Multivariate image processing: methods and applications

Ch. COLLET - J. CHANUSSOT - K. CHEHDI -

version 1.5 March 24th 2009

Table des matières

Chapitre 1

Multivariate image processing : methods and applications

Chapitre 2

Wavelet transform for the denoising of multivariate images

2.1. Introduction

An increasing attention is being paid to multispectral images for a great number of applications (medicine, agriculture, archeology, forestry, coastal management, remote sensing . . .) because many features of the underlying scene have unique spectral characteristics that become apparent in imagery when viewing combinations of its different components. Hence, in satellite imaging, a better analysis of the nature of the materials covering the surface of the earth is achieved [LAN 00]. Typically, multispectral imaging systems employ radiometers as acquisition instruments which operate in different spectral channels. Each one delivers a digital image in a small range of the visible or non visible wavelengths. As a result, the spectral components form a multicomponent image corresponding to a single sensed area. Usually, satellites have three to a dozen of radiometers. Multispectral sensors offer a valuable advantage over color aerial photographs, thanks to their ability to record reflected light in the near infrared domain. Near infrared is the most sensitive spectral domain used to map vegetation canopy properties [GUY 90]. There are several families of on-board multispectral radiometers in the different satellite systems. The first example is SPOT 3 which has two High Resolution Visible imaging systems (HRV1 and HRV2). Each HRV is designated to operate in two sensing modes : a 10 m resolution "Panchromatic" (P) mode over the range [0.5, 0.73] μm and a 20 m resolution multispectral mode. For the multispectral mode, the first channel is associated with the range [0.5, 0.59] μm, the second channel with the range [0.61, 0.78] μm and the third one with the range [0.79, 0.89]

Chapitre rédigé par C. CHAUX, A. BENAZZA-BENYAHIA, J.-C. PESQUET, L. DUVAL.

μm. The SPOT family provides a service continuity with the upgraded satellite SPOT 4 (launched on March 1998) and SPOT 5 (launched on May 2002). In addition to the former 3 channels, SPOT 4 and SPOT 5 imaging systems gather images in a fourth channel corresponding to a short wave infrared spectral range ([1.58, 1.75] μm). The fourth channel was introduced in order to allow early observations of plant growth.

Another well-known family of multispectral satellite imaging systems is the set of Thematic Mapper instruments with the launch of Landsat 1 in 1972. Since April 1999, Landsat 7 carries the Enhanced Thematic Mapper Plus (ETM+) sensors which are similar to the Thematic Mapper sensors with additional features. An ETM+ Landsat scene is formed by 7 spectral components at a 30 m spatial resolution (except in the thermal band with a spatial resolution of 60 m) and a panchromatic image with 15 m pixel resolution. Recently, commercial satellites like Ikonos and Quickbird have provided very high resolution images. For instance, Ikonos 4 (resp. Quickbird) collects data with a level of detail of 4 m (resp. 2.4 or 2.8 m) in 4 spectral ranges (blue, green, red, and near infrared).

Despite the dramatical technological advances in terms of spatial and spectral resolutions of the radiometers, data still suffer from several degradations. For instance, the sensor limited aperture, aberrations inherent to optical systems and mechanical vibrations create a blur effect in remote sensing images [JAI 89]. In optical remote sensing imagery, there are also many noise sources. Firstly, the number of photons received by each sensor during the obturation time may fluctuate around its average implying a photon noise. A thermal noise may be caused by the electronics of the recording and the communication channels during the data downlinking. Intermittent saturations of any detector in a radiometer may give rise to an impulsive noise whereas a structured periodic noise is generally caused by interferences between electronic components. Detector striping (resp. banding) are consequences of calibration differences among individual scanning detectors (resp. from scan-to-scan). Besides, component-to-component misregistration may occur : corresponding pixels in different components are not systematically associated with the same position on the ground.

As a result, it is mandatory to apply deblurring, denoising and geometric corrections to the degraded observations in order to fully exploit the information they contain. In this respect, it is used to distinguish between on-board and on-ground processing. Indeed, on-board procedures should simultaneously fulfill real-time constraints and low mass memory requirements. The involved acquisition bit-rates are high (especially for very high resolution missions) and hence, they complicate the software implementation of enhancement processing. This is the reason why ASIC (Application-Specific Integrated Circuit) hardware circuits are employed. Such on-board circuits enable very basic processing since they present a lower performance than ground-based ones. For instance, Landsat ETM+ raw data are corrected for scan line direction and band alignment only. No radiometric or geometric correction is applied. Consequently, most of the efforts for enhancing the data are performed after their reception at the terrestrial stations. In this context, denoising is a delicate task since it aims at attenuating the noise level while maintaining the significant image features. Generally, the focus is

put on additive Gaussian noise. In this respect, many works have been carried out concerning single-component images. The pionnering ones were based on linear spatial filters and nonlinear ones [JAI 89, PIT 90]. In parallel to these efforts, a gain in performance can be achieved by attenuating the noise in a transform domain in which the image representation yields a tractable statistical modelling. The seminal work of Donoho has shown the potentialities of the Wavelet Transform (WT) for reducing a Gaussian additive noise thanks to its sparsity and decorrelation properties [DON 93]. As a consequence, several wavelet-based image denoising methods were investigated. The objective of this work is to give an overview of the most relevant on-ground wavelet-based noise reduction methods devoted to multicomponent images. Two approaches can be considered. The first one consists of independently applying any monochannel noise reduction method to each component. Although its principle is simple, this approach suffers from a serious drawback as it does not account for the cross-component dependences. This has motivated the development of an alternative approach in which the noisy components are jointly processed. In broad outline, it is also possible to classify all the denoising methods (whatever they are componentwise or multivariate ones) into non Bayesian and Bayesian methods. For the latter category, a prior distribution model is adopted for the unknown image.

This chapter is organized as follows. Notations and the observation model are presented in Section 2.2. Section 2.3 is a concise overview on wavelet transforms and filter banks. Componentwise and multichannel denoising methods are presented in Section 2.4 : a wide panel of approaches is tackled (wavelet-based, Bayesian estimation, . . .). Finally, some comparisons are drawn in Section 2.5 before concluding the chapter with Section 2.6.

2.2. Observation model

This section is devoted to the characterization of multichannel satellite images.

2.2.1. *Observed images*

The "clean" unknown multicomponent image $(s^{(1)}(\mathbf{k}), \ldots, s^{(B)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$, where $\mathbb{K} \subset \mathbb{Z}^2$ is a set of spatial indices, consists of B images corresponding to B spectral bands captured by B sensors. The notation $s^{(b)}(\mathbf{k})$ thus designates the intensity value of the pixel at position $\mathbf{k}$ in the b-th image component as represented in Fig. 2.1.

To better figure out the multichannel context, Fig. 2.2 displays 6 components of a Landsat 7 image. As can be noticed, image components share common structures, some details being present in specific spectral bands only. This phenomenon can be explained by the fact that some sensors are, for example, better able to capture vegetation whereas others are calibrated for soils. This is one of multispectral image specific properties that should be taken into account in the design of processing methods.
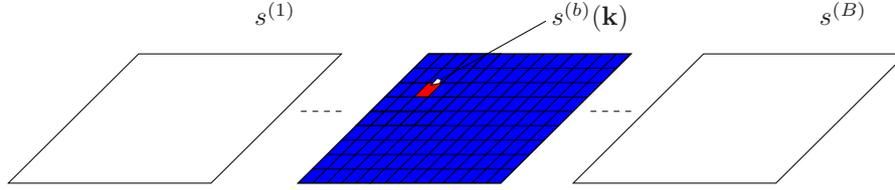
Figure 2.1. *In red, a pixel at spatial position* **k** *in the b-th image component.*

Unfortunately, the observed images denoted by $(r^{(1)}(\mathbf{k}))_{\mathbf{k} \in \mathbb{K}}$, ..., $(r^{(B)}(\mathbf{k}))_{\mathbf{k} \in \mathbb{K}}$ are subject to various degradations which are detailed in the next section.

2.2.2. *Degradation model*

The observed images are corrupted by noises coming from different sources [LAN 86] : atmospheric, sensor detector/preamplifier and quantization. In spite of the various statistical distributions of these noise sources, the global noise present in acquired data can be realistically modelled by an additive zero-mean spatially white Gaussian noise [COR 03, ABR 02] thus leading to the following model :

$$\forall b \in \{1, \ldots, B\}, \forall \mathbf{k} \in \mathbb{K}, \qquad r^{(b)}(\mathbf{k}) = s^{(b)}(\mathbf{k}) + n^{(b)}(\mathbf{k}). \qquad (2.1)$$

Following a multivariate approach, we define the unknown vector signal $\mathbf{s}$, the vector noise $\mathbf{n}$ and the vector observation $\mathbf{r}$, as

$$\forall \mathbf{k} \in \mathbb{K}, \qquad \begin{cases} \mathbf{s}(\mathbf{k}) & \triangleq & [s^{(1)}(\mathbf{k}), \ldots, s^{(B)}(\mathbf{k})]^\top \\ \mathbf{n}(\mathbf{k}) & \triangleq & [n^{(1)}(\mathbf{k}), \ldots, n^{(B)}(\mathbf{k})]^\top \\ \mathbf{r}(\mathbf{k}) & \triangleq & [r^{(1)}(\mathbf{k}), \ldots, r^{(B)}(\mathbf{k})]^\top \end{cases}$$

and, consequently, Equation (2.1) can be reexpressed into a more concise form as

$$\forall \mathbf{k} \in \mathbb{K}, \qquad \mathbf{r}(\mathbf{k}) = \mathbf{s}(\mathbf{k}) + \mathbf{n}(\mathbf{k}) \qquad (2.2)$$

where $\mathbf{n}$ is an i.i.d. zero-mean Gaussian multivariate noise with covariance matrix $\mathbf{\Gamma}^{(\mathbf{n})} \in \mathbb{R}^{B \times B}$. This matrix can take different forms, three of which will catch our attention :

1) When the noise is uncorrelated from a component to another with the same variance σ^2 in each channel, the matrix takes the following form : $\mathbf{\Gamma}_1^{(\mathbf{n})} = \sigma^2 \, \mathbf{I}_B$, where $\mathbf{I}_B$ denotes the identity matrix of size $B \times B$.

2) When the noise is uncorrelated with various noise levels in the spectral bands, we have $\mathbf{\Gamma}_2^{(\mathbf{n})} = \text{Diag}(\sigma_1^2, \ldots, \sigma_B^2)$.
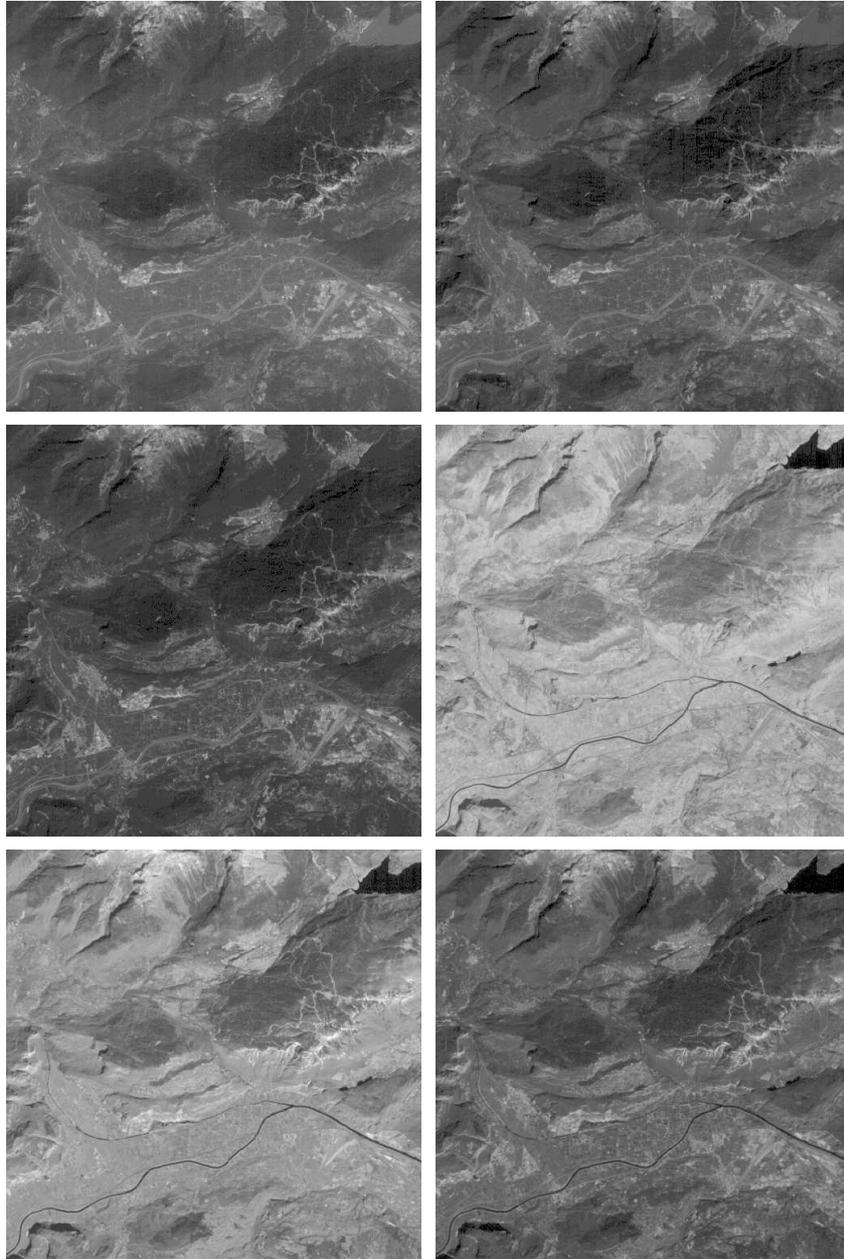
Figure 2.2. 6 *components of a Landsat* 7 *satellite image.*

