)}), \quad (22)$$

the splitting approach consists of decomposing the operator T in operators $(T_i)_{1 \leq i \leq I}$ such that $T_i \circ T_i^* = \chi_i \text{Id}$. We subsequently assume that $(\mathbb{I}_i)_{1 \leq i \leq I}$ is a partition of $\{1, \dots, N\}$ in nonempty sets. For every $i \in \{1, \dots, I\}$, let N_i be the number of elements in \mathbb{I}_i and let $\Upsilon_i: \mathbb{R}^{N_i} \rightarrow]0, +\infty[: (\xi^{(n)})_{n \in \mathbb{I}_i} \mapsto \sum_{n \in \mathbb{I}_i} \varphi_n(\xi^{(n)})$. Assume that $\varphi \circ T = \sum_{i=1}^I \Upsilon_i \circ T_i$ where T_i is the linear operator from \mathcal{G} to \mathbb{R}^{N_i} associated with the matrix

$$\begin{bmatrix} t_{n_1}^\top \\ \vdots \\ t_{n_{N_i}}^\top \end{bmatrix}$$

with $\mathbb{I}_i = \{n_1, \dots, n_{N_i}\}$. The following assumption plays a prominent role in the splitting approach:

Assumption 3.3 *For all $i \in \{1, \dots, I\}$, $(t_n)_{n \in \mathbb{I}_i}$ is a family of orthogonal having the same norm $\chi_i^{1/2}$ where $\chi_i > 0$.*

Consider AA when the frame representation is associated with a union of $I = \mu$ orthogonal bases. where the regularization function r is separable. The minimization problem (15) can be rewritten as

$$\hat{y} = \arg \min_y d(Hy, z) + \sum_{i=1}^{\mu} r_i(F_i y). \quad (23)$$

where, for every $i \in \{1, \dots, I\}$, $F_i^* \circ F_i = \text{Id}$ and $r_i \in \Gamma_0(\mathbb{R}^{K/I})$. By considering this form of Problem (15), we are able to compute the proximity operators associated with each $\text{prox}_{r_i \circ F_i}$ by (21). Algorithm 2 can thus be applied to find a solution.

3.2.3 Iterative approach

Consider now the duality principle²³ to compute $\text{prox}_{r \circ F}(y)$ ($\forall y \in \mathbb{R}^L$) for a frame which is non necessarily tight. By using the definition of the proximity operator given in (17), the primal problem consists of finding

$$\min_{p \in \mathbb{R}^L} \frac{1}{2} \|y - p\|^2 + r(Fp) \quad (24)$$

and the associated dual problem is then

$$\min_{v \in \mathbb{R}^K} \frac{1}{2} \|y - F^*v\|^2 + r^*(v) \quad (25)$$

where $r^* \in \Gamma_0(\mathbb{R}^K)$ represents the conjugate of r . This formulation is a particular case of the criterion proposed by Combettes et al.¹¹ and can be solved by using the Splitting Dual-Primal algorithm¹¹. In Algorithm 3, under the assumption that $\sum_{n \in \mathbb{N}} \|a_n\| \leq +\infty$ and $\sum_{n \in \mathbb{N}} \|b_n\| \leq +\infty$, the sequence $(v_n)_{n \in \mathbb{N}}$ converges to a solution v of the Dual problem (25) and $\text{prox}_{r \circ F}(y) = y - F^*v$.

Algorithm 3 Splitting Dual-Primal algorithm to compute $\text{prox}_{r \circ F}(y)$

- 1: Select $\epsilon \in]0, \min\{1, 1/\bar{\mu}\}[$
 - 2: Set $v_0 \in \mathbb{R}^K$
 - 3: **for** $n \in \mathbb{N}$ **do**
 - 4: $y_n = y - F^*v_n + b_n$
 - 5: $\gamma_n \in [\epsilon, 2/\bar{\mu} - \epsilon]$
 - 6: $\lambda_n \in [\epsilon, 1]$
 - 7: $v_{n+1} = v_n + \lambda_n(\gamma_n F y_n - \gamma_n \text{prox}_{\frac{1}{\gamma_n} r}(v_n + F y_n) + a_n)$
 - 8: **end for**
-

