# A nonlinear Stein based estimator for multichannel image denoising

Caroline Chaux, Laurent Duval, Amel Benazza-Benyahia, Jean-Christophe Pesquet

# A Nonlinear Stein Based Estimator for Multichannel Image Denoising

Caroline Chaux, *Member IEEE*, Laurent Duval, *Member IEEE*,
Amel Benazza-Benyahia, and Jean-Christophe Pesquet, *Senior Member IEEE*

**Abstract**

The use of multicomponent images has become widespread with the improvement of multisensor systems having increased spatial and spectral resolutions. However, the observed images are often corrupted by an additive Gaussian noise. In this paper, we are interested in multichannel image denoising based on a multiscale representation of the images. A multivariate statistical approach is adopted to take into account both the spatial and the inter-component correlations existing between the different wavelet subbands. More precisely, we propose a new parametric nonlinear estimator which generalizes many reported denoising methods. The derivation of the optimal parameters is achieved by applying Stein's principle in the multivariate case. Experiments performed on multispectral remote sensing images clearly indicate that our method outperforms conventional wavelet denoising techniques.

**Index Terms**

Multicomponent image, multichannel noise, denoising, multivariate estimation, block estimate, Stein's principle, nonlinear estimation, thresholding, $M$-band wavelet transform, dual-tree wavelet transform, frames.

C. Chaux and J.-C. Pesquet are with the Université Paris-Est, Institut Gaspard Monge and CNRS-UMR 8049, 77454 Marne-la-Vallée Cedex 2, France. E-mail: {caroline.chaux,jean-christophe.pesquet}@univ-paris-est.fr, phone: +33 1 60 95 77 39.

L. Duval is with the Institut Français du Pétrole (IFP), Technology, Computer Science and Applied Mathematics Division, 92500 Rueil-Malmaison, France. E-mail: laurent.duval@ifp.fr, phone: +33 1 47 52 61 02.

A. Benazza-Benyahia is with the Unité de Recherche en Imagerie Satellitaire et ses Applications (URISA), École Supérieure des Communications de Tunis, Route de Raoued 3.5 Km, 2083 Ariana, Tunisia. E-mail: benazza.amel@supcom.rnu.tn, phone: +216 71 857 000 Ext. 1027.

## I. INTRODUCTION

Many real world images are contaminated by noise during their acquisition and/or transmission. In particular, multichannel imaging is prone to quality degradation due to the imperfectness of the sensors often operating in different spectral ranges [1], [2]. In order to alleviate the influence of such disturbing artifacts on subsequent analysis procedures, denoising appears as a crucial initial step in multicomponent image enhancement. In this context, attention has been paid to developing efficient denoising methods. Usually, the noise removal problem is considered as a regression problem. The challenge thus resides in finding realistic statistical models which lead to both efficient and tractable denoising approaches. To this respect, linearly transforming the signal from the spatial domain to a more suitable one may drastically improve the denoising performance. The rationale for such a transformation is the observation that some representations possessing good energy concentration and decorrelation properties tend to simplify the statistical analysis of many natural images. For instance, the Discrete Wavelet Transform (DWT) constitutes a powerful tool for image denoising [3], [4]. The DWT, computed for each channel/component separately, usually yields "larger" coefficients for signal features and "smaller" ones for noise since it forms an unconditional basis for several classes of regular signals [5]. For monochannel signals or images, the seminal work of Donoho and Johnstone has shown that a mere wavelet coefficient thresholding constitutes a simple yet effective technique for noise reduction [6]. Based on Stein's Unbiased Risk Estimator (SURE), they have proposed the SUREshrink technique [7]. Subsequently, several extensions of their work have been performed, *e.g.* in [8]–[11]. Recently, the denoising problem in the wavelet domain has gained more attention in the case of multichannel images. Indeed, the increasing need for multicomponent images in several applications such as medical imaging and remote sensing has motivated a great interest in designing tractable denoising methods dedicated to this kind of images. Componentwise processing can be performed for each modality, but a joint denoising should be preferred in order to exploit the cross-channel similarities in an efficient way [12]. The problem of a joint estimation in the wavelet domain has been formulated in [13]. More precisely, the use of joint threshold estimators was investigated in two situations: overcomplete representations of a noisy image [14] and multiple observations of the same image [13]. A scale-adaptive wavelet thresholding was designed for multichannel images in the case of an i.i.d. (independent identically distributed) Gaussian vector noise whose components are independent and have the same variance [15]. In a Bayesian framework, several prior models have been considered such as *multivariate* Bernoulli-Gaussian ones [16]. A generalized Gaussian distribution was also considered for modelling the marginal distribution of each subband in each channel and a simple shrinkage was applied

depending on the local spectral activity [17]. A vector-based least-square approach was also investigated in the wavelet domain [18]. Recently, the application of Stein's principle [19]–[21] in the multivariate case has motivated the design of a nonlinear estimator in [22]. In this paper, links existing between the proposed nonlinear estimator and Bayesian approaches were discussed. In particular, the structure of the estimator was motivated by a multivariate Bernoulli-Gaussian model reflecting the sparseness of the wavelet representation as well as the statistical dependencies existing between the different components. We point out that the form of the estimator in [22] is not the same as the one proposed in this paper. In particular, the estimator in [22] does not involve any thresholding operation. Moreover, the estimator does not allow to take into account spatial dependencies but only those existing between the multichannel data at a given position.

In parallel to these works, the idea of performing a joint spatial denoising of the coefficients, rather than using a conventional term-by-term processing, has emerged in statistics. This idea, stemming from an incentive for capturing statistical dependences between spatial neighboring wavelet coefficients, was first investigated for single component images in both non-Bayesian and Bayesian cases [23], [24]. A successful extension was also carried out in the case of multichannel images by considering hybrid (spectral and spatial) neighborhoods [25].

In this paper, we aim at building a new estimator allowing to take into account the various correlations existing in multichannel image data. This estimator also provides a unifying framework for several denoising methods proposed in the literature. More precisely, our contributions are the following.

• The method applies to any vector-valued data embedded in a multivariate Gaussian noise. As illustrated later on, many examples of such multivariate contexts (inter-component, spatial and inter-scale) can be found. They naturally include multivariate denoising obtained with vectors of samples sharing the same spatial position in different channels.

• The estimator can be computed in any image representation domain. For instance, in addition to wavelet domains, usually considered in conventional denoising methods, we propose to exploit more general frame decompositions such as the dual-tree wavelet transform [26], [27].

• The computation of the estimated value can be performed with the help of various observations. Again, our method includes most of the reported estimation methods acting in that way. Furthermore, it offers a great flexibility in the choice of these auxiliary data.

• The form of the proposed estimator is quite general. More precisely, we focus on deriving thresholding estimators including an exponent parameter and a linear part. Optimal parameters are derived from Stein's principle.

• The denoising approach allows to handle any covariance matrix between the multichannel noise components.

Notwithstanding its generality, the proposed approach remains tractable and compares quite favorably with state-of-the-art methods. The paper is organized as follows. In Section II, we present the relevant background and introduce notations for a general formulation of the estimator, based on the concept of Reference Observation Vector. In Section III, we describe the proposed multivariate nonlinear estimator. In Section IV, we give the specific form taken by this new estimator for multichannel images decomposed by a wavelet transform or an $M$-band dual-tree wavelet transform. In Section V, experimental results are given for remote sensing images showing that the proposed estimator outperforms existing ones and some concluding remarks are drawn in Section VI.

Throughout this paper, the following notations will be used: let $M$ be an integer greater than or equal to 2, $\mathbb{N}_M = \{0, \ldots, M-1\}$ and $\mathbb{N}_M^\star = \{1, \ldots, M-1\}$; $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{R}_+$ are the sets of integers, reals and positive reals; $\lceil . \rceil$ denotes rounding towards the immediate upper integer. Besides, $\widehat{a}$ denotes the Fourier transform of a function $a$, $(\delta_m)_{m\in\mathbb{Z}}$ is the Kronecker sequence (equal to 1 if $m = 0$ and 0 otherwise), $(f)_+ = f$ if $f > 0$ and 0 otherwise, and $\mathbb{1}\{A\} = 1$ if condition $A$ is true and 0 otherwise.

## II. Background

### A. General formulation of the multichannel estimator

In multisensor imaging, $B$ vectors of observed data samples $(r^{(1)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$, …, $(r^{(B)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$, are provided where $B$ is the number of effective sensors and $\mathbb{K}$ is a set of spatial indices ($\mathbb{K} \subset \mathbb{Z}^2$). Generally, these data correspond to noisy realizations of $B$ unknown signals $(s^{(1)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$, …, $(s^{(B)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$, respectively. Subsequently, our task will consist in devising methods to reduce the noise present in the observations. Two alternatives can be envisaged in this context. On the one hand, a monochannel approach builds an estimator $\overset{\wedge}{s}{}^{(b)}(\mathbf{k})$ of $s^{(b)}(\mathbf{k})$ only from the observations $(r^{(b)}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$, for each channel $b \in \{1, \ldots, B\}$. On the other hand, a multivariate technique attempts to estimate $s^{(b)}(\mathbf{k})$ by accounting not only for the individual data set $\{r^{(b)}(\mathbf{k})\}_{\mathbf{k}\in\mathbb{K}}$, but also for the remaining ones $\{r^{(1)}(\mathbf{k})\}_{\mathbf{k}\in\mathbb{K}}$, …, $\{r^{(b-1)}(\mathbf{k})\}_{\mathbf{k}\in\mathbb{K}}$, $\{r^{(b+1)}(\mathbf{k})\}_{\mathbf{k}\in\mathbb{K}}$, …, $\{r^{(B)}(\mathbf{k})\}_{\mathbf{k}\in\mathbb{K}}$.