3) Finally, a non-diagonal $\mathbf{\Gamma}_3^{(\mathbf{n})}$ matrix accounts for cross-channel correlations be-tween co-located noise samples. One can choose $\mathbf{\Gamma}_3^{(\mathbf{n})} = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \rho \\ \vdots & \ddots & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}$,

where $\rho \in (0, 1]$ is the correlation factor between two different noise components.

Our objective is thus to perform a multispectral image denoising under the considered assumptions. In this respect, we will see that the use of a multiscale linear transform such as the wavelet decomposition may be of great use.

2.3. An overview of discrete wavelets and multiscale transforms

2.3.1. *Historical remarks*

A discrete 1D signal r with location index k can be classically written as the following linear expansion :

$$\sum_{k \in \mathbb{Z}} r(k)\delta_k \,,$$

where δ represents the discrete (Kronecker) delta sequence located at 0. While this representation yields optimal sample location, corresponding to the canonical basis, it lacks in providing insights to the inherent signal structure which are beneficial to further processing. Due to the approximate linear nature of many physical processes, signal processing techniques have endeavored to employ a wealth of other suitable linear signal representations, such as the Fourier transform. Under some technical assumptions, the Fourier transform yields the following expansion of the signal :

$$\forall \nu \in [0, 1), \qquad R(\nu) = \sum_{k \in \mathbb{Z}} r(k) \exp(-\imath 2\pi k\nu).$$

Although widely used, the Fourier transform does not however allows us to enlighten the time behaviour of the signal. For signals possessing a certain regularity as well as singularities, localized representations called wavelets have generated a tremendous interest in the past 20 years. The story of wavelets is actually older, since it originated a century ago in a famous paper by A. Haar [HAA 10], who considered decompositions of functions into uniformly convergent series. It is generally considered that discrete wavelets, in their modern form, have emerged in the 1980's. Most of the related works have been nicely gathered in [HEI 06]. We follow here a derivation of wavelet representations based on filter banks, based on the pioneering work by Croisier *et al.* [CRO 76].

2.3.2. *2-band and M-band filter banks*

We consider square summable sequences $(h_m[k])_{k \in \mathbb{Z}}$ with $m \in \{0, \ldots, M-1\}$ representing the impulse responses of M filters. We often characterize each filter by its frequency response H_m, which is the Fourier transform of its impulse response. The basic building block for a 2-band filter bank based decomposition is illustrated in Fig. 2.3 : the digital signal $r(k)$ is decomposed into two frequency bands by a set of filters H_0 et H_1 followed by a decimation by a factor of 2, leading to a pair of coefficient sequences $(r_{1,0}(k))_{k \in \mathbb{Z}}$ and $(r_{1,1}(k))_{k \in \mathbb{Z}}$. The set of analysis filters H_0 and H_1 with its associated decimators is called an analysis filter bank. A reconstructed signal $\tilde{r}$ is obtained from $r_{1,0}$ and $r_{1,1}$ after a factor of 2 upsampling operator followed by filtering through synthesis filters $\widetilde{H}_0$ et $\widetilde{H}_1$ and summation.
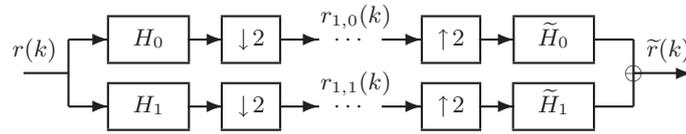
Figure 2.3. *Analysis/synthesis 2-band filter bank.*

The overall construction satisfies the Perfect Reconstruction property (PR) when the signals $\tilde{r}$ and r are equal (eventually up to an integer delay and a non-zero multiplicative factor, which can be incorporated in the filter coefficients). Such a property is verified for non trivial filter families (i.e. including filters with several delays) [SMI 84], whose properties are summarized for instance in [MEY 90, COH 92, DAU 92, MAL 08]. A traditional example is given by the Haar analysis filter bank with analysis filters H_0 and H_1 of length 2 :

$$\big(h_0[0], h_0[1]\big) = \frac{1}{\sqrt{2}}(1,\ 1)$$

$$\big(h_1[0], h_1[1]\big) = \frac{1}{\sqrt{2}}(-1,\ 1).$$

and synthesis filters $\widetilde{H}_0$ and $\widetilde{H}_1$ obtained by symmetry from H_0 and H_1 around the time origin.

Due to the relatively strong constraints imposed on the four filters H_0, H_1, $\widetilde{H}_0$ and $\widetilde{H}_1$ to satisfy the PR property, some authors have proposed a more general structure, named M-band filter banks, based on two sets of $M \geq 2$ analysis and synthesis filter banks, represented in Fig. 2.4.
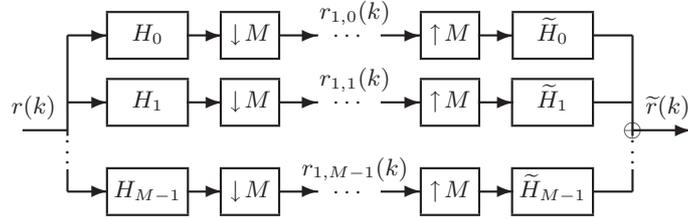
Figure 2.4. *Analysis/synthesis M-band filter bank.*

Similarly, the PR property may be obtained from an appropriate choice of analysis and synthesis filters [VAI 87], with improved flexibility in the design of the filters, since the 2-band case now represents a special instance of M-band filter banks. Moreover, the latter encompasses a large class of standard linear transforms, *e.g.* block transforms.

2.3.3. *Filter bank based multiresolution analysis*
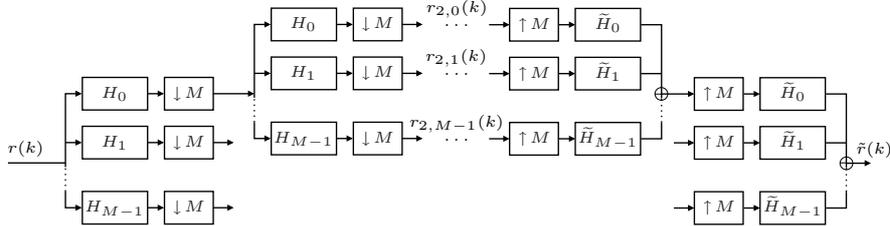
Figure 2.5. *2-level M-band wavelet analysis/synthesis wavelet decomposition.*

A multiresolution analysis of a signal consists of a decomposition where the signal is represented at different scales, allowing us to seize more easily its fine to coarse structures. It has been proved especially useful in signal recovery (denoising, deconvolution and reconstruction) as well as in data compression. A practical multiresolution analysis is obtained by cascading the basic analysis filter bank block. For the generic M-band filter bank case, assume that H_0 and H_{M-1} are a low-pass and a high-pass filter respectively, whereas $H_1, \ldots, H_{M-2}$ are band-pass filters. The low-pass filtering by H_0 followed by decimation yields a first subsampled approximation of the original signal, which may be further decomposed by the same filter bank, as represented in Fig. 2.5. The band-pass and high-pass branches yield subsampled versions of the signal details in different frequency bands, complementary to the low-pass approximation.

From the continuous-time viewpoint, such a multiresolution analysis can be studied in the space $L^2(\mathbb{R})$ of square integrable functions. Successive iterations of the basic M-band filter bank on the low-pass output is interpreted as approximations at resolution level j. The approximation spaces correspond to a decreasing sequence of nested subspaces $(V_j)_{j \in \mathbb{Z}}$ of $L^2(\mathbb{R})$, associated with one scaling function (or father wavelet) $\psi_0 \in L^2(\mathbb{R})$. The multiresolution analysis then corresponds to projections of the continuous-time signal onto subspaces $(W_{j,m})_{j \in \mathbb{Z}, m \in \{1,...,M-1\}}$, associated with $(M-1)$ mother wavelets $\psi_m \in L^2(\mathbb{R})$, $m \in \{1, \ldots, M-1\}$ [STE 93]. These functions are solutions of the following scaling equations :

$$\forall m \in \{0, \ldots, M-1\}, \qquad \frac{1}{\sqrt{M}} \psi_m \left(\frac{t}{M} \right) = \sum_{k=-\infty}^{\infty} h_m[k] \psi_0(t-k).$$

For $m \in \{0, \ldots, M-1\}$, $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$, define the family of functions

$$\psi_{j,m,k}(t) = M^{-j/2} \psi_m(M^{-j}t - k).$$

Then, under orthogonality conditions, we can write :

$$r(t) = \sum_{m \in \{1,...,M-1\}} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} r_{j,m}(k) \psi_{j,m,k}(t)$$

where

$$r_{j,m}(k) = \langle r, \psi_{j,m,k} \rangle$$

and $\langle \cdot, \cdot \rangle$ denotes the standard inner product of $L^2(\mathbb{R})$. The latter expansion is called an M-band wavelet decomposition of r onto the orthonormal wavelet basis $\{ \psi_{j,m,k}, (j,k) \in \mathbb{Z}^2, m \in \{1, \ldots, M-1\} \}$. For more insight on the continuous-time wavelet decomposition, we refer to [MAL 08, FLA 98].

2.3.4. *2D extension*

For simplicity, we only consider separable two-dimensional wavelet transforms which constitute a direct extension of the 1D case. The image is processed in two steps : the filter bank is applied successively to the image rows and columns. Consequently, the obtained 2D wavelets are equal to the tensor product of the 1D wavelets and define a $L^2(\mathbb{R}^2)$ basis. Applying such a transform to multicomponent images consists of applying the 2D transform on each channel b giving rise to the following coefficients : $\forall b \in \{1, ..., B\}, \forall \mathbf{m} = (m_1, m_2) \in \{0, ..., M-1\}^2, \forall j \in \mathbb{Z}$ and $\forall \mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$,

$$r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \langle\langle r^{(b)}, \psi_{j,m_1,k_1} \psi_{j,m_2,k_2} \rangle\rangle$$

where $\langle\langle \cdot, \cdot \rangle\rangle$ denotes the standard inner product of $L^2(\mathbb{R}^2)$. The separable property of the transform allows us to obtain a directional analysis of images, separating the horizontal, vertical and "diagonal" directions.

