



# Wavelet-based Image Deconvolution and Reconstruction

Nelly Pustelnik, Amel Benazza-Benhayia, Yuling Zheng, Jean-Christophe Pesquet

## ► To cite this version:

Nelly Pustelnik, Amel Benazza-Benhayia, Yuling Zheng, Jean-Christophe Pesquet. Wavelet-based Image Deconvolution and Reconstruction. Wiley Encyclopedia of Electrical and Electronics Engineering, 2016. hal-01164833v4

**HAL Id: hal-01164833**

**<https://hal.science/hal-01164833v4>**

Submitted on 21 Oct 2016 (v4), last revised 24 Aug 2017 (v5)

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

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

# Wavelet-based Image Deconvolution and Reconstruction

Nelly Pustelnik, Amel Benazza-Benhayia, Yuling Zheng, and Jean-Christophe Pesquet<sup>‡</sup>

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Image formation models . . . . .	4
1.2	Ill-posedness, ill-conditioning, and regularization . . . . .	8
1.3	Methodology for solving inverse problems . . . . .	10
1.3.1	Bayesian interpretation . . . . .	10
1.3.2	Fidelity to data . . . . .	11
1.3.3	Regularization in the image domain . . . . .	12
<b>2</b>	<b>Wavelet transforms</b>	<b>14</b>
2.1	Motivation . . . . .	14
2.2	Filter banks . . . . .	14
2.3	Filter bank based multiresolution analysis . . . . .	16
2.4	2D extension . . . . .	17
<b>3</b>	<b>Frames</b>	<b>17</b>
3.1	Definition of a frame . . . . .	17

---

<sup>\*</sup>N. Pustelnik (Corresponding author) is with ENS Lyon, Laboratoire de Physique, UMR CNRS 5672, F69007 Lyon, France (e-mail: nelly.pustelnik@ens-lyon.fr). A. Benazza-Benhayia is with COSIM Lab., SUP'COM, Carthage Univ., Cité Technologique des Communications, Tunisia (e-mail: benazzaamel@gmail.com). Y. Zheng and J.-C. Pesquet are with Université Paris-Est, LIGM, UMR CNRS 8049, 77454 Marne-la-Vallée, France (e-mails: zhengyuling1988@gmail.com, jean-christophe.pesquet@univ-paris-est.fr).

<sup>‡</sup>This work was supported by the Agence Nationale de la Recherche under grant GRAPHSIP ANR-14-CE27-0001.

3.2	Related operators . . . . .	18
3.3	Examples of frames . . . . .	18
3.4	Frame decomposition theorem . . . . .	19
3.5	Sparsity criterion . . . . .	19
3.6	Decomposition algorithms . . . . .	20
3.7	Designing a frame . . . . .	20
3.8	Why frames in inverse problems? . . . . .	21
3.9	Regularization in the frame domain . . . . .	21
3.10	Image reconstruction in the frame domain . . . . .	23
<b>4</b>	<b>Variational approaches using frames</b>	<b>23</b>
4.1	Background . . . . .	23
4.2	Analysis versus synthesis formulation . . . . .	25
<b>5</b>	<b>Optimization methods</b>	<b>27</b>
5.1	Synthesis formulation in the presence of Gaussian noise . . . . .	27
5.1.1	Step-size parameter . . . . .	28
5.1.2	Uniqueness of the solution . . . . .	29
5.1.3	Proximity operator . . . . .	29
5.2	Analysis formulation in the presence of Gaussian noise . . . . .	31
5.2.1	Forward-backward algorithm . . . . .	31
5.2.2	Douglas-Rachford algorithm . . . . .	32
5.2.3	PPXA+ . . . . .	33
5.2.4	Primal-dual schemes . . . . .	33
5.3	Additional constraints . . . . .	34
5.4	Frame-based strategies in the presence of Poisson noise . . . . .	35
5.4.1	Variance stabilization transform . . . . .	36

5.4.2	Dealing with the exact Kullback-Leibler divergence . . . . .	36
<b>6</b>	<b>Bayesian methods</b>	<b>37</b>
6.1	Bayesian formulation in the presence of Gaussian noise . . . . .	38
6.2	Markov Chain Monte Carlo methods . . . . .	38
6.3	Variational Bayesian approximation . . . . .	39
<b>7</b>	<b>Stein-based approaches</b>	<b>44</b>
7.1	Denoising methods . . . . .	44
7.1.1	Minimum mean square error . . . . .	44
7.1.2	Stein's estimate of $\text{MSE}_{\text{tr}}$ and $\text{MSE}_{\text{im}}$ . . . . .	45
7.1.3	Reconstruction in an orthonormal transform domain . . . . .	46
7.1.4	Reconstruction in a redundant representation . . . . .	47
7.2	Stein-based approaches for deconvolution . . . . .	50
7.3	Stein-based approaches for reconstruction . . . . .	50
<b>8</b>	<b>Blind deconvolution</b>	<b>52</b>
<b>9</b>	<b>Conclusion</b>	<b>53</b>

## Abstract

Image deconvolution and reconstruction are inverse problems which are encountered in a wide array of applications. Due to the ill-posedness of such problems, their resolution generally relies on the incorporation of prior information through regularizations, which may be formulated in the original data space or through a suitable linear representation. In this article, we show the benefits which can be drawn from frame representations, such as wavelet transforms. We present an overview of recovery methods based on these representations: (i) variational formulations and non-smooth convex optimization strategies, (ii) Bayesian approaches, especially Monte Carlo Markov Chain methods and variational Bayesian approximation techniques, and (iii) Stein-based approaches. A brief introduction to blind deconvolution is also provided.

# 1 Introduction

Image deconvolution and reconstruction belong to the class of inverse problems. They consist of recovering, from observed data, a signal/image which is the most “similar” to the original one. This constitutes a difficult task since the observed data are often degraded by various physical processes (both linear and nonlinear) during their acquisition, storage, or transmission, and they are subject to uncertainties related to the presence of random noises and the fact that the image itself is unknown. Generally, it is possible to describe such problems by the following generic observation model:

$$z = \mathcal{D}_\alpha(A\bar{y}), \quad (1)$$

where

- $z = (z_j)_{1 \leq j \leq M} \in \mathbb{R}^M$  is the vector containing the observed values, often corresponding to an image of size  $M = M_1 \times M_2$ ,
- $\bar{y} = (\bar{y}_i)_{1 \leq i \leq N} \in \mathbb{R}^N$  is the vector consisting of the (unknown) values of the original image of size  $N = N_1 \times N_2$  arranged in a lexicographic order,
- $A = (A_{j,i})_{1 \leq j \leq M, 1 \leq i \leq N} \in \mathbb{R}^{M \times N}$  is the matrix associated to a linear degradation operator,
- $\mathcal{D}_\alpha: \mathbb{R}^M \mapsto \mathbb{R}^M$  models other degradations such as nonlinear ones or the effect of the noise, parameterized by  $\alpha$ .

Many practical situations can be described by (??). In what follows we briefly discuss some of the most frequently encountered ones.

## 1.1 Image formation models

For many image modalities such as optical remote sensing imaging and microscopy, the observation model reduces to the linear additive noise model:

$$z = A\bar{y} + b \quad (2)$$

where  $A$  is a blurring operator corresponding to a square matrix ( $M = N$ ) and  $b$  is a vector of realizations of a zero-mean noise with variance  $\alpha$ , which is often Gaussian distributed. A motion between the scene and the camera, the defocus of an optical imaging system, lense imperfections, and atmospheric turbulences lead to a blur in the acquired images. In such cases, we say that an image restoration problem has to be solved. When  $A$  reduces to an identity matrix, we deal with the special case of a denoising problem. Several models of blurring operators  $A$  can be found in the literature [?]. If the blur is spatially invariant, the product  $A\bar{y}$  can be expressed as a convolution of the original image  $\bar{y}$  with the Point Spread Function (PSF)  $a \in \mathbb{R}^{Q_1 \times Q_2}$  of the sensor [?]. More precisely, under zero-end conditions, the unknown image  $\bar{y}$  and the kernel are considered as zero outside their respective domains  $[0, N_1 - 1] \times [0, N_2 - 1]$  and  $[0, Q_1 - 1] \times [0, Q_2 - 1]$ . Then, extended

image  $\bar{y}_e$  and kernel  $a_e$  of size  $M_1 \times M_2$  are defined by zero padding of  $\bar{y}$  and  $a$ , respectively. By using a two-dimensional indexing of the image intensity values, this can be expressed as

$$\bar{y}_{e,i_1,i_2} = \begin{cases} \bar{y}_{i_1,i_2} & \text{if } 0 \leq i_1 \leq N_1 - 1 \text{ and } 0 \leq i_2 \leq N_2 - 1 \\ 0 & \text{if } N_1 \leq i_1 \leq M_1 - 1 \text{ and } N_2 \leq i_2 \leq M_2 - 1, \end{cases} \quad (3)$$

$$a_{e,i_1,i_2} = \begin{cases} a_{i_1,i_2} & \text{if } 0 \leq i_1 \leq Q_1 - 1 \text{ and } 0 \leq i_2 \leq Q_2 - 1 \\ 0 & \text{if } Q_1 \leq i_1 \leq M_1 - 1 \text{ and } Q_2 \leq i_2 \leq M_2 - 1, \end{cases} \quad (4)$$

where  $M_1 \geq N_1 + Q_1 - 1$  and  $M_2 \geq N_2 + Q_2 - 1$ . Similarly, an extended discrete noise term  $b_e$  is obtained. Finally, the degradation model can be expressed as follows:

$$z_{j_1,j_2} = \sum_{i_1=0}^{M_1-1} \sum_{i_2=0}^{M_2-1} a_{e,j_1-i_1,j_2-i_2} \bar{y}_{e,i_1,i_2} + b_{e,j_1,j_2}, \quad (5)$$

where  $j_1 \in \{0, \dots, M_1 - 1\}$  and  $j_2 \in \{0, \dots, M_2 - 1\}$ . Then, for each row of index  $j_1 \in \{0, \dots, M_1 - 1\}$  of  $a_e$ , a Toeplitz matrix  $\tilde{A}_{j_1}$  with  $M_2$  columns is constructed:

$$\tilde{A}_{j_1} = \begin{bmatrix} a_{e,j_1,0} & a_{e,j_1,M_2-1} & a_{e,j_1,M_2-2} & \dots & a_{e,j_1,1} \\ a_{e,j_1,1} & a_{e,j_1,0} & a_{e,j_1,M_2-1} & \dots & a_{e,j_1,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{e,j_1,M_2-1} & a_{e,j_1,M_2-2} & a_{e,j_1,M_2-3} & \dots & a_{e,j_1,0} \end{bmatrix} \in \mathbb{R}^{M_2 \times M_2}. \quad (6)$$

Each matrix  $\tilde{A}_{j_1}$  is called a circulant matrix since each row is obtained by a circular shift of the previous one and, the first row corresponds to a circular shift of the last one. By stacking the rows of  $(z_{j_1,j_2})_{0 \leq j_1 < M_1, 0 \leq j_2 < M_2}$ , (??) is finally obtained where  $A$  has been partitioned into  $M_1^2$  partitions of size  $M_2 \times M_2$ :

$$A = \begin{bmatrix} \tilde{A}_0 & \tilde{A}_{M_1-1} & \tilde{A}_{M_1-2} & \dots & \tilde{A}_1 \\ \tilde{A}_1 & \tilde{A}_0 & \tilde{A}_{M_1-1} & \dots & \tilde{A}_2 \\ \tilde{A}_2 & \tilde{A}_1 & \tilde{A}_0 & \dots & \tilde{A}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{A}_{M_1-1} & \tilde{A}_{M_1-2} & \tilde{A}_{M_1-3} & \dots & \tilde{A}_0 \end{bmatrix}. \quad (7)$$

As the structure of  $A$  with respect to its blocks  $(\tilde{A}_{j_1})_{0 \leq j_1 < M_1}$  is circulant,  $A$  is a block-circulant matrix [?].

An example of remote sensing image degraded by a uniform blur and corrupted by a white zero-mean Gaussian noise is displayed in Fig. ??.<sup>1</sup> Several values for the blurring kernel size and the variance have been considered. In these experiments, the noise level is evaluated with a signal-to-noise-ratio (SNR) measure defined as  $10 \log_{10}(\frac{\text{variance of } \bar{y}}{\text{variance of noise}})$ . The object edges are all the more smoothed as the kernel size is becoming larger, while sharper details are also harder to distinguish when the noise variance  $\alpha$  increases.

However, in some practical contexts such as in astronomy, the blur may be spatially variant. This property (known as anisoplanatism) is mainly due to the optical distortions of the mirrors which vary in the field [?]. Spatially variant blur can also be produced by external phenomena such as atmospheric turbulences in ground-based astronomy [?] or the use of adaptive optics [?].

<sup>1</sup><https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html>

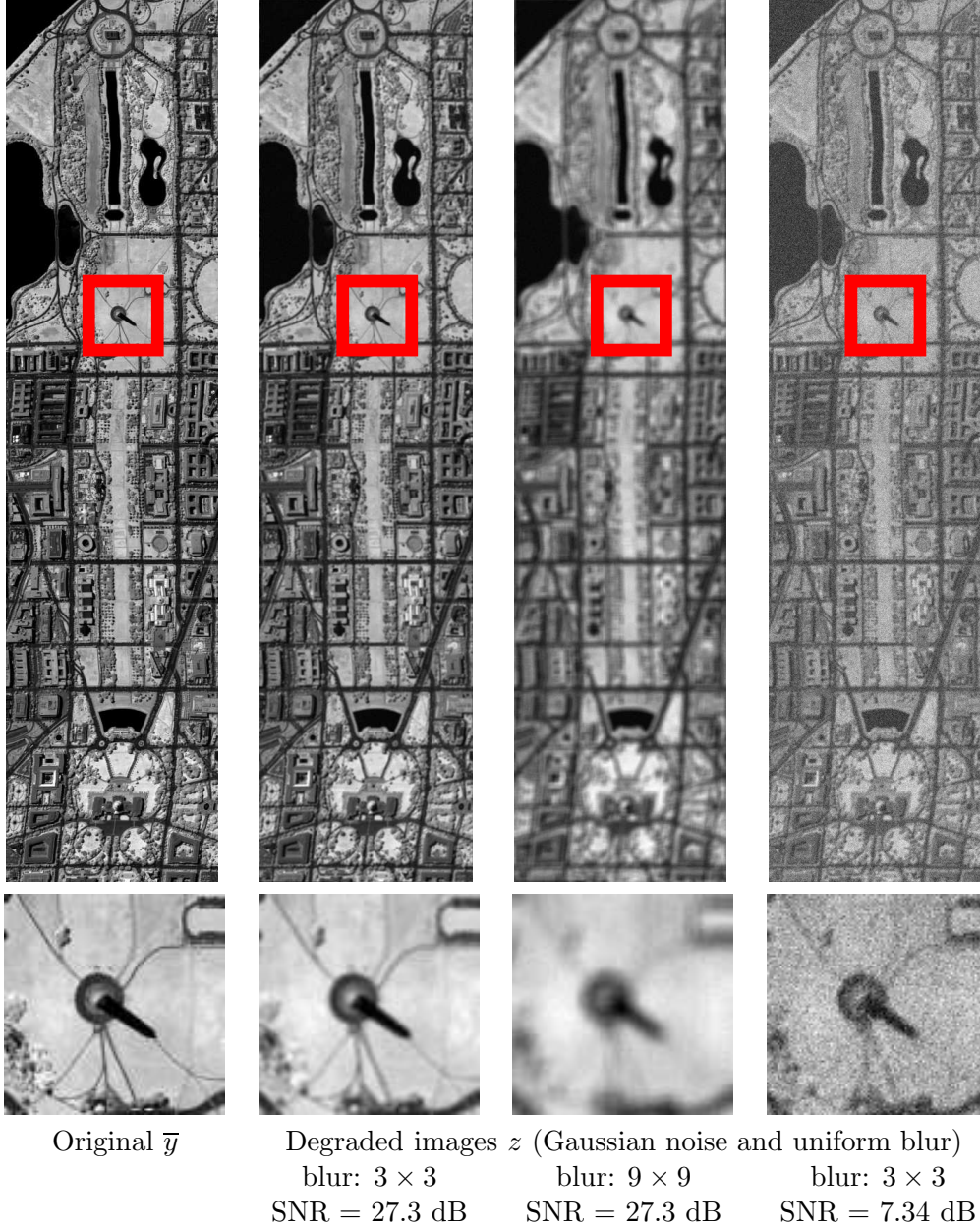


Figure 1: Degradation of a remote sensing image of size  $N = 1280 \times 307$ . The second row focuses on the red area.

In low-photon level imagery techniques, a more accurate model for the acquired data is given by a Poisson random process since the latter is a counting process suitable to express the distribution of the number of photons received by the sensor in a given time interval. In this case, each scalar observation  $z_j$  with  $j \in \{1, \dots, M\}$  can be viewed as a realization of a Poisson random variable of

intensity  $\lambda_j$  (expected value) given by

$$\lambda_j = \alpha \sum_{i=1}^N A_{j,i} \bar{y}_i, \quad (8)$$

where  $A$  and  $\bar{y}$  are assumed to have nonnegative components. The variance of observation  $z_j$  is then equal to  $\lambda_j$  and thus depends on the intensity of  $A\bar{y}$ . Confocal laser scanning microscopy, a successful technique for 3D imaging of biological specimens is a low-photon count imagery technique since the amount of light detected by the photomultiplier is significantly reduced by the confocal pinhole [?]. It also involves a linear out-of-focus blurring operator due to the diffraction limited nature of optical microscopy [?].

Note that a recent trend is to resort to more realistic noise models. In this respect, attention is paid to Poisson-Gaussian probabilistic models in several areas such as astronomy [?], microscopy [?, ?], and medical engineering [?]. This is motivated by the fact that the Poisson component reflects the photon-counting during the acquisition step whereas the Gaussian one corresponds to the thermal noise induced by the electronics of the imaging system (typically CCD sensors).

In image recovery applications, the linear operator  $A$  is not always a blur operator corresponding to the PSF of the instrument. In a wide range of tomography problems, the matrix  $A$  indeed corresponds to a projection matrix. In such a case, we say that an image reconstruction problem needs to be solved. A two-dimensional visual representation of the raw data is usually provided under the form of a sinogram: each column corresponds to the number of particles detected by a line (or a tube) of sensors in a given orientation. For instance, Positron Emission Tomography (PET) involves a projection operator  $A$  having fewer rows than columns ( $M < N$ ) in the presence of Poisson noise. Generally, PET data are acquired directly into sinograms. Fig. ?? provides two examples of sinograms for two values of the noise scale parameter  $\alpha$ . The smaller  $\alpha$ , the stronger the degradation is, as it can be observed in Figs ??(b) and ??(c).

Image reconstruction techniques can also be employed when some information is missing in the observed image, due to a low resolution acquisition (super-resolution problem), a subsampling (interpolation problem), or some occultation process (inpainting problem). In compressed sensing, the objective is to exactly recover the image (assumed to have some sparsity properties) from a very small number of data. Then, the matrix  $A$  often models a random measurement matrix with  $M \ll N$  and the noise is assumed to be additive and Gaussian [?]. Application of this technique to Parallel Magnetic Resonance Imaging (PMRI) is described in [?] and the references therein.

## 1.2 Ill-posedness, ill-conditioning, and regularization

The goal is to find an estimate  $\hat{y}(z) \in \mathbb{R}^N$  of the “clean” image  $\bar{y}$  from the measurements  $z$ . The solution is considered as supervised when  $A$  and the statistics of the noise are known, otherwise it is unsupervised or blind. In the latter case, the identification of the degradation characteristics is carried out prior to or in conjunction with the estimation procedure [?, ?]. In what follows, we mainly focus on supervised strategies.



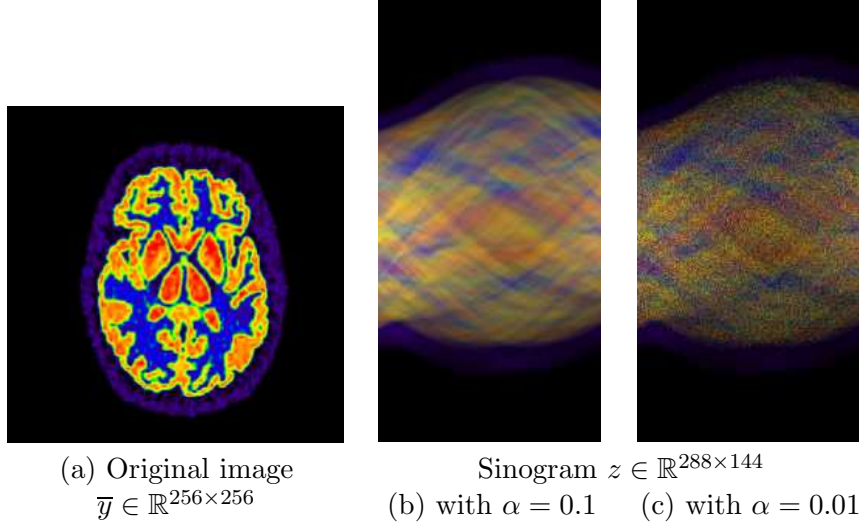


Figure 2: Example of degradation encountered in positron emission tomography.  $\bar{y}$  models the activity in the Zubal brain phantom [?]. Hot color is associated with high activity. Two sinograms are presented for different level of Poisson noise. Small scaling parameter  $\alpha$  means high degradation. The Poisson noise is signal dependent, thus areas with high intensity (warm colors) are more degraded than cool color areas (blue-black areas).

It must be emphasized that restoration techniques differ from image enhancement methods in that they account for a quantitative image formation model specifying the underlying degradations. In other words, image restoration/reconstruction constitute methodologies exploiting prior information on the degradation process.

Let us first assume that the image formation process is noise free. The problem  $z = A\bar{y}$  is said to be well-posed if it fulfills the Hadamard conditions [?] namely:

1. existence of a solution, i.e. the range  $\text{ran } A$  of  $A$  is equal to  $\mathbb{R}^M$ ,
2. uniqueness of the solution, i.e. the nullspace  $\ker A$  of  $A$  is equal to  $\{0\}$ ,
3. stability of the solution  $\hat{y}$  relatively to the observation i.e.

$$(\forall (z, z') \in (\mathbb{R}^M)^2) \quad \|z - z'\| \rightarrow 0 \quad \Rightarrow \quad \|\hat{y}(z) - \hat{y}(z')\| \rightarrow 0.$$

The existence condition means that each element of  $z \in \mathbb{R}^M$  is the image of an element  $y \in \mathbb{R}^N$ . The uniqueness condition means that, if  $\hat{y}(z)$  and  $\hat{y}'(z)$  are two solutions, then they are necessarily equal since  $\hat{y}(z) - \hat{y}'(z)$  belongs to  $\ker A$ . The stability condition allows us to ensure that a small perturbation of the observed image leads to a slight variation of the recovered image. In the following, we will relax the notation  $\hat{y}(z)$  as  $\hat{y}$ . Several situations may arise in practice.

- In the case of a square matrix with full rank  $r = M = N$ , a solution always exists as  $\text{ran } A = \mathbb{R}^N$ . Moreover, it is unique as  $A$  is injective. However,  $A$  may be ill-conditioned.

If  $(\eta_i)_{1 \leq i \leq r}$  denote the singular values of  $A$  sorted in a descending order, a high value of its condition number (the ratio  $\eta_1/\eta_N$ ) indicates a bad conditioning of  $A$ . A common example consists of a convolution operator  $A$  whose frequency response tends to vanish in some frequency band (typically a low-pass filter).

To illustrate this concept, we consider the image formation described by (??). In the case of an invertible matrix  $A$ , a naive solution consists of applying the inverse degradation model to the observation  $z$  that is

$$\hat{y} = A^{-1}(A\bar{y} + b) = \bar{y} + A^{-1}b. \quad (9)$$

In the case of the convolution degradation operator (??),  $A$  is block-circulant and, it may be diagonalized by the 2D Discrete Fourier Transform (DFT) matrix, which drastically reduces the computation cost of the inversion when the dimensions of  $A$  are power of 2 thanks to the fast Fourier transform algorithm [?, ?].

However, if  $A$  is ill-conditioned (whatever it is block-circulant or not), the inverse filtered noise  $A^{-1}b$  may become very large so that its effect becomes of paramount importance. Thus, the inverse filtering amplifies the noise leading to an irregular image.

Fig. ?? displays three degraded versions of the remote sensing image presented in Fig. ??. Three configurations are displayed. Fig. ??-(a) illustrates the image degraded by a blur only, Fig. ??-(b) presents the situation where the original image is degraded by a blur and a small noise while in Fig. ??-(c) the noise is stronger. It clearly appears that inverse filtering is not robust even when the noise is light but it works perfectly when no noise degrades the image.

- If  $A$  is not invertible, it is possible to compute its generalized (or pseudo) inverse. More precisely, if the recovery is good enough, the degraded version  $A\hat{y}$  of the solution  $\hat{y}$  can be expected to be close to the observed vector  $z$ . In other words, the Euclidean distance between  $A\hat{y}$  and  $z$  should be minimized:

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} \|Ay - z\|^2. \quad (10)$$

Thus, the problem reduces to a least squares problem [?] and, two configurations have to be studied. If  $A$  has a rank  $r = M < N$ , the problem is under-determined. Then,  $A$  is a “wide” matrix and the second Hadamard condition is not fulfilled. There exists an affine space of solution in the direction of  $\ker A$  and, very often, the solution with minimum norm is retained:

$$\hat{y} = A^\top (AA^\top)^{-1}z. \quad (11)$$

The generalized inverse of  $A$  is thus equal to  $A^\top (AA^\top)^{-1}$ .