Thus, one of the simplest relevant denoising approach consists in calculating the estimated value $\overset{\wedge}{s}{}^{(b)}(\mathbf{k})$ of $s(\mathbf{k})$ as

$$\overset{\wedge}{s}{}^{(b)}(\mathbf{k}) = f(r^{(b)}(\mathbf{k})) \tag{1}$$

where $f$ is a scalar function defined on the real line. For instance, a shrinkage function can be used, possibly involving some threshold value. Such a technique is commonly used in regression, when outliers

have to be removed in order to improve the representativity of the fit [28]. Although $r^{(b)}(\mathbf{k})$ does not necessarily depend on other observed samples, for structured signal or image analysis, neighboring samples often present some correlations. Consequently, an improvement can be expected if $\hat{s}^{(b)}(\mathbf{k})$ is calculated with the help of a subset $\mathcal{R}_{\text{ref}}^{(b)}(\mathbf{k})$ of observed sample locations. Average or median filtering [29, p. 243–245] are examples where the estimated sample depends on its neighborhood. As a result, a more general estimation rule is:

$$\hat{s}^{(b)}(\mathbf{k}) = f\big((r^{(b)}(\mathbf{k}'))_{\mathbf{k}' \in \mathcal{R}_{\text{ref}}^{(b)}(\mathbf{k})}\big). \tag{2}$$

With underlying Markovian assumptions, the context set $\{r^{(b)}(\mathbf{k}')\}_{\mathbf{k}' \in \mathcal{R}_{\text{ref}}^{(b)}(\mathbf{k})}$ can be restricted to a limited number of values around the sample location $\mathbf{k}$. These values can be gathered in a vector $\bar{\mathbf{r}}^{(b)}(\mathbf{k})$ which will be designated as the *Reference Observation Vector* (ROV). We have then

$$\hat{s}^{(b)}(\mathbf{k}) = f(\bar{\mathbf{r}}^{(b)}(\mathbf{k})). \tag{3}$$

The multivariate case can also be described by such a formula if we allow the ROV to contain additional samples from the remaining channels in order to exploit the inter-component statistical dependencies.

Another degree of freedom lies in the choice of a suitable domain for data representation. While virtually any transform can be chosen, special attention has been paid to multiscale transforms. For example, if a decomposition onto an $M$-band wavelet basis ($M \geq 2$) [30] is performed, the observed images are represented by coefficients $r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ defined at resolution level $j \geq 1$ and subband index $\mathbf{m} \in \mathbb{N}_M^2$ and the corresponding ROV will be denoted $\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$. Since the noise is usually less correlated than the data, the DWT is applied in order to provide a sparser representation of the data of interest, before further analysis [3], [4]. The goal becomes to generate estimates $\hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ of the unknown wavelet coefficients $s_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ of the original images:

$$\hat{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = f(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})). \tag{4}$$

Then, the inverse DWT is applied to the estimated coefficients in order to reconstruct the estimated signal $\hat{s}^{(b)}(\mathbf{k})$ in the spatial domain. In the literature concerning denoising, two key issues have been addressed. The first one lies in the definition of the ROV. The second one concerns the choice of an appropriate function $f$ or, in other words, a suitable expression of the estimator. In the next subsection, we give a brief overview of the main ROVs proposed until now.

### *B. Reported ROVs in the DWT domain*

Popular componentwise methods operating in the DWT domain are Visushrink [5] and SUREshrink [7]. They both employ a very basic ROV reduced to a scalar value:

$$\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = r_{j,\mathbf{m}}^{(b)}(\mathbf{k}). \tag{5}$$

Similarly to what can be done in the spatial domain, the wavelet coefficients can also be processed by *block* rather than individually, again in a mono-channel way [23], [24], [31]–[33]. The main motivation for this technique is to exploit the spatial similarities between neighboring coefficients in a given subband. The introduction of $d-1$ spatial neighbors $\mathbf{k}_1$ ,..., $\mathbf{k}_{d-1}$ of the current sample indexed by $\mathbf{k}$ in the ROV allows to take into account the spatial dependencies:

$$\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), r_{j,\mathbf{m}}^{(b)}(\mathbf{k}_1), \ldots, r_{j,\mathbf{m}}^{(b)}(\mathbf{k}_{d-1})]^\top. \tag{6}$$

For higher dimensional data, the ROV may also consist of coefficients sharing similar orientations, possibly within different scales [34]. Another generalization of the scalar case takes into account the inter-scale similarities between the current coefficient and the homologous ones defined at other scales. Based on empirical observations in image compression [35], it has been proposed to use the current coefficient ancestors at coarser scales $j+1$, $j+2$, ..., $j_{\mathrm{m}}$ eventually up to the coarsest level $J$ [36]–[38]: the ROV $\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ thus includes the corresponding $j_{\mathrm{m}} - j + 1$ coefficients at location $\mathbf{k}$, in subband $\mathbf{m}$, at resolution level $j$.

In the case of multicomponent data, additional samples borrowed from the different channels can be included in the ROVs, as shown in [34], [39] for color image as well as for multispectral image denoising. Basically, the *inter-component* correlations can be taken into account through the following ROV [22]:

$$\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [r_{j,\mathbf{m}}^{(1)}(\mathbf{k}), \ldots, r_{j,\mathbf{m}}^{(B)}(\mathbf{k})]^\top. \tag{7}$$

Such a ROV includes all the coefficients of all channels at the same spatial location, in the same subband $\mathbf{m}$ and at the same resolution level $j$. In [25], a more sophisticated multicomponent ROV $\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ has been defined which combines both spatial *and* multichannel neighbors. As particular cases, such an ROV encompasses the ROV in (7) and, also the ROV in (6). In addition, the ROV may include coefficients from different subbands.

A final potential extension of the ROVs is related to the choice of the transform. Indeed, it has been long observed that a decomposition onto a wavelet basis suffers from a lack of shift-invariance as well as a poor directionality, resulting in denoising artifacts at low signal to noise ratios. A simple way for alleviating

these problems is to use a frame decomposition built from a union of wavelet bases. In particular, a number of papers [40]–[42] have demonstrated significant improvements in scalar shrinkage when resorting to a translation-invariant wavelet representation. The latter can be viewed as a decomposition onto a union of shifted versions of a unique wavelet basis. $M$-band dual-tree wavelet decompositions [27] constitute another example of a union of 2 (resp. 4) wavelet bases in the real (resp. complex) case. The corresponding mother wavelets are then derived from the first one by Hilbert transforms, which results in an improved directional analysis. For such frame decompositions, one can extend the notion of ROV to include samples produced by the different wavelet basis decompositions operating in parallel. These facts will be further developed to motivate the application of the general estimator proposed in this paper to an $M$-band dual-tree wavelet frame [27].

### *C. A unifying framework for shrinkage functions*

In the aforementioned works, the estimation is often performed by shrinkage, so exploiting the sparseness of the representation. The most well-known method was proposed in the pioneering works of Donoho and Johnstone [5]. The estimating function $f$ is then given by

$$f(r_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = \mathrm{sign}(r_{j,\mathbf{m}}^{(b)}(\mathbf{k}))\max\{|r_{j,\mathbf{m}}^{(b)}(\mathbf{k})| - \lambda, 0\} \tag{8}$$

for a soft thresholding with threshold value $\lambda \geq 0$, where $\mathrm{sign}(\cdot)$ is the signum function. Equivalently, by using the ROV in (5), the estimating function can be expressed as

$$f(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = \left(\frac{|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})| - \lambda}{|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}\mathbf{k})|}\right)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k}). \tag{9}$$

Some works [43] have focused on the improvement of the scalar shrinkage rule, yielding for instance smoother functions such as the garrote shrinkage based on [44], which is defined as:

$$f(\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = \left(\frac{|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^2 - \lambda}{|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})|^2}\right)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k}). \tag{10}$$

Several authors have proposed vector-like generalizations to the scalar shrinkage. Cai and Silverman [23], have proposed a block estimator which takes into account the energy of the neighboring coefficients in each subband, as expressed as in (6). This estimator dominates the maximum likelihood estimator when the block size is greater than 2. This method, named "NeighBlock", consists in applying the following shrinkage rule:

$$\hat{\mathbf{s}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = \left(\frac{\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^2 - \bar{\lambda}d\sigma^2}{\|\bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^2}\right)_+ \mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \tag{11}$$

where $\bar{\lambda} > 0$, $d$ is the number of components in the ROV, $\mathbf{r}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is a subpart of the ROV, $\overset{\wedge}{\mathbf{s}}{}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is the associated vector of estimated values, $\|.\|$ denotes the classical Euclidean norm of $\mathbb{R}^d$ and $\sigma^2$ denotes the noise variance. Such a function is clearly reminiscent of the scalar garrote shrinkage defined in (10). Based on an asymptotic minimax study, Cai and Silverman suggested appropriate values for $\bar{\lambda}$ and $d$. They considered both overlapping and non-overlapping variants of this approach. In particular, the so-called "NeighCoeff" method corresponds to the case when $\overset{\wedge}{\mathbf{s}}{}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ reduces to a scalar estimate. Then, the corresponding estimating function is:

$$f(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = \left( \frac{\|\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^2 - \bar{\lambda} d \sigma^2}{\|\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^2} \right)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k}). \tag{12}$$