2.3.5. *Other related representations*

Nevertheless, wavelets suffer from some drawbacks : the first one is a lack of shift invariance whose potential shift-variant edge artifacts are not desirable in applications like denoising. Another drawback of decompositions onto wavelet bases is that they provide a relatively rough directional analysis. Tools that improve the representation of geometric information like textures and edges, and preserve them during processing are thus required. Consequently, during the last decade, many authors proposed more sophisticated representation tools called frames having exact or approximate shift-invariance properties and/or better taking into account geometrical image features. One such frames is simply obtained by dropping the decimation step in the previous filter bank structures, so leading to an undecimated wavelet transform [COI 95, PES 96] which has a redundancy equal to the number J of considered resolution levels. Note that such overcomplete wavelet representations can be built by considering the union of M^J shifted wavelet bases. In this case, cycle spinning denoising techniques may be used which consist of estimating the signal in each basis and averaging the resulting M^J estimates. In order to reduce the computational cost of these decompositions or to better capture geometrical features, other frame representations have been designed. These decompositions provide a local, multiscale, directional analysis of images and they often have a limited redundancy [COI 92, CAN 06, DO 05, MAL 09, CHA 06].

2.3.6. *Related model*

An M-band orthonormal discrete wavelet decomposition over J resolution levels is performed to the observation field $r^{(b)}$ for each channel b. This decomposition produces $M^2 - 1$ wavelet subband sequences $r_{j,\mathbf{m}}^{(b)}$, $\mathbf{m} \in \{0, ..., M-1\}^2 \setminus \{(0,0)\}$, each of size $L_j \times L_j$, at every resolution level j and an additional approximation sequence $r_{J,\mathbf{0}}^{(b)}$ of size $L_J \times L_J$, at the coarsest resolution level J (to simplify our presentation, we consider square images).

On the one hand, the linearity of the Discrete Wavelet Transform (DWT) yields (see. Fig. 2.6) :
$$\forall \mathbf{k} \in \mathbb{K}_j, \quad \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) = \mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) + \mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) \tag{2.3}$$
where $\mathbb{K}_j = \{0, \ldots, L_j - 1\}^2$ and

$$\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \triangleq [s_{j,\mathbf{m}}^{(1)}(\mathbf{k}), \ldots, s_{j,\mathbf{m}}^{(B)}(\mathbf{k})]^\top,$$

$$\mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) \triangleq [n_{j,\mathbf{m}}^{(1)}(\mathbf{k}), \ldots, n_{j,\mathbf{m}}^{(B)}(\mathbf{k})]^\top,$$

$$\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \triangleq [r_{j,\mathbf{m}}^{(1)}(\mathbf{k}), \ldots, r_{j,\mathbf{m}}^{(B)}(\mathbf{k})]^\top.$$

On the other hand, the orthonormality of the DWT preserves the spatial whiteness of $\mathbf{n}_{j,\mathbf{m}}$. More specifically, it is easily shown that the latter field is an i.i.d. $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma}^{(\mathbf{n})})$ random vector process.

A final required assumption is that the random vectors $(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))_{k \in \mathbb{K}}$ are identically distributed for any given value of $(j, \mathbf{m})$.
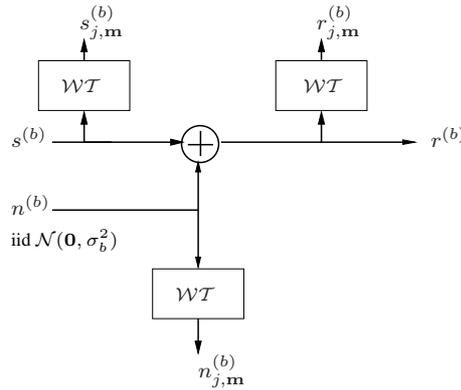
Figure 2.6. *Considered model in the wavelet transform domain.*

2.4. A survey of the most relevant univariate and multivariate denoising methods

Our objective is to build an estimator $\overset{\wedge}{\mathbf{s}}$ of the multichannel image $\mathbf{s}$ from the degraded observation $\mathbf{r}$. The estimating function is denoted by $\mathbf{f}$ and, we have thus $\overset{\wedge}{\mathbf{s}} = \mathbf{f}(\mathbf{r})$. In the present case, as shown in Section 2.2.2, we have to deal with Gaussian noise removal. This is a multivariate estimation problem since the original multichannel image is composed of $B \in \mathbb{N}^*$ components $s^{(b)}$ of size $L \times L$, with $b \in \{1, \ldots, B\}$. Different denoising techniques are presented below. We briefly describe Fourier domain methods and then we focus our attention on wavelet-based methods. But first and foremost, we present the general context we adopt for all the methods operating in the wavelet domain.

2.4.1. *Context in the wavelet domain*

In the wavelet domain, by using the notations defined in Section 2.3, the degradation model (2.2) becomes (2.3). Actually, we consider the more flexible situation where an observation sequence $(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))_{\mathbf{k} \in \mathbb{K}_j}$ of d-dimensional real-valued vectors with $b \in \{1, \ldots, B\}$ and $d > 1$, is given by

$$\forall \mathbf{k} \in \mathbb{K}_j, \qquad \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \overline{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) + \overline{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

and $(\bar{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$ is a zero-mean spatially white Gaussian noise with covariance matrix $\mathbf{\Gamma}^{(\bar{\mathbf{n}}^{(b)})}$, which is assumed to be invertible. The three above vectors will be taken of the form :

$$\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \begin{bmatrix} r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \\ \tilde{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \end{bmatrix}, \quad \bar{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \begin{bmatrix} s_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \\ \tilde{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \end{bmatrix}, \quad \bar{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \begin{bmatrix} n_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \\ \tilde{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \end{bmatrix}$$

where $\tilde{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$, $\tilde{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ and $\tilde{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ are random vectors of dimension $d-1$. These vectors may for example correspond to neighboring variables of the associated scalar variables $r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$, $s_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ and $n_{j,\mathbf{m}}^{(b)}(\mathbf{k})$. In this context, our objective is to estimate $s_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ using the observation sequence $(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$. The vector $\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is called the *Reference Observation Vector* (ROV) from which the following estimate is built :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad \hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = f_{j,\mathbf{m}}^{(b)}\big(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big).$$

Explicit choices of the ROV sequence $(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$ are detailed in the next paragraphs.

2.4.2. *Popular componentwise methods*

A first strategy for denoising a multichannel image is to perform a componentwise processing without taking into account any statistical dependence existing between the channels.

2.4.2.1. *Frequency domain*

A very popular method operating in the frequency domain is the Wiener filter [WIE 49]. This filter is designed so as to minimize the mean square error $\mathsf{E}[|\hat{s}{}^{(b)}(\mathbf{k}) - s^{(b)}(\mathbf{k})|^2]$, for every $b \in \{1, \ldots, B\}$, under the assumption that $s^{(b)}$ is a wide-sense stationary random field. The frequency response of this filter reads

$$\forall \boldsymbol{\nu} \in [0,1)^2, \quad H^{(b)}(\boldsymbol{\nu}) = \frac{S_{s^{(b)}}(\boldsymbol{\nu})}{S_{s^{(b)}}(\boldsymbol{\nu}) + \sigma_b^2}$$

where $S_{s^{(b)}}$ denotes the power spectrum density of $s^{(b)}$ and σ_b is the standard deviation of the noise in channel b. One of the main drawbacks of this method is that it requires the a priori knowledge of the power spectrum density or an empirical estimation of it. Note that a multicomponent version of the Wiener filter has been derived in [ANG 91] by using the multi-input multi-output 2D filter minimizing $\mathsf{E}[\|\overset{\wedge}{\mathbf{s}}(\mathbf{k}) - \mathbf{s}(\mathbf{k})\|^2]$,[1] so taking into account the spectrum density matrix of $\mathbf{s}$.

---

1. $\|.\|$ denotes the classical Euclidean norm of $\mathbb{R}^B$.

However, it may appear more useful to solve the problem in the wavelet transform domain [ATK 03], where we can take advantage of a space-frequency representation of the images. Indeed, noise coefficients are usually distributed over small wavelet coefficients, whereas signal coefficients are concentrated on high magnitude ones. In addition, in [DON 94], Donoho and Jonhstone showed that a both simple and efficient approach for noise removal is available, through wavelet thresholding.

2.4.2.2. *Visushrink*

Visushrink [DON 93] is a componentwise method, which means that the ROV reduces to a scalar :

$$\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = r_{j,\mathbf{m}}^{(b)}(\mathbf{k}).$$

Two kinds of thresholdings are usually employed :

– hard thresholding :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad f_{j,\mathbf{m}}^{(b)}\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \begin{cases} r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) & \text{if } |r_{j,\mathbf{m}}^{(b)}(\mathbf{k})| > \chi^{(b)} \\ 0 & \text{otherwise.} \end{cases}$$

– soft thresholding :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad f_{j,\mathbf{m}}^{(b)}\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \text{sign}(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})) \max\{|r_{j,\mathbf{m}}^{(b)}(\mathbf{k})| - \chi^{(b)}, 0\}$$

$$= \Big(\frac{|r_{j,\mathbf{m}}^{(b)}(\mathbf{k})| - \chi^{(b)}}{|r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|}\Big)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \tag{2.4}$$

where $\text{sign}(\cdot)$ is the signum function.

These two shrinkage rules are illustrated in Fig. 2.7. The problem here is to find the best threshold value $\chi^{(b)} > 0$. In [DON 93], the authors have derived the so-called universal threshold $\chi^{(b)} = 2\sigma_b\sqrt{\log(L)}$ which relies on the fact that the maximum values of any set of L^2 independent random variables identically distributed as $\mathcal{N}(0, \sigma_b^2)$ are smaller than the proposed threshold $\chi^{(b)}$ with a high probability [MAL 08, p. 556].

2.4.3. *Extension to block-based method*

In order to take into account correlations between wavelet coefficients, some authors have proposed to apply a block shrinkage. More precisely, in [CAI 01], it is proposed to exploit the spatial dependences, which corresponds to the following choice of the ROV :

$$\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), r_{j,\mathbf{m}}^{(b)}(\mathbf{k} - \mathbf{k}_1), \dots, r_{j,\mathbf{m}}^{(b)}(\mathbf{k} - \mathbf{k}_{d-1})]^\top$$

where $\mathbf{k}_1, \dots, \mathbf{k}_{d-1}$ allow us to define the neighborhood of interest for the pixel $\mathbf{k}$.
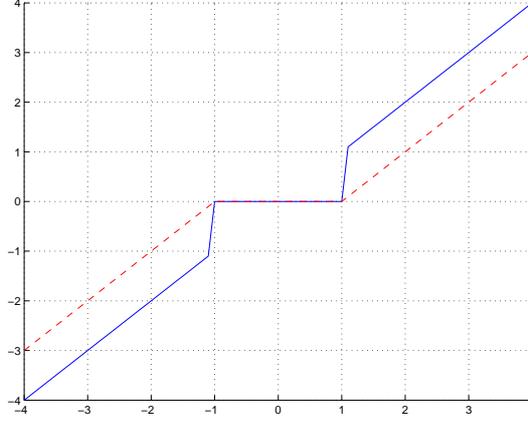
Figure 2.7. *Hard (continuous line) and soft (dashed line) thresholdings.*

The associated shrinkage rule named "NeighBlock" is given by

$$\overset{\curlywedge(b)}{s}_{j,\mathbf{m}}(\mathbf{k}) = \left(\frac{\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^2 - \bar{\chi} d \sigma_b^2}{\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^2} \right)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

where $\bar{\chi} > 0$ and d is the number of components in the ROV.