3.3 Frame constant

Another issue related to the forward-backward algorithm for SA, consists of computing the norm $\|HF^*\|^2$. The problem is clearly illustrated in Algorithm 1 (SA) where the step-size γ_n is inversely proportional to this norm. Consider the objective function (16) for which $f = d(\cdot, z)$ is α -Lipschitz differentiable. By definition of Lipschitz differentiability,

$$(\forall (y_1, y_2) \in (\mathbb{R}^M)^2) \quad \|\nabla f(y_1) - \nabla f(y_2)\| \leq \alpha \|y_1 - y_2\| \quad (26)$$

In Section 3.1.2 the function $f_S \in \Gamma_0(\mathbb{R}^K)$ has been introduced such that $f_S = f \circ H \circ F^*$. The gradient of f_S is defined as

$$(\forall x \in (\mathbb{R}^K)), \quad \nabla f_S(x) = FH^*(\nabla f(HF^*x)) \quad (27)$$

which yields to

$$(\forall (x_1, x_2) \in (\mathbb{R}^K)^2) \quad \|\nabla f_S(x_1) - \nabla f_S(x_2)\| \leq \alpha \|HF^*\|^2 \|x_1 - x_2\|. \quad (28)$$

Consequently f_S is $\alpha \|HF^*\|^2$ -Lipschitz differentiable which explains why γ_n was chosen inversely proportional to $\alpha \|HF^*\|^2$ in Algorithm 1 (SA). Most of the time, the norm $\|HF^*\|^2$ is not easy to compute and the convergence rate depends on the Lipschitz bound. To overcome this difficulty, we propose to use Algorithm 4 which converges to $\|HF^*\|^2$.

The rationale of this algorithm is as follows. Let $B = HF^*$ and perform the eigenvalue decomposition $U\Lambda U^*$ of the matrix associated with the positive semi-definite operator B^*B , where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_K\}$ and $U = [u_1, \dots, u_K] \in \mathbb{R}^{K \times K}$ is an orthogonal matrix. Moreover, $\|HF^*\|^2 = \lambda_{i_0}$ where $i_0 \in \arg \max_{1 \leq i \leq K} \lambda_i$. Thus, let $x_0 \in \mathbb{R}^K$ which does not belong to an eigenspace of B^*B , we can write,

$$\frac{\|B^n x_0\|^2}{\|B^{n-1} x_0\|^2} = \frac{\sum_{i=1}^K \lambda_i^n |\langle x_0, u_i \rangle|^2}{\sum_{i=1}^K \lambda_i^{n-1} |\langle x_0, u_i \rangle|^2}, \quad (29)$$

which yields to

$$\lim_{n \rightarrow +\infty} \frac{\|B^n x_0\|^2}{\|B^{n-1} x_0\|^2} = \lambda_{i_0} = \|HF^*\|^2. \quad (30)$$

Algorithm 4 Frame constant computation

Select randomly $x_0 \in \mathbb{R}^K$, set $\rho_0 = \epsilon + 1$, $n = 1$ and $\rho_n = 1$

while $\frac{|\rho_n - \rho_{n-1}|}{\rho_n} \geq \epsilon$ **do**

 Set $x_n = B^* B x_{n-1}$ where $B = HF^*$

 Set $\rho_n = \frac{\|x_n\|}{\|x_{n-1}\|}$

end while

After convergence take $\|HF^*\|^2 = \rho_n$

4. NUMERICAL ILLUSTRATIONS

In this section, a deconvolution problem is addressed with uniform blur H of size 5×5 and additive Gaussian noise of standard deviation $\sigma = 8$. A comparison will be made between the performance of AA and SA in image deblurring, using in each case the appropriate tools to optimize the related optimality criterion. Concerning the frame representation, two examples will be considered: regularization using GenLOT^{24, 25} transform and the union of two orthonormal bases. Since the blur operator H is ill-conditioned, the considered deblurring problem is ill-posed. A regularization should then be used to stabilize the solution. Although many choices for the regularization function can be made, a classical ℓ_1 regularization will be used as a penalty term in this paper. This means that a Laplace distribution will be retained as the prior to model the wavelet frame coefficients.

4.1 Comparison of AA/SA Regularization using GenLOT transform

In this section, the main goal is to compare the performance of AA and SA, when using the appropriate proximity algorithms to optimize the related optimality criterion.