In case of a rank  $r = N < M$ ,  $A$  is a “tall” matrix and the first Hadamard condition is not fulfilled. So, there is no exact solution if  $z \notin \text{Im } A$ . The least squares problem however has a unique solution given by

$$\hat{y} = (A^\top A)^{-1}A^\top z. \quad (12)$$

The generalized inverse of  $A$  is then equal to  $(A^\top A)^{-1}A^\top$ .

Note that in both cases, the generalized inverse filter often leads to irregular solutions as the inverse filter does.

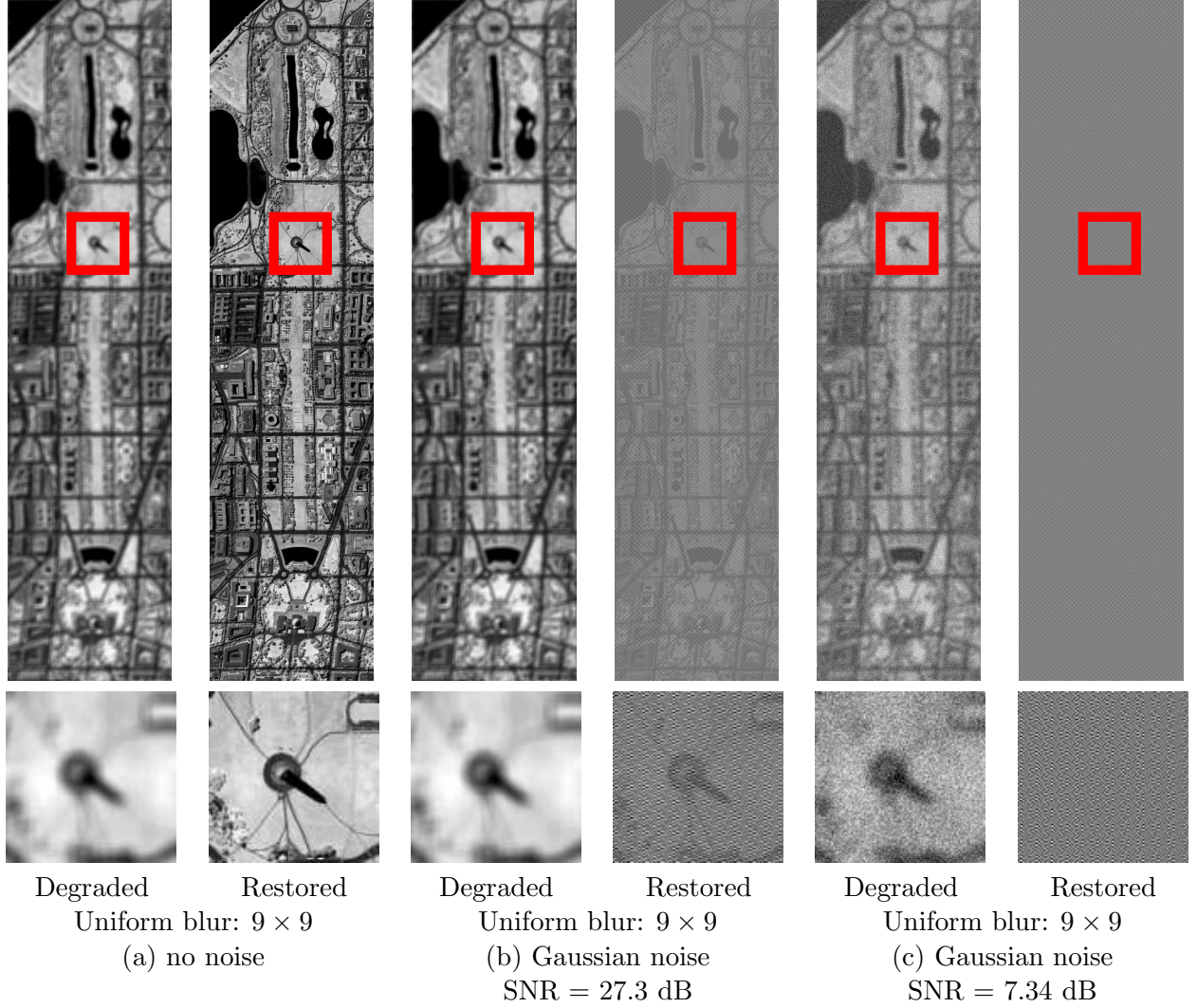


Figure 3: Three examples of image restoration with inverse filtering for a similar blur and different level of noise. The inverse filtering is not robust to noise, even if the noise level is small.

In conclusion, whatever the dimensions and the rank of  $A$  are, the need to stabilize the solution or to guarantee its uniqueness motivates the following alternative problem formulation:

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} \|Ay - z\|^2 + \chi\phi(y) \quad (13)$$

where  $\phi: \mathbb{R}^N \mapsto ]-\infty, +\infty]$  denotes the regularization term [?] and,  $\chi > 0$  is the regularization parameter that adjusts the tradeoff between the data fidelity and the degree of regularity.

### 1.3 Methodology for solving inverse problems

This subsection aims at describing some commonly used regularized criteria for solving inverse problems. Their definition requires prior knowledge about (i) the noise and the linear degradation operator, and (ii) the reference image  $\bar{y}$  (piecewise constant image, texture image,  $\dots$ ).

#### 1.3.1 Bayesian interpretation

To better cope with the criterion to be minimized for solving the generic inverse problem corresponding to (??), it seems suitable to resort to a Bayesian formulation. To this end,  $\bar{y}$  and  $z$  are considered as realizations of random vectors  $\bar{Y}$  et  $Z$ . The estimation can be performed based on the Maximum A Posteriori (MAP) strategy. The objective is to find the estimate  $\hat{y}$  which maximizes the posterior probability distribution  $\mu_{\bar{Y}|Z=z}$ . Thanks to Bayes theorem,  $\mu_{\bar{Y}|Z=z}(y)$  can be expressed in terms of the likelihood function  $\mu_{Z|\bar{Y}=y}(z)$ , the prior distribution  $\mu_{\bar{Y}}(y)$ , and the marginal distribution  $\mu_Z(z)$ , here assumed to be nonzero. Therefore, the problem reduces to find  $\hat{y}$  such that

$$\begin{aligned} \hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmax}} \mu_{\bar{Y}|Z=z}(y) &\Leftrightarrow \hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmax}} \mu_{Z|\bar{Y}=y}(z) \frac{\mu_{\bar{Y}}(y)}{\mu_Z(z)} \\ &\Leftrightarrow \hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmax}} \mu_{Z|\bar{Y}=y}(z) \mu_{\bar{Y}}(y). \end{aligned} \quad (14)$$

The monotonicity of the logarithm function allows us to rewrite (??) as follows

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} \left( -\log \mu_{Z|\bar{Y}=y}(z) - \log \mu_{\bar{Y}}(y) \right). \quad (15)$$

The first term in (??) quantifies the fidelity to the observed data  $z$  and, depends on the underlying degradation model. The second one is related to the prior information on the original image  $\bar{y}$ . Thus, if  $\psi/(2\tilde{\alpha})$  with  $\psi: \mathbb{R}^N \mapsto ]-\infty, +\infty]$  and  $\tilde{\alpha} > 0$ , and  $\tilde{\chi}\phi$  with  $\phi: \mathbb{R}^N \mapsto ]-\infty, +\infty]$  and  $\tilde{\chi} > 0$  denote potentials associated with the likelihood and the prior distribution, respectively (up to some possible additive constants), the optimization task amounts to find

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} \frac{1}{2\tilde{\alpha}} \psi(y) + \tilde{\chi} \phi(y). \quad (16)$$

#### 1.3.2 Fidelity to data

When  $\mathcal{D}_\alpha$  models a white Gaussian noise with variance  $\alpha = \sigma^2$ , the likelihood  $\mu_{Z|\bar{Y}=y}(z)$  is the following Gaussian function:

$$\mu_{Z|\bar{Y}=y}(z) = \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} e^{-\frac{\|Ay-z\|^2}{2\sigma^2}}. \quad (17)$$

Hence, the data-dependent term corresponding to the neg-logarithm of  $\mu_{Z|\bar{Y}=y}(z)$  is quadratic:

$$-\log \mu_{Z|\bar{Y}=y}(z) \propto \frac{1}{2\sigma^2} \|Ay - z\|^2 = \frac{1}{2\sigma^2} \psi(y). \quad (18)$$

Combining (??) and (??) leads to the minimization problem (??) where  $\chi = 2\sigma^2\tilde{\chi}$ .

In a similar way, when  $\mathcal{D}_\alpha$  describes a Poisson noise model with scale parameter  $\alpha$ , the likelihood is given by

$$\mu_{Z|\bar{Y}=y}((z_j)_{1 \leq j \leq M}) = \prod_{j=1}^M \frac{\exp(-\alpha(Ay)_j)}{z_j!} (\alpha(Ay)_j)^{z_j}. \quad (19)$$

where  $(z_j)_{1 \leq j \leq M}$  are nonnegative integer values. Therefore, the data-dependent term is related to the discrete generalized Kullback-Leibler divergence [?] denoted by  $D_{\text{GKL}}$ :

$$-\log \mu_{Z|\bar{Y}=y}((z_j)_{1 \leq j \leq M}) \propto \sum_{j=1}^M -z_j \log(\alpha(Ay)_j) + \alpha(Ay)_j \propto D_{\text{GKL}}(z, \alpha Ay). \quad (20)$$

At this point, it is important to note that it is much more difficult to account for Poissonian statistics than Gaussian ones since the data-dependent term is not finite for every value of  $y$ , but only when, for every  $j \in \{1, \dots, M\}$ ,  $(Ay)_j > 0$ . In addition, this term does not have a bounded second-order derivative on its domain, which may limit the choice of methods to minimize Criterion (??) in the Poissonian case. These difficulties will be discussed in Section ??.

Note also that if no penalization is introduced, the estimation reduces to a mere maximum likelihood estimation:

$$\hat{y} \in \text{Argmax}_{y \in \mathbb{R}^N} \log \mu_{Z|Y=y}((z_j)_{1 \leq j \leq M}). \quad (21)$$

In the case of Poisson statistics,  $\hat{y}$  cancels the derivative of the Poisson data fidelity term given by (??). Therefore, its  $N$  components  $(\hat{y}_j)_{1 \leq j \leq N}$  should satisfy the following condition:

$$(\forall j \in \{1, \dots, N\}) \quad \frac{1}{\alpha \sum_{i=1}^M A_{i,j}} \sum_{i=1}^M A_{i,j} \frac{z_i}{\sum_{j'=1}^N A_{i,j'} \hat{y}_{j'}} = 1. \quad (22)$$

This nonlinear equation cannot be solved in a straightforward manner. In this respect, the iterative Expectation-Maximisation (EM) is used [?, Annexe I]. It is very often called the Lucy-Richardson algorithm in the fields of astronomy and microscopy [?, ?] and Expectation-Maximisation Maximum Likelihood (EM-ML) in PET applications [?, ?, ?]. A nonlinear iterative method is then adopted to determine a sequence of approximations  $(\hat{y}^{[n]})_{n \in \mathbb{N}}$  of the target image:

$$(\forall j \in \{1, \dots, N\}) \quad \hat{y}_j^{[n+1]} = \frac{\hat{y}_j^{[n]}}{\alpha \sum_{i=1}^M A_{i,j}} \sum_{i=1}^M A_{i,j} \frac{z_i}{\sum_{j'=1}^N A_{i,j'} \hat{y}_{j'}^{[n]}}. \quad (23)$$

It is easy to check that the iterates  $\hat{y}^{[n]}$  are never negative-valued provided that the initial estimate is not. Under some assumptions, the sequence of vectors  $(\hat{y}^{[n]})_{n \in \mathbb{N}}$  converges towards a maximizer of the Poissonian likelihood [?, ?]. Nevertheless, this algorithm is known to have a slow convergence rate. In order to alleviate this drawback, fast versions such as the Ordered Subsets Expectation Maximization (OSEM) [?] have been developed. Anyway, because of the lack of regularization, the image recovery process is not very stable and may be dominated by noise. Consequently, resorting to regularization is recommended.

In what follows, we will present some of the most frequently used regularization terms.

### 1.3.3 Regularization in the image domain

A first example of regularization corresponds to the Tikhonov one [?] for which the likelihood is assumed to be Gaussian (see (??)) and the penalty term has the following form:

$$\phi(y) = \|\Lambda y\|_2^2 \quad (24)$$

with  $\Lambda \in \mathbb{R}^{N \times N}$ . Cancelling the gradient of the criterion leads to the following optimal solution:

$$\hat{y} = (A^\top A + \tilde{\chi} \Lambda^\top \Lambda)^{-1} A^\top z, \quad (25)$$

provided that  $A^\top A + \tilde{\chi} \Lambda^\top \Lambda$  is invertible. As  $\hat{y}$  is linear with respect to  $z$ , it is possible to interpret  $\hat{y}$  as the output of a linear recovery filter. Moreover, it is possible to express the solution in terms of the generalized Singular Value Decomposition (SVD) of  $(A, \Lambda)$  [?]. However, for large-scale problems, the direct matrix inversion is not always feasible and alternative approaches have to be followed. For instance, if both  $A$  and  $\Lambda$  are block-circulant, it is possible to diagonalize them by DFT matrices [?, ?]. Hence, the solution is sought in the Fourier transform domain with a reduced computational load.

Many choices are possible for the linear operator  $\Lambda$  and, this degree of freedom makes Tikhonov regularization very flexible. For instance, if  $\Lambda = \text{Id}$ , solutions with low energy are promoted. In practice,  $\Lambda$  often corresponds to a high-pass filter operator such as a gradient or Laplacian operator. A weighted sum of the first and second order derivative operators may also be considered. In this way, a smooth solution with limited high frequency energy is searched while keeping the difference  $Ay - z$  small, and the parameter  $\tilde{\chi}$  controls the smoothness of the solution. Indeed, in the case of a Gaussian likelihood, if  $\tilde{\chi}$  decreases, the solution tends to be irregular and close to the inverse filter output, whereas if  $\tilde{\chi}$  is large, the solution is approaching zero.

An alternative choice consists of choosing  $\Lambda$  such that

$$\Lambda^\top \Lambda = \sigma^2 R_{\bar{Y}}^{-1} \quad (26)$$

where  $R_{\bar{Y}}$  is the autocorrelation matrices of  $\bar{Y}$  defined as

$$R_{\bar{Y}} = \mathbb{E}[\bar{Y} \bar{Y}^\top]. \quad (27)$$

The related recovery filter is known as the parametric Wiener filter. If  $\tilde{\chi} = 1/2$ , it reduces to the standard Wiener filter [?] which is defined as the linear filter that minimizes the mean square error between the estimate and the target image.

Tikhonov regularization leads to estimators  $\hat{y}$  that are linear with respect to the observation  $z$ . Despite their simple closed form, these estimators may suffer from some limitations. For instance, the over-smoothing of significant edges may lead to an unacceptable quality of the recovered images [?]. To alleviate such shortcomings, resorting to nonquadratic regularization appears as an appealing alternative. Nonquadratic functions of the image derivatives can also be introduced, which often outperform the quadratic ones [?, ?, ?, ?]. Entropy and divergence based regularizations are also employed in order to promote energy concentration [?]. One of the most popular regularizations proposed by Rudin, Osher, and Fatemi (ROF) [?] is related to total variation. More precisely, let  $\Omega$  be an open set of  $\mathbb{R}^2$  and let  $W^{1,1}(\Omega)$  denote the Sobolev space of summable functions defined

on  $\Omega$  such that their derivatives in the sense of distributions with an order less than or equal to 1 are also summable. The total variation is then defined as

$$(\forall y \in W^{1,1}(\Omega)), \quad \text{tv}(y) = \int_{\Omega} \|Dy(w)\|_2 \, dw \quad (28)$$

where  $Dy(w) \in \mathbb{R}^2$  is the gradient of  $y$  at location  $w$ . In finite dimension, several definitions of the total variation are possible depending on the choice of the discrete differentiation operator. The most employed one [?] is expressed as

$$(\forall y \in \mathbb{R}^{N_1 \times N_2}), \quad \phi(y) = \text{tv}(y) = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \sqrt{((\nabla_1 y)_{i_1,i_2})^2 + ((\nabla_2 y)_{i_1,i_2})^2} \quad (29)$$

with

$$(\nabla_1 y)_{i_1,i_2} = \begin{cases} y_{i_1+1,i_2} - y_{i_1,i_2} & \text{if } i_1 < N_1 \\ 0 & \text{if } i_1 = N_1 \end{cases} \quad (30)$$

and

$$(\nabla_2 y)_{i_1,i_2} = \begin{cases} y_{i_1,i_2+1} - y_{i_1,i_2} & \text{if } i_2 < N_2 \\ 0 & \text{if } i_2 = N_2. \end{cases} \quad (31)$$

The popularity of the total variation penalization may be accounted for by its good denoising performance especially if the underlying image contains large uniform regions (cartoon-like image). However, for natural images, this method exhibits staircasing effects [?]. Another difficulty raised by this regularization is that it does not correspond to a proper prior distribution  $\mu_Z$  if a Bayesian viewpoint is adopted. Note that one may also think of employing nonconvex penalizations [?], which however may introduce difficulties in the solution of the associated minimization problem, e.g. the presence of spurious local minimas.

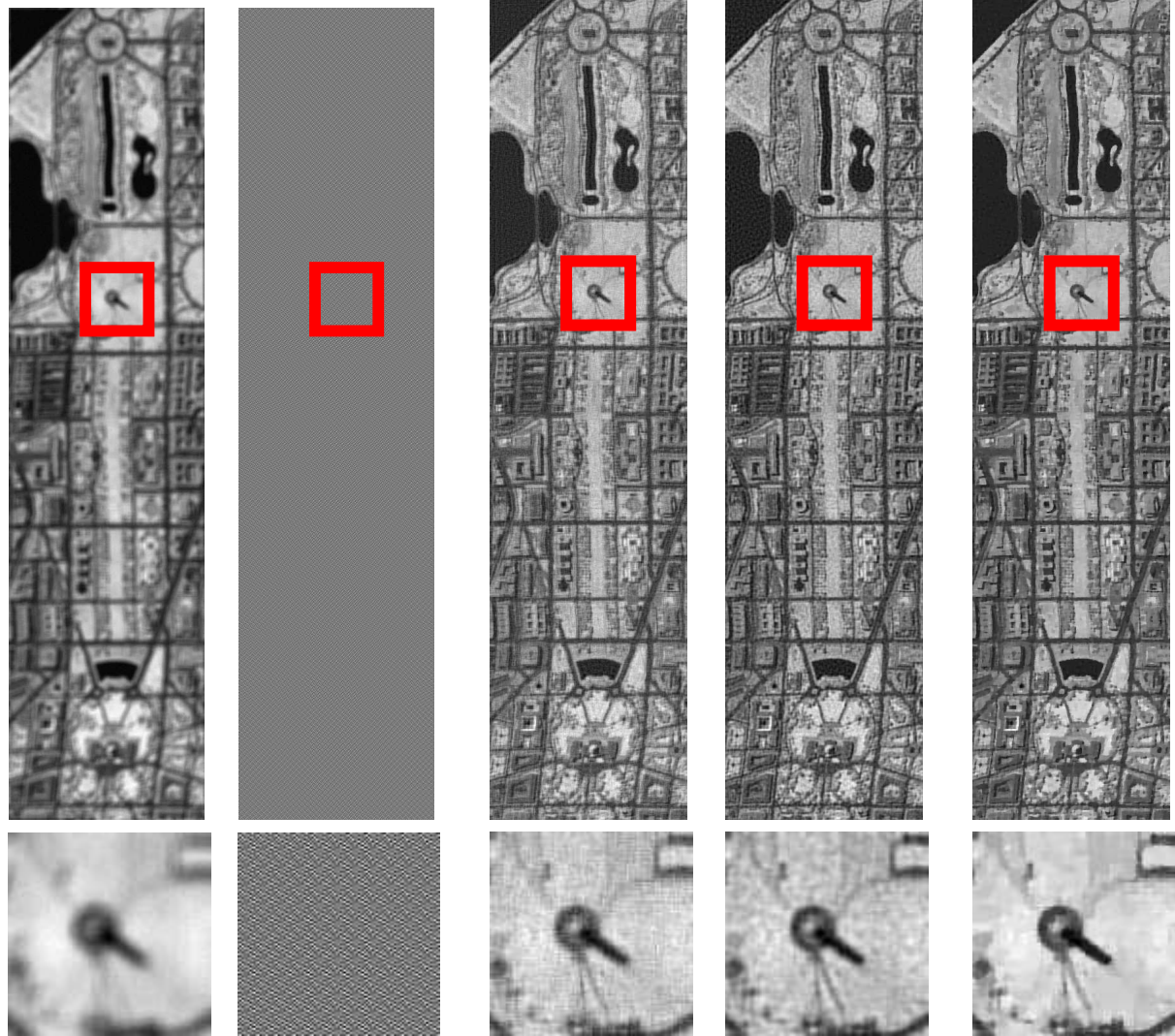
So far, the aforementioned methods perform the reconstruction in the original data space. Changing the data description through a representation leading to insightful coefficients is another powerful strategy to regularize inverse problems. A great deal of attention was paid to perform reconstruction on coefficients resulting from wavelet transforms or frame decompositions. In what follows, the concepts of wavelet transform and frame representations will be presented.

## 2 Wavelet transforms

### 2.1 Motivation

Finding a representation of  $\bar{y}$  leading to coefficients summarizing its intrinsic properties, while allowing the noise to be easily separated from the useful information has been the topic of much attention in signal and image processing. Linear transforms are the simplest tools for generating such coefficients. Traditionally, decompositions onto orthonormal bases such as discrete Fourier or cosine transforms [?] have been employed because they lead to a concentration of the energy into few coefficients. However, their good frequency localization properties are achieved at the expense of a poor spatial localization. In order to reach a tradeoff between both the spatial and frequency





(a) Degraded Uniform blur  $9 \times 9$  Gaussian noise (b) Inverse filtering (c) Quadratic regularisation  $\Lambda = \text{Id}$  (d)  $\Lambda$  Laplacian (e) Total variation regularization

Figure 4: Illustration of image reconstruction with different regularizations in the image domain. (a) Degraded image (SNR = 27.3 dB), (b) Inverse filtering, (c) Tikhonov regularization with  $\Lambda = \text{Id}$ , (d) Tikhonov regularization with  $\Lambda$  modeling the Laplacian operator, (e) total variation regularization.

localizations, the Wavelet Transform (WT) has been introduced [?, ?, ?] and many wavelet-like basis decompositions have also been proposed such as wavelet packets [?] or modulated lapped transforms [?],... As will be further discussed, WT has proven to be a very powerful tool for inverse problems in image processing. This section is devoted to a brief review of the wavelet transforms. More details can be found in [?].



## 2.2 Filter banks

A multiresolution analysis of a 1D discrete-time signal  $(r(k))_{k \in \mathbb{Z}}$  corresponds to a decomposition where the signal is represented at different scales in order to analyze its fine to coarse structures. In practice, it is generated by applying an M-band filter bank consisting of analysis and synthesis filter banks as depicted in Fig. ??.

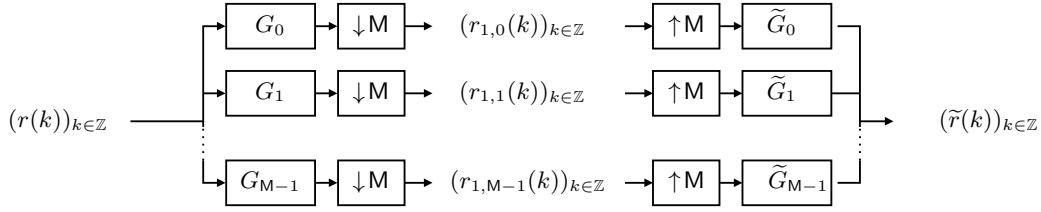


Figure 5: Analysis/synthesis M-band filter bank.

Recall that an analysis M-band filter bank involves M filters with frequency responses  $G_0, \dots, G_{M-1}$ . Filters  $G_0$  and  $G_{M-1}$  are respectively low-pass and high-pass; whereas  $G_1, \dots, G_{M-2}$  are band-pass filters. The set of filters  $\tilde{G}_0, \dots, \tilde{G}_{M-1}$  with its associated upsampling operators is called a synthesis filter bank. The low-pass filtering of  $(r(k))_{k \in \mathbb{Z}}$  by  $G_0$  followed by decimation generates a first subsampled approximation  $(r_{1,0}(k))_{k \in \mathbb{Z}}$  of the original signal. The band-pass and high-pass branches provide subsampled detail signals, in different frequency bands which are complementary to the low-pass approximation. A reconstructed signal  $(\tilde{r}(k))_{k \in \mathbb{Z}}$  is obtained from  $(r_{1,0}(k))_{k \in \mathbb{Z}}, \dots, (r_{1,M-1}(k))_{k \in \mathbb{Z}}$  after upsampling by a factor M, a filtering through the synthesis filters, and a summation.

All the involved filters should satisfy some constraints in order to guarantee the equality between  $\tilde{r}(k)$  and  $r(k)$ , at each time  $k$ , (possibly up to an integer delay and a non-zero multiplicative factor, which can be incorporated into the filter coefficients) [?].

Furthermore, the smoothed version  $(r_{1,0}(k))_{k \in \mathbb{Z}}$  can be further decomposed by the same filter bank, in an iterative manner, as shown in Fig. ??.