In [ŞEN 02], interscale dependencies have been exploited by defining the following ROV :

$$\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), r_{j+1,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{M} \rceil), \dots, r_{J,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{M^{J-j}} \rceil)]^\top.$$

The associated estimator called "bivariate shrinkage" is defined by :

$$\overset{\curlywedge(b)}{s}_{j,\mathbf{m}}(\mathbf{k}) = \left(\frac{\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\| - \frac{\sqrt{3}\sigma_b^2}{\sigma_{s^{(b)}}}}{\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|} \right)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

where $\sigma_{s^{(b)}} > 0$. It can be derived by a Maximum A Posteriori (MAP) rule by considering as a prior model for the wavelet coefficients the non-Gaussian bivariate probability density function

$$p(s_{j,\mathbf{m}}^{(b)}(\mathbf{k}), s_{j+1,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{2} \rceil)) \propto \exp \left(- \frac{\sqrt{3}}{\sigma_{s^{(b)}}} \sqrt{|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^2 + |s_{j+1,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{2} \rceil)|^2} \right).$$

Note that interscale dependencies are also taken into account in [SCH 04] where a multivalued image wavelet thresholding is performed. Other works developed Bayesian estimation procedures imposing a prior on the noise-free data.

2.4.4. *Bayesian multichannel approaches*

As previously mentioned, Bayesian approaches require a prior data statistical modeling.

2.4.4.1. *Bernoulli-Gaussian priors*

A Bernoulli-Gaussian (BG) prior is an appropriate model to reflect the sparsity of the wavelet representation of natural images [ABR 98, LEP 99]. With this statistical model, some authors derived Bayesian estimates [BEN 03, ELM 05].

Let us first present the method proposed in [BEN 03, ELM 05]. In each subband of index $(j, \mathbf{m})$, the probability distribution $p_{j,\mathbf{m}}$ of the coefficients $(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$ can be written as follows :

$$\forall \mathbf{u} \in \mathbb{R}^B, \quad p_{j,\mathbf{m}}(\mathbf{u}) = (1 - \epsilon_{j,\mathbf{m}})\delta(\mathbf{u}) + \epsilon_{j,\mathbf{m}}\, g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}(\mathbf{u})$$

where $g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}$ denotes the $\mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}})$ multivariate normal probability density function. The mixture parameter $\epsilon_{j,\mathbf{m}}$ corresponds to the probability that a coefficient vector $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})$ contains useful information. In order to avoid degenerated MAP estimates, it is used to couple the multivariate prior model with hidden random variables $q_{j,\mathbf{m}}(\mathbf{k})$. The sequence $(q_{j,\mathbf{m}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}_j}$ is an i.i.d. binary sequence of random variables defining the following conditional densities : for every $\mathbf{k} \in \mathbb{K}_j$,

$$p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 0) = \delta(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})),$$
$$p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 1) = g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))$$

with $\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1) = \epsilon_{j,\mathbf{m}} \in [0, 1]$. In practice, the hyperparameters $\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}$ and $\epsilon_{j,\mathbf{m}}$ related to the BG priors can be estimated by a moment method or an Expectation Maximization (EM) technique. As the noise is Gaussian, the following conditional probability densities are easily derived : for every $\mathbf{k} \in \mathbb{K}_j$,

$$\begin{cases} p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 0) & = & g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{n})}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) \\ p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 1) & = & g_{\mathbf{0},\boldsymbol{\Gamma}^{(\mathbf{n})}+\boldsymbol{\Gamma}^{(\mathbf{s})}_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})). \end{cases}$$

Thus, a two-step estimation procedure can be used for noise removal :

1) For every $\mathbf{k} \in \mathbb{K}_j$, the estimate $\hat{q}_{j,\mathbf{m}}(\mathbf{k})$ of $q_{j,\mathbf{m}}(\mathbf{k})$ is set to 1 if :

$$\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 0 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) < \mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})),$$

otherwise $\hat{q}_{j,\mathbf{m}}(\mathbf{k})$ is set to 0. This implies that :

$$\hat{q}_{j,\mathbf{m}}(\mathbf{k}) = \begin{cases} 1 & \text{if } \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})^{\top}\mathbf{M}_{j,\mathbf{m}}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) > \chi_{j,\mathbf{m}}, \\ 0 & \text{otherwise} \end{cases}$$

where $\mathbf{M}_{j,\mathbf{m}}$ is the semi-definite positive matrix :

$$\mathbf{M}_{j,\mathbf{m}} = (\mathbf{\Gamma}^{(\mathbf{n})})^{-1} - (\mathbf{\Gamma}_{j,\mathbf{m}}^{(\mathbf{s})} + \mathbf{\Gamma}^{(\mathbf{n})})^{-1},$$

and the threshold $\chi_{j,\mathbf{m}}$ is defined by

$$\chi_{j,\mathbf{m}} = 2\ln\left(\frac{1 - \epsilon_{j,\mathbf{m}}}{\epsilon_{j,\mathbf{m}}}\right) + \ln\left(\frac{\mid \mathbf{\Gamma}_{j,\mathbf{m}}^{(\mathbf{s})} + \mathbf{\Gamma}^{(\mathbf{n})}\mid}{\mid \mathbf{\Gamma}^{(\mathbf{n})}\mid}\right)$$

where $|\mathbf{A}|$ denotes the determinant of matrix $\mathbf{A}$.

2) On the one hand, if $\hat{q}_{j,\mathbf{m}}(\mathbf{k}) = 0$, it is expected that the related observation is dominated by the noise, according to the definition of the hidden variables. Hence, it is natural to set $\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k}) = \mathbf{0}$. On the other hand, if $\hat{q}_{j,\mathbf{m}}(\mathbf{k}) = 1$, the Bayesian estimate of $\mathbf{s}_{j,\mathbf{m}}$ minimizing a quadratic cost is computed. It corresponds to the a posteriori conditional mean. The posterior distribution is Gaussian as the bivariate distribution of $(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))$ is Gaussian when $q_{j,\mathbf{m}}(\mathbf{k}) = 1$. It is easy to check that :

$$\forall \mathbf{k} \in \mathbb{K}_j, \quad \mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), q_{j,\mathbf{m}}(\mathbf{k}) = 1] = \mathbf{Q}_{j,\mathbf{m}}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})$$

where

$$\mathbf{Q}_{j,\mathbf{m}} \overset{\triangle}{=} \mathbf{\Gamma}_{j,\mathbf{m}}^{(\mathbf{s})}(\mathbf{\Gamma}_{j,\mathbf{m}}^{(\mathbf{s})} + \mathbf{\Gamma}^{(\mathbf{n})})^{-1}.$$

It appears that the estimator amounts to a shrinkage rule that performs a tradeoff between a linear estimation in the sense of a minimum mean square error and a hard thresholding.

An alternate approach to this two-step estimation procedure is the use of the a posteriori conditional mean which, for every $\mathbf{k} \in \mathbb{K}_j$, can be expressed as

$$\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})] = \mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), q_{j,\mathbf{m}}(\mathbf{k}) = 1]\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))$$

since $p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), q_{j,\mathbf{m}}(\mathbf{k}) = 0) = \delta(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}))$. Besides, we can write :

$$\begin{aligned}\forall \mathbf{k} \in \mathbb{K}_j, \ \mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1 \mid \mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) &= \frac{p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \mid q_{j,\mathbf{m}}(\mathbf{k}) = 1)\mathsf{P}(q_{j,\mathbf{m}}(\mathbf{k}) = 1)}{p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))} \\ &= \frac{\epsilon_{j,\mathbf{m}}\theta_1(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))}{\epsilon_{j,\mathbf{m}}\theta_1(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) + (1 - \epsilon_{j,\mathbf{m}})\theta_0(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))} \\ &\overset{\triangle}{=} \gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\end{aligned}$$

with the following definitions :

$$\theta_0 \overset{\triangle}{=} g_{\mathbf{0},\mathbf{\Gamma}^{(\mathbf{n})}}, \quad \theta_1 \overset{\triangle}{=} g_{\mathbf{0},\mathbf{\Gamma}_{j,\mathbf{m}}^{(\mathbf{s})} + \mathbf{\Gamma}^{(\mathbf{n})}}.$$

The optimal mean-square Bayesian estimate can be easily deduced :

$$\forall \mathbf{k} \in \mathbb{K}, \qquad \overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k}) = \gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathbf{Q}_{j,\mathbf{m}}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}). \qquad (2.5)$$

Note that in [ELM 05], interscale dependencies are taken into account in addition to cross-channel ones.

2.4.4.2. *Laplacian mixture model*

A Bayesian componentwise approach was proposed by Pižurica and Philips [PIŽ 06]. They have described a simple way for applying it to a multicomponent image, when the noise is componentwise decorrelated ($\mathbf{\Gamma^{(n)}} = \mathbf{\Gamma_1^{(n)}} = \sigma^2 \, \mathbf{I}_B$). For each component at each resolution level and in each oriented subband, the principle is to consider a mixture of two truncated Generalized Gaussian (GG) (also called generalized Laplacian) distributions where a Bernoulli random variable controls the switching between the central part of the distribution and its tails. More precisely, for each component b, the authors have considered as a prior the GG distribution :

$$\forall u \in \mathbb{R}, \quad p_{j,\mathbf{m}}^{(b)}(u) = \frac{\left(\lambda_{j,\mathbf{m}}^{(b)}\right)^{1/\beta_{j,\mathbf{m}}^{(b)}}}{2\Gamma(1/\beta_{j,\mathbf{m}}^{(b)})} \exp(-\lambda_{j,\mathbf{m}}^{(b)}|u|^{\beta_{j,\mathbf{m}}^{(b)}})$$

where $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}dt$ is the Gamma function, $\lambda_{j,\mathbf{m}}^{(b)} > 0$ is the scale parameter and $\beta_{j,\mathbf{m}}^{(b)} > 0$ is the shape parameter. It is worth pointing out that the prior hyperparameters can be easily estimated from the fourth moments of the noisy coefficients $r_{j,\mathbf{m}}^{(b)}$ [SIM 96]. Then, they have defined a signal of interest as a noise-free coefficient which exceeds a given threshold $T_{j,\mathbf{m}}^{(b)}$. To estimate the signal of interest from the noisy observations, they have introduced a sequence of Bernoulli variables $q_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ associated with the two hypotheses $\mathcal{H}_0$ "the noise-free signal is not of interest" and $\mathcal{H}_1$ "the noise-free signal is of interest" :

$$\mathcal{H}_0 : |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| \leq T_{j,\mathbf{m}}^{(b)} \qquad \text{and} \qquad \mathcal{H}_1 : |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| > T_{j,\mathbf{m}}^{(b)}.$$

In other words, if $q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1$, the coefficient $s_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is of interest. Therefore, it is possible to compute $\mathsf{P}\left(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\right)$:

$$\mathsf{P}\left(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\right) = 1 - \Gamma_{\mathrm{inc}}\left(\lambda_{j,\mathbf{m}}^{(b)}(T_{j,\mathbf{m}}^{(b)})^{\beta_{j,\mathbf{m}}^{(b)}}, 1/\beta_{j,\mathbf{m}}^{(b)}\right)$$

where Γ_{inc} is the incomplete Gamma function. Hence, the following conditional probabilities can be easily derived :

$$p(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0) = \begin{cases} C_0 \exp(-\lambda_{j,\mathbf{m}}^{(b)}|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^{\beta_{j,\mathbf{m}}^{(b)}}) & \text{if } |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| \leq T_{j,\mathbf{m}}^{(b)} \\ 0 & \text{otherwise} \end{cases}$$

$$p\left(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\right) = \begin{cases} 0 & \text{if } |s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| \leq T_{j,\mathbf{m}}^{(b)} \\ C_1 \exp\left(-\lambda_{j,\mathbf{m}}^{(b)}|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^{\beta_{j,\mathbf{m}}^{(b)}}\right) & \text{otherwise} \end{cases}$$

where C_0 and C_1 are normalizing constants. For multivalued images, the authors proposed to exploit a local information from the different channels by defining the band

activity indicator $z_{j,\mathbf{m}}$ as the average over the B channels of the magnitudes of the B homologous noisy coefficients :

$$z_{j,\mathbf{m}}(\mathbf{k}) = \frac{1}{B} \sum_{b'=1}^{B} |r_{j,\mathbf{m}}^{(b')}(\mathbf{k})|.$$