Based on the noise model and prior distribution, the optimality criteria in (15) and (16) may be reexpressed as:

- AA:

$$\hat{y} = \arg \min_y \frac{1}{2\sigma^2} \|Hy - z\|^2 + \kappa \|Fy\|_1. \quad (31)$$

- SA:

$$\hat{y} = F^* \arg \min_x \left[\frac{1}{2\sigma^2} \|HF^*x - z\|^2 + \kappa \|x\|_1 \right]. \quad (32)$$

To minimize this criterion, the forward-backward algorithm (see Algorithm 1 for AA and SA) was retained using $f = \frac{1}{2\sigma^2} \|\cdot - z\|^2$ and $r = \|\cdot\|_1$. The main difficulty when dealing with AA is to compute the proximity operator of $r \circ F$. To perform this task, Algorithm 3 was used as detailed in Section 3.2.3.

However, when SA is adopted, the main difficulty is to evaluate the Lipschitz constant of the gradient of f_S .

In fact, since the gradient of f is $\frac{1}{\sigma^2}$ -Lipschitz, f_S is then $\frac{1}{\sigma^2} \|HF^*\|^2$ -Lipschitz differentiable. The difficulty is therefore to compute $\|HF^*\|^2$, which may be achieved using Algorithm 4 introduced in Section 3.3.

The comparison in this section is mainly based on the performance of AA and SA in terms of visual restoration quality and Signal to Noise Ratio defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\bar{y}\|}{\|\bar{y} - \hat{y}\|} \right),$$

where \bar{y} is the known reference image and \hat{y} is the estimated one. Tests are conducted on two standard 256×256 images: *cameraman* and *barbara*.

Figure 1 shows the reference cameraman image (a), the degraded one (b) and restored ones using AA (c) and SA (d). We notice that visually, the two restored images are quite similar. In terms of SNR, AA gives slightly better restored images with 0.2 dB of SNR improvement over SA.



Figure 1. Original image (a), degraded one (b) and restored images using AA (c) and SA (d).

Based on this observation, one might say that AA outperforms SA in restoration quality using the same frame representation and optimization algorithm. However, when looking at restoration results for the *barbara* image in Figure 2, we notice that, for this image, AA and SA lead to the same performance.

Moreover, complementary tests illustrated in Table 1 using other frame representations (here the translation invariant wavelet transform²⁶) showed that SA may outperform AA. Consequently, one can conclude that the performance of AA and SA may depend on the frame representation, as well as on the image itself.

Table 1. SNR evaluation for image deblurring using translation invariant wavelet transform.

	SNR (dB)		
	Degraded	AA	SA
cameraman	16.60	18.90	19.26
barbara	17.47	19.25	19.49

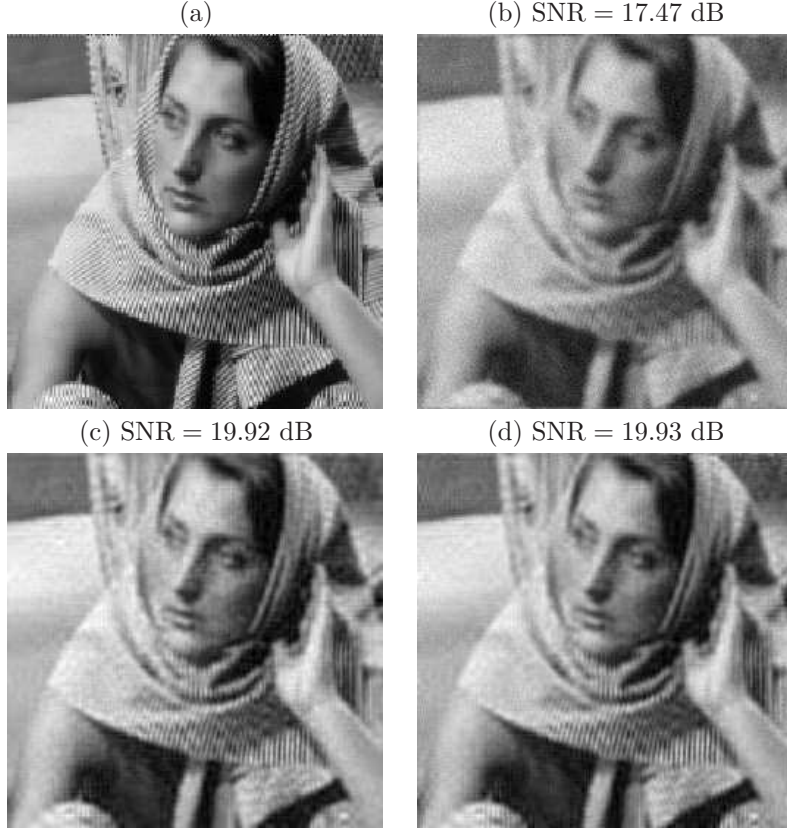


Figure 2. Original image (a), degraded one (b) and restored images using AA (c) and SA (d).