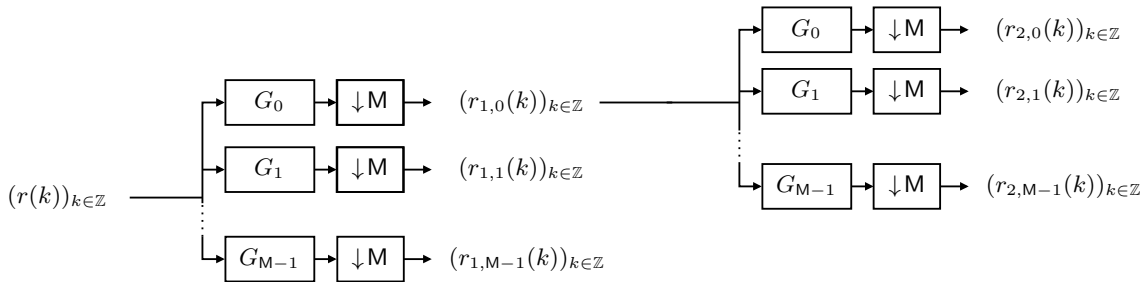


Figure 6: 2-stage M-band wavelet analysis wavelet decomposition.

## 2.3 Filter bank based multiresolution analysis

Such a filter-bank decomposition can also be considered as a multiresolution analysis of the space  $\mathbb{L}^2(\mathbb{R})$  of square integrable (i.e. finite energy) functions from a continuous-time viewpoint. Successive iterations of the basic  $M$ -band analysis filter bank on the low-pass output result in an approximation at resolution level  $j$ . The approximation spaces constitute a decreasing sequence of nested subspaces of  $\mathbb{L}^2(\mathbb{R})$ , associated with a scaling function  $\psi_0 \in \mathbb{L}^2(\mathbb{R})$ . The computation of the WT reduces to projections of the continuous-time signal onto subspaces characterized by  $(M-1)$  mother wavelets  $\psi_m \in \mathbb{L}^2(\mathbb{R})$ ,  $m \in \{1, \dots, M-1\}$  [?]. These functions fullfil the following conditions called the *scaling equations*:

$$(\forall m \in \{0, \dots, M-1\})(\forall t \in \mathbb{R}) \quad \frac{1}{\sqrt{M}}\psi_m\left(\frac{t}{M}\right) = \sum_{k=-\infty}^{\infty} g_m(k)\psi_0(t-k) \quad (32)$$

where  $(g_m(k))_{k \in \mathbb{Z}}$  is the impulse response of filter  $G_m$ . For every  $m \in \{0, \dots, M-1\}$ ,  $j \in \mathbb{Z}$ , and  $k \in \mathbb{Z}$ , let us define scaled and shifted versions of the scaling function and mother wavelets:

$$(\forall t \in \mathbb{R}) \quad \psi_{j,m,k}(t) = M^{-j/2}\psi_m(M^{-j}t - k). \quad (33)$$

Under orthonormality conditions, it is possible to show that

$$(\forall t \in \mathbb{R}) \quad r(t) = \sum_{m \in \{1, \dots, M-1\}} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} r_{j,m}(k) \psi_{j,m,k}(t), \quad (34)$$

where

$$r_{j,m}(k) = \int_{\mathbb{R}} \psi_{j,m,k}(t) r(t) dt. \quad (35)$$

Thus, any signal  $r$  in  $\mathbb{L}^2(\mathbb{R})$  can be linearly expanded onto the orthonormal wavelet basis  $\{\psi_{j,m,k}, (j, k) \in \mathbb{Z}^2, m \in \{1, \dots, M-1\}\}$ .

## 2.4 2D extension

For the sake of simplicity, the extension to 2D wavelet transforms is often handled in a separable way. More precisely, the analysis filter bank is applied successively to the image rows and columns. This amounts to define the 2D wavelets as a tensor product of the 1D wavelets, so yielding a basis of  $\mathbb{L}^2(\mathbb{R}^2)$ . As a result of an  $M$ -band orthonormal discrete wavelet decomposition over  $J$  resolution levels, an approximation subband at the coarsest resolution level  $J$  is obtained as well as  $J \times (M^2 - 1)$  wavelet subbands oriented into the horizontal, vertical and “diagonal” directions at different scales. Although the wavelet transform has been found to be an efficient multiscale representation of the information, much efforts have been paid to improve the representation of geometric informations (namely textures and edges) in order to preserve them during processing. More precisely, the interest has moved towards more general representations known as frames (in the sense of Hilbert space theory), that are redundant. This appeal of frames can be mainly explained by the flexibility of these overcomplete representations to capture local features of the signal and by their quasi shift-invariance properties. The next section is dedicated to a presentation of the concept of frame transforms and their most important properties that highlight their interest for solving inverse problems.

### 3 Frames

#### 3.1 Definition of a frame

The first definition of frames was given by Duffin and Schaeffer in their study on non harmonic Fourier series [?]. Let  $\mathbb{K}$  denote a nonempty subset of  $\mathbb{N}$  and, let  $\mathcal{G}$  denote a separable Hilbert space endowed with an inner product  $\langle \cdot, \cdot \rangle$  and a norm  $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ . A family  $(e_k)_{k \in \mathbb{K}}$  of  $\mathcal{G}$  is a frame if there exist two constants  $\underline{\nu}$  and  $\overline{\nu}$  in  $]0, +\infty[$  such as

$$(\forall y \in \mathcal{G}) \quad \underline{\nu} \|y\|^2 \leq \sum_{k \in \mathbb{K}} |\langle y, e_k \rangle|^2 \leq \overline{\nu} \|y\|^2. \quad (36)$$

The constants  $\underline{\nu}$  and  $\overline{\nu}$  are called lower and upper frame bounds, respectively. Thus, the frame concept covers both orthonormal bases ( $\underline{\nu} = \overline{\nu} = 1$ ), and overcomplete sets of vectors. Note that, in finite dimension, it can be shown that a frame reduces to a spanning set of  $\mathcal{G}$ , and  $\text{card}[\mathbb{K}] \geq \dim[\mathcal{G}]$  [?] and the upper bound in (??) is always met.

#### 3.2 Related operators

The associated frame analysis operator  $F$  is the injective bounded linear operator defined as

$$F: \mathcal{G} \rightarrow \ell^2(\mathbb{K}): y \mapsto (\langle y, e_k \rangle)_{k \in \mathbb{K}}. \quad (37)$$

where  $\ell^2(\mathbb{K})$  designates the sequences  $(\xi_k)_{k \in \mathbb{K}}$  such that  $\sum_{k \in \mathbb{K}} |\xi_k|^2 < +\infty$ .  $F$  also called the Bessel map associated with  $(e_k)_{k \in \mathbb{K}}$ . Its adjoint  $F^*$  is the frame synthesis (or reconstruction) operator defined as

$$F^*: \ell^2(\mathbb{K}) \rightarrow \mathcal{G}: (\xi_k)_{k \in \mathbb{K}} \mapsto \sum_{k \in \mathbb{K}} \xi_k e_k. \quad (38)$$

It can be shown that  $F^*$  is a surjective bounded linear operator [?]. By composing  $F$  and  $F^*$ , the frame operator  $S = F^*F$  is obtained:

$$S: \mathcal{G} \rightarrow \mathcal{G}: y \mapsto \sum_{k \in \mathbb{K}} \langle y, e_k \rangle e_k. \quad (39)$$

If we can choose  $\underline{\nu} = \overline{\nu} = \nu$ , the frame  $(e_k)_{k \in \mathbb{K}}$  is said to be tight. In this case, the following equality holds:

$$F^*F = \nu \text{Id}. \quad (40)$$

The family  $(e_k)_{k \in \mathbb{K}}$  is an orthonormal basis of  $\mathcal{G}$  if and only if  $F^{-1} = F^*$ .

#### 3.3 Examples of frames

A union of  $\nu$  orthonormal bases is the simplest example of a tight frame. Curvelets [?, ?] also constitute an example of tight frames of  $\mathcal{G} = \mathbb{L}^2(\mathbb{R}^2)$ . A real (resp. complex) dual-tree wavelet

decomposition is the union of two (resp. four) orthonormal wavelet bases [?, ?]. Under some conditions, contourlets [?] also are tight frames. From a historical viewpoint, Gabor frames [?, ?] have played an important role in many inverse problems.

### 3.4 Frame decomposition theorem

It can be shown that  $S$  is an isomorphism with inverse mapping  $S^{-1}$ . The family of vectors  $(\tilde{e}_k)_{k \in \mathbb{K}}$  defined as

$$(\forall k \in \mathbb{K}) \quad \tilde{e}_k = S^{-1} e_k \quad (41)$$

constitutes a frame of  $\mathcal{G}$  called the dual frame of  $(e_k)_{k \in \mathbb{K}}$  with frame bounds  $\overline{\nu}^{-1}$  and  $\underline{\nu}^{-1}$ .

It is also possible to prove that  $FS^{-1}$  is the frame analysis operator of the dual frame  $(\tilde{e}_k)_{k \in \mathbb{K}}$ . Furthermore, we have:

$$(\forall y \in \mathcal{G}) \quad y = \sum_{k \in \mathbb{K}} \langle y, e_k \rangle \tilde{e}_k, \quad (42)$$

$$= \sum_{k \in \mathbb{K}} \langle y, \tilde{e}_k \rangle e_k. \quad (43)$$

The last expression can also be rewritten as

$$y = Dc \quad (44)$$

where

- $D$  can be viewed as a (possibly infinite dimensional) matrix whose columns are the vectors  $(e_k)_{k \in \mathbb{K}}$ ,
- $c$  is a vector whose  $k$ -component is the coefficient  $\langle y, \tilde{e}_k \rangle$ .

Thus, every vector  $y$  in  $\mathcal{G}$  corresponds to a linear combination of the frame elements. Although a similar property holds for a basis, the difference is that for a frame, a vector  $c$  allowing Relation (??) to be satisfied is generally not uniquely defined and it can thus be replaced by another one than  $(\langle y, \tilde{e}_k \rangle)_{k \in \mathbb{K}}$ . This degree of freedom is of crucial interest in inverse problems. Indeed, the determination of the coefficients can be guided by a criterion that fits the desired properties of the signal.

At this point, it is worthwhile to note that solving an inverse problem using a frame representation requires to address key practical issues:

1. the definition of a criterion for the determination of the coefficients of the sought image;
2. the design of the frame itself;
3. the computation of the coefficients associated with an element of  $\mathcal{G}$  for a given frame.

### 3.5 Sparsity criterion

When the family  $(e_k)_{k \in \mathbb{K}}$  is linearly dependent (i.e. the frame is *overcomplete*), we have seen that the vector of coefficients  $c$  is not uniquely defined. Therefore, it is possible to select coefficients that meet a certain condition, e.g. a sparsity assumption. In other words, the signal is considered as a combination of few atoms and, a sparse representation is promoted with a majority of null (or quasi-null) coefficients. Ideally, the problem can be recast as the following constrained optimization problem:

$$\text{minimize } \|c\|_0 \text{ under the constraint } y = Dc \quad (45)$$

where  $\|c\|_0$  denotes the number of nonzero components of  $c$ . This is however an NP-hard problem, which is intractable [?]. In finite dimension, a full search strategy indeed requires to test all the  $k$ -uplets that can be extracted from  $c$  for  $k = 1, \dots, \text{card}[\mathbb{K}]$ . The number of such  $k$ -uplets amounts to

$$\binom{\text{card}[\mathbb{K}]}{k} = \frac{(\text{card}[\mathbb{K}])!}{k!(\text{card}[\mathbb{K}] - k)!}. \quad (46)$$

When  $\text{card}[\mathbb{K}]$  increases, the number of  $k$ -uplets to be tested increases as  $\text{card}[\mathbb{K}]^k$ . The cost is prohibitive and the problem becomes infeasible.

Therefore, the sparsity measure  $\|c\|_0$  is often replaced by the  $\ell^2$  or  $\ell^1$  norm since these two criteria are mathematically tractable. The last one especially is very popular since it can be shown that, under some technical assumptions [?] it gives an exact solution to Problem (??). The resulting optimization problem reads:

**Problem 3.1** Minimize  $\|c\|_1 = \sum_{k \in \mathbb{K}} |c_k|$  subject to the constraint  $y = Dc$ .

### 3.6 Decomposition algorithms

Two strategies can be considered for solving Problem (??): sequential and parallel approaches. Sequential approaches consist of greedy methods that aim at choosing one vector of the frame at a time. More precisely, the frame vector that “best” approximates the signal is found and, the related approximation is computed by selecting iteratively new vectors according to some criterion such as the best match to the residual. In contrast, in parallel strategies, all the vectors of the frame are selected at the initial step and some of them are discarded according to some optimization process. Matching pursuit [?] and orthogonal matching pursuit [?] are examples of greedy algorithms, whereas basis pursuit algorithms [?, ?] correspond to parallel decomposition methods.

### 3.7 Designing a frame

The elements of the chosen frame are expected to reflect the properties of the signal to be analyzed. Therefore, the choice of a frame is a delicate task that has a strong impact on the final performance of the reconstruction/restoration. Several strategies have been developed in order to construct frames [?]. The general guidelines are as follows.

- Strategies that promote a low computational cost employ pre-specified frames (e.g. overcomplete wavelets, curvelets, contourlets, steerable wavelet filters, short-time Fourier transforms) that are not signal-adapted. Tight frames are often preferred as they can be easily pseudo-inverted. The performance is satisfactory as long as the resulting representations are sparse.
- Frames can also be built by exploiting prior knowledge on the signal to be analyzed [?].
- It is also possible to construct atoms of the frame through a learning on a training set of reference signals that are assumed to be similar to the signals to be analyzed [?, ?, ?].
- Recent procedures based on dictionary learning make it possible to build the frame from the signal to be analyzed, but their computational cost is often high [?].

### 3.8 Why frames in inverse problems?

In inverse problems, it has been observed that it is more convenient to consider overcomplete frames rather than bases since a higher number of atoms leads to a more flexible linear model with an expected reduced number of nonzero terms for matching the signal. Wavelet bases are well-adapted to approximate piecewise smooth functions. But, in the case of an image containing more complex structures, their lack of shift invariance [?, ?] may lead to reconstruction artifacts, e.g. ringing artifacts near edges.

Overcomplete frame decompositions lead to more precise approximations. Nevertheless, since the family  $(e_k)_{k \in \mathbb{K}}$  is linearly dependent, as already mentioned, the set of coefficients is no longer unique. Therefore, a sparsity criterion needs to be employed in order to concentrate the information in the data on a relatively few and robustly estimated coefficients.

### 3.9 Regularization in the frame domain

As in Problem ??, a common choice for a regularization using an analysis frame is the  $\ell^1$ -norm applied to the frame coefficients, i.e.,

$$\phi(y) = g(Fy) = \|Fy\|_1 = \sum_{k \in \mathbb{K}} |(Fy)_k|.$$

A statistical motivation for this choice is that a prior probability density function proportional to  $\exp(-\|F \cdot\|_1)$  appears to provide a good approximate model for the frame coefficients, by noticing that the negative log-likelihood associated with this distribution leads to the  $\ell^1$ -norm. When the frame is a basis, this distribution reduces to an i.i.d. Laplacian model (*cf.* Figs ?? (b,c,d)). Other choices for  $g$  can be considered in order to better capture inter/intrascale behaviours [?].

A close relation can be established between the  $\ell^1$ -norm applied to the image gradient (total variation) and the undecimated wavelet transform involving a Haar filter [?]. However, due to the flexibility offered by frames and their ability to generate multiscale analyses, the use of such a transform may yield a better performance than total variation for natural images [?].

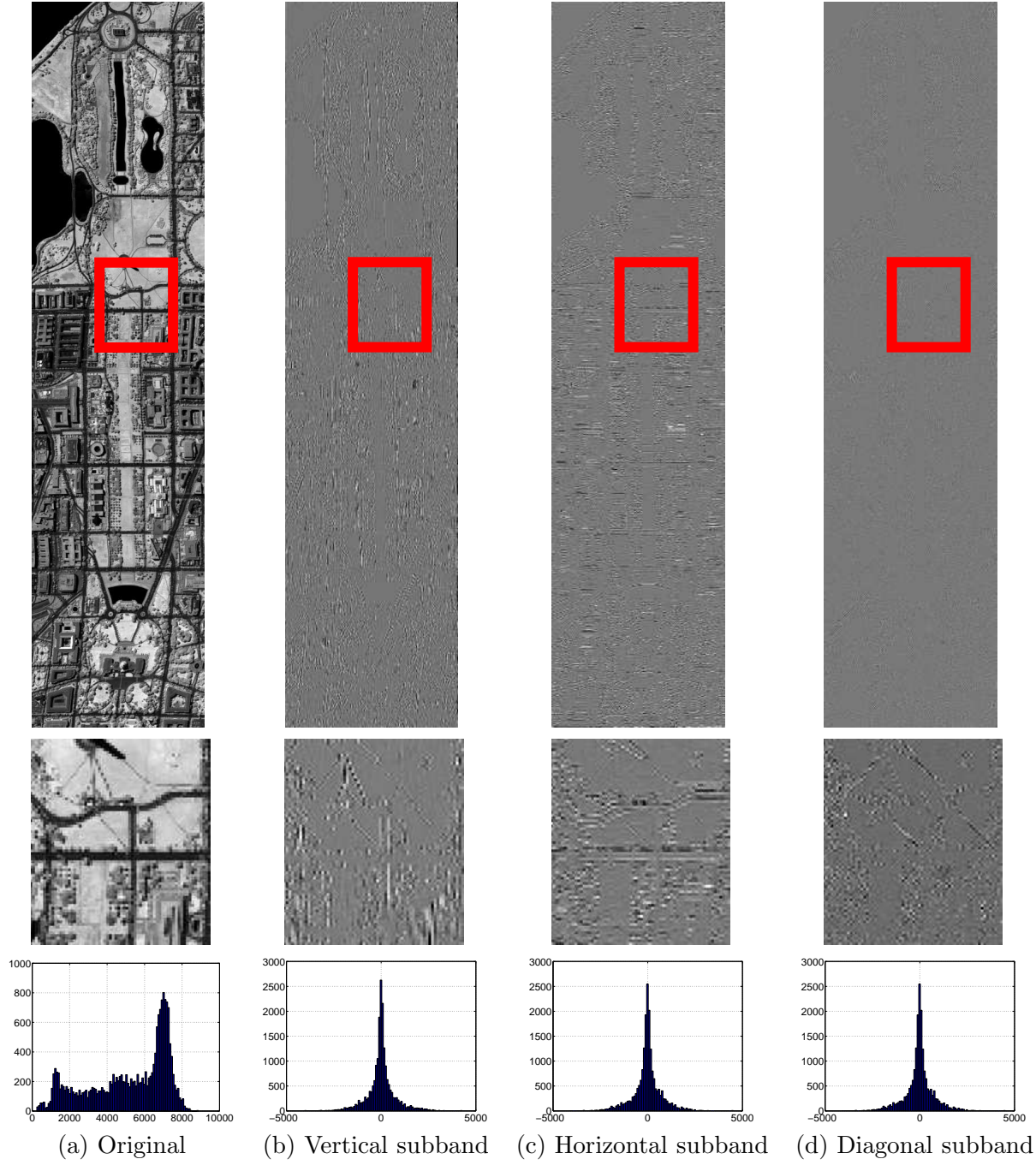


Figure 7: Example of 2D wavelet representation. Zoom associated with the red area is displayed on the second row and the histogram associated to each subband is presented on the third row.

### 3.10 Image reconstruction in the frame domain

If a synthesis frame is employed, the degradation model (??) becomes

$$z = \mathcal{D}_\alpha(AF^*\bar{x}) \quad (47)$$

where  $\bar{x} \in \mathbb{R}^K$  are the frame coefficients of the reference image  $\bar{y}$  such as  $\bar{y} = F^*\bar{x}$ . In what follows, we adopt the following notation:  $\mathbb{K} = \{1, \dots, K\}$ ,  $\mathcal{G} = \mathbb{R}^N$ , and  $K \geq N$ . The objective is to recover the frame coefficients  $\hat{x}$  that are the closest to  $\bar{x}$ , from the observations  $z$ .

At this point, it is worth noting that for a given observation model (reflected by given  $\mathcal{D}_\alpha$  and  $A$ ), the differences between the reported methods stem from the combination of several choices:

- the employed frame,
- the adopted objective function that the estimate  $\hat{x}$  should minimize, in particular the choice of the possible regularization functions,
- the optimization method used to compute the optimal solution.

In Fig. ??, we display several reconstruction results based on wavelet basis or redundant frame representations. The frame used in our experiment is a dual-tree wavelet decomposition [?, ?]. The analysis and synthesis results are presented. These two configurations will be further discussed in the next section.

In the next section, we will describe the most important methods operating on the frame coefficients deriving from these multiple choices.

## 4 Variational approaches using frames

### 4.1 Background

In a variational framework, the restored image  $\hat{y} \in \mathbb{R}^N$  is sought as a minimizer of a criterion modeling the prior knowledge on the problem. The most intuitive criterion is composed of a data fidelity term and a regularization term. Formally, if we denote by  $\psi: \mathbb{R}^N \rightarrow ]-\infty, +\infty[$  the data fidelity term and by  $g: \mathbb{R}^K \rightarrow ]-\infty, +\infty[$  the regularization term, the variational approach consists of finding

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} \psi(y) + g(Fy). \quad (48)$$

As already explained, the data term ensures that the signal  $A\hat{y}$  is close to the observed data  $z$ . This closeness can be measured in terms of a squared  $\ell^2$ -norm or by using a divergence such as the Kullback-Leibler one. The choice of this data term is often motivated by knowledge about the statistical properties of the noise (cf. Section ??). To avoid an irregular solution due to noise, the second term imposes some smoothness on the solution. This regularity can be measured in terms



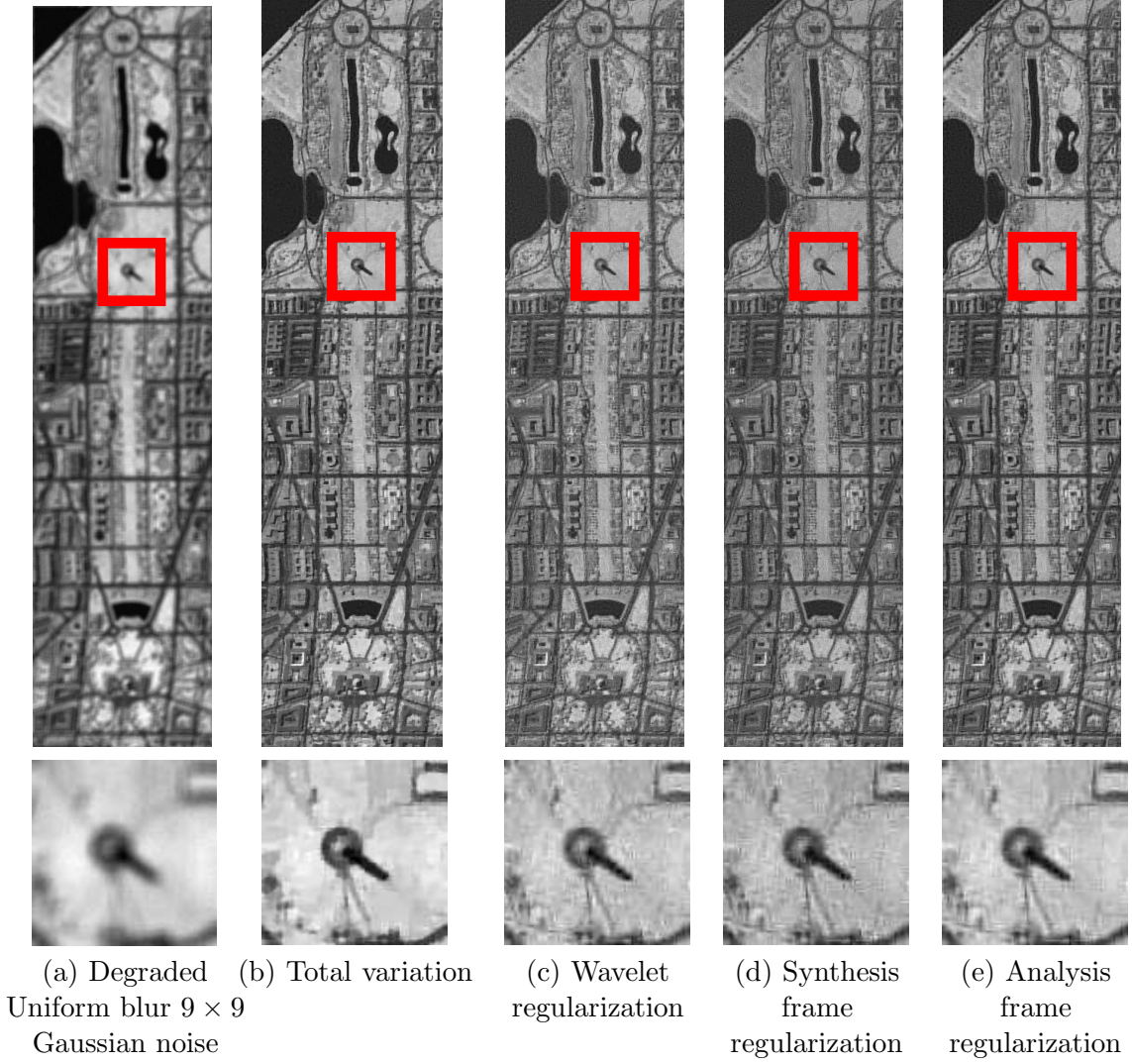


Figure 8: Illustration of image reconstruction with different regularization based on wavelet and frames. A restoration result based on total variation is also presented in order to easily compared the restoration results. (a) Degraded image (SNR = 27.3 dB), (b) total variation regularization, (c) wavelet basis regularization using *symmlets* of order 3 with 3 decomposition levels, (d)/(e) synthesis/analysis frame based regularization using *symmlets* of order 3 with 3 decomposition levels.

of small variations of the gradient. This leads to  $g(Fy) \propto \text{tv}(y)$ , in which case the equivalent of  $F$  is the gradient operator with  $K = 2N$  and  $g$  is a sum of Euclidean norms of two-dimensional vectors (see Section ??). The regularity can also be expressed in terms of sparsity of the image by choosing  $g(Fy) \propto \|y\|_1$  ( $K = N$ ). In a probably more realistic way, the sparsity can be imposed on the coefficients in a suitable general frame by setting  $g(Fy) \propto \|Fy\|_1$ . These choices have been discussed in Section ??.