In the meantime, Şendur and Selesnick [45] introduced a Bayesian approach allowing to model inter-scale dependencies between two consecutive levels. These authors consequently formulated the problem in the 2-band wavelet domain. In their approach, the ROV is given by $\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [r_{j,\mathbf{m}}^{(b)}(\mathbf{k}), r_{j+1,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{2} \rceil)]^\top$, $r_{j+1,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{2} \rceil)$ being the "parent" of $r_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ (at the next coarser resolution). By considering as a prior model the non-Gaussian bivariate probability density function

$$p(s_{j,\mathbf{m}}^{(b)}(\mathbf{k}), s_{j+1,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{2} \rceil)) \propto \exp\Big( - \frac{\sqrt{3}}{\sigma_s} \sqrt{\big|s_{j,\mathbf{m}}^{(b)}(\mathbf{k})\big|^2 + \big|s_{j+1,\mathbf{m}}^{(b)}(\lceil \frac{\mathbf{k}}{2} \rceil)\big|^2} \Big), \qquad \sigma_s > 0 \tag{13}$$

the following Maximum A Posteriori (MAP) estimator was derived:

$$f(\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})) = \left( \frac{\|\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\| - \frac{\sqrt{3}\sigma^2}{\sigma_s}}{\|\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|} \right)_+ r_{j,\mathbf{m}}^{(b)}(\mathbf{k}) \tag{14}$$

where the noise variance is again denoted by $\sigma^2$.

More recently, in the context of signal restoration problems, Combettes and Wajs [46] have studied the properties of proximity operators corresponding to the solutions of some convex regularization problems. In particular, an interpretation of one of their results is the following. Let us adopt a Bayesian formulation by assuming that the vector $\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ is a noisy observation of the vector $\mathbf{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ of multichannel coefficients at location $\mathbf{k}$, embedded in white Gaussian noise with variance $\sigma^2$. Further assume that the vectors $\mathbf{s}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ are independent of the noise, mutually independent and have a prior distribution proportional to $\exp(-\overline{\lambda}\|\cdot\|)$ with $\overline{\lambda} > 0$. The MAP estimation of $\mathbf{s}_{j,\mathbf{m}}^{(b)}$ is found by solving the optimization problem:

$$\min_{\mathbf{u} \in \mathbb{R}^B} \overline{\lambda}\|\mathbf{u}\| + \frac{1}{2\sigma^2}\|\mathbf{u} - \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|^2. \tag{15}$$

It is shown in [46] that the minimizer of the MAP criterion is

$$\overset{\wedge}{\mathbf{s}}{}_{j,\mathbf{m}}^{(b)} = \Big( \frac{\|\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\| - \overline{\lambda}\sigma^2}{\|\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})\|} \Big)_+ \overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}). \tag{16}$$

The three previous block-thresholding estimators have been derived from different perspectives and they have also been applied in different ways. However, it is possible to describe them through a general shrinkage factor $\eta_\lambda(\|\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}\|^\beta)$, where

$$\forall \tau \in \mathbb{R}_+, \qquad \eta_\lambda(\tau) = \Big(\frac{\tau-\lambda}{\tau}\Big)_+ \tag{17}$$

and $\beta > 0$ and $\lambda \geq 0$ take specific values in each of the aforementioned block estimators. We also remark that this generalized shrinkage obviously encompasses the soft and garrote thresholdings provided in (9) and (10).

## III. Proposed nonlinear estimator

### A. Notations

We will now propose a more general adaptive estimator that can be applied in any representation domain. We will therefore drop the indices $j$ and $\mathbf{m}$ and we will consider the general situation where an observation sequence $(\overline{\mathbf{r}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{Z}^2}$ of $d$-dimensional real-valued vectors ($d \in \mathbb{N}$, $d > 1$) is defined as

$$\forall \mathbf{k} \in \mathbb{Z}^2, \qquad \overline{\mathbf{r}}(\mathbf{k}) = \overline{\mathbf{s}}(\mathbf{k}) + \overline{\mathbf{n}}(\mathbf{k}), \tag{18}$$

where $(\overline{\mathbf{n}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{Z}^2}$ is a $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma}^{(\overline{\mathbf{n}})})$ noise and $(\overline{\mathbf{s}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{Z}^2}$ is an identically distributed second-order random sequence which is independent of $(\overline{\mathbf{n}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{Z}^2}$. We will assume that the covariance matrix $\mathbf{\Gamma}^{(\overline{\mathbf{n}})}$ is invertible. These random vectors are decomposed as

$$\overline{\mathbf{r}}(\mathbf{k}) = \begin{bmatrix} r(\mathbf{k}) \\ \tilde{\mathbf{r}}(\mathbf{k}) \end{bmatrix}, \qquad \overline{\mathbf{s}}(\mathbf{k}) = \begin{bmatrix} s(\mathbf{k}) \\ \tilde{\mathbf{s}}(\mathbf{k}) \end{bmatrix}, \qquad \overline{\mathbf{n}}(\mathbf{k}) = \begin{bmatrix} n(\mathbf{k}) \\ \tilde{\mathbf{n}}(\mathbf{k}) \end{bmatrix} \tag{19}$$

where $r(\mathbf{k})$, $s(\mathbf{k})$ and $n(\mathbf{k})$ are scalar random variables. We aim at estimating the first component $s(\mathbf{k})$ of the vector $\overline{\mathbf{s}}(\mathbf{k})$ using an observation sequence $(\overline{\mathbf{r}}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$ where $\mathbb{K}$ is a finite subset of $\mathbb{Z}^2$. We recall that, although (18) does not introduce an explicit dependence between $s(\mathbf{k})$ and the vector $\tilde{\mathbf{r}}(\mathbf{k})$ of the last $d-1$ components of $\overline{\mathbf{r}}(\mathbf{k})$, such a statistical dependence may exist, due to the dependence between the components of $\overline{\mathbf{s}}(\mathbf{k})$ themselves. The estimated sequence will be denoted by $(\hat{s}(\mathbf{k}))_{\mathbf{k}\in\mathbb{K}}$.

### B. Form of the adaptive estimator

In order to gain more flexibility in the denoising procedure, the following generalized form of shrinkage estimate will be considered:

$$\hat{s}(\mathbf{k}) = \eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta)\ \mathbf{q}^\top \overline{\mathbf{r}}(\mathbf{k}), \tag{20}$$

where the function $\eta_\lambda(\cdot)$ is given by (17) with $\lambda \geq 0$, $\beta > 0$ and $\mathbf{q} \in \mathbb{R}^d$. The vector $\mathbf{q}$ corresponds to a linear parameter. We notice, in particular, that if the threshold value $\lambda$ is set to zero, the considered estimator reduces to $\hat{s}(\mathbf{k}) = \mathbf{q}^\top \bar{\mathbf{r}}(\mathbf{k})$. This shows that linear estimators constitute a subset of the considered class of estimators. In addition, by an appropriate choice of the vector $\mathbf{q}$, estimators consisting of a preliminary decorrelation of the data followed by a thresholding step also appear as special cases of the proposed estimator. Note that, in conventional multichannel data analysis, it is customary to decorrelate the data before processing. The most common examples are fixed channel conversions (like those from stereo to mono in sound processing or from RGB to luminance/chrominance components in color image or video processing). When the data modalities are less standardized (for instance in satellite imaging), adaptive methods such as the Karhunen-Loève transform or Independent Component Analysis (ICA) [47] can be used. The latter adaptive transforms can also be performed in the wavelet domain, *e.g.* in each subband.

Furthermore, in order to limit the computational complexity in the implementation of the estimator, it can be useful to constrain the vector $\mathbf{q}$ to belong to some vector subspace of reduced dimension $d' \leq d$. Let $\mathbf{P} \in \mathbb{R}^{d \times d'}$ be the matrix whose column vectors form a basis of this subspace. We have then $\mathbf{q} = \mathbf{P}\mathbf{a}$ where $\mathbf{a} \in \mathbb{R}^{d'}$. As a simple example, by choosing

$$\mathbf{P} = \begin{bmatrix} \mathbf{I}_{d'} \\ \mathbf{O} \end{bmatrix}$$

where $\mathbf{I}_{d'}$ denotes the identity matrix of size $d' \times d'$, we see that we only introduce in the estimator a linear combination of the first $d'$ components of the vector $\bar{\mathbf{r}}(\mathbf{k})$. In summary, the proposed form of the estimator is parameterized by $\lambda$, $\beta$ and $\mathbf{a}$ for a given choice of $\mathbf{P}$.

Our objective is to find the optimal parameters that minimize the quadratic risk defined as $R(\lambda, \beta, \mathbf{a}) = \mathsf{E}[|s(\mathbf{k}) - \hat{s}(\mathbf{k})|^2]$, for a predefined value of $\mathbf{P}$. It is easy to show that the risk reads:

$$\begin{aligned} R(\lambda, \beta, \mathbf{a}) &= \mathsf{E}[|s(\mathbf{k}) - \hat{s}(\mathbf{k})|^2] \\ &= \mathsf{E}[|s(\mathbf{k})|^2] + \mathsf{E}[|\eta_\lambda(\|\bar{\mathbf{r}}(\mathbf{k})\|^\beta)\mathbf{a}^\top \mathbf{P}^\top \bar{\mathbf{r}}(\mathbf{k})|^2] - 2\mathsf{E}[\eta_\lambda(\|\bar{\mathbf{r}}(\mathbf{k})\|^\beta)\mathbf{a}^\top \mathbf{P}^\top \bar{\mathbf{r}}(\mathbf{k}) s(\mathbf{k})]. \end{aligned} \quad (21)$$

The minimization of the risk is not obvious for any observation model. Indeed, since the $s(\mathbf{k})$ are unknown, it seems impossible to express the rightmost term $\mathsf{E}[\eta_\lambda(\|\bar{\mathbf{r}}(\mathbf{k})\|^\beta)\mathbf{a}^\top \mathbf{P}^\top \bar{\mathbf{r}}(\mathbf{k}) s(\mathbf{k})]$. However, in the case of a *Gaussian* noise, it is possible to apply an extension of Stein's principle [19] for deriving an explicit expression. In the next subsection, we will state and prove such an extended Stein's formula.