4.2 Regularization using AA

When a union of two orthonormal bases ($\mu = 2$) is considered (for example Symmlets of length 8 and Daubechies of length 6), the duality approach (Section 3.2.3) and the splitting approach (Section 3.2.2) can be used to deal with $r \circ F$ when considering AA. The goal of this section is to demonstrate the good convergence rate behaviour of the splitting approach for this particular frame representation. To do this, we consider Criterion (31) and detail the proposed PPRA-based algorithms for each approach and their associated convergence rates.

The first approach consists of an internal loop of Splitting Dual-Primal algorithm in the Parallel Proximal one which yields to Algorithm 5.

The second approach splits Criterion (31) as follows:

$$\hat{y} = \arg \min_y \frac{1}{2\sigma^2} \|Hy - z\|^2 + \kappa \|F_1 y\|_1 + \kappa \|F_2 y\|_1. \quad (33)$$

where F_1 (resp. F_2) is associated to Symmlet (resp. Daubechies) decomposition. The proximity operator of each function can be expressed explicitly and thus the Parallel Proximal algorithm can be used. Steps are detailed in Algorithm 6.

Algorithms 5 and 6 converge to a solution to (31). Figure 3 illustrates the convergence rate of both approaches in iteration number and CPU time. As it can be noticed, Algorithm 5 converges faster than 6 in iteration number; this is related to the number of proximity operators to compute in each step (2 for Algorithm 5 and 3 for Algorithm 6). However, in terms of CPU time, the internal loop strongly penalizes the convergence rate of Algorithm 5 as it can be observed in Figure 3 (right); this result emphasizes the interest in using a splitting approach for frames which can be decomposed in a union (with few terms) of

Algorithm 5 PPRA with an intern loop of Splitting Dual-Primal algorithm

- 1: Choose $\epsilon \in]0, +\infty[$ small enough
- 2: Let $\gamma \in]0, +\infty[$
- 3: Let $\omega_1 = 1/2$ and $\omega_2 = 1/2$
- 4: Set $u_{1,0} = u_{2,0} = z$ and $y_0 = \sum_{j=1}^2 \omega_j u_{j,0}$.
- 5: **for** $n \in \mathbb{N}$ **do**
- 6: Set $p_{1,n} = (\text{Id} + \frac{\gamma}{\omega_1 \sigma^2} H^* H)^{-1}(u_{1,n} + \frac{\gamma}{\omega_1 \sigma^2} H^* z)$
- 7: Set $v_0 = u_{2,n}$
- 8: Set $m = 0$
- 9: **while** $\|v_{m+1,n} - v_{m,n}\| > \epsilon \|v_{m,n}\|$ **do**
- 10: Set $\pi_{m,n} = u_{2,n} - F^* v_{m,n}$
- 11: Let $\tilde{\gamma}_{m,n} \in]0, 1[$
- 12: Let $\tilde{\lambda}_{m,n} \in]0, 1[$
- 13: Set $v_{m+1,n} = v_{m,n} + \tilde{\lambda}_{m,n} \tilde{\gamma}_{m,n} (F \pi_{m,n} - \text{prox}_{\frac{\gamma \kappa}{\omega_2 \tilde{\gamma}_{m,n}} \|\cdot\|_1}(\frac{v_{m,n}}{\tilde{\gamma}_{m,n}} + F \pi_{m,n}))$
- 14: **end while**
- 15: $p_{2,n} = \pi_{m,n}$
- 16: $p_n = \sum_{j=1}^2 \omega_j p_{j,n}$
- 17: $\lambda_n \in]0, 2[$
- 18: $u_{1,n+1} = u_{1,n} + \lambda_n (2 p_n - y_n - p_{1,n})$
- 19: $u_{2,n+1} = u_{2,n} + \lambda_n (2 p_n - y_n - p_{2,n})$
- 20: $y_{n+1} = y_n + \lambda_n (p_n - y_n)$
- 21: **end for**