Although this form of criterion is frequently encountered in the literature, one can incorporate

as many constraints as informations available about the signal to estimate. For instance, for an 8 bit-coded image, it is natural to impose a constraint on the dynamic range of the image, by enforcing the variable  $y$  to belong to the hypercube  $C = [0, 255]^N$ . The associated constrained minimizer reads

$$\hat{y} \in \underset{y \in C}{\operatorname{Argmin}} \psi(y) + g(Fy). \quad (49)$$

Incorporating such a constraint can be equivalently formulated as adding an indicator function<sup>2</sup> to the criterion:

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\operatorname{Argmin}} \psi(y) + g(Fy) + \iota_C(y). \quad (50)$$

The interest of wavelet transforms or more generally frames is to provide an insightful representation of an image in order to highlight some of its features. However, the information available directly on data, in the spatial domain, such as the presence of sharp edges is also useful to perform an efficient recovery. For this reason, it is important to note that some functionals should be applied directly on the pixels of the image and some others on the frame coefficients, possibly using different frames. The resulting variational problem can be formulated in a general setting [?] that consists of finding  $\hat{y}$  such that

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\operatorname{Argmin}} \sum_{i=1}^I f_i(y) + \sum_{j=1}^J g_j(Fy), \quad (51)$$

where, for every  $i \in \{1, \dots, I\}$ ,  $f_i: \mathbb{R}^N \rightarrow ]-\infty, +\infty]$  and, for every  $j \in \{1, \dots, J\}$ ,  $g_j: \mathbb{R}^K \rightarrow ]-\infty, +\infty]$ . The functions  $(f_i)_{1 \leq i \leq I}$  and  $(g_j)_{1 \leq j \leq J}$  can take the  $+\infty$  value in order to deal with constraints, as it is the case for the indicator function in (??). Additional technical assumptions often adopted on the involved functions is that they are proper (at least finite in one point), convex, and that they are lower-semicontinuous.<sup>3</sup>

## 4.2 Analysis versus synthesis formulation

The criterion (??) is called an *analysis formulation* of the variational problem in contrast with another formulation, called *synthesis formulation*, where one estimates  $\hat{y} = F^* \hat{x}$  such that

$$\hat{x} \in \underset{x \in \mathbb{R}^K}{\operatorname{Argmin}} \sum_{i=1}^I f_i(F^*x) + \sum_{j=1}^J g_j(x). \quad (52)$$

The solutions resulting from the analysis and synthesis formulations only match for some specific choices of  $F$ ,  $(f_i)_{1 \leq i \leq I}$ , and  $(g_j)_{1 \leq j \leq J}$ . The results given in the following were partly derived in [?, ?].

**Property 4.1** *If  $F$  corresponds to a representation onto an orthonormal basis, then the analysis formulation (??) and the synthesis formulation (??) are equivalent.*

<sup>2</sup>The indicator function  $\iota_C$  of a set  $C$  is such that  $\iota_C(y) = 0$  if  $y \in C$  and  $\iota_C(y) = +\infty$  otherwise.

<sup>3</sup>A function  $f: \mathbb{R}^N \rightarrow ]-\infty, +\infty]$  is lower-semicontinuous if its epigraph  $\{(x, \zeta) \in \mathbb{R}^N \times \mathbb{R} \mid f(x) \leq \zeta\}$  is a closed set.

*Proof.* This property is related to the fact that  $y = F^*x \Leftrightarrow x = Fy$ , when  $F$  corresponds to a decomposition onto an orthonormal basis.  $\square$

**Property 4.2** *The analysis formulation (??) is a particular case of the synthesis formulation (??).*

*Proof.* By definition of a frame,  $F^*$  is surjective. Consequently, for every  $y \in \mathbb{R}^N$ , there exists an element  $x$  in  $\mathbb{R}^K$  such that  $y = F^*x$  and (??) can be rewritten as

$$\hat{y} \in \underset{y=F^*x}{\text{Argmin}} \sum_{i=1}^I f_i(F^*x) + \sum_{j=1}^J g_j(F^*x), \quad (53)$$

that is

$$\hat{y} = F^*\hat{x} \quad \text{with} \quad \hat{x} \in \underset{x}{\text{Argmin}} \sum_{i=1}^I f_i(F^*x) + \sum_{j=1}^J h_j(x), \quad (54)$$

where, for every  $j \in \{1, \dots, J\}$ ,  $h_j = g_j(F^*\cdot)$ .  $\square$

**Property 4.3** *Let  $F$  be a non bijective tight frame analysis operator. If, for every  $j \in \{1, \dots, J\}$ ,  $g_j$  can be written as a sum of functions  $h_{j,1}: \text{Im}F \rightarrow ]-\infty, +\infty]$  and  $h_{j,2}: \ker F^* \rightarrow ]-\infty, +\infty]$ , i.e.,*

$$(\forall (y, x_\perp) \in \mathbb{R}^N \times \ker F^*) \quad g_j(Fy + x_\perp) = h_{j,1}(Fy) + h_{j,2}(x_\perp) \quad (55)$$

*where  $\ker F^*$  is the nullspace of  $F^*$ , and if, for every  $u \in \ker F^*$ ,  $h_{j,2}(u) \geq h_{j,2}(0)$ , then the analysis formulation (??) and the synthesis formulation (??) are equivalent.*

*Proof.* Since  $F$  is not bijective,  $\ker F^* \neq \{0\}$ . Then, every  $x \in \mathbb{R}^K$  can be decomposed as a sum of an element belonging to the range of  $F$  and an element belonging to the nullspace of  $F^*$  since  $\ker F^* = (\text{Im}F)^\perp$ , i.e.

$$(\forall x \in \mathbb{R}^K) (\exists (y, x_\perp) \in \mathbb{R}^N \times \ker F^*) \quad x = Fy + x_\perp. \quad (56)$$

Consequently,  $F^*x = F^*Fy$ , which shows that solving (??) is equivalent to

$$\underset{\substack{y \in \mathbb{R}^N \\ x_\perp \in \ker F^*}}{\text{minimize}} \sum_{i=1}^I f_i(F^*Fy) + \sum_{j=1}^J g_j(Fy + x_\perp). \quad (57)$$

As  $F$  is a tight frame operator,  $F^*F = \nu \text{Id}$  with  $\nu > 0$ . This yields

$$\underset{\substack{y \in \mathbb{R}^N \\ x_\perp \in \ker F^*}}{\text{minimize}} \sum_{i=1}^I \tilde{h}_i(y) + \sum_{j=1}^J g_j(Fy + x_\perp), \quad (58)$$

where, for every  $i \in \{1, \dots, I\}$ ,  $\tilde{h}_i = h_i(\nu \cdot)$ . By using now the separability assumption on the functions  $(g_j)_{1 \leq j \leq J}$  as expressed by (??), the problem becomes

$$\begin{cases} \underset{y \in \mathbb{R}^N}{\text{minimize}} & \sum_{i=1}^I \tilde{h}_i(y) + \sum_{j=1}^J h_{j,1}(Fy) \\ \underset{x_\perp \in \ker F^*}{\text{minimize}} & \sum_{j=1}^J h_{j,2}(x_\perp). \end{cases} \quad (59)$$

Moreover, since we have assumed that, for every  $u \in \ker F^*$ ,  $h_{j,2}(u) \geq h_{j,2}(0)$ , the problem reduces to

$$\underset{y \in \mathbb{R}^N}{\text{minimize}} \quad \sum_{i=1}^I \tilde{h}_i(y) + \sum_{j=1}^J h_{j,1}(Fy), \quad (60)$$

which corresponds to a frame analysis formulation.  $\square$

## 5 Optimization methods

Finding a solution to the analysis or synthesis formulations (??)-(??) requires the design of an optimization strategy that mainly depends on the number of involved functions, their nature (smooth or not), and the form of the involved matrices, in particular the linear degradation operator  $A$  (diagonal matrix, circulant matrix,...) or the frame matrix  $F$  (tight frame or not). In this section, we provide a quick overview of the main algorithmic solutions proposed in the literature in the context of frame-based image recovery.

### 5.1 Synthesis formulation in the presence of Gaussian noise

One of the most commonly used criterion for solving inverse problem involving a wavelet-based regularization is the following:

$$\hat{x} \in \underset{x \in \mathbb{R}^K}{\text{Argmin}} \quad \|AF^*x - z\|^2 + \chi\|x\|_1 \quad (61)$$

where  $\chi > 0$ .

The difficulty of finding an efficient numerical solution stems from the nonsmooth behaviour of the penalization term. Due to this fact, classical gradient descent based methods cannot be applied directly. A basic solution could be to employ a subgradient based strategy [?] but it is well-known that such an approach suffers from slow convergence rate. Another class of approaches consists of solving a smooth relaxation of this criterion by replacing the  $\ell^1$ -norm with a differentiable approximation such as the Huber function [?, ?], with the drawback of limiting the achieved sparsity.

The first works providing a suitable algorithm for solving exactly (??), when  $F$  models a wavelet transform, were derived by Figueiredo *et al.* [?] by means of an expectation-maximization algorithm and by Daubechies *et al.* [?] by means of convex analysis tools. Even though the derivations of the algorithm differs, both lead to an iteration consisting of a gradient descent step followed by a soft-thresholding. This iteration has been incorporated into a proximal formalism and generalized by Combettes and Wajs in [?]. For our purpose, the latter framework mainly allows us to provide convergence guarantees with less restrictive conditions on the step-size parameter and also to deal with a wider class of possibly nonsmooth regularization terms. This algorithmic solution known as *forward-backward* (FB) can handle frame representations as explained by Chaux *et al.* [?]. The iterations of the FB algorithm in the context of frames are given below:

Set  $\gamma_n \in \mathbb{R}$  such that  $\inf_{n \in \mathbb{N}} \gamma_n > 0$  and  $\sup_{n \in \mathbb{N}} \gamma_n < \frac{1}{\|AF^*\|^2}$ .  
Set  $\lambda_n \in ]0, 1]$  such that  $\inf_{n \in \mathbb{N}} \lambda_n > 0$ .  
Initialize  $x^{[0]} \in \mathbb{R}^K$ .  
For  $n = 0, 1 \dots$   
 $\quad x^{[n+1]} = x^{[n]} + \lambda_n (\text{prox}_{\gamma_n \chi \|\cdot\|_1}(x^{[n]} - 2\gamma_n F A^\top (AF^* x^{[n]} - z)) - x^{[n]})$

where  $\text{prox}_{\gamma_n \chi \|\cdot\|_1}$  denotes the soft-thresholding with threshold value  $\gamma_n \chi$  that is

$$(\forall u = (u_k)_{1 \leq k \leq K}) \quad \text{prox}_{\gamma_n \chi \|\cdot\|_1}(u) = (\text{sign}(u_k) \max\{0, |u_k| - \gamma_n \chi\})_{1 \leq k \leq K}. \quad (62)$$

and  $\text{sign}$  denotes the signum function. The sequence  $(x^{[n]})_{n \in \mathbb{N}}$  generated by this algorithm is guaranteed to converge to a solution to (??). In the literature, this algorithm often specialized to the case when, for every  $n \in \mathbb{N}$ ,  $\lambda_n = 1$  and  $\gamma_n \in ]0, \frac{1}{2\|AF^*\|^2}]$ , is called Iterative Soft-Thresholding Algorithm (ISTA) [?, ?],<sup>4</sup> although convergence results actually allow the stepsize to be increased almost by a factor 2, resulting in a faster convergence. Some variants of the algorithm have also been developed in order to accelerate its convergence, the most famous ones being the Two-step Iterative Shrinkage-Thresholding (TwIST) [?] and Fast ISTA (FISTA) [?, ?]. These algorithms take advantage of subspace acceleration, the update at iteration  $n$  making use of the previous iterations  $x^{[n]}$  and  $x^{[n-1]}$ , so leading to improved worst case bounds in terms of functional convergence rates. Another useful way of improving the convergence behaviour is to employ preconditioning strategies which are reminiscent of quasi-Newton methods [?].

### 5.1.1 Step-size parameter

In practice, the convergence speed of the forward-backward algorithm is highly related to the choice of the stepsizes  $(\gamma_n)_{n \in \mathbb{N}}$ . It is often observed that the closer to the upper-bound  $1/\|AF^*\|^2$  these parameter are, the faster the solution is reached. It is thus important to compute accurately  $\|AF^*\|^2$ , the value of which is rarely known a priori. The computation of this norm is achieved by the following arguments. Let us perform the eigenvalue decomposition of the matrix  $B^\top B$  where  $B = AF^*$ , i.e.

$$B^\top B = U \Lambda U^\top$$

where  $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_K\}$  and where  $U = [u_1, \dots, u_K] \in \mathbb{R}^{K \times K}$  is an orthogonal matrix. By definition of the norm,  $\|B\|^2 = \|AF^*\|^2 = \lambda_{i_0}$  where  $i_0 \in \underset{1 \leq i \leq K}{\text{Argmax}} \lambda_i$ . If  $x^{[0]}$  denote a vector in  $\mathbb{R}^K$  which does not lie in the nullspace of  $B$ , we can write,

$$(\forall n \in \mathbb{N}^*) \quad \frac{\|B^n x^{[0]}\|^2}{\|B^{n-1} x^{[0]}\|^2} = \frac{\sum_{i=1}^K \lambda_i^n |\langle x^{[0]}, u_i \rangle|^2}{\sum_{i=1}^K \lambda_i^{n-1} |\langle x^{[0]}, u_i \rangle|^2}, \quad (63)$$

which leads to

$$\lim_{n \rightarrow +\infty} \frac{\|B^n x^{[0]}\|^2}{\|B^{n-1} x^{[0]}\|^2} = \lambda_{i_0} = \|AF^*\|^2. \quad (64)$$

Based on (??), a sequence  $(\rho^{[n]})_{n \in \mathbb{N}}$  converging to  $\|AF^*\|^2$  can be built as follows:

---

<sup>4</sup>It is also sometimes known as the proximal-gradient algorithm or the thresholded Landweber algorithm.

Choose randomly  $x^{[0]} \in \mathbb{R}^K$ .  
Set  $\rho^{[0]} = 0$ ,  $\rho^{[1]} = 1$ , and  $n = 1$ .  
Set  $\epsilon > 0$ .  
While  $|\rho^{[n]} - \rho^{[n-1]}| \geq \epsilon \rho^{[n]}$   

$$\left[ \begin{array}{l} x^{[n]} = FA^\top AF^* x^{[n-1]} \\ \rho^{[n]} = \frac{\|x^{[n]}\|}{\|x^{[n-1]}\|} \\ n \leftarrow n + 1. \end{array} \right.$$

In the literature, this algorithm is known as the *power iteration*.

### 5.1.2 Uniqueness of the solution

The set of solutions to (??) is reduced to a singleton if  $AF^*$  is an injective transform. Another way to ensure uniqueness is to add a squared  $\ell^2$ -norm penalization in the criterion, by looking for

$$\hat{x} = \underset{x \in \mathbb{R}^K}{\operatorname{argmin}} \|AF^*x - z\|^2 + \epsilon \|x\|^2 + \chi \|x\|_1 \quad (65)$$

with  $\epsilon > 0$ . The additional term can be interpreted as a regularization  $g(x) = \epsilon \|x\|^2 + \chi \|x\|_1$  that is sometimes called an *elastic net* penalization [?]. Then, the iterations of FB become

$$x^{[n+1]} = x^{[n]} + \lambda_n (\operatorname{prox}_{\gamma_n \chi \|\cdot\|_1} ((1 - 2\gamma_n \epsilon)x^{[n]} - 2\gamma_n FA^\top (AF^*x^{[n]} - z)) - x^{[n]}).$$

Since the solution is unique, the choice of the initial value  $x^{[0]}$  does not impact the asymptotic behaviour of the algorithm. Usual choices are  $x^{[0]} = 0$  or  $x^{[0]} = Fz$  (when  $M = N$ ).

### 5.1.3 Proximity operator

In the forward-backward algorithm, we have introduced the operator  $\operatorname{prox}_{\gamma_n \chi \|\cdot\|_1}$  for which a closed form expression has been given. However, the notation  $\operatorname{prox}$  is much more informative because it refers to the proximity operator, a mathematical tool possessing numerous interesting properties from an optimization standpoint [?]. The proximity operator associated with a convex, lower-semicontinuous, and proper function  $h: \mathbb{R}^\kappa \rightarrow ]-\infty, +\infty]$  at a point  $u \in \mathbb{R}^\kappa$  is defined as the unique minimizer of  $h + \|\cdot - u\|^2$ , i.e.

$$\begin{aligned} \operatorname{prox}_h : \mathbb{R}^\kappa &\rightarrow \mathbb{R}^\kappa \\ u &\mapsto \underset{v \in \mathbb{R}^\kappa}{\operatorname{argmin}} \frac{1}{2} \|v - u\|^2 + h(v). \end{aligned} \quad (66)$$

This operator generalizes the definition of the projection onto a nonempty closed convex set  $C \subset \mathcal{H}$  since the proximity operator of the indicator function  $\iota_C$  of  $C$  reduces to this projection. Numerous closed form expressions are listed in [?, ?, ?, ?] such as for the proximity operators associated with the negative log-likelihood of the generalized Gaussian distribution, exponential distribution,

gamma distribution, chi distribution, uniform distribution, and more generally, of a large class of log-concave distributions [?]. Here, we recall the expression of the proximity operator associated with the negative log-likelihood of the generalized Gaussian distribution for an exponent value  $p \in \{1, 4/3, 3/2, 2, 3, 4\}$ . The graph of the resulting proximity operator is plotted in Fig. ??.

**Example 5.1** [?] Let  $\chi > 0$ ,  $p \in [1, +\infty[$  and  $\varphi: \mathbb{R} \rightarrow ]-\infty, +\infty]: \xi \mapsto \chi|\xi|^p$ . Then, for every  $\xi \in \mathbb{R}$ ,

$$\text{prox}_{\varphi}\xi = \begin{cases} \text{sign}(\xi) \max\{|\xi| - \chi, 0\} & \text{if } p = 1 \\ \xi + \frac{4\chi}{3 \cdot 2^{1/3}} ((\epsilon - \xi)^{1/3} - (\epsilon + \xi)^{1/3}) & \text{where } \epsilon = \sqrt{\xi^2 + 256\chi^3/729} \text{ if } p = \frac{4}{3} \\ \xi + \frac{9\chi^2 \text{sign}(\xi)}{8} \left(1 - \sqrt{1 + \frac{16|\xi|}{9\chi^2}}\right) & \text{if } p = \frac{3}{2} \\ \frac{\xi}{1+2\chi} & \text{if } p = 2 \\ \text{sign}(\xi) \frac{\sqrt{1+12\chi|\xi|}-1}{6\chi} & \text{if } p = 3 \\ \left(\frac{\epsilon+\xi}{8\chi}\right)^{1/3} - \left(\frac{\epsilon-\xi}{8\chi}\right)^{1/3} & \text{where } \epsilon = \sqrt{\xi^2 + 1/(27\chi)} \text{ if } p = 4. \end{cases} \quad (67)$$

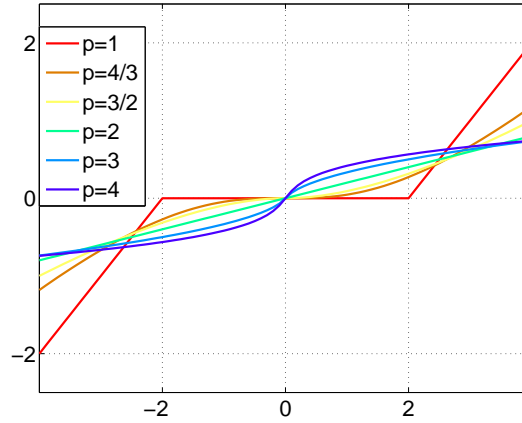


Figure 9: Illustration of  $\text{prox}_{\chi|\cdot|^p}$  with  $p \in \{1, 4/3, 3/2, 2, 3, 4\}$  and  $\chi = 2$ . The figure displays the proximity operator value as a function of its argument. This clearly shows that various behaviours are obtained depending on the value of  $p$ .

The previous formulas become useful in practice when they are combined with a result stated below that is applicable to functions which are separable in an orthonormal basis. This assumption is often valid when modeling prior information in an orthonormal wavelet transform or simply in a canonical basis.

**Property 5.2** [?, Proposition 2.10] Let  $h: \mathbb{R}^N \rightarrow ]-\infty, +\infty]: u \mapsto \sum_{k=1}^N \varphi_k(\langle u, e_k \rangle)$  where  $(e_k)_{1 \leq k \leq N}$  denotes an orthonormal basis of  $\mathbb{R}^N$  and  $(\varphi_k)_{1 \leq k \leq N}$  are convex, lower-semicontinuous, and proper functions from  $\mathbb{R}$  to  $]-\infty, +\infty]$ . Then

$$(\forall u \in \mathbb{R}^N) \quad \text{prox}_h u = \sum_{k=1}^N (\text{prox}_{\varphi_k} \langle u, e_k \rangle) e_k.$$

For example, if  $(e_k)_{1 \leq k \leq N}$  is the canonical basis of  $\mathbb{R}^N$ ,

$$(\forall u = (u_k)_{1 \leq k \leq N} \in \mathbb{R}^N) \quad \text{prox}_{\chi \|\cdot\|_1} u = (\text{prox}_{\chi|\cdot|} u_k)_{1 \leq k \leq N},$$

which combined with Example ?? with  $p = 1$  leads to (??). Similar arguments yield closed form expressions of  $\text{prox}_{\chi \|\cdot\|_p}$ , when  $p \in \{4/3, 3/2, 2, 3, 4\}$ .

Another proximity operator commonly encountered when dealing with group sparsity or with total variation regularizations is the operator associated with the 2-dimensional Euclidean norm, which can be generalized to higher dimension.

**Example 5.3** [?] Let  $\mu > 0$ , and let

$$\varphi: \mathbb{R}^2 \rightarrow \mathbb{R}: (\xi_1, \xi_2) \mapsto \mu \sqrt{|\xi_1|^2 + |\xi_2|^2}.$$

We have then, for every  $(\xi_1, \xi_2) \in \mathbb{R}^2$ ,

$$\text{prox}_\varphi(\xi_1, \xi_2) = \begin{cases} \left(1 - \frac{\mu}{\sqrt{|\xi_1|^2 + |\xi_2|^2}}\right)(\xi_1, \xi_2), & \text{if } \sqrt{|\xi_1|^2 + |\xi_2|^2} > \mu; \\ (0, 0), & \text{otherwise.} \end{cases} \quad (68)$$

## 5.2 Analysis formulation in the presence of Gaussian noise

We now focus on the resolution of the analysis counterpart of Problem (??) which consists of finding

$$\hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} \|Ay - z\|^2 + \chi \|Fy\|_1. \quad (69)$$

### 5.2.1 Forward-backward algorithm

When applied to (??), the form of the FB algorithm becomes:

Set  $\gamma_n \in \mathbb{R}$  such that  $\inf_{n \in \mathbb{N}} \gamma_n > 0$  and  $\sup_{n \in \mathbb{N}} \gamma_n < \frac{1}{\|A\|^2}$ .  
Set  $\lambda_n \in ]0, 1]$  such that  $\inf_{n \in \mathbb{N}} \lambda_n > 0$ .  
Initialize  $y^{[0]} \in \mathbb{R}^N$ .  
For  $n = 0, 1 \dots$   
     $y^{[n+1]} = y^{[n]} + \lambda_n (\text{prox}_{\gamma_n \chi \|F \cdot\|_1} (y^{[n]} - 2\gamma_n A^\top (Ay^{[n]} - z)) - y^{[n]})$

The main difficulty in the use of these iterations stems from the proximal computation step. Indeed, the proximity operator of a function composed with a linear operator has a closed form expression only in a few specific cases, as explained below.



**Property 5.4** [?, proposition 11] *Let  $\mathcal{H}$  and  $\mathcal{G}$  be two real Hilbert spaces of finite dimensions, let  $f$  be a convex, lower-semicontinuous, and proper function from  $\mathcal{G}$  to  $] -\infty, +\infty]$ , let  $T: \mathcal{H} \rightarrow \mathcal{G}$  be a linear operator and let  $T^*$  be its adjoint. We assume that  $TT^* = \tau \text{Id}$  with  $\tau \in ]0, +\infty[$ . Then,*

$$\text{prox}_{f \circ T} = \text{Id} + \chi^{-1} T^* \circ (\text{prox}_{\tau f} - \text{Id}) \circ T. \quad (70)$$

Coming back to the computation of  $\text{prox}_{\gamma n \chi \|F \cdot\|_1}$ , the application of the above formula requires the property  $FF^* = \tau \text{Id}$  to be satisfied, which is only guaranteed for orthonormal bases (with  $\tau = 1$ ). We recall that tight frames satisfy  $F^*F = \nu \text{Id}$  but  $FF^*$  is generally not equal to  $\nu \text{Id}$  (see (??)).