### C. Stein's formula

*Proposition 1:* Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous, almost everywhere differentiable function such that:

$$\forall \boldsymbol{\theta} \in \mathbb{R}^d, \lim_{\|\mathbf{t}\| \to +\infty} f(\mathbf{t}) \exp\Big( - \frac{(\mathbf{t}-\boldsymbol{\theta})^\top (\boldsymbol{\Gamma}^{(\overline{\mathbf{n}})})^{-1} (\mathbf{t}-\boldsymbol{\theta})}{2} \Big) = 0; \tag{22}$$

$$\mathsf{E}[|f(\overline{\mathbf{r}}(\mathbf{k}))|^2] < +\infty \quad \text{and} \quad \mathsf{E}\Big[\Big\| \frac{\partial f(\overline{\mathbf{r}}(\mathbf{k}))}{\partial \overline{\mathbf{r}}(\mathbf{k})} \Big\|\Big] < +\infty. \tag{23}$$

Then,

$$\mathsf{E}[f(\overline{\mathbf{r}}(\mathbf{k}))s(\mathbf{k})] = \mathsf{E}[f(\overline{\mathbf{r}}(\mathbf{k}))r(\mathbf{k})] - \mathsf{E}\Big[\frac{\partial f(\overline{\mathbf{r}}(\mathbf{k}))}{\partial \overline{\mathbf{r}}(\mathbf{k})}\Big]^\top \mathsf{E}[\overline{\mathbf{n}} n]. \tag{24}$$

*Proof:* Let $\mathbf{T} : \mathbb{R}^d \to \mathbb{R}^d$ be a continuous, almost everywhere differentiable function such that

$$\forall \boldsymbol{\theta} \in \mathbb{R}^d, \lim_{\|\mathbf{t}\| \to +\infty} \mathbf{T}(\mathbf{t}) \exp\Big( - \frac{(\mathbf{t}-\boldsymbol{\theta})^\top (\boldsymbol{\Gamma}^{(\overline{\mathbf{n}})})^{-1} (\mathbf{t}-\boldsymbol{\theta})}{2} \Big) = \mathbf{0}; \tag{25}$$

$$\mathsf{E}[\|\mathbf{T}(\overline{\mathbf{r}}(\mathbf{k}))\|^2] < +\infty \quad \text{and} \quad \mathsf{E}\Big[\Big\| \frac{\partial \mathbf{T}(\overline{\mathbf{r}}(\mathbf{k}))}{\partial \overline{\mathbf{r}}^\top(\mathbf{k})} \Big\|_{\mathrm{F}}\Big] < +\infty. \tag{26}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. In this multivariate context, Stein's principle [19] can be expressed as

$$\mathsf{E}[\mathbf{T}(\overline{\mathbf{r}}(\mathbf{k}))\overline{\mathbf{s}}^\top(\mathbf{k})] = \mathsf{E}[\mathbf{T}(\overline{\mathbf{r}}(\mathbf{k}))\overline{\mathbf{r}}^\top(\mathbf{k})] - \mathsf{E}\Big[\frac{\partial \mathbf{T}(\overline{\mathbf{r}}(\mathbf{k}))}{\partial \overline{\mathbf{r}}^\top(\mathbf{k})}\Big] \boldsymbol{\Gamma}^{(\overline{\mathbf{n}})}. \tag{27}$$

Eq. (24) follows by choosing $\mathbf{T} : \mathbf{t} \mapsto [f(\mathbf{t}), 0, \ldots, 0]^\top$ and focusing on the top-left element of matrix $\mathsf{E}[\mathbf{T}(\overline{\mathbf{r}}(\mathbf{k}))\overline{\mathbf{s}}^\top(\mathbf{k})]$. ■

### D. Risk expression

We define the function $f : \mathbf{u} \mapsto \eta_\lambda(\|\mathbf{u}\|^\beta)\, \mathbf{a}^\top \mathbf{P}^\top \mathbf{u}$. It is easy to check that this function $f$ satisfies the conditions of Prop. 1. Consequently, the last term can be calculated thanks to (24). This yields

$$\mathsf{E}[s(\mathbf{k}) f(\overline{\mathbf{r}}(\mathbf{k}))] = \mathsf{E}[r(\mathbf{k}) f(\overline{\mathbf{r}}(\mathbf{k}))] - \mathsf{E}\Big[\frac{\partial f\big(\overline{\mathbf{r}}(\mathbf{k})\big)}{\partial \overline{\mathbf{r}}(\mathbf{k})}\Big]^\top \boldsymbol{\Gamma}^{(\overline{\mathbf{n}},n)} \tag{28}$$

where $\boldsymbol{\Gamma}^{(\overline{\mathbf{n}},n)} = \mathsf{E}[\overline{\mathbf{n}}(\mathbf{k}) n(\mathbf{k})]$. We then have

$$\begin{aligned}
\frac{\partial f\big(\overline{\mathbf{r}}(\mathbf{k})\big)}{\partial \overline{\mathbf{r}}(\mathbf{k})} &= \mathbf{a}^\top \mathbf{P}^\top \overline{\mathbf{r}}(\mathbf{k}) \frac{\partial \eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta)}{\partial \overline{\mathbf{r}}(\mathbf{k})} + \eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta) \frac{\partial \mathbf{a}^\top \mathbf{P}^\top \overline{\mathbf{r}}(\mathbf{k})}{\partial \overline{\mathbf{r}}(\mathbf{k})} \\
&= \frac{\lambda\beta}{\|\overline{\mathbf{r}}(\mathbf{k})\|^{\beta+1}} \mathbb{1}\{\|\,\overline{\mathbf{r}}(\mathbf{k})\|^\beta > \lambda\} \frac{\partial \|\overline{\mathbf{r}}(\mathbf{k})\|}{\partial \overline{\mathbf{r}}(\mathbf{k})}\, \overline{\mathbf{r}}^\top(\mathbf{k}) \mathbf{P}\mathbf{a} + \eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta) \mathbf{P}\mathbf{a} \\
&= \eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta) \mathbf{P}\mathbf{a} + \lambda \overline{\mathbf{r}}(\mathbf{k}) \boldsymbol{\xi}^\top(\mathbf{k}) \mathbf{P}\mathbf{a}
\end{aligned} \tag{29}$$

where $\boldsymbol{\xi}(\mathbf{k}) = \mathbb{1}\{\|\,\overline{\mathbf{r}}(\mathbf{k})\|^\beta > \lambda\}\, \beta\, \overline{\mathbf{r}}(\mathbf{k}) / \|\,\overline{\mathbf{r}}(\mathbf{k})\|^{\beta+2}$. This leads to the following expression of the risk:

$$R(\lambda, \beta, \mathbf{a}) = \mathsf{E}[|r(\mathbf{k}) - f(\overline{\mathbf{r}}(\mathbf{k}))|^2] + 2\mathsf{E}[\eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta)] \mathbf{a}^\top \mathbf{P}^\top \boldsymbol{\Gamma}^{(\overline{\mathbf{n}},n)} + 2\lambda \mathbf{a}^\top \mathbf{P}^\top \mathsf{E}[\boldsymbol{\xi}(\mathbf{k}) \overline{\mathbf{r}}^\top(\mathbf{k})] \boldsymbol{\Gamma}^{(\overline{\mathbf{n}},n)} - \sigma^2 \tag{30}$$

where $\sigma^2 = \mathsf{E}[|n(\mathbf{k})|^2]$.

We will now look for parameters $\lambda$, $\beta$ and $\mathbf{a}$ that minimize the risk expression (30) for a given choice of $\mathbf{P}$.