Consequently, the local minimum mean square estimator is

$$\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})]$$
$$= \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})\big)\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1]$$
$$+ \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})\big)\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0]$$

by assuming that the coefficients $(r_{j,\mathbf{m}}^{(b')}(\mathbf{k}))_{1 \leq b' \leq B}$ are independent conditionally to $\mathcal{H}_0$ or $\mathcal{H}_1$. The sparseness of the wavelet representation allows us to consider that the second term takes very low values and to approximate $\mathsf{E}[s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1]$ by $r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$. Then, the following estimate (called ProbShrink) is derived for each subband $(j, \mathbf{m})$ and channel b :

$$\overset{\wedge (b)}{s}_{j,\mathbf{m}}(\mathbf{k}) = \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1|r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})\big)r_{j,\mathbf{m}}^{(b)}(\mathbf{k}).$$

After some manipulations, the explicit expression of the ProbShrink estimate is :

$$\overset{\wedge (b)}{s}_{j,\mathbf{m}}(\mathbf{k}) = \frac{\eta\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\xi\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\mu}{1 + \eta\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\xi\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big)\mu}r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

where

$$\eta\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \frac{p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}{p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)}$$

$$\xi\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big) = \frac{p\big(z_{j,\mathbf{m}}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}{p\big(z_{j,\mathbf{m}}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)}$$

$$\mu = \frac{\mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}{1 - \mathsf{P}\big(q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)}.$$

In practice, it is used to set $T_{j,\mathbf{m}}^{(b)} = \sigma$ and the computation of the conditional densities $p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)$ and $p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)$ can be deduced from $p\big(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})| q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 1\big)$ and $p\big(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = 0\big)$. Indeed, we have : $\forall \ell \in$

$\{0, 1\}$,

$$p\big(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \ell\big)$$

$$= \int_{\mathbb{R}} g_{0,\sigma^2}(r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) - s_{j,\mathbf{m}}^{(b)}(\mathbf{k}))p\big(s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|q_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \ell\big)ds_{j,\mathbf{m}}^{(b)}(\mathbf{k}),$$

where g_{0,σ^2} denotes the $\mathcal{N}(0, \sigma^2)$ probability density. For the statistical characterization of the band activity indicator $z_{j,\mathbf{m}}$, it is assumed that the coefficients $(r_{j,\mathbf{m}}^{(b')}(\mathbf{k}))_{1 \leq b' \leq B}$ are all distributed according to either $\mathcal{H}_0$ or $\mathcal{H}_1$. Therefore, the conditional density of $z_{j,\mathbf{m}}(\mathbf{k})$ given $q_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is the B times convolution of the conditional densities of $|r_{j,\mathbf{m}}^{(b')}(\mathbf{k})|$ given $q_{j,\mathbf{m}}^{(b)}(\mathbf{k})$.

2.4.4.3. *Gaussian scale mixture model*

Inspired by the work for monochannel images [POR 03], Scheunders and de Backer [SCH 07] have assumed that the unknown signal vector $\mathbf{s}_{j,\mathbf{m}}$ can be expressed as follows :

$$\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) = \sqrt{z_{j,\mathbf{m}}(\mathbf{k})}\mathbf{u}_{j,\mathbf{m}}(\mathbf{k})$$

where $\mathbf{u}_{j,\mathbf{m}}(\mathbf{k})$ is a zero-mean normal vector $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma}^{(\mathbf{u}_{j,m})})$ and $z_{j,\mathbf{m}}(\mathbf{k})$ is a positive scalar random variable independent of $\mathbf{u}_{j,\mathbf{m}}(\mathbf{k})$. The prior probability density function of $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})$ can be considered as a Gaussian Scale Mixture (GSM) as we have :

$$p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})) = \int_0^\infty p(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|z_{j,\mathbf{m}}(\mathbf{k}))p(z_{j,\mathbf{m}}(\mathbf{k}))dz_{j,\mathbf{m}}(\mathbf{k}).$$

Indeed, $p(z_{j,\mathbf{m}}(\mathbf{k}))$ corresponds to a mixing density and $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|z_{j,\mathbf{m}}(\mathbf{k}) \sim \mathcal{N}(\mathbf{0}, z_{j,\mathbf{m}}(\mathbf{k})\mathbf{\Gamma}^{(\mathbf{u}_{j,m})})$. The posterior mean estimate $\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})]$ can then be expressed as :

$$\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})] =$$

$$\int_0^\infty p(z_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})]dz_{j,\mathbf{m}}(\mathbf{k}).$$

On the one hand, we note that the involved conditional posterior mean in the right hand side of the previous equation corresponds to the Wiener estimate :

$$\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), z_{j,\mathbf{m}}(\mathbf{k})] = z_{j,\mathbf{m}}(\mathbf{k})\mathbf{\Gamma}^{(\mathbf{u}_{j,m})}\big(z_{j,\mathbf{m}}(\mathbf{k})\mathbf{\Gamma}^{(\mathbf{u}_{j,m})} + \mathbf{\Gamma}^{(\mathbf{n})}\big)^{-1}\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}).$$

On the other hand, Bayes rule allows us to express $p(z_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))$ as :

$$p(z_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})) \propto p(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})|z_{j,\mathbf{m}}(\mathbf{k}))p(z_{j,\mathbf{m}}(\mathbf{k})).$$

In practice, in order to calculate this conditional probability, specifying the distribution of $z_{j,\mathbf{m}}(\mathbf{k})$ is required. In the case of monochannel images, Portilla *et al.* [POR 03] have proposed the (improper) Jeffrey's prior :

$$p(z_{j,\mathbf{m}}(\mathbf{k})) \propto 1/z_{j,\mathbf{m}}(\mathbf{k}).$$

In [SCH 07], another alternative has been considered when a single component noise-free image y is available, which is employed as an ancillary variable. It is assumed that the joint distribution of $(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}), y_{j,\mathbf{m}}(\mathbf{k}))$ is a GSM. Consequently, it is found that the probability of $\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})$ conditioned on $y_{j,\mathbf{m}}(\mathbf{k})$ is a GSM. Then, by adopting the same strategy as previously (development of the conditional posteriori means and application of Bayes rule), it has been shown that the Bayesian estimate $\mathsf{E}[\mathbf{s}_{j,\mathbf{m}}(\mathbf{k})|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), y_{j,\mathbf{m}}(\mathbf{k})]$ can be explicitly calculated.

2.4.5. *Variational approaches*

In wavelet-based variational approaches, the wavelet coefficients $(\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}})_{j,\mathbf{m}}$ of the denoised multichannel image $\overset{\wedge}{\mathbf{s}}$ are obtained by minimizing the objective function :

$$(\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}} \mapsto \frac{1}{2}\sum_{\mathbf{k}\in\mathbb{K}}\big((\mathbf{u}(\mathbf{k})-\mathbf{r}(\mathbf{k}))\big)^{\top}\mathbf{Q}^{-1}\big((\mathbf{u}(\mathbf{k})-\mathbf{r}(\mathbf{k}))\big) + h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) \quad (2.6)$$

where $\mathbf{u}$ designates a generic multispectral image with B channels of size $L \times L$ and $(\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}$ is its wavelet representation, $\mathbf{Q}$ is a definite positive matrix of size $B \times B$ and h is some appropriate function. The first quadratic term represents a data fidelity measure with respect to the observation model. The function h usually corresponds to a penalty term (also called regularization term) used to incorporate prior information about the original image $\mathbf{s}$. In the particular case when $\mathbf{Q} = \mathbf{\Gamma}^{(\mathbf{n})}$ and $\exp(-h(\cdot))$ is (up to a multiplicative constant) a probability density function, this approach amounts to a MAP approach where h is the potential associated with the prior distribution adopted for the wavelet coefficients of $\mathbf{s}$.

Some classical choices for h are :

$- h = i_C$ where i_C is the indicator function of a closed convex set C ($i_C\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = 0$ if $(\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}} \in C$ and $+\infty$ otherwise). This function imposes a hard constraint on the solution (e.g. the positivity of its pixel values).

$- h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = \lambda \sum_{b=1}^{B}\sum_{\mathbf{k}\in\mathbb{K}}\|\nabla u^{(b)}(\mathbf{k})\|_{2}^{2}$ where $\nabla u^{(b)}$ is a discrete version of the gradient of $u^{(b)}$, $\|\cdot\|_2$ is the Euclidean norm of $\mathbb{R}^2$ and λ is a positive factor. This corresponds to a Tikhonov regularization [TIK 63], which serves to ensure the smoothness of the denoised image.

$- h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = \lambda \sum_{b=1}^{B}\sum_{\mathbf{k}\in\mathbb{K}}\|\nabla u^{(b)}(\mathbf{k})\|_{2}$ which can be viewed as a discrete version of a Total Variation (TV) penalization [RUD 92, TEB 98]. A multivariate version of this function was also proposed in [BLO 98] (see also [AUJ 06] for extensions).

$-h\big((\mathbf{u}_{j,\mathbf{m}})_{j,\mathbf{m}}\big) = \sum_{b=1}^{B} \sum_{(j,\mathbf{m})} \lambda_{j,\mathbf{m}}^{(b)} \sum_{\mathbf{k}\in\mathbb{K}_j} |u_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^{\beta_{j,\mathbf{m}}^{(b)}}$ where the exponents $\beta_{j,\mathbf{m}}^{(b)}$ and the scaling parameters $\lambda_{j,\mathbf{m}}^{(b)}$ are positive. This function is the potential of an independent Generalized Gaussian distribution for the wavelet coefficients [CHA 00, ANT 02]. Under some assumptions, it can also be interpreted as a regularity measure in terms of a Besov norm [CHA 98, LEP 01]. A particular case of interest is when the exponents are all equal to 1. The resulting ℓ^1 norm is used to promote the sparsity of the wavelet representation of the solution, by setting many of its coefficients to zero [TRO 06].

Note that a composite form consisting of a sum of several of the above functions can also be considered. Each of these penalizations may indeed present its own advantages and drawbacks. For example, the Tikhonov regularization tends to oversmooth image edges whereas the TV penalization may introduce visual staircase effects.
Another point to be emphasized is that the solution to the minimization of (2.6) can be obtained in different manners, according to the choice of h. Sometimes, an explicit expression of $\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}$ can be derived. This arises, in particular, for Tikhonov regularization or, when $\mathbf{Q}$ is a diagonal matrix and a GG potential is used with $\beta_{j,\mathbf{m}}^{(b)} \in \{1, 4/3, 3/2, 2, 3, 4\}$ [CHA 07]. In other cases, iterative algorithms must be applied to compute $\overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}$. When a hard constraint is imposed, the constraint set C often is decomposed as the intersection of elementary closed convex sets and iterative techniques of projection onto each of these sets can be applied [COM 96, COM 03, COM 04]. One of such approaches is the well-known Projections Onto Convex Sets (POCS) algorithm. Provided that h is a convex function, other iterative convex optimization algorithms can be employed to bring efficient numerical solutions to the considered large-size variational problems (see [CHA 07, COM 07, COM 08] and references therein). These optimization methods are also useful when a redundant frame representation of the data is used instead of an orthonormal one.