### 5.2.2 Douglas-Rachford algorithm

An alternative way for solving (??) consists of introducing an auxiliary variable  $u \in \mathbb{R}^K$  so that

$$\hat{y} \in \underset{(y,u) \in \mathbb{R}^N \times \mathbb{R}^K}{\text{Argmin}} \|Ay - z\|^2 + \chi \|u\|_1 \quad \text{subject to} \quad u = Fy, \quad (71)$$

or equivalently

$$\hat{y} \in \underset{(y,u) \in \mathbb{R}^N \times \mathbb{R}^K}{\text{Argmin}} \|Ay - z\|^2 + \chi \|u\|_1 + \iota_V(u, y), \quad (72)$$

where we have introduced the indicator function of the vector space

$$V = \{(y, u) \in \mathbb{R}^N \times \mathbb{R}^K \mid Fy - u = 0\}. \quad (73)$$

This criterion can be efficiently minimized with the so-called Douglas-Rachford algorithm [?, ?] that requires to compute the proximity operator of each term involved in the criterion. The proximity operator associated with the  $\ell^1$ -norm has been given in the previous section. The proximity operator of the data fidelity term takes a closed form expression that is

$$(\forall y \in \mathbb{R}^N) \quad \text{prox}_{\gamma \|A \cdot - z\|_2^2} y = (\text{Id} + 2\gamma A^\top A)^{-1} (y + 2\gamma A^\top z), \quad \gamma > 0,$$

where the inversion can be computed efficiently if  $A$  is a block-circulant operator (see section ??), and, by definition of the projection onto a vector space

$$(\forall (y, u) \in \mathbb{R}^N \times \mathbb{R}^K) \quad P_V(y, u) = (y, u) - (F^*v, -v) \quad \text{with} \quad v = (FF^* + \text{Id})^{-1}(Fy - u).$$

Following the matrix inversion lemma (Sherman-Morrison-Woodbury formula) and assuming that  $F$  models a tight frame, the inversion reduces to  $(FF^* + \text{Id})^{-1} = \text{Id} - \frac{1}{1+\nu} FF^*$ . The resulting Douglas-Rachford algorithm is given below:

Set  $\gamma \in ]0, +\infty[$ .  
Set  $\lambda_n \in \mathbb{R}$  such that  $\inf_{n \in \mathbb{N}} \lambda_n > 0$  and  $\sup_{n \in \mathbb{N}} \lambda_n < 2$ .  
Initialize  $y^{[0]} \in \mathbb{R}^N$  and  $u^{[0]} \in \mathbb{R}^K$   
For  $n = 0, 1 \dots$   

$$\begin{cases} y^{[n+\frac{1}{2}]} = (\text{Id} + 2\gamma A^\top A)^{-1} (y^{[n]} + 2\gamma A^\top z) \\ u^{[n+\frac{1}{2}]} = \text{prox}_{\gamma \chi \|\cdot\|_1} u^{[n]} \\ (y^{[n+1]}, u^{[n+1]}) = (y^{[n]}, u^{[n]}) + \lambda_n (P_V(2y^{[n+\frac{1}{2}}] - y^{[n]}, 2u^{[n+\frac{1}{2}]} - u^{[n]}) - (y^{[n+\frac{1}{2}}], u^{[n+\frac{1}{2}}])) \end{cases}$$

Under some technical assumptions, the sequence  $(y^{[n]})_{n \in \mathbb{N}}$  can be shown to converge to a solution to (??).

### 5.2.3 PPXA+

The trick of introducing an auxiliary variable is also exploited in related approaches such as the Alternating Direction Methods of Multipliers (ADMM) [?, ?, ?], also called Split Bregman iteration by some authors [?],<sup>5</sup> the Simultaneous Direction Method of Multipliers (SDMM) [?], or in the Parallel ProXimal Algorithm (PPXA+) [?, ?]. Such algorithms are instrumental in many recent works devoted to image restoration which involve tight frames and a block-circulant degradation matrix.

The common point in all these methods is that the inversions of  $A^\top A$  and  $F^* F$  are performed in a kind of joint manner. Here, we describe the iterations corresponding to PPXA+ [?], the spirit remains the same for the other aforementioned algorithms.

Set  $(\gamma_1, \gamma_2)$  in  $]0, +\infty[^2$ .  
Initialize  $v^{[0]} \in \mathbb{R}^M$ ,  $v^{[-1/2]} \in \mathbb{R}^M$ ,  $u^{[0]} \in \mathbb{R}^K$ , and  $u^{[-1/2]} \in \mathbb{R}^K$ .  
For  $n = 0, 1 \dots$

$$\left[ \begin{array}{l} y^{[n+1]} = (\gamma_1 A^\top A + \gamma_2 F^* F)^{-1} (\gamma_1 A^\top v^{[n]} + \gamma_2 F^* u^{[n]}) \\ v^{[n+\frac{1}{2}]} = v^{[n-\frac{1}{2}]} + A(2y^{[n+1]} - y^{[n]}) - v^{[n]} \\ u^{[n+\frac{1}{2}]} = u^{[n-\frac{1}{2}]} + F(2y^{[n+1]} - y^{[n]}) - u^{[n]} \\ v^{[n+1]} = \frac{1}{2+\gamma_1} (\gamma_1 v^{[n+\frac{1}{2}]} + 2z) \\ u^{[n+1]} = \text{prox}_{\frac{\lambda}{\gamma_2} \|\cdot\|_1} (u^{[n+\frac{1}{2}]}). \end{array} \right.$$

Under some technical assumptions, the sequence  $(y^{[n]})_{n \in \mathbb{N}}$  generated by PPXA+ converges to a solution to (??). In particular, the matrix  $\gamma_1 A^\top A + \gamma_2 F^* F$  needs to be invertible. Note that, similarly to the Douglas-Rachford algorithm, this inversion usually involves large size matrices and can thus be performed only for some specific choices of  $A$  and  $F$  (e.g. block-circulant degradation matrix and tight frame). Another interesting property of PPXA+ is that it can be employed for solving more sophisticated problems than (??), especially when hybrid regularizations are used [?]. In this case, the proximity operators of the different regularization terms can be computed in parallel.

### 5.2.4 Primal-dual schemes

For a wide array of image restoration problems,  $A$  does not model a block-circulant matrix, for instance in tomography applications. Moreover, the tight frame assumption may appear restrictive when overcomplete dictionaries are employed [?]. In order to perform image recovery in such a

---

<sup>5</sup>Note that ADMM can be viewed as an application of the Douglas-Rachford algorithm to the dual optimization problem.

general setting, recent algorithms in the image restoration literature have been proposed [?, ?, ?, ?, ?]. These algorithms are based on a primal-dual formulation and often make use of the FB algorithm for finding the associated saddle point. The iterations may thus combine gradient descent steps and proximal ones dealing with nonsmooth terms. The iterations derived from [?] are as follows:

Set  $(\sigma, \tau) \in ]0, +\infty[^2$  such that  $1/\tau - \sigma\|L\|^2 \geq \|A\|^2$   
Initialize  $y^{[0]} \in \mathbb{R}^M$  and  $u^{[0]} \in \mathbb{R}^K$ .  
For  $n = 0, 1 \dots$   

$$\begin{cases} y^{[n+1]} = y^{[n]} - 2\tau A^*(Ay^{[n]} - z) - \tau\sigma F^*u^{[n]} \\ v^{[n]} = u^{[n]} + F(2y^{[n+1]} - y^{[n]}) \\ u^{[n]} = v^{[n]} - \text{prox}_{\frac{\chi}{\sigma}\|\cdot\|_1}(v^{[n]}) \end{cases}$$

Once again, under some technical assumptions, the sequence  $(y^{[n]})_{n \in \mathbb{N}}$  converges to a solution to (??).

### 5.3 Additional constraints

In many practical situations, informations on the dynamic range of the pixel intensity values are available. The analysis formulation in the presence of data degraded by Gaussian noise becomes:

$$\text{Find } \hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} \|Ay - z\|^2 + \chi\|Fy\|_1 + \iota_C(y) \quad (74)$$

where  $C$  is a nonempty closed convex subset of  $\mathbb{R}^N$ . Incorporating such a constraint can be done cleverly by replacing  $\text{prox}_{h+\iota_C}$  with  $P_C \circ \text{prox}_h$  where  $h$  models either the data fidelity term or the regularization term. However, such an argument can be employed only in a few situations. We recall next the main conditions for the validity of this approach:

**Property 5.5** [?] *Let  $(e_k)_{1 \leq k \leq N}$  be an orthonormal basis of  $\mathbb{R}^N$  and let  $(\varphi_k)_{1 \leq k \leq N}$  be a family of convex, lower-semicontinuous, and proper functions from  $\mathbb{R}$  to  $] -\infty, +\infty]$ . Let  $h$  be a separable function such that*

$$h: \mathbb{R}^N \rightarrow ]-\infty, +\infty]: u \mapsto \sum_{k=1}^N \varphi_k(\langle u, e_k \rangle). \quad (75)$$

*Let  $C$  be a separable convex set such that*

$$C = \bigcap_{k \in \{1, \dots, N\}} \{u \in \mathbb{R}^N \mid \langle u, e_k \rangle \in C_k\} \quad (76)$$

*where  $(C_k)_{1 \leq k \leq N}$  are closed intervals of  $\mathbb{R}$  such that  $(\forall k \in \{1, \dots, N\}) C_k \cap \text{dom } \varphi_k \neq \emptyset$ . Then,*

$$(\forall u \in \mathbb{R}^N) \quad \text{prox}_{h+\iota_C} u = \sum_{k=1}^N \pi_k e_k \quad (77)$$

where

$$(\forall k \in \{1, \dots, N\}) \quad \pi_k = \begin{cases} \inf C_k & \text{if } \text{prox}_{\varphi_k} \langle u, e_k \rangle < \inf C_k \\ \sup C_k & \text{if } \text{prox}_{\varphi_k} \langle u, o_k \rangle > \sup C_k \\ \text{prox}_{\varphi_k} \langle u, e_k \rangle & \text{otherwise.} \end{cases} \quad (78)$$

For example, this property is applicable if  $C$  corresponds to a box constraint, and  $A$  is the identity matrix (denoising problem). In other cases, we need to resort to splitting algorithms designed for minimizing a sum of a finite number of functions such as PPXA+, SDMM or primal-dual algorithms. For more details regarding these algorithms, the readers should refer to [?, ?, ?, ?, ?].

#### 5.4 Frame-based strategies in the presence of Poisson noise

As described in Section ??, when the data are corrupted with Poisson noise, the criterion we aim at minimizing usually involves the generalized Kullback-Leibler divergence  $D_{\text{GKL}}$ . Here is the associated analysis  $\ell^1$ -minimization problem:

$$\text{Find } \hat{y} \in \underset{y \in \mathbb{R}^N}{\text{Argmin}} D_{\text{GKL}}(z, \alpha A y) + \chi \|F y\|_1 \quad (79)$$

where

$$D_{\text{GKL}}(z, \alpha A y) = \sum_{i=1}^M \psi_i((A y)_i)$$

and, for every  $i \in \{1, \dots, M\}$ ,

$$(\forall v \in \mathbb{R}) \quad \psi_i(v) = \begin{cases} \alpha v - z_i + z_i \ln \left( \frac{z_i}{\alpha v} \right) & \text{if } z_i > 0 \text{ and } v \in ]0, +\infty[, \\ \alpha v & \text{if } z_i = 0 \text{ and } v \in [0, +\infty[, \\ +\infty & \text{otherwise,} \end{cases}$$

with  $\chi > 0$ . The generalized Kullback-Leibler divergence is differentiable on its domain, but it does not have a Lipschitzian gradient. More precisely, for every  $i \in \mathbb{I} = \{i \in \{1, \dots, M\} \mid z_i > 0\}$ , the second-order derivative of  $\psi_i$  is given by

$$(\forall v \in ]0, +\infty[) \quad \psi_i''(v) = \frac{z_i}{v^2}, \quad (80)$$

and it is thus unbounded close to 0.

Since the FB algorithm requires the data fidelity term to be Lipschitz differentiable, it cannot be used for solving (??). In the literature, two classes of alternative solutions have been proposed to circumvent this difficulty. The first one consists of approximating the data fidelity terms with a Lipschitz differentiable function, while the second approach is based on the computation of the proximity operator associated with the Kullback-Leibler divergence.

#### 5.4.1 Variance stabilization transform

The fact that Poisson data have components with variance proportional to the target signal may account for the lack of Lipschitz differentiability of the gradient of their negative log-likelihood. Hence, one may look for transformations allowing the variance to be stabilized. The most common variance stabilization transform is the Anscombe transform [?] which is equivalent to replace the potential  $\psi_i$  ( $i \in \{1, \dots, N\}$ ) of the Poisson distribution by

$$(\forall v \in \mathbb{R}) \quad \tilde{\psi}_i(v) = \begin{cases} \frac{1}{2} \left( 2\sqrt{\alpha v + \frac{3}{8}} - z_i \right)^2 & \text{if } v \in [0, +\infty[ \\ +\infty & \text{otherwise.} \end{cases} \quad (81)$$

The resulting frame synthesis formulation reads

$$\text{Find } \hat{x} \in \underset{x \in \mathbb{R}^K}{\text{Argmin}} \sum_{i=1}^M \tilde{\psi}_i((AF^*x)_i) + \chi\|x\|_1 \quad (82)$$

where the data fidelity term is a convex function having a Lipschitzian gradient [?]. We can then refer to the previous section and employ the FB algorithm to build a sequence that converges to a solution to (??). When a frame analysis formulation is adopted, the primal-dual algorithm presented in Section ?? can be used in order to solve the associated optimization problem. Note that other choices for  $(\tilde{\psi}_i)_{1 \leq i \leq N}$  have been proposed in literature (see for instance [?]).

#### 5.4.2 Dealing with the exact Kullback-Leibler divergence

In [?], the authors noticed that the proximity operator associated with the negative log-likelihood of the Poisson distribution is related to the gamma distribution and they provided the expression of this operator. More precisely, for every  $\gamma > 0$  and  $i \in \{1, \dots, M\}$ , we have [?, Example 30]

$$(\forall v \in \mathbb{R}) \quad \text{prox}_{\gamma\psi_i} v = \frac{v - \gamma\alpha + \sqrt{|v - \gamma\alpha|^2 + 4\gamma z_i}}{2}. \quad (83)$$

It results that any optimization algorithm relying on the proximity operator of these functions can be employed. We now give the iterations corresponding to PPXA+ because of their simplicity. However, any primal-dual scheme can be employed in order to provide an algorithm working for any type of degradation operator  $A$  and frame  $F$ . The reader can refer to [?, ?, ?] for an overview of various algorithms reported in the literature for solving optimization problems involving the Kullback-Leibler divergence.

Set  $(\gamma_1, \gamma_2)$  in  $]0, +\infty[^2$ .  
Initialize  $v^{[0]} \in \mathbb{R}^M$ ,  $v^{[-1/2]} \in \mathbb{R}^M$ ,  $u^{[0]} \in \mathbb{R}^K$ , and  $u^{[-1/2]} \in \mathbb{R}^K$ .

For  $n = 0, 1 \dots$

$$\left[ \begin{array}{l} y^{[n+1]} = (\gamma_1 A^\top A + \gamma_2 F^* F)^{-1} (\gamma_1 A^\top v^{[n]} + \gamma_2 F^* u^{[n]}) \\ v^{[n+\frac{1}{2}]} = v^{[n-\frac{1}{2}]} + A(2y^{[n+1]} - y^{[n]}) - v^{[n]} \\ u^{[n+\frac{1}{2}]} = u^{[n-\frac{1}{2}]} + F(2y^{[n+1]} - y^{[n]}) - u^{[n]} \\ v^{[n+1]} = (\text{prox}_{\psi_i/\gamma_1}(v_i^{[n+\frac{1}{2}]})_{1 \leq i \leq M} \\ u^{[n+1]} = \text{prox}_{\frac{\chi}{\gamma_2} \|\cdot\|_1}(u^{[n+\frac{1}{2}]}). \end{array} \right.$$

## 6 Bayesian methods

The previous sections have been focused on the computation of an estimate of the unknown image by minimizing a criterion. From a Bayesian perspective, the solution corresponds to the Maximum A Posteriori estimate. Other Bayesian estimator may however be more relevant. In particular, much effort has been devoted to the posterior expectation, which is also known as the minimum mean square error (MMSE) estimate. For instance, in [?], the MMSE is computed for image denoising by employing a Gaussian Scale Mixture distribution to model local spatial neighborhoods of wavelet coefficients. This prior model is extended to multicomponent image denoising in [?]. In a similar manner, in [?], a MMSE is derived from a mixture of two truncated generalized Gaussian distributions. The MMSE is expressed by

$$\hat{y} = \int_{\mathbb{R}^N} y \mu_{\overline{Y}|Z=z}(y) dy \quad (84)$$

where the posterior probability density  $\mu_{\overline{Y}|Z=z}(y)$  can be expressed as follows (see Section 1.3.1):

$$\mu_{\overline{Y}|Z=z}(y) = \frac{\mu_{Z|\overline{Y}=y}(z) \mu_{\overline{Y}}(y)}{\mu_Z(z)}. \quad (85)$$

Here, the normalization constant  $\mu_Z(z)$  is given by

$$\mu_Z(z) = \int_{\mathbb{R}^N} \mu_{Z|\overline{Y}=y}(z) \mu_{\overline{Y}}(y) dy. \quad (86)$$

Nonetheless, when the signal dimensionality  $N$  is large, the integral in (??) as well as that in (??) are often computationally intractable. To tackle this problem, two types of methods are generally used: stochastic simulation methods and methods based on analytical approximations. The most important class of stochastic simulation methods consists of Markov Chain Monte Carlo (MCMC) methods [?, ?] which generate an asymptotically exact approximation of the posterior distribution through a set of samples. The statistical characteristics of unknown variables, such as their posterior expectation, can then be estimated by empirical means computed from these samples.

Even though MCMC methods can deliver accurate estimates, they often suffer from a high computational cost. Another class of approaches which are generally more efficient consists of determining analytical approximations of the true posterior distribution. These approximations are chosen to have simpler forms than the original distribution. Therefore, the mean values of the approximate distributions are much easier to compute and can be used to approximate the posterior expectation estimate. This section is devoted to present these two classes of approaches. For simplicity, our presentation will be focused on the synthesis formulation in the presence of Gaussian noise.

## 6.1 Bayesian formulation in the presence of Gaussian noise

In Section 5, the most common  $\ell^1$ -norm regularization based minimization problem has been addressed. This problem can be interpreted as computing the mode of the posterior distribution expressed as

$$(\forall x \in \mathbb{R}^K) \quad \mu_{\bar{X}|Z=z}(x) \propto \exp(-\varphi(x)) \quad (87)$$

with

$$\varphi(x) = \frac{\|AF^*x - z\|^2}{2\sigma^2} + \tilde{\chi}\|x\|_1. \quad (88)$$

Note that the normalization constant (partition function) of the distribution of the frame coefficients  $\mu_{\bar{X}|Z=z}$  has no closed form expression. We thus need to resort to numerical methods to compute the posterior expectation estimate.

## 6.2 Markov Chain Monte Carlo methods

The objective of stochastic simulation methods is to draw a set of samples from a target distribution  $\pi$ , which corresponds to  $\mu_{\bar{X}|Z=z}$  in our problem. Very frequently, the target distribution cannot be simulated directly. In such cases, Markov Chain Monte Carlo methods provide useful means of generating samples. The central idea of MCMC methods is to construct a Markov Chain with  $\pi$  as the stationary distribution, i.e. the limiting distribution of the Markov Chain. In the following, we briefly illustrate the principle of MCMC methods by presenting one of the most prevalent MCMC techniques: the Metropolis-Hastings (MH) algorithm.

The MH algorithm is an almost universal algorithm for generating Markov Chains with stationary distribution  $\pi$ . This algorithm has been firstly introduced by Metropolis et al. [?] and then generalized in a more statistical setting by Hastings [?]. The generic MH algorithm can be described as follows:

Initialize  $x^{[0]} \in \mathbb{R}^K$ .  
 For  $n = 0, 1 \dots$   
   Given  $x^{[n]}$ , generate  $x^*$  from a *proposal distribution*  $q(\cdot|x^{[n]})$   
   Determine  $x^{[n+1]}$  by an accept-reject step:  
     
$$x^{[n+1]} = \begin{cases} x^* & \text{with probability } \rho(x^{[n]}, x^*) \\ x^{[n]} & \text{with probability } 1 - \rho(x^{[n]}, x^*) \end{cases}$$
  
     where  
     
$$\rho(x^{[n]}, x^*) = \min \left( \frac{\pi(x^*)}{\pi(x^{[n]})} \frac{q(x^{[n]}|x^*)}{q(x^*|x^{[n]})}, 1 \right)$$

After a so-called burn-in period, the samples generated by MH algorithm can be averaged so as to provide an estimation of the mean of  $\pi$ .

By choosing different *proposal distributions*  $q(\cdot|x^{[n]})$ , different MH algorithms are obtained. For example, if  $q(\cdot|x^{[n]})$  is a Gaussian distribution centered at  $x^{[n]}$ , we obtain the random walk MH algorithm which is one of the most widely used MCMC method. In the case when  $q(\cdot|x^{[n]})$  does not depend on  $x^{[n]}$ , we get another special case: the independent MH algorithm. These algorithms are easily implementable and have low computational cost per iteration. However, they may require a large number of iterations to converge. Recently, several faster variants of MH algorithms have also been proposed, e.g. the Metropolis Adjusted Langevin Algorithms [?] and the Hamiltonian Monte Carlo algorithms [?]. The interested reader is referred to [?] for a general overview of MCMC methods.

As we can see from the above algorithm, MH algorithms do not require knowledge about the normalization constant of  $\pi$  since this constant cancels in the computation of the acceptance ratio  $\rho$ . Consequently, they can be used to draw samples from distributions with unknown partition functions which is the case of the posterior distribution  $\mu_{\bar{X}|Z=z}$ . For more details about sampling the posterior distribution with frame representations, the interested reader can refer to [?, ?, ?, ?].

### 6.3 Variational Bayesian approximation

Surrogates to the previous methods are provided by analytical approximation methods, especially the Variational Bayesian Approximation (VBA) techniques [?, ?], which establish a compromise between the approximation accuracy and the computational load. The central idea of VBA is to approximate the original posterior distribution  $\mu_{\bar{X}|Z=z}$  by a separable density:

$$(\forall x \in \mathbb{R}^K) \quad q_{\bar{X}}(x) = \prod_{\ell=1}^L q_{\bar{X}_\ell}(x_\ell) \quad (89)$$

where the components of  $\bar{X} = (\bar{X}_\ell)_{1 \leq \ell \leq L}$  have been partitioned into  $L$  disjoint subsets of elements.

This approximate density should be as close as possible to the original posterior distribution in the sense of minimizing a measure of dissimilarity between them. A natural choice for the



dissimilarity measure between two probability densities is the Kullback-Leibler ( $\mathcal{D}_{\text{KL}}$ ) divergence defined as

$$\mathcal{D}_{\text{KL}}(q_{\overline{X}}, \mu_{\overline{X}|Z=z}) = \int_{\mathbb{R}^K} q_{\overline{X}}(x) \log \frac{q_{\overline{X}}(x)}{\mu_{\overline{X}|Z=z}(x)} dx. \quad (90)$$

The  $\mathcal{D}_{\text{KL}}$  divergence is always non-negative and it vanishes only at  $q_{\overline{X}} = \mu_{\overline{X}|Z=z}$ .

The optimal separable density  $\hat{q}_{\overline{X}}$  can be obtained by solving

$$\hat{q}_{\overline{X}} \in \text{Argmin } \mathcal{D}_{\text{KL}}(q_{\overline{X}}, \mu_{\overline{X}|Z=z}) \quad \text{subject to } (??). \quad (91)$$

Note that this optimization problem is formulated in a functional space. An optimality condition for  $q_{\overline{X}}$  to be a solution to (??) is [?]:

$$(\forall \ell \in \{1, \dots, L\}) \quad q_{\overline{X}_\ell}(x_\ell) = \frac{1}{\kappa_\ell} \exp \left( \langle \log \mu_{\overline{X}, Z}(x, z) \rangle_{q_{(\overline{X}_i)_{i \neq \ell}}} \right) \quad (92)$$

where  $\kappa_\ell \in ]0, +\infty[$  is a normalization constant,  $\mu_{\overline{X}, Z}$  denotes the joint distribution which is equal to the product of the likelihood function and the prior distribution, and  $\langle \cdot \rangle_q$  represents the expected value with respect to the distribution  $q$ , e.g. here,  $\langle \ln \mu_{\overline{X}, Z}(x, z) \rangle_{q_{(\overline{X}_i)_{i \neq \ell}}} = \int \ln \mu_{\overline{X}, Z}(x, z) \prod_{i \neq \ell} (q_{\overline{X}_i}(x_i) dx_i)$ .

Due to the implicit relations existing between the densities  $(q_{\overline{X}_\ell})_{1 \leq \ell \leq L}$  in (??), an analytical expression of an optimal solution  $\hat{q}_{\overline{X}}$  does not exist except for extremely simple cases. In practice, these distributions are determined in an iterative way by following a block-alternating optimization strategy. This amounts to updating one of the separable components  $(q_{\overline{X}_\ell})_{1 \leq \ell \leq L}$  while fixing the others. In a generic fashion, classical VBA approaches can be summed up as follows:

Set initial distributions  $(q_{\overline{X}_\ell}^{[0]})_{1 \leq \ell \leq L}$ .  
 For  $n = 0, 1 \dots$   
   Determine  $q_{\overline{X}_1}^{[n+1]}$  using (??)  
    $\vdots$   
   Determine  $q_{\overline{X}_L}^{[n+1]}$  using (??).