### *E. Determination of the parameter* $\mathbf{a}$

We first aim at calculating the value of $\mathbf{a}$ that minimizes the risk (30). By noticing that the risk is a quadratic convex function of $\mathbf{a}$, the minimization can be performed by differentiating w.r.t. $\mathbf{a}$ and then finding $\mathbf{a}^*(\lambda,\beta)$ such that $\partial R/\partial \mathbf{a}\big(\lambda,\beta,\mathbf{a}^*(\lambda,\beta)\big) = 0$. It readily follows that

$$\mathbf{a}^*(\lambda,\beta) = \big(\mathbf{P}^\top \mathsf{E}[\eta_\lambda^2(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta)\overline{\mathbf{r}}(\mathbf{k})\overline{\mathbf{r}}^\top(\mathbf{k})]\mathbf{P}\big)^{-1}\mathbf{P}^\top\big(\mathsf{E}[\eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta) r(\mathbf{k})\overline{\mathbf{r}}(\mathbf{k})] - \mathsf{E}[\eta_\lambda(\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta)]\mathbf{\Gamma}^{(\overline{\mathbf{n}},n)} - \lambda\,\mathsf{E}[\boldsymbol{\xi}(\mathbf{k})\overline{\mathbf{r}}^\top(\mathbf{k})]\mathbf{\Gamma}^{(\overline{\mathbf{n}},n)}\big). \quad (31)$$

### *F. Determination of the parameters* $\lambda$ *and* $\beta$

Starting from (30), the risk $R(\lambda,\beta,\mathbf{a})$ can be re-expressed as $R(\lambda,\beta,\mathbf{a}) = \mathsf{E}[\rho_{\lambda,\beta,\mathbf{a}}(\mathbf{k})]$ where

$$\rho_{\lambda,\beta,\mathbf{a}}(\mathbf{k}) = \alpha_2(\mathbf{k})\lambda^2 + \alpha_1(\mathbf{k})\lambda + \alpha_0(\mathbf{k}) \quad (32)$$

and

$$\alpha_0(\mathbf{k}) = r^2(\mathbf{k}) - \sigma^2 + \mathbb{1}\{\|\,\overline{\mathbf{r}}(\mathbf{k})\|^\beta > \lambda\}\mathbf{a}^\top\mathbf{P}^\top\Big(2\,\mathbf{\Gamma}^{(\overline{\mathbf{n}},n)} + \big(\mathbf{a}^\top\mathbf{P}^\top\overline{\mathbf{r}}(\mathbf{k}) - 2\,r(\mathbf{k})\big)\,\overline{\mathbf{r}}(\mathbf{k})\Big)$$

$$\alpha_1(\mathbf{k}) = 2\mathbf{a}^\top\mathbf{P}^\top\Big(\frac{\big(r(\mathbf{k}) - \mathbf{a}^\top\mathbf{P}^\top\overline{\mathbf{r}}(\mathbf{k})\big)\,\overline{\mathbf{r}}(\mathbf{k}) - \mathbf{\Gamma}^{(\overline{\mathbf{n}},n)}}{\|\overline{\mathbf{r}}(\mathbf{k})\|^\beta} + \beta\,\overline{\mathbf{r}}(\mathbf{k})\frac{\overline{\mathbf{r}}^\top(\mathbf{k})\mathbf{\Gamma}^{(\overline{\mathbf{n}},n)}}{\|\overline{\mathbf{r}}(\mathbf{k})\|^{\beta+2}}\Big)\mathbb{1}\{\|\,\overline{\mathbf{r}}(\mathbf{k})\|^\beta > \lambda\}$$

$$\alpha_2(\mathbf{k}) = \mathbb{1}\{\|\,\overline{\mathbf{r}}(\mathbf{k})\|^\beta > \lambda\}\frac{\big(\mathbf{a}^\top\mathbf{P}^\top\overline{\mathbf{r}}(\mathbf{k})\big)^2}{\|\overline{\mathbf{r}}(\mathbf{k})\|^{2\beta}}.$$

In practice, under standard mixing assumptions for $(\overline{\mathbf{n}}(\mathbf{k}))_{k\in\mathbb{Z}^2}$ and $(\overline{\mathbf{s}}(\mathbf{k}))_{k\in\mathbb{Z}^2}$ [48], $R(\lambda,\beta,\mathbf{a})$ can be estimated via an empirical average $\overset{\wedge}{R}(\lambda,\beta,\mathbf{a})$ computed over $\mathbb{K}$, provided that the data length $K = \mathrm{card}(\mathbb{K})$ is large enough. Following a procedure similar to the search implemented for the SUREshrink estimator, we will subsequently determine optimal values of $\lambda$ and $\beta$ for this consistent risk estimate. More precisely, the norms of the ROVs $(\|\overline{\mathbf{r}}(\mathbf{k})\|)_{\mathbf{k}\in\mathbb{K}}$ are first sorted in descending order, so that $\|\overline{\mathbf{r}}(\mathbf{k}_1)\| \geq \|\overline{\mathbf{r}}(\mathbf{k}_2)\| \geq \ldots \geq \|\overline{\mathbf{r}}(\mathbf{k}_K)\|$. To study the variations of $\overset{\wedge}{R}(\lambda,\beta,\mathbf{a})$ w.r.t. $\lambda$, we consider the case when $\lambda \in I_{i_0}$ with $i_0 \in \{1,\ldots,K+1\}$ and

$$I_{i_0} = \begin{cases} \big[\|\overline{\mathbf{r}}(\mathbf{k}_1)\|^\beta, \infty\big) & \text{if } i_0 = 1 \\ \big[\|\overline{\mathbf{r}}(\mathbf{k}_{i_0})\|^\beta, \|\overline{\mathbf{r}}(\mathbf{k}_{i_0-1})\|^\beta\big) & \text{if } i_0 \in \{2,\ldots,K\} \\ \big[0, \|\overline{\mathbf{r}}(\mathbf{k}_K)\|^\beta\big) & \text{if } i_0 = K+1. \end{cases} \quad (33)$$

On the interval $I_{i_0}$, the risk estimate then takes the following form:[1]

$$\overset{\curlywedge}{R}(\lambda,\beta,\mathbf{a}) = \frac{1}{K}\Big(\sum_{i=1}^{i_0-1}\rho_{\lambda,\beta,\mathbf{a}}(\mathbf{k}_i) + \sum_{i=i_0}^{K}\rho_{\lambda,\beta,\mathbf{a}}(\mathbf{k}_i)\Big) \tag{34}$$

$$= \frac{1}{K}\Big(\lambda^2\sum_{i=1}^{i_0-1}\alpha_2(\mathbf{k}_i) + \lambda\sum_{i=1}^{i_0-1}\alpha_1(\mathbf{k}_i) + \sum_{i=1}^{i_0-1}\alpha_0(\mathbf{k}_i) + \sum_{i=i_0}^{K}r^2(\mathbf{k}_i) - (K+1-i_0)\sigma^2\Big). \tag{35}$$

In other words, $\overset{\curlywedge}{R}(\lambda,\beta,\mathbf{a})$ is a piecewise second-order polynomial function of $\lambda$. Assume now that $i_0 \in \{2,\ldots,K\}$. For given values of $\beta$ and $\mathbf{a}$, the minimum over $\mathbb{R}$ of the polynomial in (35) is reached at

$$\tilde{\lambda}_{i_0}(\beta,\mathbf{a}) = -\frac{\sum_{i=1}^{i_0-1}\alpha_1(\mathbf{k}_i)}{2\sum_{i=1}^{i_0-1}\alpha_2(\mathbf{k}_i)}. \tag{36}$$

The minimum over $\left[\|\overline{\mathbf{r}}(\mathbf{k}_{i_0})\|^\beta, \|\overline{\mathbf{r}}(\mathbf{k}_{i_0-1})\|^\beta\right]$ of the estimated risk is therefore given by

$$\lambda^*_{i_0}(\beta,\mathbf{a}) = \begin{cases} \|\overline{\mathbf{r}}(\mathbf{k}_{i_0-1})\|^\beta & \text{if } \tilde{\lambda}_{i_0}(\beta,\mathbf{a}) \geq \|\overline{\mathbf{r}}(\mathbf{k}_{i_0-1})\|^\beta \\ \tilde{\lambda}_{i_0}(\beta,\mathbf{a}) & \text{if } \tilde{\lambda}_{i_0}(\beta,\mathbf{a}) \in I_{i_0} \\ \|\overline{\mathbf{r}}(\mathbf{k}_{i_0})\|^\beta & \text{if } \tilde{\lambda}_{i_0}(\beta,\mathbf{a}) < \|\overline{\mathbf{r}}(\mathbf{k}_{i_0})\|^\beta. \end{cases} \tag{37}$$

The minimizers $\lambda^*_1(\beta,\mathbf{a})$ and $\lambda^*_{K+1}(\beta,\mathbf{a})$ of the estimated risk over $I_1$ and $I_{K+1}$ can be found in a similar way. The global minimizer $\lambda^*(\beta,\mathbf{a})$ of the estimated risk is subsequently computed as

$$\lambda^*(\beta,\mathbf{a}) = \arg\min_{(\lambda^*_{i_0}(\beta,\mathbf{a}))_{1\leq i_0\leq K+1}} \overset{\curlywedge}{R}(\lambda^*_{i_0}(\beta,\mathbf{a}),\beta,\mathbf{a}). \tag{38}$$

To determine the optimal value $\beta^*(\mathbf{a})$ of the exponent $\beta$, we can then proceed to an exhaustive search over a set $\mathcal{V}$ of possible values for this parameter by choosing

$$\beta^*(\mathbf{a}) = \arg\min_{\beta\in\mathcal{V}} \overset{\curlywedge}{R}(\lambda^*(\beta,\mathbf{a}),\beta,\mathbf{a}). \tag{39}$$

In our experiments, it was observed that a restricted set of a few search values is sufficient to get good results.

### *G. Iterative optimization algorithm*

The optimal expression of the vector $\mathbf{a}$ is derived in a closed form in Section III-E as a function of the parameters $\lambda$ and $\beta$. In this way, the optimization problem simply reduces to the determination of the latter two parameters. On the other hand, given $\mathbf{a}$, a procedure for determining the optimal values of

[1]We adopt here the convention $\sum_{i=1}^{0}\cdot = \sum_{i=K+1}^{K}\cdot = 0$.