Algorithm 6 PPRA for splitting criterion

- 1: Let $\gamma \in]0, +\infty[$
- 2: Set $\omega_1 = 1/2$ and $\omega_2 = \omega_3 = 1/4$
- 3: Set $u_{1,0} = u_{2,0} = u_{3,0} = z$ and $y_0 = \sum_{j=1}^3 \omega_j u_{j,0}$.
- 4: **for** $n \in \mathbb{N}$ **do**
- 5: $p_{1,n} = (\text{Id} + \gamma/(\omega_1 \sigma^2) H^* H)^{-1}(u_{1,n} + \gamma/(\omega_1 \sigma^2) H^* z)$
- 6: $p_{2,n} = F_1^* \text{prox}_{(\gamma \kappa)/\omega_2 \|\cdot\|_1} F_1 u_{2,n}$
- 7: $p_{3,n} = F_2^* \text{prox}_{(\gamma \kappa)/\omega_3 \|\cdot\|_1} F_2 u_{3,n}$
- 8: $p_n = \sum_{j=1}^3 \omega_j p_{j,n}$
- 9: $\lambda_n \in]0, 2[$
- 10: $u_{1,n+1} = u_{1,n} + \lambda_n (2 p_n - y_n - p_{1,n})$
- 11: $u_{2,n+1} = u_{2,n} + \lambda_n (2 p_n - y_n - p_{2,n})$
- 12: $u_{3,n+1} = u_{3,n} + \lambda_n (2 p_n - y_n - p_{3,n})$
- 13: $y_{n+1} = y_n + \lambda_n (p_n - y_n)$
- 14: **end for**

orthonormal bases. However, the splitting approach cannot be used for any kind of frames, whereas $\text{prox}_{r \circ F}$ can always be computed by using the Splitting Dual Primal algorithm.

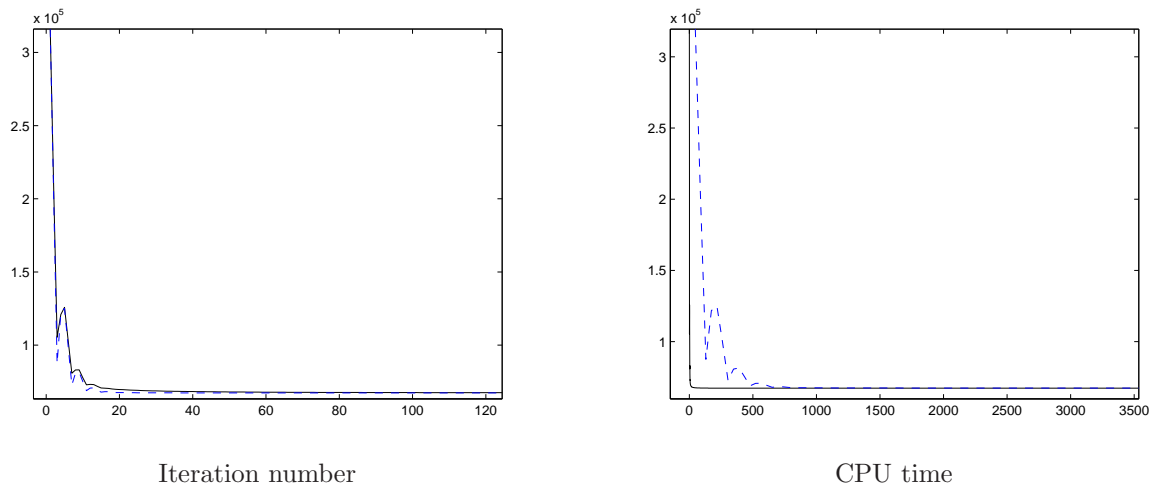


Figure 3. Convergence rate between Algorithm 5 (dashed line) and Algorithm 6 (continuous line) in the case of a union of orthonormal bases.

5. CONCLUSION

In this paper, we discussed several aspects of convex optimization tools and algorithms that may be used when dealing with inverse problems with frame representations. A particular attention has been paid to iterative proximal algorithms and related parameters which may need to be carefully chosen. Using these algorithms, two competing approaches (analysis versus synthesis) have been investigated. An application to an image deblurring problem showed that the performance of both methods depends on the used frame and even on the image itself. Consequently, it is quite difficult to conclude about the superiority of one approach over the other, in a general context.