Nonetheless, the above procedure cannot be easily implemented since the direct optimization of non parametric distributions is not a trivial task. Due to this fact, conjugate priors are generally needed in practice. A prior distribution is said to be conjugate to the likelihood if the corresponding posterior distribution belongs to the same family of distributions as the prior one. The use of conjugate distributions ensures that the optimal approximate distributions belong to the same known families of distributions as the prior ones, e.g. the Gaussian family, the gamma family,... In such cases, the optimization of distributions can be performed by optimizing their parameters, which leads to an easy block-coordinate descent algorithm, similarly to the previous one. Recently, more efficient algorithms have been proposed based on a gradient descent-like method and subspace optimization techniques (see [?, ?] for more details).

Let us now turn our attention to the application of VBA to the posterior distribution given by (??). A difficulty in this context stems from the fact that the Laplace prior distribution on  $\bar{X}$  (corresponding to an  $\ell^1$ -norm regularization term) is not conjugate to the likelihood which is Gaussian. A common way of alleviating this problem is to resort to the hierarchical representation of the Laplace prior [?, ?]. More specifically, for every  $k \in \{1, \dots, K\}$ , the  $k$ -th component  $\bar{X}_k$  of  $\bar{X}$  can be viewed as a zero-mean Gaussian variable with variance  $U_k$ , this hidden random variable following an exponential distribution as specified below:

$$(\forall x_k \in \mathbb{R}) \quad \mu_{\bar{X}_k|U_k=u_k}(x_k) = \frac{1}{\sqrt{2\pi u_k}} \exp\left(-\frac{x_k^2}{2u_k}\right) \quad (93)$$

$$(\forall u_k \in ]0, +\infty[) \quad \mu_{U_k}(u_k) = \frac{\tilde{\chi}^2}{2} \exp\left(-\frac{\tilde{\chi}^2}{2} u_k\right). \quad (94)$$

With this hierarchical representation, using the Bayes theorem, we can express the joint posterior distribution of the unknown vector  $\bar{X}$  and the hidden variable  $U = (U_1, \dots, U_K)^\top$ :

$$\begin{aligned} (\forall x \in \mathbb{R}^K)(\forall u \in ]0, +\infty[^K) \\ \mu_{\bar{X}, U|Z=z}(x, u) &\propto \mu_{Z|\bar{X}=x}(z) \prod_{k=1}^K \mu_{\bar{X}_k|U_k=u_k}(x_k) \mu_{U_k}(u_k) \\ &\propto \frac{1}{(2\pi\sigma^2)^{\frac{M}{2}}} \exp\left(-\frac{\|AF^*x - z\|^2}{2\sigma^2}\right) \prod_{k=1}^K \frac{1}{\sqrt{2\pi u_k}} \exp\left(-\frac{x_k^2}{2u_k}\right) \frac{\tilde{\chi}^2}{2} \exp\left(-\frac{\tilde{\chi}^2}{2} u_k\right). \end{aligned} \quad (95)$$

This posterior distribution is only known up to a normalization constant, but we can approximate this distribution by using VBA approaches. To apply VBA, the first step consists of making more precise the separability hypothesis. The chosen separability hypothesis should make the approximate solution given by (??) tractable. For example, we can here consider a fully separable form:

$$q_{\bar{X}, U}(x, u) = \prod_{k=1}^K q_{\bar{X}_k}(x_k) q_{U_k}(u_k). \quad (96)$$

Then, we apply the iterative VBA optimization procedure described above. Considering the full separability hypothesis, we thus alternate between determinations of  $(q_{\bar{X}_k})_{1 \leq k \leq K}$  and  $(q_{U_k})_{1 \leq k \leq K}$ .

- Determination of  $(q_{\bar{X}_k})_{1 \leq k \leq K}$ :

For every  $k \in \{1, \dots, K\}$ , since  $\mu_{\bar{X}_k|U_k=u_k}$  is a Gaussian probability density function which is conjugate to the likelihood  $\mu_{Z|\bar{X}}$ , (??) leads to the fact that  $q_{\bar{X}_k}$  is a Gaussian distribution with mean  $\zeta_k$  and variance  $v_k$  expressed as

$$v_k = \left( \frac{(FA^\top AF^*)_{k,k}}{\sigma^2} + \langle u_k^{-1} \rangle_{q_{U_k}} \right)^{-1}, \quad (97)$$

$$\zeta_k = \frac{v_k}{\sigma^2} \left( (FA^\top z)_k - \sum_{i \neq k} (FA^\top AF^*)_{k,i} \zeta_i \right), \quad (98)$$

where  $\langle u_k^{-1} \rangle_{q_{U_k}}$  represents the expectation of  $U_k^{-1}$  with respect to  $q_{U_k}$ .

- Determination of  $(q_{U_k})_{1 \leq k \leq K}$ :

By using (??), for every  $k \in \{1, \dots, K\}$ , we can obtain:

$$(\forall u_k \in ]0, +\infty[) \quad q_{U_k}(u_k) \propto \frac{1}{\sqrt{u_k}} \exp \left( -\frac{\tilde{\chi}^2}{2} u_k - \frac{\langle x_k^2 \rangle_{q_{\bar{X}_k}}}{2u_k} \right), \quad (99)$$

where  $\langle x_k^2 \rangle_{q_{\bar{X}_k}} = \zeta_k^2 + v_k$ . Therefore,  $q_{U_k}$  is a generalized inverse Gaussian distribution. As shown in (??) and (??), the determination of the parameters of  $q_{\bar{X}_k}$  requires the knowledge of the expectation  $\langle u_k^{-1} \rangle_{q_{U_k}}$ . By using properties of generalized inverse Gaussian distributions [?], the expectation  $\langle u_k^{-1} \rangle_{q_{U_k}}$  is given by

$$\langle u_k^{-1} \rangle_{q_{U_k}} = \frac{\tilde{\chi}}{\sqrt{\zeta_k^2 + v_k}}. \quad (100)$$

Alternatively, this formula can be derived by using the hierarchical representation of the Laplace distribution (details can be found in [?]).

As a result, the obtained algorithm alternates between the iterative computation of  $(v_k, \zeta_k)_{1 \leq k \leq K}$  by using (??) and (??) and the determination of the expectation of  $(U_k^{-1})_{1 \leq k \leq K}$  by using (??). In the first step, the variables  $(\langle u_k^{-1} \rangle_{q_{U_k}})_{1 \leq k \leq K}$  computed at the previous iterations are used, which are then updated from the modified values of  $(v_k, \zeta_k)_{1 \leq k \leq K}$ . The procedure can be summed up as follows:

```

Set initial values  $(\zeta_k^{[0]})_{1 \leq k \leq K}$ ,  $(v_k^{[0]})_{1 \leq k \leq K}$  and  $(\langle u_k^{-1} \rangle_{q_{U_k}}^{[0]})_{1 \leq k \leq K}$ 
For  $n = 0, 1 \dots$ 
    For  $k = 1, \dots, K$ 
        Determine  $v_k^{[n+1]}$  and  $\zeta_k^{[n+1]}$  using (??) and (??)
    For  $j = 1, \dots, K$ 
        Determine  $\langle u_k^{-1} \rangle_{q_{U_k}}^{[n+1]}$  using (??).

```

Finally, the posterior expectation estimate can be approximated by  $\zeta = (\zeta_1, \dots, \zeta_K)^\top$  generated at convergence. We show in Fig. ?? a reconstruction result obtained by a VBA based reconstruction approach.

In VBA techniques, separable distributions are adopted to approximate the original posterior distribution. In the literature, there exists another type of approaches which compute approximate marginal posterior distributions by minimizing the Bethe free energy under milder assumptions. The interested reader can refer to [?, ?] for more details on such approaches.

Finally, it is important to note that one of the main advantages of Bayesian approaches is that they allow hyperparameters, e.g.  $\tilde{\chi}$ , to be estimated, by considering them as part of the vector of unknown variables and introducing some (possibly non informative) prior on them.

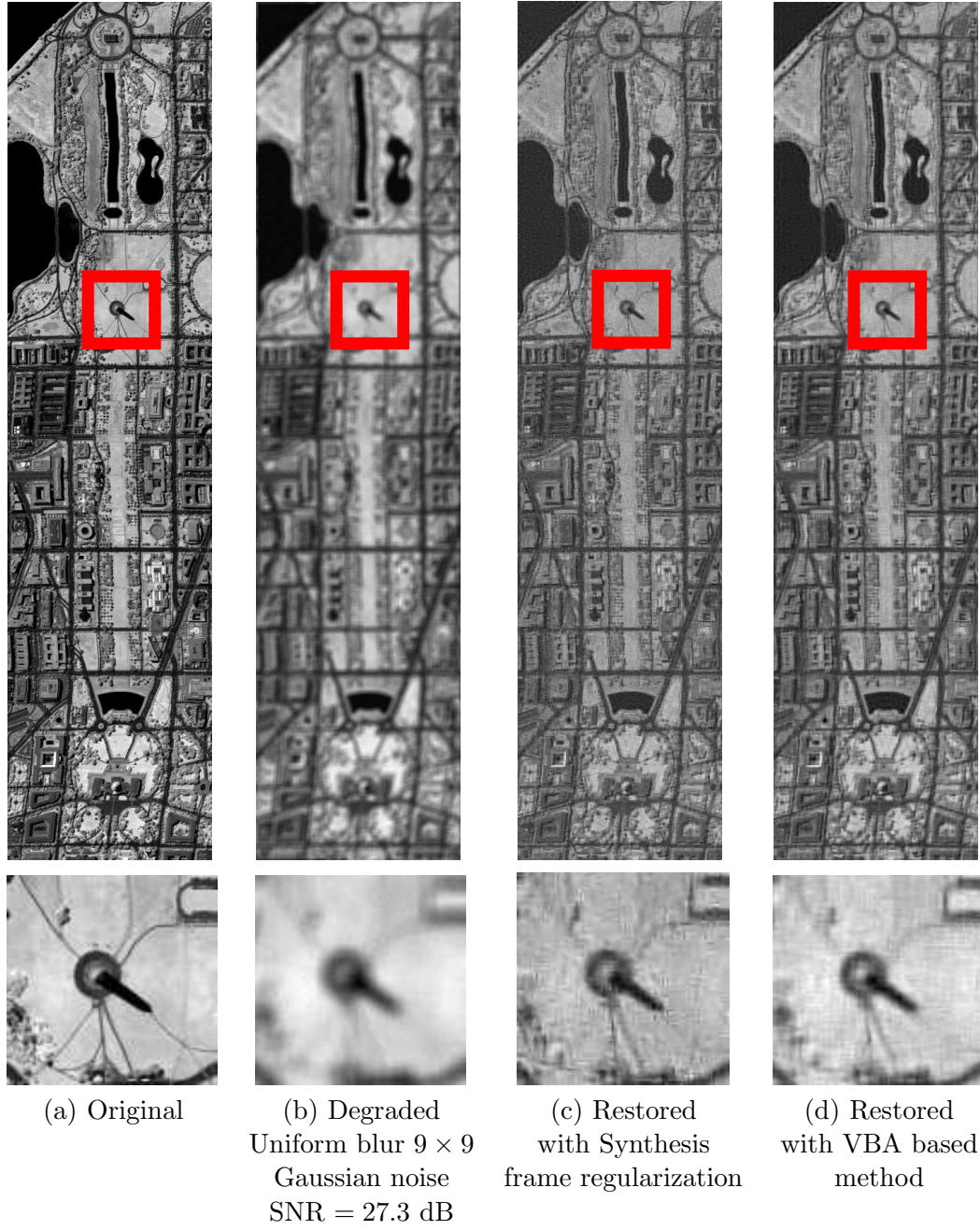


Figure 10: Illustration of image restoration with a VBA based algorithm using Laplace prior distribution on the coefficients of a dual-tree wavelet decomposition with *symmlets* of order 3 over 3 resolution levels.

## 7 Stein-based approaches

### 7.1 Denoising methods

#### 7.1.1 Minimum mean square error

Different possible optimality criteria can be considered for evaluating the quality of a recovery method. As mentioned in Section ??, a customary one is the Mean Square Error (MSE). In this case, two alternative measures should be distinguished. The first one consists of measuring the MSE in the spatial domain:

$$\text{MSE}_{\text{im}} = \mathbb{E}[\|\bar{y} - \hat{y}\|^2]. \quad (101)$$

In the second strategy, the MSE is expressed in the transform domain:

$$\text{MSE}_{\text{tr}} = \mathbb{E}[\|\bar{x} - \hat{x}\|^2] \quad (102)$$

where  $\bar{x}$  (resp.  $\hat{x}$ ) is the vector of wavelet/frame coefficients of  $\bar{y}$  (resp.  $\hat{y}$ ). As mentioned in [?],  $\text{MSE}_{\text{im}}$  reduces to  $\text{MSE}_{\text{tr}}$  only if the decomposition is performed onto an orthonormal basis. Consequently, in the general case of a frame decomposition, the estimate that optimizes  $\text{MSE}_{\text{tr}}$  does not necessarily minimize  $\text{MSE}_{\text{im}}$ . In other words, methods that aim at minimizing  $\text{MSE}_{\text{tr}}$  in non-orthonormal frame representations are suboptimal. Hence, it is mandatory to distinguish between the reconstruction in an orthonormal basis and the reconstruction in a non-orthonormal basis or a redundant frame.

Furthermore, in a stochastic setting, it is easy to check that the estimate that minimizes  $\text{MSE}_{\text{tr}}$  is the conditional posterior mean:

$$\hat{x} = \mathbb{E}[\bar{x}|x_z] \quad (103)$$

where  $x_z$  is the vector of frame coefficients related to the observation vector  $z$ . However, in practice, it is difficult to calculate  $\hat{x}$  since in general, there is no explicit expression of the posterior distribution and the computation of the related integral may also be tedious.

Moreover, from the following expressions of these quadratic risks:

$$\text{MSE}_{\text{im}} = \mathbb{E}[\|\bar{y}\|^2] + \mathbb{E}[\|\hat{y}\|^2] - 2\mathbb{E}[\bar{y}^\top \hat{y}] \quad (104)$$

and

$$\text{MSE}_{\text{tr}} = \mathbb{E}[\|\bar{x}\|^2] + \mathbb{E}[\|\hat{x}\|^2] - 2\mathbb{E}[\bar{x}^\top \hat{x}], \quad (105)$$

it can be observed that the terms  $\mathbb{E}[\bar{y}^\top \hat{y}]$  and  $\mathbb{E}[\bar{x}^\top \hat{x}]$  play a prevalent role. However, a main difficulty arising in the minimization of either  $\text{MSE}_{\text{im}}$  or  $\text{MSE}_{\text{tr}}$  is that these terms depend on  $\bar{y}$  (resp.  $\bar{x}$ ), which are unknown.

In the case of the observation model associated to (??) with an additive Gaussian noise, this difficulty can be circumvented by resorting to an estimator  $\mathcal{E}_{\text{im}}$  of  $\text{MSE}_{\text{im}}$ , which is computed only from the observed data  $z$ . In a similar way, building a statistical estimate  $\mathcal{E}_{\text{tr}}$  of  $\text{MSE}_{\text{tr}}$  also is an appealing solution. These estimations are made possible thanks to a result by Stein [?] leading to an *unbiased estimator* of the risk under quite general conditions.

### 7.1.2 Stein's estimate of $\text{MSE}_{\text{tr}}$ and $\text{MSE}_{\text{im}}$

We assume that the vector of  $K$  coefficients  $x_z$  is partitioned into  $L$  subvectors  $(x_z)_1, \dots, (x_z)_L$ . For each  $\ell \in \{1, \dots, L\}$ ,  $\mathbb{K}_\ell$  denotes the set of indexes of the components of  $(x_z)_\ell$  and  $K_\ell = \text{card}[\mathbb{K}_\ell]$  is its cardinality.. The overall number of frame coefficients is thus  $K = \sum_{\ell=1}^L K_\ell$ . A similar partitioning is also adopted for  $\bar{x}$ . A basic example of such a subvector in the case of a wavelet decomposition is a subband of coefficients at a given scale and orientation. For color or multispectral images, the coefficients in the same spectral channel in the same subband may be gathered in the same subset. Furthermore, an estimating function  $\Theta_\ell$  is applied to the components in each subset  $\mathbb{K}_\ell$  with  $\ell \in \{1, \dots, L\}$ . In this case, the quadratic risk  $\text{MSE}_{\text{tr}}$  can be expressed as the sum of  $L$  partial MSE associated with each subset  $\mathbb{K}_\ell$ , i.e.,

$$\text{MSE}_{\text{tr}} = \sum_{\ell=1}^L \mathbb{E}[\|(\bar{x})_\ell - (\hat{x})_\ell\|^2] \quad (106)$$

where, for every  $\ell \in \{1, \dots, L\}$ ,  $(\hat{x})_\ell = \Theta_\ell((x_z)_\ell)$ .

In order to find an estimate of the MSE, some technical assumptions are required. More precisely, it is assumed that each estimating function  $\Theta_\ell$  is a continuous and almost everywhere differentiable function satisfying the following technical assumptions:<sup>6</sup>

1. For every  $u \in \mathbb{R}^{K_\ell}$ ,

$$\lim_{\|w\| \rightarrow +\infty} \Theta_\ell(u + w) \exp(-w^\top \mathbf{\Gamma}_\ell^{-1} w) = 0, \quad (107)$$

where  $\mathbf{\Gamma}_\ell$  is the covariance matrix of the Gaussian noise coefficient vector in subset  $\ell$ ;

2.  $\mathbb{E}[\|\Theta_\ell((x_z)_\ell)\|^2] < +\infty$  and  $\mathbb{E}[\|\nabla \Theta_\ell((x_z)_\ell)\|_1] < +\infty$ , where  $\nabla \Theta_\ell((x_z)_\ell)$  is the Jacobian matrix of the estimating function at  $(x_z)_\ell$ . The Jacobian matrix is defined as

$$(\forall u_\ell \in \mathbb{R}^{K_\ell}) \quad \nabla \Theta_\ell(u_\ell) = \frac{\partial \Theta_\ell(u_\ell)}{\partial u_\ell^\top}. \quad (108)$$

Then, Stein's formula [?] for an additive Gaussian noise allows us to get an expression of the risk which no longer depends on the original image:

$$\mathbb{E}[(\hat{x})_\ell (\bar{x})_\ell^\top] = \mathbb{E}[(\hat{x})_\ell (x_z)_\ell^\top] - \mathbb{E}[\nabla \Theta_\ell((x_z)_\ell)] \mathbf{\Gamma}_\ell. \quad (109)$$

Consequently, by dropping the mathematical expectation in (??) and by noting that  $\mathbb{E}[(\bar{x})_\ell^\top (\hat{x})_\ell] = \text{tr}(\mathbb{E}[(\bar{x})_\ell (\hat{x})_\ell^\top])$ , an unbiased estimate  $\mathcal{E}_{\text{tr}}$  of  $\text{MSE}_{\text{tr}}$  can be derived [?]:

$$\mathcal{E}_{\text{tr}} = \sum_{\ell=1}^L \|(\hat{x})_\ell - (x_z)_\ell\|^2 - \text{tr}(\mathbf{\Gamma}_\ell) + 2\text{tr}(\nabla \Theta_\ell((x_z)_\ell) \mathbf{\Gamma}_\ell). \quad (110)$$

---

<sup>6</sup>More general weak differentiability assumptions can be made [?].

Under the same assumptions, it is also possible to define an unbiased estimate  $\mathcal{E}_{\text{im}}$  of  $\text{MSE}_{\text{im}}$  thanks to Stein's principle [?]. By assuming that the noise is spatially white with zero-mean and variance  $\alpha$ , we get

$$\mathcal{E}_{\text{im}} = \|\hat{y} - z\|^2 - N\alpha + 2 \sum_{\ell=1}^L \text{tr}(\nabla \Theta_{\ell}((x_z)_{\ell}) \tilde{\Gamma}_{\ell}), \quad (111)$$

where, for every  $\ell \in \{1, \dots, L\}$ ,  $\tilde{\Gamma}_{\ell}$  is the noise cross-correlation matrix given by

$$\tilde{\Gamma}_{\ell} = \alpha(\langle e_k, \tilde{e}_{k'} \rangle)_{(k,k') \in \mathbb{K}_{\ell}^2} \quad (112)$$

where  $(e_k)_{1 \leq k \leq K}$  (resp.  $(\tilde{e}_k)_{1 \leq k \leq K}$ ) designate the elements of the frame employed to decompose (resp. reconstruct) the image (for example, the latter frame may be the dual frame of the former as explained in Section ??). Since, for every  $\ell \in \{1, \dots, L\}$ ,

$$\Gamma_{\ell} = \alpha(\langle e_k, e_{k'} \rangle)_{(k,k') \in \mathbb{K}_{\ell}^2}, \quad (113)$$

we recover that  $\mathcal{E}_{\text{tr}} = \mathcal{E}_{\text{im}}$  when an orthonormal decomposition is employed.

It is often assumed that, for every  $\ell \in \{1, \dots, L\}$ , each estimating function  $\Theta_{\ell}$  is parametrized by a vector parameter  $\theta_{\ell}$ . Consequently, the problem reduces to find the value of the vector  $\theta_{\ell}$  that minimizes the estimate of the quadratic risk. In what follows, we will describe the most relevant denoising functions reported until now. As mentioned before, we will distinguish between the noise removal in an orthonormal transform domain and the one in a more general frame.

### 7.1.3 Reconstruction in an orthonormal transform domain

Pioneering noise reduction methods operating in an orthonormal WT have been developed by Donoho and Johnstone [?]. They have first proposed the *VisuShrink* method, which exploits the fact that the coefficients  $\bar{x}$  are sparse, unlike those of the noise. Indeed, it has been proved that, with a high probability, the maximum values of any set of  $N$  independent and identically distributed zero-mean Gaussian random variables with variance  $\alpha$  (namely, the noise coefficients) do not exceed the “universal” threshold  $\eta_{\text{univ}} = \sqrt{2\alpha \log(N)}$  for large values of  $N$  [?]. Obviously, a very simple way to discard the noise coefficients is to shrink the coefficients  $((x_z)_{\ell,k})_{1 \leq k \leq K_{\ell}}$  of vector  $(x_z)_{\ell}$  in each subband of index  $\ell$ . To this end, a hard-thresholding can be employed as a shrinkage operator:

$$(\forall \ell \in \{1, \dots, K_{\ell}\}) \quad (\hat{x}_{\text{hard}})_{\ell,k} = \begin{cases} (x_z)_{\ell,k} & \text{if } |(x_z)_{\ell,k}| > \eta_{\ell} \\ 0 & \text{otherwise} \end{cases} \quad (114)$$

or a soft-thresholding can be used:

$$\begin{aligned} (\forall \ell \in \{1, \dots, K_{\ell}\}) \quad (\hat{x}_{\text{soft}})_{\ell,k} &= \text{sign}((x_z)_{\ell,k}) \max\{|(x_z)_{\ell,k}| - \eta_{\ell}, 0\} \\ &= \text{prox}_{\eta_{\ell}|\cdot|}((x_z)_{\ell,k}), \end{aligned} \quad (115)$$

where  $\eta_{\ell} = \eta_{\text{univ}}$ . The soft-thresholding may be preferred to the hard one since it is continuous. Indeed, hard-thresholding can be sensitive to small changes in the data due to its discontinuity. In

practice, the achieved results are not always acceptable: in particular, Gibbs-like phenomena may appear in the vicinity of discontinuities due to the high value of the universal threshold. Actually, this flaw can be explained by the fact that the shrinkage rule is near-optimal in terms of a minimax strategy, but not in terms of the MSE. To remedy this drawback of VisuShrink, several solutions have been investigated. For instance, in [?, ?], a hybrid scheme was reported where the thresholded coefficients are employed to design a wavelet-based Wiener filtering in each subband. Recently, another strategy was described in [?]. More precisely, the fact that Stein's lemma does not apply to the hard-thresholding estimator as it is a discontinuous function, motivated the design of a consistent risk estimate called Stein CONSistent Risk Estimator (SCORE). Even it is biased, it allows the threshold to be automatically adjusted. Nevertheless, the most well-known alternative to VisuShrink was presented in [?], where the threshold value  $\eta_\ell$  is adjusted for each subband index  $\ell \in \{1, \dots, L\}$  so as to minimize the partial Stein's Unbiased Risk Estimator (SURE):

$$\sum_{\ell=1}^{K_\ell} \mathbb{E}[(\widehat{x}_{\text{soft}})_{\ell,k} - (x_z)_{\ell,k}]^2 + 2\alpha \mathbb{E}[(\widehat{x}_{\text{soft}})'_{\ell,k}].$$

where  $(\widehat{x}_{\text{soft}})'_{\ell,k}$  designates the derivative of the soft-thresholding operator calculated at  $(x_z)_{\ell,k}$  (which is defined everywhere except at  $\pm\eta_\ell$ ). The resulting denoising method was coined *SUREShrink*. It has been noticed that when the noise level is very high, the noise coefficients have a dominant contribution to the partial risk, especially when the original signal is very sparse. In this case, the estimation of the partial risk is not reliable and, it is preferred to resort to a robust estimate within a hybrid scheme [?]. More precisely, when an estimate of the energy of the clean data is below a given threshold, the risk estimate is judged as unreliable and VisuShrink is employed instead of SUREShrink for the concerned subband. It is also worth pointing out that Stein-based risk estimates can be employed to search the best data representation (wavelet packets or local cosine basis) [?].

Moreover, the weak differentiability of the estimate required by Stein's principle has enabled several families of parameterized estimating functions to be considered. For instance, for processing multicomponent images such as multispectral images, in [?], a Bayesian framework involving Bernoulli-Gaussian priors was initially adopted to build a multivariate estimate that takes into account the spectral cross-component similarities. However, as the values of the hyperparameters result from statistical inference exploiting a prior model, they may be suboptimal. An alternative solution thus consists of searching for the parameters that minimize the Stein-based estimate of the quadratic risk.