$\lambda$ and $\beta$ is described in Section III-F. In order to get optimized values of the estimator parameters, we therefore propose to apply the following iterative optimization approach:

1) Initialization: Fix $\mathbf{P}$ and $\mathcal{V}$. Set the iteration number $p = 1$ and $\mathbf{a}^{(0)} = [1, 0, \ldots, 0]^\top \in \mathbb{R}^{d'}$
2) Iteration $p$
   a) Set $\beta^{(p)} = \beta^*(\mathbf{a}^{(p-1)})$ and $\lambda^{(p)} = \lambda^*(\beta^{(p)}, \mathbf{a}^{(p-1)})$ as described in Section III-F.
   b) Set $\mathbf{a}^{(p)} = \mathbf{a}^*(\lambda^{(p)}, \beta^{(p)})$ using (31) where the expectations are replaced by sample estimates.
3) Set $p \leftarrow p + 1$ and goto step 2 until convergence.
4) Return the optimized values $(\lambda^{(p)}, \beta^{(p)}, \mathbf{a}^{(p)})$ of the parameters.

We point out that, although we were not able to prove the convergence of the optimized parameters, the generated sequence $(\overset{\curlywedge}{R}(\lambda^{(p)}, \beta^{(p)}, \mathbf{a}^{(p)}))_p$ is a decreasing convergent sequence. This means that the generated parameters at each iteration of the algorithm allow to decrease the risk value.

## IV. Multicomponent wavelet denoising

Our objective here is to apply the nonlinear estimator developed in the previous section to noise reduction in degraded multicomponent images by considering wavelet-based approaches. The original multichannel image is composed of $B \in \mathbb{N}^*$ components $s^{(b)}$ of size $L \times L$, with $b \in \{1, \ldots, B\}$. Each image component $s^{(b)}$ is corrupted by an additive noise $n^{(b)}$, which is assumed independent of the images of interest. Consequently, we obtain the following noisy observation field $r^{(b)}$ defined by:

$$\forall \mathbf{k} \in \mathbb{K}, \qquad r^{(b)}(\mathbf{k}) = s^{(b)}(\mathbf{k}) + n^{(b)}(\mathbf{k}), \tag{40}$$

where $\mathbb{K} = \{1, \ldots, L\}^2$. Following a multivariate approach, we define:

$$\forall \mathbf{k} \in \mathbb{K}, \qquad \left\{ \begin{array}{lcl} \mathbf{s}(\mathbf{k}) & \triangleq & [s^{(1)}(\mathbf{k}), \ldots, s^{(B)}(\mathbf{k})]^\top \\ \mathbf{n}(\mathbf{k}) & \triangleq & [n^{(1)}(\mathbf{k}), \ldots, n^{(B)}(\mathbf{k})]^\top \\ \mathbf{r}(\mathbf{k}) & \triangleq & [r^{(1)}(\mathbf{k}), \ldots, r^{(B)}(\mathbf{k})]^\top \end{array} \right. . \tag{41}$$

Obviously, the observation model (40) can be rewritten as $\forall \mathbf{k} \in \mathbb{K}, \quad \mathbf{r}(\mathbf{k}) = \mathbf{s}(\mathbf{k}) + \mathbf{n}(\mathbf{k})$. In many optical systems, the noise stems from a combination of photonic and electronic noises cumulated with quantization errors. Subsequently, we will assume that the noise vector process $\mathbf{n}$ is zero-mean iid Gaussian with covariance matrix $\mathbf{\Gamma}^{(\mathbf{n})}$. In [1] and [2], this was shown to constitute a realistic assumption for satellite systems. It is worth noticing that a non diagonal matrix $\mathbf{\Gamma}^{(\mathbf{n})}$ indicates that inter-component correlations exist between co-located noise samples.

Hereafter, we will use two decompositions. The first one consists in a critically decimated $M$-band wavelet transform whereas the second one, corresponds to an $M$-band dual-tree wavelet decomposition we recently proposed [27], which permits a directional analysis of images.

### *A. $M$-band wavelet basis estimation*

*1) Model:* We first consider an $M$-band orthonormal discrete wavelet transform (DWT) [30] over $J$ resolution levels applied, for each channel $b$, to the observation field $r^{(b)}$. This decomposition produces $M^2-1$ wavelet subband sequences $r^{(b)}_{j,\mathbf{m}}$, $\mathbf{m} \in \mathbb{N}^2_M \setminus \{(0,0)\}$, each of size $L_j \times L_j$ (where $L_j = L/M^j$)[2], at every resolution level $j$ and an additional approximation sequence $r^{(b)}_{J,\mathbf{0}}$ of size $L_J \times L_J$, at resolution level $J$.

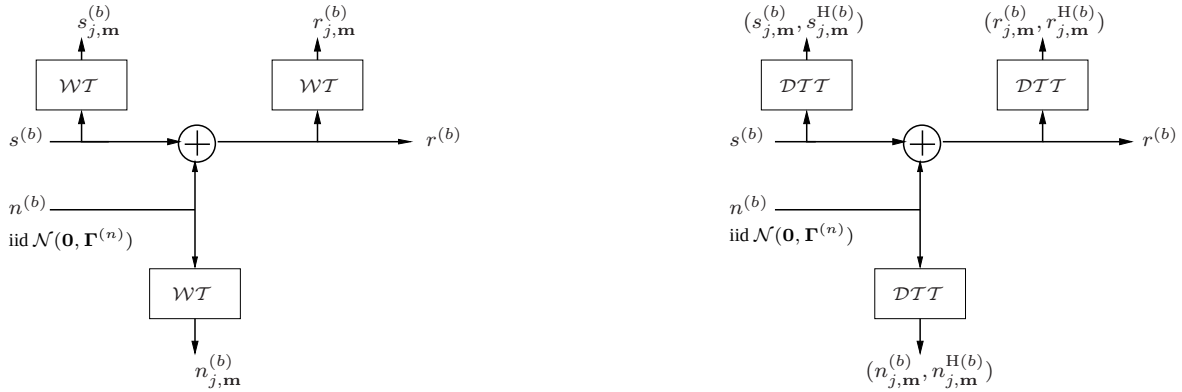


Fig. 1. Considered models in the wavelet transform domain (left) and in the dual-tree transform domain (right).

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

$$\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) \stackrel{\triangle}{=} [s^{(1)}_{j,\mathbf{m}}(\mathbf{k}), \ldots, s^{(B)}_{j,\mathbf{m}}(\mathbf{k})]^\top,$$
$$\mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) \stackrel{\triangle}{=} [n^{(1)}_{j,\mathbf{m}}(\mathbf{k}), \ldots, n^{(B)}_{j,\mathbf{m}}(\mathbf{k})]^\top,$$
$$\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) \stackrel{\triangle}{=} [r^{(1)}_{j,\mathbf{m}}(\mathbf{k}), \ldots, r^{(B)}_{j,\mathbf{m}}(\mathbf{k})]^\top.$$

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. $\mathcal{N}(\mathbf{0}, \mathbf{\Gamma^{(n)}})$ 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 $(j, \mathbf{m})$.

[2]For simplicity, $L$ is assumed to be divisible by $M^J$.

*2) Associated estimator:* As described in Section III, our estimator can be directly applied to the $M$-band DWT coefficients. As in conventional approaches, the approximation coefficients (*i.e.* $j = J$ and $\mathbf{m} = (0,0)$) are kept untouched. The parameters $\lambda_{j,\mathbf{m}}$, $\beta_{j,\mathbf{m}}$ and $\mathbf{q}_{j,\mathbf{m}}$ can be determined adaptively, for every subband $(j,\mathbf{m})$ and every component $b$. In this case, the ROV can be scalar, spatial, inter-component or combined spatial/inter-component. More detailed examples will be given in Section V.

### *B. $M$-band dual-tree wavelet frame estimation*

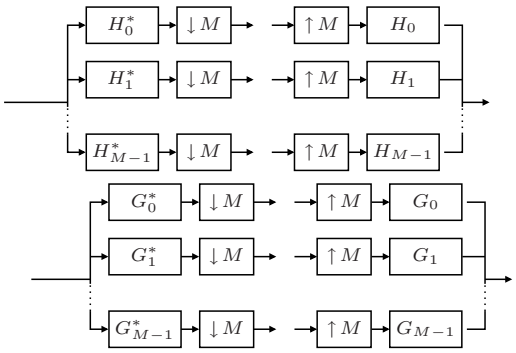


Fig. 2. Pair of analysis/synthesis $M$-band para-unitary filter banks.

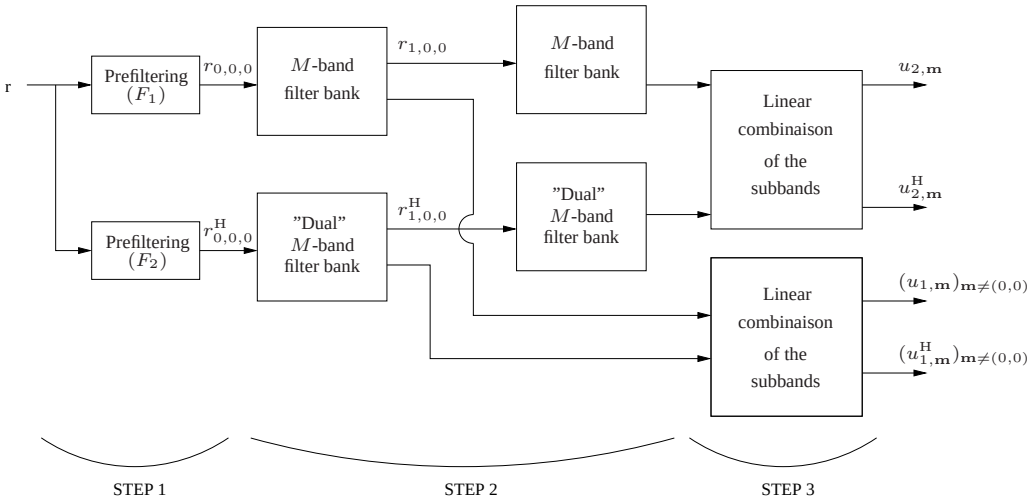


Fig. 3. Dual-tree $2D$.