REFERENCES

- [1] Mallat, S., [*A wavelet tour of signal processing*], Academic Press, San Diego, CA, USA, 2nd ed. (1998).
- [2] Guerrero-Colon, J., Mancera, L., and Portilla, J., “Image restoration using space-variant gaussian scale mixtures in overcomplete pyramids,” *IEEE Trans. on Image Proc.* **17**, 27–41 (Jan. 2008).
- [3] Pesquet, J.-C., Benazza-Benyahia, A., and Chaux, C., “A SURE approach for digital signal/image deconvolution problems,” *IEEE Trans. on Signal Proc.* (2009). To appear.
- [4] Vonesch, C. and Unser, M., “A fast multilevel algorithm for wavelet-regularized image restoration,” *IEEE Trans. on Image Proc.* **18**, 509–523 (March 2009).
- [5] Candès, E., Demanet, L., Donoho, D., and Ying, L., “Fast discrete curvelet transforms,” *Multiscale Modeling and Simulation* **5**, 861–899 (Mar. 2006).
- [6] Selesnick, I. W., Baraniuk, R. G., and Kingsbury, N. G., “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine* **22**, 123–151 (Nov. 2005).
- [7] Mallat, S., “Geometrical grouplets,” *Appl. and Comp. Harm. Analysis* **26**, 161–180 (Mar. 2009).
- [8] Elad, M., Milanfar, P., and Ron, R., “Analysis versus synthesis in signal priors,” *Inverse Problems* **23**, 947–968 (June 2007).
- [9] Carlván, M., Weiss, P., Blanc-Féraud, L., and Zerubia, J., “Algorithme rapide pour la restauration d’image rgularise sur les coefficients d’ondelettes,” *Proc. GRETSI* (Sept., 8-11 2009).
- [10] Nesterov, Y., “Gradient methods for minimizing composite objective function,” *CORE Discussion paper 76* (2007).

- [11] Combettes, P. L., Dung, D., and Cong Vu, B., “Dualization of signal recovery problems,” (2009). arXiv:0907.0436v2.
- [12] Chaux, C., Combettes, P. L., Pesquet, J.-C., and Wajs, V. R., “A variational formulation for frame-based inverse problems,” *Inverse Problems* **23**, 1495–1518 (June 2007).
- [13] Combettes, P. L. and Pesquet, J.-C., “A proximal decomposition method for solving convex variational inverse problems,” *Inverse Problems* **24** (Dec. 2008).
- [14] Moreau, J. J., “Proximité et dualité dans un espace hilbertien,” *Bull. Soc. Math. France* **93**, 273–299 (1965).
- [15] Tikhonov, A., “Tikhonov regularization of incorrectly posed problems,” *Soviet Mathematics Doklady* **4**, 1624–1627 (1963).
- [16] Robert, C. and Casella, G., [*Monte Carlo statistical methods*], Springer (New York, 2004).
- [17] Weiss, P., Blanc-Fraud, L., and Aubert, G., “Efficient schemes for total variation minimization under constraints in image processing,” *SIAM journal on Scientific Computing* **31**, 2047–2080 (2008).
- [18] Daubechies, I., Defrise, M., and Mol, C. D., “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics* **57**, 1413–1457 (Aug. 2004).
- [19] Combettes, P. L. and Pesquet, J.-C., “A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,” *IEEE Journal of Selected Topics in Signal Processing* **1**, 564–574 (December 2007).
- [20] Bauschke, H. H. and Combettes, P. L., “A Dykstra-like algorithm for two monotone operators,” *Pacific Journal of Optimization* (2008).
- [21] Moreau, J. J., “Fonctions convexes duales et points proximaux dans un espace hilbertien,” *C. R. Acad. Sci.* **255**, 2897–2899 (1962).
- [22] Combettes, P. L. and Wajs, V. R., “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling and Simulation* **4**, 1168–1200 (2005).
- [23] Zălinescu, C., [*Convex analysis in general vector spaces*], World Scientific (2002).
- [24] de Queiroz, R. L., Nguyen, T. Q., and Rao, K. R., “The genlot: generalized linear-phase lapped orthogonal transform,” *IEEE Trans. on Signal Proc.* **40**, 497–507 (1996).
- [25] Gauthier, J., Duval, L., and Pesquet, J.-C., “Optimization of synthesis oversampled complex filter banks,” *IEEE Trans. on Signal Proc.* (2009). To appear.
- [26] Coifman, R. and Donoho, D., “Translation-invariant de-noising,” in [*Wavelets and Statistics*], Antoniadis, A. and Oppenheim, G., eds., *Lecture Notes in Statistics* **103**, 125–150, Springer, New York, NY, USA (1995).