#### 7.1.4 Reconstruction in a redundant representation

A substantial gain in noise removal may be achieved by using overcomplete representations [?, ?]. For instance, the use of a translation invariant wavelet transform [?, ?] leads to a significant reduction of Gibbs artifacts [?] even when the risk estimate  $\mathcal{E}_{\text{tr}}$  is minimized instead of the spatial domain one  $\mathcal{E}_{\text{im}}$ . This behaviour can be explained as follows: by considering a large number of translated versions of the wavelet basis functions, it is possible to reduce the mismatch between their features and those of the signal of interest. In [?], Raphan and Simoncelli have investigated this issue for more general frame representations. For the sake of clarity, we present their rationale by only focusing on the case of tight analysis frames corresponding to the union of  $\nu > 1$



orthonormal bases of the same type, where the basis functions correspond to  $\nu$  shifted versions of those corresponding to a given orthonormal basis. As in the previous section, let us assume that these representations are structured in subbands  $(\mathbb{K}_\ell)_{1 \leq \ell \leq L}$ , where every subband  $(\mathbb{K}_{o,\ell})_{1 \leq \ell \leq L}$  in the overcomplete representation has a size  $\nu$  times larger than that of the related one generated by an orthonormal decomposition ( $K = \nu N$ ). We will also assume that the coefficients in each basis have the same marginal statistics for a given subband. Therefore, if a pointwise estimating function  $\vartheta_\ell: \mathbb{R} \rightarrow \mathbb{R}$  is applied to each coefficient in subband  $m$  of the orthonormal representations, it may appear natural to choose the same estimating function for the frame representation. Then, if  $\text{MSE}_{\text{tr},o}$  and  $\text{MSE}_{\text{tr}}$  denote the mean square errors involved in the non-redundant and the overcomplete representations, we have

$$\begin{aligned} \text{MSE}_{\text{tr},o} &= \sum_{\ell=1}^L \sum_{k=1}^{\text{card}[\mathbb{K}_{o,\ell}]} \mathbb{E} \left[ (x_{o,\ell,k} - \vartheta_\ell((x_z)_{\ell,k}))^2 \right] \\ &= \nu^{-1} \sum_{\ell=1}^L \sum_{k=1}^{K_\ell} \mathbb{E} \left[ (x_k - \vartheta_\ell((x_z)_{\ell,k}))^2 \right] \\ &= \nu^{-1} \text{MSE}_{\text{tr}} \end{aligned} \quad (116)$$

since, for every  $\ell \in \{1, \dots, L\}$ ,  $K_\ell = \nu \text{card}[\mathbb{K}_{o,\ell}]$ . Furthermore, by considering the dual synthesis frame which is also tight, is easy to show that

$$\|\bar{y} - \hat{y}\|^2 \leq \nu^{-1} \sum_{\ell=1}^L \|(\bar{x})_\ell - \hat{x}_\ell\|^2. \quad (117)$$

This implies that

$$\text{MSE}_{\text{im}} \leq \nu^{-1} \text{MSE}_{\text{tr}} = \text{MSE}_{\text{tr},o}. \quad (118)$$

Since  $\text{MSE}_{\text{tr},o} = \text{MSE}_{\text{im},o}$ , it can be deduced that

$$\text{MSE}_{\text{im}} \leq \text{MSE}_{\text{im},o}. \quad (119)$$

This equation reflects the benefit which can be drawn from removing the noise in the transform domain by using a redundant frame representation. This advantage has motivated the development of several efficient noise removal methods.

For instance, in [?], the chosen structure for the estimator corresponds to a linear combination of nonlinear elementary functions for which the optimization of  $\mathcal{E}_{\text{im}}$  can be easily handled. Note that such an idea was already formulated in the case of orthonormal basis representations in [?]. More precisely, in each subband  $\ell$ , the estimator is a Linear Expansion of  $I_\ell \in \mathbb{N}^*$  Thresholding (LET) functions  $\vartheta_{\ell,i}$  applied to the components of  $(x_z)_\ell$ :

$$(\forall k \in \{1, \dots, K_\ell\}) \quad \hat{x}_{\ell,k} = \sum_{i=1}^{I_\ell} a_{\ell,i} \vartheta_{\ell,i}((x_z)_{\ell,k}) \quad (120)$$

where  $(a_{\ell,i})_{1 \leq i \leq I_\ell}$  are weighting factors. The functions  $(\vartheta_{\ell,i})_{1 \leq i \leq I_\ell}$  must be weakly differentiable. Usually, only a small number  $I_\ell$  of them is necessary and their forms are selected so as to generate estimators ranging from a linear function to a smooth approximation of a hard-threshold by varying the values of the weighting factors. In [?], both orthonormal and redundant representations were considered with a linear expansion of two functions ( $I_\ell = 2$ ). In a similar way, Raphan and

Simoncelli have employed linear expansions of 4 localized bumps [?]. In the image domain, the denoised image is expressed as follows:

$$\hat{y} = \sum_{\ell=1}^L \sum_{i=1}^{I_\ell} a_{\ell,i} \beta_{\ell,i} \quad (121)$$

where, for every  $\ell \in \{1, \dots, L\}$  and  $i \in \mathbb{K}_\ell$ ,

$$\beta_{\ell,i} = \sum_{k=1}^{L_\ell} \vartheta_{\ell,i}((x_z)_{\ell,k}) \tilde{e}_k \quad (122)$$

can be viewed as a reconstruction of the  $m$ -th subband information using the estimating function  $\vartheta_{m,i}$ . The estimation problem then amounts to the search of the values of the weights  $a_{\ell,i}$  that minimize  $\mathcal{E}_{\text{im}}$ . Because of the linear dependence of the estimator on these parameters, this optimization reduces to the resolution of a linear system of equations.

The approach was improved in [?] by exploiting the similarities existing between the coefficients of the same orientation at different scales. To this end, a predictor is first designed to approximate the magnitude of the children wavelet coefficients (in the next finer scale) of any wavelet coefficient. Then, this predictor serves to separate noisy coefficients between weakly and highly affected ones in the denoising procedure. It was observed that the interscale version of SURE-LET outperforms conventional interscale methods.

Extensions of SURE-LET to the case of multicomponent images using non redundant representation [?] or overcomplete ones [?] were developed, yielding significant gains compared with state-of-the-art redundant componentwise approaches.

Note that most of the aforementioned denoising methods proceed in a term-by-term manner. Another improvement can be achieved by accounting for the spatial dependencies between the coefficients. Then, the coefficients are not processed individually, but blockwise. The idea of a joint spatial denoising of the coefficients has been firstly explored in both non-Bayesian and Bayesian cases [?, ?]. A further improvement arises by choosing the block size and the threshold value at each resolution level by minimizing the SURE [?]. In the case of multispectral images, a wide class of parameterized estimators operating on hybrid (spectral and spatial) neighborhoods of overcomplete representations have been employed to capture simultaneously the cross-component and spatial dependencies. The parameters related to the estimators are set to minimize  $\mathcal{E}_{\text{tr}}$  by exploiting Stein's principle [?, ?].

Note that the computation of the derivative of the estimator involved in SURE may be a difficult task if the chosen estimator has no explicit form as is the case with iterative approaches (variational, PDE-based, and Bayesian methods). To extend the applicability of SURE-based approaches, a Monte-Carlo method viewed as a black-box numerical scheme has been introduced in [?].

So far, Stein's formula is restricted to the case of a spatially white Gaussian noise. Eldar has derived a generalization of SURE (GSURE) for multivariate exponential noise distributions [?]. Moreover, specific SURE-like denoising methods have been developed for non-Gaussian noise models [?, ?]. For instance, in the presence of non-additive Poisson noise, a Haar-wavelet domain Poisson unbiased risk estimate can be derived and a soft-thresholding operator can be determined so as to minimize this estimated risk [?]. Further refinements have been performed by generalizing the

concept of SURE-LET. In [?], the objective is to denoise squared-magnitude magnetic resonance images that follow a noncentral chi-square distribution with two degrees of freedom. The noise is a non additive Rician noise. To this end, an unbiased risk estimate is also built and employed to optimize pointwise estimators for undecimated wavelet transform coefficients by using LETs.

## 7.2 Stein-based approaches for deconvolution

The application of Stein’s principle is not restricted to denoising but can also be successfully applied to image deblurring in the presence of Gaussian noise [?, ?, ?, ?, ?]. More precisely, in [?], the contribution is twofold as a wide class of noise models and linearly degraded observations are addressed. A variational framework is adopted where the Tikhonov regularization parameter is estimated thanks to Stein’s formula. In [?], in the case of a full-rank matrix  $A$  and under normality assumptions on the noise, a recursive estimation of the risk is performed coupled with a forward-backward algorithm employed to restore the image. For a pre-selected number of iterations, the threshold value is tuned in order to optimize the Generalized SURE (GSURE). Then, for the resulting threshold, the GSURE is also used to set the overall number of iterations. This recursive strategy was also generalized to the case of a rank-deficient matrix  $A$ , the value of the threshold being updated at each iteration [?]. In [?], an unbiased estimate of the projected risk is carried out for  $\ell^1$ -regularized inverse problems, handling both synthesis and analysis sparsity priors. When  $A$  is rank-deficient, the MSE is estimated with respect to a projected version of  $\bar{x}$  onto the orthogonal of the nullspace of  $A$ . The approach has been extended to address more general models involving multiple unknown parameters by using a Stein Unbiased Gradient estimator of the Risk (SUGAR) [?]. In [?], the same kind of approach as the one presented in [?, ?] is adopted and, an analytical expression of the variance of the proposed quadratic risk empirical estimate is derived. This variance evaluation makes it possible to design a reliable SURE-LET strategy for image restoration in an overcomplete representation. In [?], a SURE-LET approach is also developed to perform a regularized deconvolution followed by noise reduction. As in [?], a linear combination of thresholding functions is used for noise removal, but the latter one involves a weighted sum of Wiener filters corresponding to various regularization constants. These parameters are not optimized numerically, only the weights of the linear combination being computed so as to minimize the estimated quadratic risk. The advantage of this restoration method is that it leads to the resolution of a linear system of equations having a low computational complexity. In Figure ??, we present the reconstruction results obtained with the approach developed in [?].

## 7.3 Stein-based approaches for reconstruction

The scope of Stein-based approaches has been further extended in order to address image reconstruction problems. To this end, it is required to compute the Jacobian matrix of the estimation function involving the reconstruction operator. Most of the reported works have been focused on such computation for specific reconstruction algorithms under an additive Gaussian noise. In this respect, attention has been paid to both non-iterative reconstruction methods and iterative ones with synthesis or analysis formulations. For instance, in [?] a non-iterative SURE-LET based approach is applied to reconstruct parallel magnetic resonance images. In [?, ?], the minimization

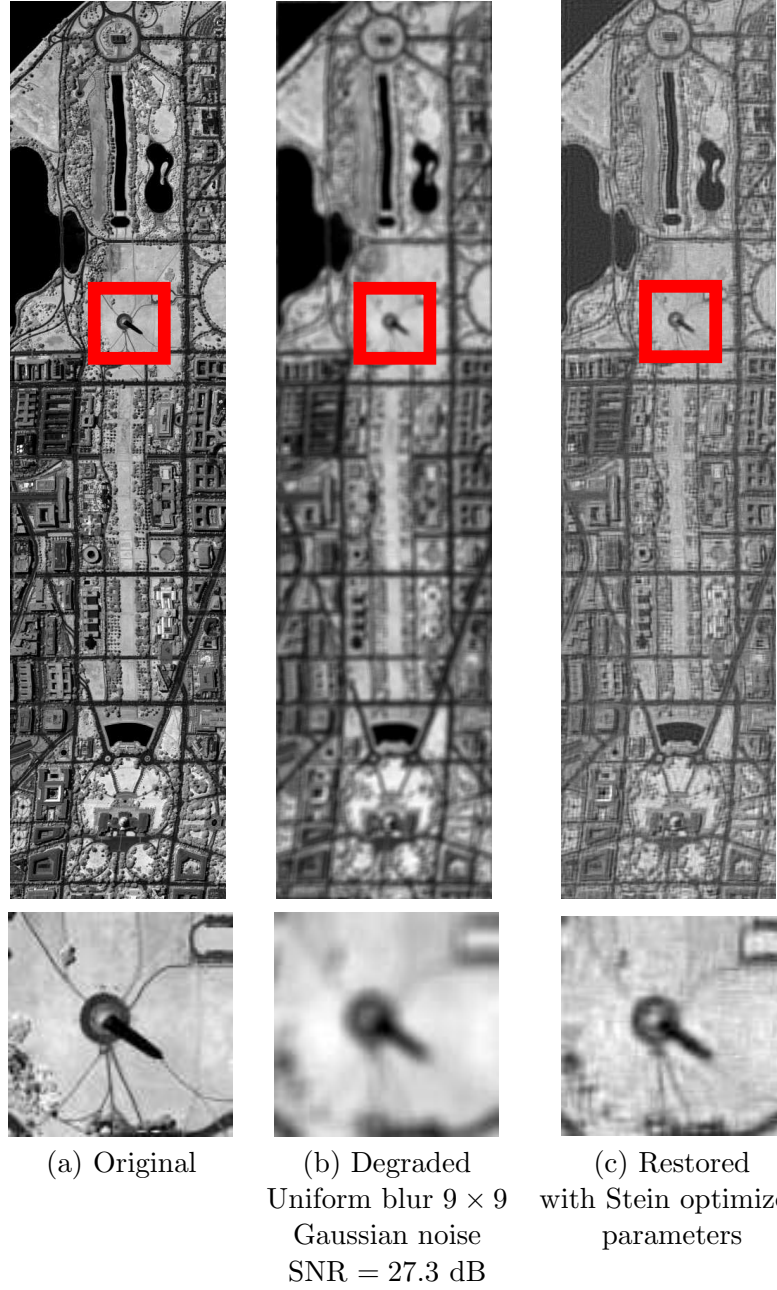


Figure 11: Illustration of image restoration with the Stein approach proposed in [?].

of SURE aims at estimating the hyperparameter of a non-quadratic regularization term corresponding to a sparse prior. In [?], the computation is carried out for two families of iterative reconstruction algorithms: the standard iterative reweighted least-squares method [?, ?] and, the ADMM algorithm. These reconstruction algorithms offer the advantage of encompassing several analysis- and synthesis-type regularizers. Two SURE-type measures are considered for these algorithms. Experimental results concerning magnetic resonance image reconstruction with a total

variation and an analysis-type  $\ell^1$ -regularization indicate that the SURE-based strategy achieves a near-MSE optimal performance.

Approaches based on set theoretic estimation [?] constitute another family of image reconstruction methods. Their principle is to consider that the recovered image lies in the intersection of a number of constraint sets. These constraint sets model prior information about an admissible solution, e.g. positivity, support, smoothness, consistency with noise characteristics,... If these sets are closed convex sets in a Hilbert space, it is possible to resort to efficient convex optimization algorithms. However, the main drawback of such approaches is that these sets must be defined *a priori*. In [?], this limitation is overcome by using Stein's principle to build novel data-adaptive convex constraints in image recovery problems.

## 8 Blind deconvolution

In practice, one often has to cope with restoration problems where the degradation filter is unknown (blind problems) or partially known (semi-blind or myopic problems). In such cases, it becomes mandatory to incorporate suitable prior information both on the sought image and the point spread function [?, ?, ?]. Indeed, in the absence of prior knowledge, intrinsic indeterminacies make the recovery problem unsolvable. For instance, if  $\bar{y} \in \mathbb{R}^N$  and  $A \in \mathbb{R}^{M \times N}$ , both  $(\bar{y}, A)$  and  $(\alpha\bar{y}, \alpha^{-1}A)$  with  $0 < \alpha \neq 1$  look like equally acceptable solutions since they produce the same degraded image (for the same noise corruption process). This phenomenon corresponds to a scaling ambiguity, but there actually exist many more ambiguities. In particular, shift ambiguities may be difficult to avoid in blind deconvolution scenarios.

Variational approaches for solving this kind of problems typically aim at minimizing a criterion consisting of the sum of three terms: a data fidelity term, a first regularization term concerning the sought image, and a second one concerning the unknown filter. When a frame analysis formulation is adopted and a quadratic data fidelity term is used, this leads to finding

$$(\hat{y}, \hat{A}) \in \underset{(y, A)}{\operatorname{Argmin}} \|Ay - z\|^2 + g(Fy) + h(A), \quad (123)$$

where  $F \in \mathbb{R}^{K \times N}$  is a frame analysis operator,  $g: \mathbb{R}^K \rightarrow ]-\infty, +\infty]$ , and  $h: \mathbb{R}^{M \times N} \rightarrow ]-\infty, +\infty]$ . Although the approach is similar to the one described in Section ??, two difficulties need to be addressed. The first one is that, even if  $g$  and  $h$  are convex functions, the objective function becomes nonconvex because of the coupling existing between variables  $y$  and  $A$  in the quadratic term. In this case, the criterion is however bi-convex in the sense that, for a given  $A$ , the objective function is a convex function of  $y$  and, for a given  $y$ , it is a convex function of  $A$ . This suggests the design of optimization algorithms which alternate between minimizations with respect to  $y$  and  $A$  [?, ?, ?]. Each of these iterations can be handled by techniques very similar to those described in Section ?. The second difficulty is related to the choice of regularization functions allowing us to alleviate the aforementioned ambiguity issues [?]. Nonsmooth convex functions  $g$  like total variation (see Section ??) or some of its extensions [?] are often employed. However, a number of works have advocated the use of nonconvex regularization functions for blind deconvolution [?, ?]. An example of such a proper choice for  $g$  is the ratio of an  $\ell^1$ -norm over an  $\ell^2$ -norm [?] or a surrogate of it [?], whereas  $h$  may be the indicator function of a convex set constraining

the operator  $A$  to satisfy some desirable properties. Alternatively, function  $g$  can be chosen to comply with a discrete Random Markov Field model [?]. Note that, due to the nonconvexity of the criterion, attention must be paid to spurious local minimas into which optimization methods may get trapped. This problem may be alleviated by making use of multiresolution approaches [?, ?]. It is also worth mentioning that variational approaches may generate a joint estimation of the image and the filter [?], but that in some approaches, one prefers to proceed in two steps [?, ?]. The aim of the first one is to provide an accurate estimation of the filter by solving a first variational problem. Then, the estimated filter is used in the second step in order to generate a higher quality restored image by employing one of the non-blind methods presented in the previous sections.

Bayesian approaches can also be applied in the context of blind deconvolution by introducing prior probabilistic knowledge on the unknown point spread function. In particular, VBA methods have been applied in [?, ?]. A parametric form of this function can be adopted (e.g. a Gaussian shape), which allows prior distributions on a small number of parameters to be introduced [?].

Note finally that SURE-type deconvolution methods have been proposed in the case of an unknown Gaussian point spread function: the PSF parameter is firstly computed to minimize an unbiased estimate of the MSE on the blurred image, then a non-blind SURE-LET deconvolution is performed [?].

## 9 Conclusion

In this paper, we have provided an overview of image recovery methods using wavelet transforms, and more generally frame representations. It has been pointed out that frames provide versatile tools for modeling prior information on the target image. Although we did not discuss these topics, similar benefits can be drawn from frames when processing 3D data, multichannel images (e.g. hyperspectral images), or video sequences.

It must be emphasized that image restoration and reconstruction is still a domain of very active research impacting a broad range of application areas (digital photography, satellite imaging, medical imaging, microscopy, spectroscopy,...). In particular, new regularizations have shown promising results in the last years. The potential of new frames has been explored, e.g. shearlets [?, ?] (see also [?]). In the meantime, generalizations of the total variation have been proposed based on a refined analysis of the contour information [?, ?], or the search for nonlocal dependencies between the pixels [?, ?], by possibly better accounting for colorimetric information [?, ?]. There has been also an increasing interest in nonconvex optimization problems. In particular, one of the main challenges is to design new criteria for better enforcing sparsity properties. In conjunction with these efforts, much work is being carried out in order to develop faster and more robust algorithms for performing optimization, simulation, and approximation tasks [?, ?, ?].

## References

- [1] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli, “A hybrid alternating proximal method for blind video restoration,” in *Proc. Eur. Signal Image Process. Conf.*, Lisbonne, Portugal, 1-5 Sept. 2014, pp. 1811–1815.
- [2] F. Abramovich, P. Besbeas, and T. Sapatinas, “Empirical Bayes approach to block wavelet function estimation,” *Comput. Statist. Data Anal.*, vol. 39, no. 4, pp. 435–451, Jun. 2002.
- [3] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representations,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [4] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE Trans. Comput.*, vol. 23, no. 1, pp. 90–93, Jan. 1974.
- [5] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
- [6] F. J. Anscombe, “The transformation, of Poisson, binomial and negative-binomial data,” *Biometrika*, vol. 35, no. 3-4, pp. 246–254, Dec. 1948.
- [7] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Variational Bayesian blind deconvolution using a total variation prior,” *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 12–26, Jan. 2009.
- [8] —, “Bayesian compressive sensing using Laplace priors,” *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
- [9] O. Batu and M. Çetin, “Parameter selection in sparsity-driven SAR imaging,” *IEEE Trans. Aero. Electron. Syst.*, vol. 47, no. 4, pp. 3040–3050, Oct. 2011.
- [10] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.
- [11] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
- [12] A. Benazza-Benyahia and J.-C. Pesquet, “Building robust wavelet estimators for multicomponent images using Stein’s principles,” *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1814–1830, Nov. 2005.
- [13] A. Benazza-Benyahia, J.-C. Pesquet, and C. Chaux, “Image denoising in the wavelet transform domain based on Stein’s principle,” in *IEEE Workshop Image Process. Theory, Tools and Applications*, Sousse, Tunisia, Nov. 24–26 2008, pp. 1–9.
- [14] A. Benichoux, E. Vincent, and R. Gribonval, “A fundamental pitfall in blind deconvolution with sparse and shift-invariant priors,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Vancouver, BC, Canada, May 26-31, 2013, pp. 6108 – 6112.

- [15] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, 1996.
- [16] J. M. Bioucas-Dias and M. A. T. Figueiredo, “A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.
- [17] T. Blu and F. Luisier, “The SURE-LET approach to image denoising,” *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778–2786, Nov. 2007.
- [18] —, “SURE-LET multichannel image denoising: interscale orthonormal wavelet thresholding,” *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 482–492, Apr. 2008.
- [19] J. Bolte, P. L. Combettes, and J.-C. Pesquet, “Alternating proximal algorithm for blind image recovery,” in *Proc. IEEE Int. Conf. Image Process.*, Hong-Kong, China, Sep. 26–29, 2010, pp. 1673–1676.
- [20] M. Born and E. Wolf, *Principles of Optics: Electromagnetic theory of propagation interference and diffraction of light*. Cambridge University Press, 1999.
- [21] K. Bredies, K. Kunisch, and T. Pock, “Total generalized variation,” *SIAM J. Imaging Sci.*, vol. 3, no. 3, pp. 492–526, 2010.
- [22] M. Burger, A. Sawatzky, and G. Steidl, “First order algorithms in variational image processing,” in *Operator Splittings and Alternating Direction Methods*, R. Glowinski, S. Osher, and W. Yin, Eds., 2014. [Online]. Available: <http://arxiv.org/abs/1412.4237>
- [23] T. T. Cai and B. W. Silverman, “Incorporating information on neighboring coefficients into wavelet estimation,” *Sankhya B*, vol. 63, pp. 127–148, 2001.
- [24] T. T. Cai and H. H. Zhou, “A data-driven block thresholding approach to wavelet estimation,” *Ann. Statist.*, vol. 37, no. 2, pp. 569–595, Apr. 2009.
- [25] P. Campisi and K. Egiazarian, Eds., *Blind Image Deconvolution: Theory and Applications*. CRC Press, 2007.
- [26] E. J. Candès and D. L. Donoho, “Recovering edges in ill-posed inverse problems: Optimality of curvelet frames,” *Ann. Stat.*, vol. 30, no. 3, pp. 784–842, Jun. 2002.
- [27] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
- [28] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [29] L. Chaàri, J.-C. Pesquet, J.-Y. Tourneret, P. Ciuciu, and A. Benazza-Benyahia, “A hierarchical Bayesian model for frame representation,” *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5560–5571, Nov. 2010.