*1) A brief overview of the decomposition:* The $M$-band real dual-tree transform (DTT) consists in performing two separable $M$-band orthonormal wavelet decompositions in parallel as illustrated by Fig. 2. The one-dimensional wavelets $(\psi_m)_{m\in\mathbb{N}^\star_M}$ corresponding to the primal tree (upper branch) are assumed

known and the "dual tree" ones $(\psi_m^{\mathrm{H}})_{m\in\mathbb{N}_M^\star}$ (used in the lower branch) are built so that they define Hilbert pairs with the primal ones. This reads in the frequency domain: $\forall m \in \mathbb{N}_M^\star,\ \widehat{\psi}_m^{\mathrm{H}}(\omega) = -\imath\,\mathrm{sign}(\omega)\widehat{\psi}_m(\omega)$. Details of construction are given in [27] and the global scheme of the decomposition is shown in Fig. 3. An important point is that the dual-tree decomposition includes a post-processing, consisting of a linear isometric combination of the primal/dual subbands (see Fig. 3). This post-processing constitutes an essential step for obtaining a directional analysis. Finally, two sets of coefficients (primal and dual ones) are obtained, which means that this representation involves a limited redundancy of a factor two.

*2) Model:* Applying this decomposition to a multichannel image having $B$ components and using similar notations to Section IV-A.1, we obtain the following coefficients for the original data, the observed ones and the noise, respectively:

- before post-processing: $(\mathbf{s}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{s}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k})), (\mathbf{r}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{r}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k})), (\mathbf{n}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{n}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}))$;
- after post-processing: $(\mathbf{v}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{v}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k})), (\mathbf{u}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{u}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k})), (\mathbf{w}_{j,\mathbf{m}}(\mathbf{k}), \mathbf{w}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}))$.

Note that a post-processing is not applied to all subbands (see [27]) as the Hilbert condition is only verified by mother wavelets. As a consequence, the linear isometric combination is not performed for subbands processed by low pass filters. More precisely, the post-processing consists of the following unitary transform of the detail coefficients: for all $\mathbf{m} \in \mathbb{N}_M^{\star 2}$,

$$\forall \mathbf{k} \in \mathbb{K}_j, \qquad \mathbf{w}_{j,\mathbf{m}}(\mathbf{k}) = \frac{1}{\sqrt{2}}\big(\mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) + \mathbf{n}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k})\big) \quad \text{and} \quad \mathbf{w}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}) = \frac{1}{\sqrt{2}}\big(\mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) - \mathbf{n}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k})\big). \quad (42)$$

Similar relations hold for the original and observed data. Furthermore, invoking the linearity property of the transform, these coefficients are related by (see. Fig. 1 (right)):

$$\begin{aligned} \forall \mathbf{k} \in \mathbb{K}_j, \quad \mathbf{r}_{j,\mathbf{m}}(\mathbf{k}) &= \mathbf{s}_{j,\mathbf{m}}(\mathbf{k}) + \mathbf{n}_{j,\mathbf{m}}(\mathbf{k}) \quad \text{and} \quad \mathbf{u}_{j,\mathbf{m}}(\mathbf{k}) = \mathbf{v}_{j,\mathbf{m}}(\mathbf{k}) + \mathbf{w}_{j,\mathbf{m}}(\mathbf{k}) \\ \mathbf{r}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}) &= \mathbf{s}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}) + \mathbf{n}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}) \qquad\qquad \mathbf{u}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}) = \mathbf{v}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}) + \mathbf{w}_{j,\mathbf{m}}^{\mathrm{H}}(\mathbf{k}). \end{aligned} \quad (43)$$

*3) Noise statistical properties:* In our recent work [49], [50], a detailed analysis of the noise statistical properties after such a dual tree decomposition has been performed. In the sequel, some of the main results we obtained are briefly summarized. Let us recall the definition of the deterministic cross-correlation function between the primal and dual wavelets: for all $(m, m') \in \mathbb{N}_M^2$,

$$\forall \tau \in \mathbb{R}, \qquad \gamma_{m,m'}(\tau) = \int_{-\infty}^{\infty} \psi_m(x)\psi_{m'}^{\mathrm{H}}(x-\tau)\,dx. \quad (44)$$

We have obtained the following expressions for the covariance fields: for all $j \in \mathbb{Z}$, $\mathbf{m} = (m_1, m_2) \in \mathbb{N}_M^2$, $\mathbf{m}' = (m_1', m_2') \in \mathbb{N}_M^2$, $\mathbf{k} = (k_1, k_2) \in \mathbb{K}_j$ and $\mathbf{k}' = (k_1', k_2') \in \mathbb{K}_j$, with the index difference $\Delta_{p_i} =$

$p_i - p'_i$, $i \in \{1,2\}$:

$$\left.\begin{array}{r} \mathsf{E}[\mathbf{n}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{n}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] \\ \mathsf{E}[\mathbf{n}^{\mathrm{H}}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{n}^{\mathrm{H}}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] \end{array}\right\} = \mathbf{\Gamma}^{(\mathbf{n})}\delta_{\Delta_{m_1}}\delta_{\Delta_{m_2}}\delta_{\Delta_{k_1}}\delta_{\Delta_{k_2}}$$

$$\mathsf{E}[\mathbf{n}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{n}^{\mathrm{H}}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] = \mathbf{\Gamma}^{(\mathbf{n})}\gamma_{m_1,m'_1}(\Delta_{k_1})\gamma_{m_2,m'_2}(\Delta_{k_2})\,.$$

It can be further noticed that, for $\mathbf{m} \neq \mathbf{0}$, the random vectors $\mathbf{n}_{j,\mathbf{m}}(\mathbf{k})$ and $\mathbf{n}^{\mathrm{H}}_{j,\mathbf{m}}(\mathbf{k})$ at a given location $\mathbf{k}$ are mutually uncorrelated.

After post-processing, the covariances of the transformed noise coefficient fields can be easily deduced from (42): for all $(\mathbf{m},\mathbf{m}') \in \mathbb{N}^{\star 2}_M$ and $(\mathbf{k},\mathbf{k}') \in \mathbb{K}^2_j$,

$$\mathsf{E}[\mathbf{w}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{w}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] = \mathsf{E}[\mathbf{n}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{n}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] + \mathsf{E}[\mathbf{n}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{n}^{\mathrm{H}}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] \tag{45}$$

$$\mathsf{E}[\mathbf{w}^{\mathrm{H}}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{w}^{\mathrm{H}}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] = \mathsf{E}[\mathbf{n}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{n}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] - \mathsf{E}[\mathbf{n}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{n}^{\mathrm{H}}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] \tag{46}$$

$$\mathsf{E}[\mathbf{w}_{j,\mathbf{m}}(\mathbf{k})(\mathbf{w}^{\mathrm{H}}_{j,\mathbf{m}'}(\mathbf{k}'))^\top] = \mathbf{0}. \tag{47}$$

In summary, noise coefficients are inter-tree correlated before the post-transform whereas after the post-transform, they are spatially correlated. This constitutes an important consequence of the post-processing stage.

*4) Associated estimator:* In the $M$-band DTT case, the primal and dual coefficients are both estimated. For each component $b \in \{1,\ldots,B\}$, the estimator reads: for the subbands which are not linearly combined ($\mathbf{m} \notin \mathbb{N}^\star_M$),

$$\overset{\curlywedge}{s}{}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) = \eta_{\lambda^{(b)}_{j,\mathbf{m}}}(\|\overline{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})\|^{\beta^{(b)}_{j,\mathbf{m}}})\ (\mathbf{q}^{(b)}_{j,\mathbf{m}})^\top \overline{\mathbf{r}}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) \tag{48}$$

$$\overset{\curlywedge}{s}{}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k}) = \eta_{\lambda^{\mathrm{H}(b)}_{j,\mathbf{m}}}(\|(\overline{\mathbf{r}}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k}))\|^{\beta^{\mathrm{H}(b)}_{j,\mathbf{m}}})\ \left(\mathbf{q}^{\mathrm{H}(b)}_{j,\mathbf{m}}\right)^\top \overline{\mathbf{r}}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k}), \tag{49}$$

and, for the combined subbands ($\mathbf{m} \in \mathbb{N}^\star_M$),

$$\overset{\curlywedge}{v}{}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) = \eta_{\lambda^{(b)}_{j,\mathbf{m}}}(\|\overline{\mathbf{u}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})\|^{\beta^{(b)}_{j,\mathbf{m}}})\ (\mathbf{q}^{(b)}_{j,\mathbf{m}})^\top \overline{\mathbf{u}}^{(b)}_{j,\mathbf{m}}(\mathbf{k}) \tag{50}$$

$$\overset{\curlywedge}{v}{}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k}) = \eta_{\lambda^{\mathrm{H}(b)}_{j,\mathbf{m}}}(\|(\overline{\mathbf{u}}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k}))\|^{\beta^{\mathrm{H}(b)}_{j,\mathbf{m}}})\ \left(\mathbf{q}^{\mathrm{H}(b)}_{j,\mathbf{m}}\right)^\top \overline{\mathbf{u}}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k}), \tag{51}$$