2.4.6. *Stein-based approaches*

As already mentioned, the main problem in wavelet coefficient thresholding is the determination of the threshold value. Furthermore, in practice, it often appears preferable not to require some prior knowledge about the original data. A solution to alleviate such problems is to invoke Stein's principle [STE 81] which can be formulated as follows :

Proposition 1 *Using the notations of Section 2.4.1, let $f_{j,\mathbf{m}}^{(b)} : \mathbb{R}^d \to \mathbb{R}$ be a continuous, almost everywhere differentiable function such that :*

$$\forall \boldsymbol{\theta} \in \mathbb{R}^d,$$

$$\lim_{\|\mathbf{t}\| \to +\infty} f_{j,m}^{(b)}(\mathbf{t}) \exp\left(-\frac{(\mathbf{t} - \boldsymbol{\theta})^\top (\boldsymbol{\Gamma}^{(\overline{\mathbf{n}}^{(b)})})^{-1}(\mathbf{t} - \boldsymbol{\theta})}{2}\right) = 0;$$

$$\mathsf{E}[|f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))|^2] < +\infty \quad and \quad \mathsf{E}\Big[\Big\|\frac{\partial f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))}{\partial \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})}\Big\|\Big] < +\infty.$$

Then,

$$\mathsf{E}[f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))s_{j,\mathbf{m}}^{(b)}(\mathbf{k})] = \mathsf{E}[f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))r_{j,\mathbf{m}}^{(b)}(\mathbf{k})] -$$

$$\mathsf{E}\Big[\frac{\partial f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))}{\partial \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})}\Big]^\top \mathsf{E}[\overline{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})n_{j,\mathbf{m}}^{(b)}(\mathbf{k})]. \quad (2.7)$$

2.4.6.1. *Expression of the quadratic risk using Stein's formula*

A standard criterion in signal/image processing is the mean square error (also called quadratic risk) defined as

$$\mathsf{E}[\|\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) - \overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k})\|^2]$$

$$= \sum_{b=1}^{B} \mathsf{E}[|s_{j,\mathbf{m}}^{(b)}(\mathbf{k}) - f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))|^2]$$

$$= \sum_{b=1}^{B} \Big(\mathsf{E}[|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^2] + \mathsf{E}[|f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))|^2] - 2\mathsf{E}[f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))s_{j,\mathbf{m}}^{(b)}(\mathbf{k})]\Big)$$

$$= \mathsf{E}[\|\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) - \overset{\wedge}{\mathbf{s}}_{j,\mathbf{m}}(\mathbf{k})\|^2] + 2\sum_{b=1}^{B} \mathsf{E}\Big[\frac{\partial f_{j,\mathbf{m}}^{(b)}(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}))}{\partial \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})}\Big]^\top \mathsf{E}[\overline{\mathbf{n}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})n_{j,\mathbf{m}}^{(b)}(\mathbf{k})].$$

The last equality has been obtained by using Stein's formula (see (2.7)). The advantage of the latter expression is that it only depends on the observed data and no longer on the unknown original data. This means that no prior information is necessary to build an estimate minimizing the quadratic risk. Another remark to be formulated is that the minimization of the risk will be performed by optimizing a limited number of variables parameterizing the estimator $f_{j,\mathbf{m}}^{(b)}$ (e.g. a threshold value). Due to the assumption of stationarity made at the end of Section 2.3, it can be noticed that the above risk is independent of $\mathbf{k}$, so that it will denoted $R_{j,\mathbf{m}}$ in the next paragraphs.

2.4.6.2. *Existing methods based on Stein's principle*

We will now present a variety of Stein-based estimators (denoted hereafter by the SURE prefix for Stein's Unbiased Risk Estimate) ranging from componentwise thresholding rules to more sophisticated block-based methods.

– SUREshrink [DON 95] is a very popular componentwise method. It thus operates separately on each channel b. It consists of a soft-thresholding as given by (2.4), where subband-dependent threshold values $\chi_{j,\mathbf{m}}^{(b)}$ are computed in order to minimize the contribution $R_{j,\mathbf{m}}^{(b)}$ of the b-th channel to the quadratric risk $R_{j,\mathbf{m}}$. More exactly, the classical sample estimate $\tilde{R}_{j,\mathbf{m}}^{(b)}$ of $R_{j,\mathbf{m}}^{(b)}$ is employed. Then, the variables $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k})|$ are sorted in descending order, so that $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_1)| \geq |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_2)| \geq \ldots \geq |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{L_j^2})|$. It can then be shown that if, for $i_0 \in \{2, \ldots, L_j^2\}$, $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0-1})| > \chi_{j,\mathbf{m}}^{(b)} \geq |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0})|$, the risk estimate $\tilde{R}_{j,\mathbf{m}}^{(b)}$ is a second-order binomial increasing function of $\chi_{j,\mathbf{m}}^{(b)}$ over the considered interval. It can be deduced that the optimal threshold value over the interval $[|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0})|, |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0-1})|)$ is $|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{i_0})|$. So, the optimal threshold value over $\mathbb{R}_+$ is found by evaluating $\tilde{R}_{j,\mathbf{m}}^{(b)}$ on the finite set of candidate values $\{|\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_1)|, \ldots, |\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{L_j^2})|, 0\}$.

– Starting from the Bayesian estimate given by (2.5), the authors in [BEN 05] derived a multivariate shrinkage estimator called SUREVECT. More precisely, the parameters $\epsilon_{j,\mathbf{m}}$ and $\mathbf{Q}_{j,\mathbf{m}}$ are directly adjusted so as to minimize $R_{j,\mathbf{m}}$. For simplicity, the matrix $\mathbf{\Gamma}_{j,\mathbf{m}}^{(s)}$ involved in the expression of the θ_1 function has not been included in the set of parameters estimated with a SURE procedure. An approximate value (like that obtained by a moment method) can be employed. It can be proved that

$$R_{j,\mathbf{m}} = \mathrm{tr}\big((\mathbf{Q}_{j,\mathbf{m}} - \mathbf{A}_{j,\mathbf{m}})\mathbf{C}_{j,\mathbf{m}}(\mathbf{Q}_{j,\mathbf{m}} - \mathbf{A}_{j,\mathbf{m}})^\top - \mathbf{A}_{j,\mathbf{m}}\mathbf{C}_{j,\mathbf{m}}\mathbf{A}_{j,\mathbf{m}}^\top\big)$$

with

$$
\begin{aligned}
\mathbf{A}_{j,\mathbf{m}} &\triangleq \mathbf{B}_{j,\mathbf{m}}\big(\mathbf{C}_{j,\mathbf{m}}\big)^{-1} \\
\mathbf{B}_{j,\mathbf{m}} &\triangleq \mathsf{E}[\gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))^\top] - \mathbf{\Gamma}\big(\mathsf{E}[\gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))]\mathbf{I}_B \\
&\quad + \mathsf{E}[\mathbf{\nabla}\gamma_{\epsilon_{j,\mathbf{m}}}(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))^\top]\big) \\
\mathbf{C}_{j,\mathbf{m}} &\triangleq \mathsf{E}[\gamma_{\epsilon_{j,\mathbf{m}}}^2(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))\mathbf{r}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}))^\top]
\end{aligned}
$$

where it has been assumed that $\mathbf{C}_{j,\mathbf{m}}$ is invertible. It is easily shown that setting $\mathbf{Q}_{j,\mathbf{m}} = \mathbf{A}_{j,\mathbf{m}}$ allows us to minimize $R_{j,\mathbf{m}}$ and that the optimal value $\epsilon_{j,\mathbf{m}}$ within $[0, 1]$ should maximize $\mathrm{tr}(\mathbf{B}_{j,\mathbf{m}}\mathbf{C}_{j,\mathbf{m}}^{-1}\mathbf{B}_{j,\mathbf{m}}^\top)$. It appears that this amounts to a mere optimization of a one-variable function which can be performed with conventional numerical optimization tools.

– More recently, we proposed an approach to take into account spatial and cross-channel dependences by using Stein's principle [CHA 08]. More precisely, the proposed estimator takes the form :

$$\hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = f_{j,\mathbf{m}}^{(b)}(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = \eta_{\chi_{j,\mathbf{m}}^{(b)}}(\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^{\alpha_{j,\mathbf{m}}^{(b)}}) \, (\boldsymbol{\rho}_{j,\mathbf{m}}^{(b)})^{\top} \bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

where $\eta_{\chi_{j,\mathbf{m}}^{(b)}}(\cdot)$ is the thresholding function given by

$$\forall \tau \in \mathbb{R}_+, \qquad \eta_{\chi_{j,\mathbf{m}}^{(b)}}(\tau) = \left(\frac{\tau - \chi_{j,\mathbf{m}}^{(b)}}{\tau} \right)_+$$

and $\chi_{j,\mathbf{m}}^{(b)} \geq 0$, $\alpha_{j,\mathbf{m}}^{(b)} > 0$ and $\boldsymbol{\rho}_{j,\mathbf{m}}^{(b)} \in \mathbb{R}^d$. The vector $\boldsymbol{\rho}_{j,\mathbf{m}}^{(b)}$ corresponds to a linear parameter. This estimator generalizes the block-based estimator proposed in [CAI 01, ŞEN 02] as well as linear estimators. The parameters are computed similarly to the SUREShrink approach (more details can be found in [CHA 08]). Although the choice of the ROV is quite flexible, examples of some ROVs corresponding to spatial and across channel neigborhoods were considered in [CHA 08].

– Another efficient approach is SURE-LET (the SURE-Linear Expansion of Threshold [LUI 08]). Then, the considered expression of the estimator is

$$f_{j,\mathbf{m}}^{(b)}(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = a_{j,\mathbf{m}}^{(b),1} \, r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) + a_{j,\mathbf{m}}^{(b),2} \left(1 - \exp\left(- \frac{(r_{j,\mathbf{m}}^{(b)}(\mathbf{k}))^8}{(\omega_{j,\mathbf{m}}^{(b)})^8} \right) \right) r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$$

where $\omega_{j,\mathbf{m}}^{(b)} > 0$ and $a_{j,\mathbf{m}}^{(b),1}$, $a_{j,\mathbf{m}}^{(b),2}$ are real-valued weighting factors. In the case when $(a_{j,\mathbf{m}}^{(b),1}, a_{j,\mathbf{m}}^{(b),2}) = (1, 0)$ a linear estimator is obtained, whereas in the case when $(a_{j,\mathbf{m}}^{(b),1}, a_{j,\mathbf{m}}^{(b),2}) = (0, 1)$ and $\omega_{j,\mathbf{m}}^{(b)} \gg 1$, a thresholding-like rule is obtained. Note that other possible choices of the nonlinear part of the estimator have been proposed in [PES 97, RAP 08]. This estimator possesses a number of attractive features :

  - the determination of the optimal values of $a_{j,\mathbf{m}}^{(b),1}$ and $a_{j,\mathbf{m}}^{(b),2}$ minimizing $R_{j,\mathbf{m}}^{(b)}$ reduces to solving a set of linear equations.

  - The use of a linear combination of more than two terms in the estimator expression can be addressed in a similar manner.

  - The approach can be extended to the multichannel case [LUI 08] (interscale dependencies are also taken into account).

  - When employing non-orthonormal representations (e.g. redundant frames), there is no equivalence between the minimization of $R_{j,\mathbf{m}}$ for each subband and the minimization of the mean square error in the space domain. However, for a SURE-LET estimator, it turns out that the minimization of the latter criterion can be easily carried out, so ensuring an optimal use of the flexibility offered by non-orthonormal representations.

2.5. Method comparisons

In this section, we will give some insights about the appropriateness of a noise removal method for a given application. Indeed, the previous survey has indicated that specifying a denoising method involves several choices such as :

  – componentwise processing or joint processing,

  – Bayesian strategy or non Bayesian approach,

  – choice of the representation,

  – computational complexity.

In what follows, we will briefly discuss these issues, some presented methods being illustrated in Fig. 2.8.

2.5.1. *Componentwise processing versus joint processing*

A great number of efficient denoising methods have been already developed for monochannel images. Hence, the temptation is strong to directly apply them to each component of a given noisy multichannel image. However, if correlations exist across the components, it is judicious to exploit them. A basic method is a two-step procedure which firstly removes such correlations through a principal component analysis or an independent component analysis and, then applies a monovariate denoising for each transformed component [GRE 88]. However, in general, it has been observed that jointly denoising the components outperforms the basic two-step method [BEN 05, LUI 08].

2.5.2. *Bayesian strategy versus non Bayesian approach*

Wavelet transforms are known to generate very compact representations of regular signals. This is an appealing property which simplifies the statistical prior modelling required in the Bayesian framework. In this respect, for both componentwise and joint processing, it is important to note that the wavelet coefficients have been modelled in different ways :

  – marginal probabilistic models for the subband coefficients,

  – joint probabilistic models accounting for interscale dependencies between the subband coefficients [ŞEN 02], [ROM 01, ELM 05] or for spatial intrascale dependencies [MAL 97], [CHA 08].

The more accurate the prior distribution, the better the performance. Concerning multivariate estimators, it has been found that Bernoulli-Gaussian and Gaussian Scale Mixtures priors behave similarly in terms of signal-to-noise ratios for natural images,

and they are clearly better than Gaussian priors [DEB 08].