- [30] L. Chaâri, N. Pustelnik, C. Chaux, and J.-C. Pesquet, “Solving inverse problems with over-complete transforms and convex optimization techniques,” in *Proc. SPIE, Wavelets XIII*, vol. 7446, San Diego, California, USA, Aug. 2-8 2009, 14p.
- [31] A. Chambolle, “An algorithm for total variation minimization and applications,” *J. Math. Imaging Vision*, vol. 20, no. 1-2, pp. 89–97, Jan. 2004.
- [32] A. Chambolle and C. Dossal, “On the convergence of the iterates of “FISTA??,” Sep. 2014. [Online]. Available: <https://hal.inria.fr/hal-01060130>
- [33] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *J. Math. Imaging Vision*, vol. 40, no. 1, pp. 120–145, May 2011.
- [34] T. F. Chan and C. K. Wong, “Convergence of the alternating minimization algorithm for blind deconvolution,” *Linear Algebra Appl.*, vol. 316, no. 1-3, pp. 259–285, Sept. 2000.
- [35] C. Chaux, A. Benazza-Benyahia, and J.-C. Pesquet, “A block-thresholding method for multispectral image denoising,” in *Proc. SPIE, Wavelets X*, vol. 5914, San Diego, CA, USA, Aug. 2005, pp. 1–H1,1–H13.
- [36] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, “A variational formulation for frame-based inverse problems,” *Inverse Problems*, vol. 23, no. 4, pp. 1495–1518, Jun. 2007.
- [37] C. Chaux, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet, “A new estimator for image denoising using a 2D Dual-Tree  $M$ -band wavelet decompositions,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 3, Toulouse, France, May 14-19 2006, pp. 249–252.
- [38] —, “A nonlinear stein based estimator for multichannel image denoising,” *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3855–3870, Aug. 2008.
- [39] C. Chaux, L. Duval, and J.-C. Pesquet, “Image analysis using a dual-tree  $M$ -band wavelet transform,” *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2397–2412, Aug. 2006.
- [40] C. Chaux, J.-C. Pesquet, and N. Pustelnik, “Nested iterative algorithms for convex constrained image recovery problems,” *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 730–762, Jun. 2009.
- [41] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.
- [42] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pesquet, “A nonlocal structure tensor-based approach for multicomponent image recovery problems,” *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5531–5544, Dec. 2014.
- [43] H. Choi and R. G. Baraniuk, “Analysis of wavelet domain wiener filters,” in *IEEE-Signal Process. Int. Symp. Time-frequency and Time-scale Anal.*, Pittsburgh, Pennsylvania, USA, Oct. 6–9 1998, pp. 613–616.
- [44] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, “A majorize-minimize subspace approach for  $\ell_2$ - $\ell_0$  image regularization,” *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 563–591, 2013.

- [45] E. Chouzenoux, A. Jezierska, J. Pesquet, and H. Talbot, “A convex approach for image restoration with exact Poisson-Gaussian likelihood,” Tech. Rep., 2015. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00922151v2>
- [46] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, “A block coordinate variable metric forward-backward algorithm,” Tech. Rep., 2013. [Online]. Available: [http://www.optimization-online.org/DB\\_HTML/2013/12/4178.html](http://www.optimization-online.org/DB_HTML/2013/12/4178.html)
- [47] A. Cohen, I. Daubechies, and P. Vial, “Wavelets on interval and fast wavelet transforms,” *Appl. Comput. Harmon. Anal.*, vol. 1, no. 1, pp. 54–81, 1993.
- [48] R. R. Coifman and D. L. Donoho, “Translation-invariant de-noising,” in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York, USA: Springer-Verlag, 1995, vol. 103, pp. 125–150.
- [49] R. R. Coifman and M. V. Wickerhauser, “Entropy-based algorithms for best basis selection,” *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 713–718, Mar. 1992.
- [50] P. L. Combettes, “The foundations of set theoretic estimation,” *Proc. IEEE*, vol. 81, no. 2, pp. 182–208, Feb. 1993.
- [51] P. L. Combettes and J.-C. Pesquet, “Wavelet-constrained image restoration,” *Int. J. Wavelets Multiresolut. Inf. Process.*, vol. 2, no. 4, pp. 371–389, Dec. 2004.
- [52] —, “A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,” *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 564–574, Dec. 2007.
- [53] —, “A proximal decomposition method for solving convex variational inverse problems,” *Inverse Problems*, vol. 24, no. 6, pp. x+27, Dec. 2008.
- [54] —, “Proximal splitting methods in signal processing,” in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York: Springer-Verlag, 2010, pp. 185–212.
- [55] —, “Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,” *Set-Valued Var. Anal.*, vol. 20, no. 2, pp. 307–330, Jun. 2011.
- [56] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.
- [57] L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Aug. 2013.
- [58] —, “Semi-local total variation for regularization of inverse problems,” in *Proc. Eur. Signal Image Process. Conf.*, Lisbon, Portugal, Sep. 1–5 2014, pp. 1806–1810.
- [59] B. R. Corner, R. M. Narayanan, and S. E. Reichenbach, “Noise estimation in remote sensing imagery using data masking,” *Internat. J. Remote Sensing*, vol. 24, no. 4, pp. 689–702, 2003.

- [60] A. Danielyan, V. Katkovnik, and K. Egiazarian, “BM3D frames and variational image deblurring,” *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1715–1728, Apr. 2012.
- [61] I. Daubechies, *Ten lectures on wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.
- [62] I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.
- [63] C. Deledalle, S. Vaiter, G. Peyré, and J. Fadili, “Stein Unbiased GrADient estimator of the Risk (SUGAR) for multiple parameter selection,” *SIAM J. Imaging Sci.*, vol. 7, no. 4, pp. 2448–2487, 2014.
- [64] C. Deledalle, S. Vaiter, G. Peyré, J. Fadili, and C. Dossal, “Unbiased risk estimation for sparse analysis regularization,” in *Proc. IEEE Int. Conf. Image Process.*, Orlando, FL, USA, Sep. 30 ?- Oct. 3 2012, pp. 3053–3056.
- [65] C.-A. Deledalle, G. Peyré, and J. Fadili, “Stein COnsistent Risk Estimator (SCORE) for hard thresholding,” in *Signal Process. Adapt. Sparse Struct. Repr.*, Lausanne, Switzerland, Jul. 8–11 2013.
- [66] A. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
- [67] M. N. Do and M. Vetterli, “The contourlet transform: an efficient directional multiresolution image representation,” *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
- [68] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, Aug. 1994.
- [69] —, “Adapting to unknown smoothness via wavelet shrinkage,” *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.
- [70] R. J. Duffin and A. C. Schaeffer, “A class of nonharmonic fourier series,” *Trans. Amer. Math. Soc.*, vol. 72, no. 2, pp. 341–366, Mar. 1952.
- [71] F.-X. Dupé, M. J. Fadili, and J.-L. Starck, “A proximal iteration for deconvolving Poisson noisy images using sparse representations,” *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 310–321, Feb. 2009.
- [72] S. Durand and M. Nikolova, “Denoising of frame coefficients using  $\ell_1$  data-fidelity term and edge-preserving regularization,” *Multiscale Model. Simul.*, vol. 6, no. 2, pp. 547–576, 2007.
- [73] J. Eckstein and D. P. Bertekas, “On the Douglas-Rachford splitting methods and the proximal point algorithm for maximal monotone operators,” *Math. Programm.*, vol. 55, no. 3, pp. 293–318, Apr. 1992.
- [74] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

- [75] M. Elad, P. Milanfar, and R. Ron, “Analysis versus synthesis in signal priors,” *Inverse Problems*, vol. 23, no. 3, pp. 947–968, Jun. 2007.
- [76] Y. C. Eldar, “Generalized SURE for exponential families: Applications to regularization,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.
- [77] H. W. End, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Dordrecht, The Netherlands: Kluwer Academic, 1996.
- [78] M. J. Fadili and J.-L. Starck, *Curvelets and Ridgelets*, in *Encyclopedia Complex. Syst. Sci.*, R. Meyers, Ed. Springer, 2009, vol. 3.
- [79] C. Févotte and S. J. Godsill, “Sparse linear regression in unions of bases via Bayesian variable selection,” *Signal Process. Lett.*, vol. 13, no. 7, pp. 441–444, Jul. 2006.
- [80] M. A. T. Figueiredo and J. M. Bioucas-Dias, “Restoration of Poissonian images using alternating direction optimization,” *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.
- [81] M. A. T. Figueiredo, J. M. Bioucas-Dias, J. P. Oliveira, and R. D. Nowak, “A new majorization-minimization algorithm and an experimental comparison with wavelet denoising,” in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, Oct. 8–11 2006, pp. 2633–2636.
- [82] M. A. T. Figueiredo and R. D. Nowak, “An EM algorithm for wavelet-based image restoration,” *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.
- [83] A. Florescu, E. Chouzenoux, P. Pesquet, J.-C. Ciuciu, and S. Ciochina, “A majorize-minimize memory gradient method for complex-valued inverse problems,” *Signal Process.*, vol. 103, pp. 285–295, Oct. 2014.
- [84] A. Foi and G. Boracchi, “Foveated self-similarity in nonlocal image filtering,” in *Proc. SPIE, Human Vision and Electronic Imaging XVII*, vol. 8291, Burlingame, CA, USA, Jan. 2012.
- [85] A. Fraysse and T. Rodet, “A measure-theoretic variational Bayesian algorithm for large dimensional problems,” *SIAM J. Imaging Sci.*, vol. 7, no. 4, pp. 2591–2622, 2014.
- [86] J. Gauthier, L. Duval, and J.-C. Pesquet, “Two denoising sure-let methods for complex oversampled subband decompositions,” in *Proc. Eur. Signal Image Process. Conf.*, Lausanne, Switzerland, Aug. 25–28 2008.
- [87] D. Geman and G. Reynolds, “Constrained restoration and the recovery of discontinuities,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
- [88] D. Geman and C. Yang, “Nonlinear image recovery with half-quadratic regularization,” *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–945, Jul. 1995.
- [89] S. Geman and D. McClure, “Bayesian image analysis: An application to single photon emission tomography,” in *Proc. Statist. Comput. Section Amer. Statist. Association*, 1985, pp. 12–18.

- [90] S. P. Ghael, A. M. Sayeed, and R. G. Baraniuk, “Improved wavelet denoising via empirical Wiener filtering,” in *Proc. SPIE, Wavelet Applications in Signal and Image Processing V*, vol. 3169, San Diego, CA, USA, Jul. 1997, pp. 389–399.
- [91] G. Gilboa and S. Osher, “Nonlocal operators with applications to image processing,” *Multi-scale Model. Simul.*, vol. 7, no. 3, pp. 1005–1028, 2009.
- [92] R. Giryes, M. Elad, and Y. Eldar, “The projected GSURE for automatic parameter tuning in iterative shrinkage methods,” *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, May 2010.
- [93] R. Glowinski and A. Marrocco, “Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de dirichlet non linéaires,” *Revue Française d’Automatique, Informatique, et Recherche Opérationnelle*, vol. 9, no. 2, pp. 41–76, 1975.
- [94] B. Gnedenko, “Sur la distribution limite du terme maximum d’une série aléatoire,” *Ann. of Math.*, vol. 44, no. 3, pp. 423–453, Jul. 1943.
- [95] T. Goldstein and S. Osher, “The split Bregman method for  $\ell_1$ -regularized problems,” *SIAM J. Imaging Sci.*, vol. 2, pp. 323–343, 2009.
- [96] G. H. Golub and C. F. Van Loan, *Matrix computations*. The Johns Hopkins University Press; 3rd edition, 1996.
- [97] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison Wesley, 1992.
- [98] I. F. Gorodnitsky and B. D. Rao, “Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm,” *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
- [99] V. K. Goyal, M. Vetterli, and N. T. Thao, “Quantized overcomplete expansions in  $\mathbb{R}^n$ : Analysis, synthesis, and algorithms,” *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 16–31, Jan. 1998.
- [100] P. Green, K. Latuszynski, M. Pereyra, and C. P. Robert, “Bayesian computation: a perspective on the current state, and sampling backwards and forwards,” *Statist. Comput.*, 2015, to appear. [Online]. Available: <http://arxiv.org/abs/1502.01148>
- [101] R. Gribonval and K. Schnass, “Dictionary identification - sparse matrix-factorization via  $\ell_1$ -minimization,” *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3523–3539, 2010.
- [102] J. Hadamard, “Sur les problèmes aux dérivées partielles et leur signification physique,” *Princeton Univ. Bull.*, vol. 13, pp. 49–52, 1902.
- [103] D. Han and D. R. Larson, “Frames, bases, and group representations,” *Mem. Amer. Math. Soc.*, vol. 147, no. 697, pp. x+94, 2000.
- [104] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

- [105] S. Häuser and G. Steidl, “Fast finite shearlet transform: a tutorial.” 2014. [Online]. Available: <http://arxiv.org/pdf/1202.1773.pdf>
- [106] C. Heil and D. Walnut, “Continuous and discrete wavelet transforms,” *SIAM Rev.*, vol. 31, pp. 628–666, 1989.
- [107] H. Hudson and R. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, Dec. 1994.
- [108] G. Huerta, “Multivariate Bayes wavelet shrinkage and applications,” *J. Appl. Statist.*, vol. 32, no. 5, pp. 529–542, 2005.
- [109] B. R. Hunt, “The application of constrained least squares estimation to image restoration by digital computer,” *IEEE Trans. Comput.*, vol. 22, no. 9, pp. 805–812, Sep. 1973.
- [110] B. Jørgensen, *Statistical properties of the generalized inverse Gaussian distribution*. New York, USA: Springer-Verlag, 1982.
- [111] M. Kaaniche, A. A. Benazza-Benyahia, B. Pesquet-Popescu, and J.-C. Pesquet, “Vector lifting schemes for stereo image coding,” *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2463–2475, 2009.
- [112] N. Komodakis and N. Paragios, “MRF-based blind image deconvolution,” in *Proc. Asian Conference on Computer Vision*, Daejeon, Korea, 5-9 Nov. 2012, pp. 361–374.
- [113] N. Komodakis and J.-C. Pesquet, “Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems,” *IEEE Signal Process. Mag.*, 2015, to appear. [Online]. Available: <http://arxiv.org/abs/1406.5429>
- [114] H. Krim, D. Tucker, S. Mallat, and D. L. Donoho, “On denoising and best signal representation,” *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2225–2238, Nov. 1999.
- [115] D. Krishnan, T. Tay, and R. Fergus, “Blind deconvolution using a normalized sparsity measure,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Colorado Springs, USA, 20-25 Jun. 2011, pp. 233–240.
- [116] F. Krzakala, A. Manoel, E. W. Tramel, and L. Zdeborova, “Variational free energies for compressed sensing,” in *Proc. IEEE Int. Symp. Information Theory*, Honolulu, HI, USA, Jun. 29 – Jul. 4 2014, pp. 1499–1503.
- [117] D. Kundur and D. Hatzinakos, “Blind image deconvolution,” *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 43–64, May 1996.
- [118] H. R. Kunsch, “Robust priors for smoothing and image restoration,” *Ann. Inst. Stat. Math.*, vol. 46, no. 1, pp. 1–19, 1994.
- [119] R. L. Lagendijk and J. Biemond, *The Handbook of Image and Video Processing*. New-York, N. J., USA: A. Bovik, Eds. Academic Press, 2005.

- [120] R. L. Lagendijk, J. Biemond, and D. E. Boeke, "Identification and restoration of noisy blurred images using the expectation maximization algorithm," *IEEE Trans. Acous., Speech Signal Process.*, vol. 38, pp. 1180–1191, Jul. 1990.
- [121] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomo.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.
- [122] H. Lantéri and C. Theys, "Restoration of astrophysical images—The case of Poisson data with additive Gaussian noise," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 15, pp. 2500–2513, 2005.
- [123] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 423–438, Apr. 2005.
- [124] S. Lefkimmiatis, A. Roussos, M. Unser, and P. Maragos, "Convex generalizations of total variation based on the structure tensor with applications to inverse problem," in *Scale-Space and Variational Methods in Comput. Vision*, vol. 7893, Graz region, Austria, Jun. 2–6 2013, pp. 48–60.
- [125] D. Leporini and J.-C. Pesquet, "Bayesian wavelet denoising: Besov priors and non-Gaussian noises," *Signal Process.*, vol. 81, no. 1, pp. 55–67, 2001.
- [126] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Miami, Florida, USA, 20–25 Jun. 2009, pp. 1964?–1971.
- [127] C. Louchet, "Modèles variationnels et bayésiens pour le débruitage d'images : de la variation totale vers les moyennes non-locales." Mathématiques Appliquées, Université Paris Descartes, 2008.
- [128] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron. J.*, vol. 79, no. 6, pp. 745–754, Jun. 1974.
- [129] F. Luisier, T. Blu, and P. J. Wolfe, "A CURE for noisy magnetic resonance images: Chi-square unbiased risk estimation," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3454–3454, Aug. 2012.
- [130] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of Poisson-corrupted images," *Signal Process.*, vol. 90, no. 2, pp. 415–427, Feb. 2010.
- [131] D. J. C. MacKay, "Ensemble learning and evidence maximization," Cavendish Laboratory, University of Cambridge, Tech. Rep., 1995.
- [132] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Model. Simul.*, vol. 7, no. 1, pp. 214?–241, 2008.
- [133] S. Mallat, *A wavelet tour of signal processing*. San Diego, USA: Academic Press, 1997.
- [134] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, pp. 3397–3415, Dec. 1993.

- [135] H. S. Malvar, *Signal Processing with Lapped Transforms*. Norwood, MA, USA: Artech House, 1992.
- [136] A. Marin, C. Chaux, J.-C. Pesquet, and P. Ciuciu, “Image reconstruction from multiple sensors using Steins principle. application to parallel MRI,” in *Proc. IEEE Int. Symp. Biomed. Imaging*, Mar. 30 – Apr. 2 2011, pp. 465 – 468.
- [137] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [138] J. G. Nagy, V. P. Pauca, R. J. Plemmons, and T. C. Torgersen, “Anisoplanatic deconvolution of adaptive optics images,” *J. Opt. Soc. Amer. A*, vol. 22, pp. 504–513, Mar. 2005.
- [139] J. G. Nagy, V. P. Pauca, R. J. Plemmons, and T. C. Torgersen, “Space-varying restoration of optical images,” *J. Opt. Soc. Amer. A*, vol. 14, pp. 3162–3174, 1997.
- [140] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Comput.*, vol. 24, no. 4, pp. 227–234, 1995.
- [141] R. M. Neal, “MCMC using Hamiltonian dynamics,” in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, Eds. Chapman & Hall / CRC Press, 2011.
- [142] F. Orieux, J.-F. Giovannelli, and T. Rodet, “Bayesian estimation of regularization and point spread function parameters for Wiener?-Hunt deconvolution,” *J. Opt. Soc. Amer. A*, vol. 27, no. 7, pp. 1593–1607, 2010.
- [143] J. Palmer, K. Kreutz-Delgado, B. D. Rao, and D. P. Wipf, “Variational EM algorithms for non-Gaussian latent variable models,” in *Adv. Neural. Inf. Process. Syst. 18*, Y. Weiss, B. Schölkopf, and J. C. Platt, Eds. MIT Press, 2006, pp. 1059–1066.
- [144] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Proc. Asilomar Conf. Signal Syst. Comput.*, vol. 1, Pacific Grove, CA, USA, Nov. 1–3 1993, pp. 40–44.
- [145] P. Paul, D. Kalamatianos, V. Duesman, and H. Huber, “Automatic quality assessment for fluorescence microscopy images,” in *Proc. IEEE Int. Conf. Bioinformatics Bioeng.*, Athens, Greece, Oct. 8–10 2008, pp. 1–6.
- [146] J. B. Pawley, *Handbook of Biological Confocal Microscopy*. New-York, NJ, USA: Plenum Press, 1996.
- [147] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tournier, A. Hero, and S. McLaughlin, “Tutorial on stochastic simulation and optimization methods in signal processing,” 2015. [Online]. Available: <http://arxiv.org/abs/1505.00273>
- [148] D. Perrone, R. Diethelm, and P. Favaro, “Blind deconvolution via lower-bounded logarithmic image priors,” in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, ser. Lecture Notes in Computer Science, vol. 8932, 2015, pp. 112–125.



- [149] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616–4632, 2009.
- [150] J.-C. Pesquet, H. Krim, and H. Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 1964–1970, Aug. 1996.
- [151] J.-C. Pesquet and D. Leporini, "A new wavelet estimator for image denoising," in *Proc. IEEE Int. Conf. Im. Proc. Appl.*, vol. 1, Dublin, Ireland, Jul. 14–17 1997, pp. 249–253.
- [152] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pac. J. Optim.*, vol. 8, no. 2, pp. 273–305, Apr. 2012.
- [153] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 654–665, Mar. 2006.
- [154] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Process.*, no. 5, pp. 545–557, May 2002.
- [155] B. T. Polyak, "Minimization of unsmooth functionals," *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 3, pp. 14–29, 1969.
- [156] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [157] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Parallel ProXimal Algorithm for image restoration using hybrid regularization," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2450–2462, Sep. 2011.
- [158] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.
- [159] S. Ramani and J. A. Fessler, "An accelerated iterative reweighted least squares algorithm for compressed sensing MRI," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Rotterdam, Netherlands, Apr. 14–17 2010, pp. 257–260.
- [160] S. Ramani, Z. Liu, J. Rosen, J.-F. Nielsen, and J. A. Fessler, "Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and SURE-based methods," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3659–3672, Aug. 2012.
- [161] S. Ramani, J. Rosen, Z. Liu, and J. A. Fessler, "Iterative weighted risk estimation for nonlinear image restoration with analysis priors," in *Proc. SPIE, Computational Imaging X*, vol. 8296, no. 1, Burlingame, California, USA, Jan. 22 2012, pp. 82 960N1–82 960N12.

- [162] S. Rangan, A. K. Fletcher, P. Schniter, and U. Kamilov, “Inference for generalized linear models via alternating directions and Bethe free energy minimization,” 2015. [Online]. Available: <http://arxiv.org/abs/1501.01797>
- [163] M. Raphan and E. P. Simoncelli, “Optimal denoising in redundant representations,” *IEEE Trans. Image Process.*, vol. 17, pp. 1342–1352, Aug. 2008.
- [164] A. Repetti, M. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet, “Euclid in a taxicab: Sparse blind deconvolution with smoothed  $\ell_1/\ell_2$  regularization,” *Signal Process. Lett.*, vol. 22, no. 5, pp. 539–543, May 2015.
- [165] W. H. Richardson, “Bayesian-based iterative method of image restoration,” *J. Opt. Soc. America*, vol. 62, no. 1, pp. 55–59, 1972.
- [166] C. P. Robert and G. Casella, *Monte-Carlo Statistical Methods*, ser. Springer Texts in Statistics. New York, USA: Springer-Verlag, 2000.
- [167] G. O. Roberts and R. L. Tweedie, “Exponential convergence of Langevin distributions and their discrete approximations,” *Bernoulli*, pp. 341–363, 1996.
- [168] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.
- [169] P. Ruiza, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, “Variational Bayesian blind image deconvolution: A review,” *Digital Signal Process.*, 2015, to appear.
- [170] P. Scheunders and S. De Backer, “Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free image as priors,” *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1865–1875, Jul. 2007.
- [171] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, “The dual-tree complex wavelet transform,” *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
- [172] I. W. Selesnick and M. Figueiredo, “Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors,” in *SPIE*, vol. 7446, 2009.
- [173] S. Setzer, G. Steidl, and T. Teuber, “Deblurring Poissonian images by split Bregman techniques,” *J. Vis. Comm. Image Repr.*, vol. 21, no. 3, pp. 193–199, Apr. 2010.
- [174] L. A. Shepp and Y. Vardi, “Maximum likelihood reconstruction in Positron Emission Tomography,” *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, Oct. 1982.
- [175] E. P. Simoncelli, *Bayesian denoising of visual images in the wavelet domain*. Springer-Verlag, 1999, vol. 141, ch. 18, pp. 291–308.
- [176] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*. New York, USA: Springer-Verlag, 2005.
- [177] M. Sonka, *Medical Image Processing and Analysis*. Bellingham, WA, USA: SPIE, 2009.

- [178] P. Steffen, P. N. Heller, R. A. Gopinath, and C. S. Burrus, "Theory of regular  $M$ -band wavelet bases," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3497–3511, Dec. 1993.
- [179] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [180] A. Tikhonov, "Tikhonov regularization of incorrectly posed problems," *Soviet Mathematics Doklady*, vol. 4, pp. 1624–1627, 1963.
- [181] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction using sparse priors," in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, USA, Oct., 8–6 2006, pp. 1261–1264.
- [182] D. Titterton, "On the iterative image space reconstruction algorithm for ECT," *IEEE Trans. Med. Imag.*, vol. 6, no. 1, pp. 52–56, Mar. 1987.
- [183] R. Tolimieri and M. An, *Time-Frequency Representations*. Boston, MA, USA: Birkhauser, 1998.
- [184] F. Šroubek and P. Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1687–1700, Apr. 2012.
- [185] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.
- [186] P. P. Vaidyanathan, "Theory and design of  $m$ -channel maximally decimated quadrature mirror filters with arbitrary  $m$ , having the perfect-reconstruction property," *IEEE Trans. Acous., Speech Signal Process.*, vol. 35, no. 4, pp. 476–492, Apr. 1987.
- [187] C. Vonesh, S. Ramani, and M. Unser, "Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 12–14 2008, pp. 665–668.
- [188] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural. Inf. Process. Syst. 12*, S. A. Solla, T. K. Leen, and K. Müller, Eds. MIT Press, 2000, pp. 855–861.
- [189] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New-York: Wiley, 1949.
- [190] F. Xue and T. Blu, "SURE-based blind Gaussian deconvolution," in *Proc. IEEE Workshop Stat. Signal Process.*, Ann Arbor, MI, USA, Aug., 5–8 2012, pp. 452–455.
- [191] F. Xue, F. Luisier, and T. Blu, "Multi-Wiener SURE-LET deconvolution," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1954–1968, 2013.
- [192] M. Yaghoobi, T. Blumensath, and M. E. Davies, "Dictionary learning for sparse approximations with the majorization method," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2178–2191, Jun. 2009.
- [193] S. Yi, D. Labate, G. R. Easley, and H. Krim, "A shearlet approach to edge analysis and detection," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 929–941, May 2009.

- [194] Y. Zheng, A. Fraysse, and T. Rodet, “Efficient variational Bayesian approximation method based on subspace optimization,” *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 681–693, 2015.
- [195] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *J. R. Statist. Soc. B*, vol. 2, no. 67, pp. 301?–320, Apr. 2005.
- [196] I. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. Gindi, and P. Hoffer, “Computerized three-dimensional segmented human anatomy,” *Med. Phys.*, vol. 21, no. 1, pp. 299–302, Feb. 1994.