It is worth pointing out that the Bayesian approach presents two main drawbacks. Firstly, there may exist a mismatch between the data and the model and, the structure of the estimator may thus be inappropriate. Secondly, the values of the hyperparameters may be suboptimal as they are derived from statistical inference exploiting the prior model. These shortcomings have made attractive non Bayesian methods. In the case of Stein-based approaches however, the choice of the estimator form is a key issue. In [BEN 05], it has been shown that SUREVECT outperforms the tested Bayesian denoising methods. Combination of nonlinear functions (LET) also allows us to define a flexible class of estimators.

2.5.3. *Choice of the representation*

Very often, the performance of a denoising method (for a given noise level) is measured in terms of the average MSE_{im} of the mean square errors $\text{MSE}_{\text{im}}^{(b)}$ in each channel b :

$$\text{MSE}_{\text{im}} = \frac{1}{B} \sum_{b=1}^{B} \text{MSE}_{\text{im}}^{(b)} \quad \text{with} \quad \text{MSE}_{\text{im}}^{(b)} = \frac{1}{L^2} \sum_{\mathbf{k} \in \mathbb{K}} |s^{(b)}(\mathbf{k}) - \hat{s}^{(b)}(\mathbf{k})|^2.$$

When the noise is removed in the WT domain, the average mean square error MSE_{tr} of the B mean square errors $\text{MSE}_{\text{tr}}^{(b)}$ in the transform domain is very often employed :

$$\text{MSE}_{\text{tr}} = \frac{1}{B} \sum_{b=1}^{B} \text{MSE}_{\text{tr}}^{(b)} \text{ with } \text{MSE}_{\text{tr}}^{(b)} = \frac{1}{L^2} \sum_{(j,\mathbf{m})} \sum_{\mathbf{k} \in \mathbb{K}_j} |s_{j,\mathbf{m}}^{(b)}(\mathbf{k}) - \hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^2.$$

Notice however that only some of the reported methods explicitly introduce the mean square error criterion in the optimization process for the related estimator.

On the one hand, it must be pointed out that the minimization of the mean square error between the wavelet coefficients is equivalent to the minimization of the mean square error in the image domain if and only if the wavelet representation is orthonormal. Consequently, when the image is decomposed on a redundant frame, the minimization of MSE_{tr} is suboptimal. On the other hand, it has been noticed by many authors that better results are often obtained with redundant decompositions. One of the main reasons for this fact is that these transforms have better translation invariance properties, which is beneficial to the reduction of Gibbs-like visual artifacts. Especially, curvelet [CAN 99] and dual-tree wavelet decompositions [SEL 05, CHA 06] have provided recent examples of multiscale transforms giving rise to redundant representations which are especially good for denoising purposes.

The suboptimality of the denoising on a redundant representation has been alleviated for LET estimators since it has been proved that it is possible to directly minimize an unbiased estimate of MSE_{im} [LUI 08].

2.5.4. *Computational complexity*

As noticed in [DEB 08], it is clear that the computational load is increased by a joint processing rather than separate ones. Besides, in Bayesian approaches, the hyperparameter estimation may imply a high computational cost. Despite their versatility, variational methods requiring iterative solutions are also time-consuming.

Altough they yield improved results, redundant representations involve a significant increase of the computational complexity when the redundancy factor is high.

2.6. Conclusions and perspectives

We have given a broad overview of multivariate image denoising methods, with a special emphasis on wavelet-based approaches, since wavelets efficiently represent image features (both textures and edges). Under different noise correlation models, a variety of transform domain estimators have been described, ranging from Wiener to Stein-based via Bayesian or variational approaches. While most of these methods have been presented for orthonormal wavelet bases, some of them can be applied to more general frames. We would also like to mention that we do not have provided an exhaustive list of the available denoising methods. For example, the reader is referred to [BUA 05] for a description of other non-wavelet based denoising approaches.

Other degradations such as blurring may also affect multichannel images, essentially due to the optical sensor. Recent works now aim at developing advanced wavelet-based methods for restoring multichannel images degraded by a blur *and* an additive noise.

(a) SNR = −2.91 dB

(b) SNR = 4.07 dB

(c) SNR = 8.33 dB

(d) SNR = 8.59 dB

(e) SNR = 8.66 dB

(f) SNR = 8.87 dB
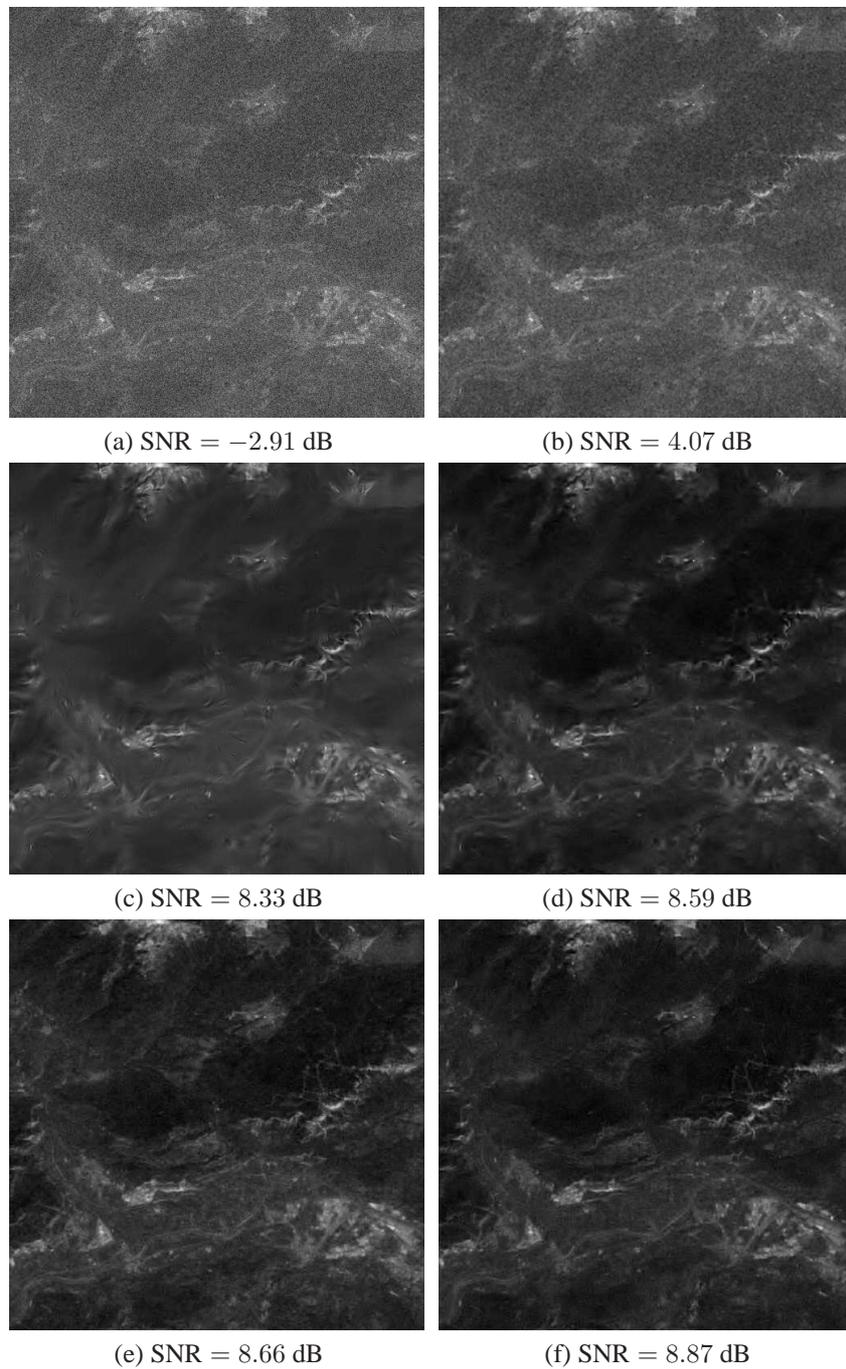
Figure 2.8. *Denoising results on the first component of a Landsat 7 satellite image (see Fig. 2.2). (a) Degraded image and restored images using (b) Wiener filter, (c) BLS-GSM [POR 03], (d) Probshrink [PIŽ 06], (e) SUREVECT [BEN 05] and (f) Block-based wavelet estimate [CHA 08]*

Chapitre 3

Bibliographie

[ABR 98] ABRAMOVICH F., SAPATINAS T., SILVERMAN B. W., « Wavelet thresholding via a Bayesian approach », *Journal of the Royal Statistical Society*, vol. 60, p. 725-749, 1998, Series B.

[ABR 02] ABRAMS M. C., CAIN S. C., « Sampling, radiometry and image reconstruction for polar and geostationary meteorological remote sensing systems », BONES P. J., FIDDY M. A., MILLANE R. P., Eds., *Proc. of SPIE, Image Reconstruction from Incomplete Data II*, vol. 4792, Seattle, WA, USA, p. 207–215, Jul. 8 2002.

[ANG 91] ANGELOPOULOS G., PITAS I., « Multichannel Wiener filters in color image restoration based on AR color image modelling », *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 4, Toronto, Canada, p. 2517–2520, Apr. 14-17 1991.

[ANT 02] ANTONIADIS A., LEPORINI D., PESQUET J.-C., « Wavelet thresholding for some classes of non-Gaussian noise », *Statistica Neerlandica*, vol. 56, n°4, p. 434–453, Dec. 2002.

[ATK 03] ATKINSON I., KAMALABADI F., JONES D. L., DO M. N., « Adaptive wavelet thresholding for multichannel signal estimation », UNSER M. A., ALDROUBI A., LAINE A. F., Eds., *Proc. of SPIE Conf. on Wavelet Applications in Signal and Image Processing X*, vol. 5207, San Diego, CA, USA, p. 28–39, Aug. 2003.

[AUJ 06] AUJOL J.-F., KANG S. H., « Color image decomposition and restoration », *Journal of Visual Communication and Image Representation*, vol. 17, n°4, p. 916–928, Aug. 2006.

[BEN 03] BENAZZA-BENYAHIA A., PESQUET J.-C., « Wavelet-based multispectral image denoising with Bernoulli-Gaussian models », *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Grado, Italy, Jun. 8-11 2003.

[BEN 05] BENAZZA-BENYAHIA A., PESQUET J.-C., « Building robust wavelet estimators for multicomponent images using Stein's principle », *IEEE Trans. on Image Proc.*, vol. 14, n°11, p. 1814–1830, Nov. 2005.

[BLO 98] BLOMGREN P. V., CHAN T. F., « Color TV : Total variation methods for restoration of vector valued images », *IEEE Trans. on Image Proc.*, vol. 7, n°3, p. 304–309, Mar. 1998.

[BUA 05] BUADES A., COLL B., MOREL J. M., « A review of image denoising algorithms, with a new one », *Multiscale Modeling and Simulation*, vol. 4, n°2, p. 490–530, 2005.

[CAI 01] CAI T. T., SILVERMAN B. W., « Incorporating information on neighboring coefficients into wavelet estimation », *Sankhya*, vol. 63, Series B., p. 127–148, 2001.

[CAN 99] CANDÈS E., DONOHO D. L., « Curvelets - a surprisingly effective nonadaptive representation for objects with edges », C. RABUT A. C., SCHUMAKER L. L., Eds., *Curves and Surfaces*, p. 105–120, Vanderbilt University Press, Nashville, TN, USA, 1999.

[CAN 06] CANDÈS E., DEMANET L., DONOHO D., YING L., « Fast discrete curvelet transforms », *Multiscale Modeling and Simulation*, vol. 5, n°3, p. 861–899, Mar. 2006.

[CHA 98] CHAMBOLLE A., DEVORE R. A., LEE N. Y., LUCIER B. J., « Nonlinear wavelet image processing : variational problems, compression, and noise removal through wavelet shrinkage », *IEEE Trans. on Image Proc.*, vol. 7, n°3, p. 319–325, Mar. 1998.

[CHA 00] CHANG S. G., YU B., VETTERLI M., « Adaptive wavelet thresholding for image denoising and compression », *IEEE Trans. on Image Proc.*, vol. 9, n°9, p. 1532–1546, Sep. 2000.

[CHA 06] CHAUX C., DUVAL L., PESQUET J.-C., « Image analysis using a dual-tree M-band wavelet transform », *IEEE Trans. on Image Proc.*, vol. 15, n°8, p. 2397–2412, Aug. 2006.

[CHA 07] CHAUX C., COMBETTES P. L., PESQUET J.-C., WAJS V. R., « A variational formulation for frame based inverse problems », *Inverse Problems*, vol. 23, p. 1495–1518, juin 2007.

[CHA 08] CHAUX C., DUVAL L., BENAZZA-BENYAHIA A., PESQUET J.-C., « A nonlinear Stein based estimator for multichannel image denoising », *IEEE Trans. on Signal Proc.*, vol. 56, n°8, p. 3855–3870, Aug. 2008.

[COH 92] COHEN A., DAUBECHIES I., FEAUVEAU J.-C., « Biorthogonal bases of compactly supported wavelets », *Comm. ACM*, vol. 45, n°5, p. 485–560, 1992.

[COI 92] COIFMAN R. R., MEYER Y., WICKERHAUSER M. V., « Wavelet analysis and signal processing », *In Wavelets and their applications*, p. 153–178, Jones and Bartlett, Boston, MA, 1992.

[COI 95] COIFMAN R., DONOHO D., « Translation-invariant de-noising », ANTONIADIS A., OPPENHEIM G., Eds., *Wavelets and Statistics*, vol. 103 de *Lecture Notes in Statistics*, p. 125–150, Springer, New York, NY, USA, 1995.

[COM 96] COMBETTES P. L., « The convex feasibility problem in image recovery », vol. 95, p. 155–270, Academic Press, NY, 1996.

[COM 03] COMBETTES P. L., « A block-iterative surrogate constraint splitting method for quadratic signal recovery », *IEEE Trans. on Signal Proc.*, vol. 51, n°7, p. 1771–1782, Jul. 2003.

[COM 04] COMBETTES P. L., PESQUET J.-C., « Wavelet-constrained image restoration », *International Journal on Wavelets, Multiresolution and Information Processing*, vol. 2, n°4,

p. 371–389, Dec. 2004.

[COM 07] COMBETTES P. L., PESQUET J.-C., « A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery », *IEEE J. Selected Topics Signal Process.*, vol. 1, n°4, p. 564–574, Dec. 2007.

[COM 08] COMBETTES P. L., PESQUET J.-C., « A proximal decomposition method for solving convex variational inverse problems », *Inverse Problems*, vol. 24, n°6, Dec. 2008, (27 pp).

[COR 03] CORNER B. R., NARAJANAN R. M., REICHENBACH S. E., « Noise estimation in remote sensing imagery using data masking », *Int. J. Remote Sensing*, vol. 24, n°4, p. 689–702, 2003.

[CRO 76] CROISIER A., ESTEBAN D., C. G., « Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques », *Int. Conf. on Inform. Science and Systems*, Patras, Greece, p. 443–446, Aug. 1976.

[DAU 92] DAUBECHIES I., *Ten Lectures on Wavelets*, CBMS-NSF, SIAM Lecture Series, Philadelphia, PA, USA, 1992.

[DEB 08] DE BACKER S., PIŽURICA A., HUYSMANS B., PHILIPS W., SCHEUNDERS P., « Denoising of multicomponent images using wavelet least-squares estimators », *Image and Vision Computing*, vol. 26, n°7, p. 1038–1051, Jul. 2008.

[DO 05] DO M. N., VETTERLI M., « The contourlet transform : an efficient directional multiresolution image representation », *IEEE Trans. on Image Proc.*, vol. 14, n°12, p. 2091–2106, Dec. 2005.

[DON 93] DONOHO D. L., « Unconditional bases are optimal bases for data compression and for statistical estimation », *Appl. and Comp. Harm. Analysis*, vol. 1, n°1, p. 100–115, Dec. 1993.

[DON 94] DONOHO D. L., JOHNSTONE I. M., « Ideal spatial adaptation by wavelet shrinkage », *Biometrika*, vol. 81, p. 425–455, Sept. 1994.

[DON 95] DONOHO D. L., JOHNSTONE I. M., « Adapting to unknown smoothness via wavelet shrinkage », *J. American Statist. Ass.*, vol. 90, p. 1200–1224, Dec. 1995.

[ELM 05] ELMZOUGHI A., BENAZZA-BENYAHIA A., PESQUET J.-C., « An interscale multivariate statistical model for MAP multicomponent image denoising in the wavelet transform domain », *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 2, Philadelphia, USA, p. 45–48, Mar. 18-23 2005.

[FLA 98] FLANDRIN P., *Time-frequency and time-scale analysis*, Academic Press, San Diego, USA, 1998.

[GRE 88] GREEN A., BERMAN M., P.SWITZER, CRAIG M. D., « A transformation for ordering multispectral data in terms of image quality with implications for noise removal », *IEEE Trans. on Geoscience and Remote Sensing*, vol. 26, n°1, p. 65–74, Jan. 1988.

[GUY 90] GUYOT G., « Optical properties of vegetation canopies », STEVEN M. D., CLARK J. A., Eds., *Applications of Remote Sensing in Agriculture*, p. 19–43, Butterworth, 1990.

[HAA 10] HAAR A., « Zur Theory der orthogalen Funktionen Systeme », *Math. Annalen*, vol. 69, p. 331–371, 1910.

[HEI 06] HEIL C., WALNUT D. F., *Fundamental papers in wavelet theory*, Princeton University Press, 2006.

[JAI 89] JAIN A. K., *Fundamentals of digital image processing*, Englewood Cliffs, NJ : Prentice Hall, 1989.

[LAN 86] LANDGREBE D. A., MALARET E., « Noise in remote-sensing systems : the effect on classification error », *IEEE Trans. on Geoscience and Remote Sensing*, vol. GE-24, n°2, p. 294–300, Mar. 1986.

[LAN 00] LANDGREBE D., « Information extraction principles and methods for multispectral and hyperspectral image data », CHEN C. H., Ed., *Information processing for remote sensing*, p. 3–38, World Scientific Publishing Co., Inc., NJ, USA, 2000.

[LEP 99] LEPORINI D., PESQUET J.-C., KRIM H., « Best basis representations based on prior statistical models », MÜLLER P., VIDAKOVIC B., Eds., *Bayesian Inference in wavelet based models*, vol. 141 de *Lecture notes in statistics*, p. 155–172, Springer, 1999.

[LEP 01] LEPORINI D., PESQUET J.-C., « Bayesian wavelet denoising : Besov priors and non-Gaussian noises », *Signal Processing*, vol. 81, p. 55–67, 2001.

[LUI 08] LUISIER F., BLU T., « SURE-LET multichannel image denoising : interscale orthonormal wavelet thresholding », *IEEE Trans. on Inform. Theory*, vol. 17, n°4, p. 482–492, April 2008.

[MAL 97] MALFAIT M., ROSSE D., « Wavelet-based image denoising using a Markov random field a priori model », *IEEE Trans. on Image Proc.*, vol. 6, n°4, p. 549–565, Apr. 1997.

[MAL 08] MALLAT S., *A wavelet tour of signal processing*, Academic Press, 3rd édition, Dec. 2008.

[MAL 09] MALLAT S., « Geometrical grouplets », *Appl. and Comp. Harm. Analysis*, vol. 26, n°2, p. 161–180, Mar. 2009.

[MEY 90] MEYER Y., *Ondelettes et opérateurs. I*, Actualités mathématiques, Hermann, Paris, France, 1990.

[PES 96] PESQUET J.-C., KRIM H., CARFANTAN H., « Time-invariant orthogonal wavelet representations », *IEEE Trans. on Signal Proc.*, vol. 44, n°8, p. 1964–1970, Aug. 1996.

[PES 97] PESQUET J.-C., LEPORINI D., « A new wavelet estimator for image denoising », *IEE Sixth Int. Conf. Im. Proc. Appl.*, vol. 1, p. 249–253, Jul. 14-17 1997.

[PIT 90] PITAS I., VENETSANOPOULOS A., *Nonlinear digital filters*, Kluwer, Dodrecht, 1990.

[PIŽ 06] PIŽURICA A., PHILIPS W., « Estimating the probability of presence of a signal of interest in multiresolution single- and multiband image denoising », *IEEE Trans. on Image Proc.*, vol. 15, n°3, p. 654–665, Mar. 2006.

[POR 03] PORTILLA J., STRELA V., WAINWRIGHT M. J., SIMONCELLI E. P., « Image denoising using scale mixtures of Gaussians in the wavelet domain », *IEEE Trans. on Image*

Proc., vol. 12, n°11, p. 1338–1351, Nov. 2003.

[RAP 08] RAPHAN M., SIMONCELLI E., « Optimal denoising in redundant representations », *IEEE Trans. on Image Proc.*, vol. 17, n°8, p. 1342–1352, Aug. 2008.

[ROM 01] ROMBERG J. K., CHOI H., BARANIUK R. G., « Bayesian tree-structured image modeling using wavelet-domain hidden Markov models », *IEEE Trans. on Image Proc.*, vol. 10, n°7, p. 1056–1068, Jul. 2001.

[RUD 92] RUDIN L. I., OSHER S., FATEMI E., « Nonlinear total variation based noise removal algorithms », *Physica D*, vol. 60, n°1-4, p. 259–268, 1992.

[SCH 04] SCHEUNDERS P., « Wavelet thresholding of multivalued images », *IEEE Trans. on Image Processing*, vol. 13, n°4, p. 411-416, April 2004.

[SCH 07] SCHEUNDERS P., DE BACKER S., « Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free image as priors », *IEEE Trans. on Image Proc.*, vol. 16, n°7, p. 1865–1872, Jul. 2007.

[SEL 05] SELESNICK I. W., BARANIUK R. G., KINGSBURY N. G., « The dual-tree complex wavelet transform », *IEEE Signal Processing Magazine*, vol. 22, n°6, p. 123–151, Nov. 2005.

[ŞEN 02] ŞENDUR L., SELESNICK I. W., « Bivariate shrinkage with local variance estimation », *Signal Processing Letters*, vol. 9, n°12, p. 438–441, Dec. 2002.

[SIM 96] SIMONCELLI E. P., ADELSON E. H., « Noise removal via Bayesian wavelet coring », *Proc. Int. Conf. on Image Processing*, vol. 1, Lausanne, Switzerland, p. 379–382, Sep. 16-19 1996.

[SMI 84] SMITH M., BARNWELL T., « A procedure for designing exact reconstruction filter banks for tree structured subband coders », *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, vol. 9, San Diego, CA, USA, p. 421–424, March 19-21 1984.

[STE 81] STEIN C., « Estimation of the mean of a multivariate normal distribution », *Annals of Statistics*, vol. 9, n°6, p. 1135–1151, 1981.

[STE 93] STEFFEN P., HELLER P. N., GOPINATH R. A., BURRUS C. S., « Theory of regular M-band wavelet bases », *IEEE Trans. on Signal Proc.*, vol. 41, n°12, p. 3497–3511, Dec. 1993.

[TEB 98] TEBOUL S., BLANC-FÉRAUD L., AUBERT G., BARLAUD M., « Variational approach for edge-preserving regularization using coupled PDE's », *IEEE Trans. on Image Proc.*, vol. 7, n°3, p. 387–397, Mar. 1998.

[TIK 63] TIKHONOV A. N., « On the solution of ill-posed problems and the method of regularization », *Dokl. Akad. Nauk SSSR*, vol. 151, p. 501–504, 1963.

[TRO 06] TROPP J. A., « Just relax : Convex programming methods for identifying sparse signals in noise », *IEEE Trans. on Inform. Theory*, vol. 52, p. 1030–1051, 2006.

[VAI 87] VAIDYANATHAN P. P., « Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having the perfect-reconstruction property », *IEEE Trans. on Acous., Speech and Signal Proc.*, vol. 35, n°4, p. 476–492, Apr. 1987.

[WIE 49] WIENER N., *Extrapolation, interpolation, and smoothing of stationary time series*, Cambridge, Technology Press of Massachusetts Institute of Technology, and New York, Wiley, 1949.