where $\overline{\mathbf{r}}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k})$ and $\overline{\mathbf{r}}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k})$ (resp. $\overline{\mathbf{u}}^{(b)}_{j,\mathbf{m}}(\mathbf{k})$ and $\overline{\mathbf{u}}^{\mathrm{H}(b)}_{j,\mathbf{m}}(\mathbf{k})$) are the ROVs for the primal and dual coefficients before (resp. after) post-transformation. Similarly to the DWT case, $(\lambda_{j,\mathbf{m}},\beta_{j,\mathbf{m}},\mathbf{q}_{j,\mathbf{m}})$ and $(\lambda^{\mathrm{H}}_{j,\mathbf{m}},\beta^{\mathrm{H}}_{j,\mathbf{m}},\mathbf{q}^{\mathrm{H}}_{j,\mathbf{m}})$ can be adaptively determined by minimizing the quadratic risk over the frame coefficients for every subband $(j,\mathbf{m})$ and every component $b$ in each tree. Furthermore, the approximation coefficients are also kept untouched. The denoised multichannel images are then obtained from the

estimated wavelet coefficients by inverting the DTT using the optimal reconstruction developed in [27]. In this case, a great flexibility exists in the choice of the ROV since the latter can be scalar, spatial, inter-component, inter-tree or combined spatial/inter-component/inter-tree as will be illustrated in the next section.

## V. Numerical results

We now provide numerical examples showing the efficiency of the proposed method. In our simulations, we consider different multichannel remote sensing images. For the sake of clarity, we only provide experimental results concerning two multispectral images. The first one designated as Tunis corresponds to a part of a SPOT3 scene depicting a urban area of the city of Tunis ($B = 3$). The second one named Trento is a Landsat Thematic Mapper image having initially seven channels. The thermal component (the sixth component) has been discarded since it is not similar to the remaining ones. Hence, the test image Trento is a $B = 6$ component image. In order to obtain reliable results from a statistical viewpoint, Monte Carlo simulations have been conducted. According to our experiments, averaging the mean square error over five noise realizations is sufficient to obtain consistent quantitative evaluations.

In the following, we discuss several topics: in particular, we compare our method with other recently proposed estimators, possibly having a multivariate structure. Then, we consider different pre-processings that can be performed on the multichannel data before applying the estimator, thus expecting improved results. The ROV being defined in a generic way in the previous section, we also study the influence of specific choices of this ROV on the denoising performance as well as the influence of the wavelet choice (considering various $M$-band filter banks). When different decompositions are performed, we set the maximum decomposition level so that the size of the approximation fields remain the same. Consequently, we decompose the images over $2$ levels for a $4$-band filter bank structure and $4$ levels for a dyadic one.

If $\sigma^{(b)}$ denotes the standard deviation of the clean multichannel component $s^{(b)}$ (of size $L \times L$) we define the initial and the final signal to noise ratios $\mathrm{SNR}_{\mathrm{initial}}^{(b)}$ and, $\mathrm{SNR}_{\mathrm{final}}^{(b)}$ in the $b$-th channel as:

$$\mathrm{SNR}_{\mathrm{initial}}^{(b)} \triangleq 10\log_{10}\left(\frac{(\sigma^{(b)})^2 L^2}{\|s^{(b)} - r^{(b)}\|^2}\right), \quad \text{and} \quad \mathrm{SNR}_{\mathrm{final}}^{(b)} \triangleq 10\log_{10}\left(\frac{(\sigma^{(b)})^2 L^2}{\|s^{(b)} - \hat{s}^{(b)}\|^2}\right). \tag{52}$$

Then, all the $B$ channel contributions are averaged into global values of the initial and final signal to noise ratio $\mathrm{SNR}_{\mathrm{initial}}$ and, $\mathrm{SNR}_{\mathrm{final}}$.

TABLE I

BRIEF DESCRIPTION OF THE TESTED METHODS.

| Acronym | Description | Ref. | Acronym | Description | Ref. |
|---|---|---|---|---|---|
| Biv. | Bivariate shrinkage method | [45] | Multivariate methods | | |
| BLS-GSM | Bayesian Least Squares (BLS) Gaussian Scale Mixture (GSM) using critically decimated DWT | [34] | ProbShrink ($. \times .$) | Multivariate method for 3-band images using critically decimated DWT and taking into account a ($. \times .$) neighborhood in each channel | [39] |
| BLS-GSM + parent | BLS-GSM using critically decimated DWT and taking into account the parent coefficient | [34] | ProbShrink red. ($. \times .$) | Multivariate method for 3-band images using undecimated DWT and taking into account a ($. \times .$) neighborhood in each channel | [39] |
| BLS-GSM red. | BLS-GSM using a full steerable pyramid (redundant transform) | [34] | Surevect | Estimator based on an extended SURE approach using a critically decimated DWT | [22] |
| Curvelets | Block estimator using curvelet transform: 7.5 times redundant | [51] | | | |

### *A. Comparison with existing methods*

We aim in this section at comparing the proposed approach with several existing denoising methods which are briefly described in Table I. Tests are performed on a $512 \times 512$ SPOT3 image of Tunis city ($B = 3$) (as some multivariate methods are limited to 3-band images) corrupted by an additive zero-mean white Gaussian noise with covariance matrix $\mathbf{\Gamma}_1^{(\mathbf{n})} = \sigma^2 \, \mathbf{I}_B$, where $\mathbf{I}_B$ denotes the identity matrix of size $B \times B$.

We first study techniques that use orthogonal wavelet transforms. We employ Daubechies wavelets of order $4$ in all the following estimators:

1) the Bivariate shrinkage, which takes into account inter-scale dependencies, the last level being processed by inverting children and parent role [45];
2) the BLS-GSM method developed in [34] including or not the parent neighborhood and considering a $3 \times 3$ spatial neighborhood;[3]
3) the ProbShrink estimator [39] for multivariate data with a $3 \times 3$ spatial neighborhood (in each channel);[4]
4) the Surevect estimator [22], which only takes into account multicomponent statistical dependencies;

[3]We use the toolbox available from Portilla's website `http://www.io.csic.es/PagsPers/JPortilla/`.

[4]We use the toolbox available from Pižurica's website `http://telin.rug.ac.be/∼sanja/`.

5) the proposed estimator where the set of values taken by $\beta_{j,\mathbf{m}}^{(b)}$ is $\mathcal{V} = \{0.5, 1, 1.5, 2\}$, the ROV is represented in Fig. 5(b). A subspace constraint is added on the vector $\mathbf{q}_{j,\mathbf{m}}^{(b)}$ so that $(\mathbf{q}_{j,\mathbf{m}}^{(b)})^\top \bar{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k})$ reduces to a linear combination of the multichannel data at the considered location and the 4 spatial nearest neighbors.

TABLE II

DENOISING RESULTS (AVERAGE VALUES COMPUTED OVER 3 CHANNELS) ON TUNIS IMAGE USING NON REDUNDANT ORTHOGONAL TRANSFORMS (SEE TAB. I) WITH DAUBECHIES WAVELETS OF ORDER 4 (LENGTH 8).

| $\sigma^2$ | $\text{SNR}_{\text{init}}$ | Biv | Probshrink ($3 \times 3$) | BLS-GSM | BLS-GSM + parent | Surevect | Proposed |
|---|---|---|---|---|---|---|---|
| 650.3 | 5.081 | 11.85 | 11.86 | 12.05 | 12.14 | 13.08 | **13.41** |
| 410.3 | 7.081 | 12.89 | 12.84 | 13.11 | 13.21 | 14.12 | **14.51** |
| 258.9 | 9.081 | 13.99 | 13.91 | 14.26 | 14.36 | 15.24 | **15.69** |
| 163.3 | 11.08 | 15.19 | 15.08 | 15.49 | 15.60 | 16.43 | **16.95** |
| 103.1 | 13.08 | 16.49 | 16.37 | 16.81 | 16.93 | 17.70 | **18.27** |
| 65.03 | 15.08 | 17.88 | 17.54 | 18.22 | 18.35 | 19.04 | **19.64** |

The obtained results are provided in Table II (the initial SNRs may be different in each channel although the noise variance is fixed). For the first three methods, denoising has been performed for each component of the multichannel data. For orthogonal wavelets, ProbShrink leads to better results when it is associated to a spatial neighborhood than when considering only the pixel value to be estimated. It performs quite similarly to the Bivariate shrinkage. The BLS-GSM estimator outperforms these two methods providing a gain of approximatively $0.2$ dB (up to $0.3$ dB by including the parent coefficient in the neighborhood). Nevertheless, the Surevect estimator brings more significant improvements and it can be observed that our method leads to even better numerical results whatever the initial noise level is. The new structure of the estimator coupled with a spatial and spectral block processing may explain such an improvement. Furthermore, the gain increases as the initial SNR increases, which is interesting in satellite imaging where the noise is often of low intensity. To be fair, we would like to mention that, although Bivariate shrinkage, Probshrink and BLS-GSM were designed for monochannel image denoising, extensions of these methods to the multivariate case could probably be envisaged.

In the monochannel case, it has been reported that the use of redundant transforms often brings noticeable improvements in denoising [51]. We subsequently compare methods that have been proved to be very efficient when combined with a redundant analysis:

1) the curvelet denoising [51] using a curvelet frame with a redundancy approximatively equal to $7.5$

and a block thresholding;[5]

2) the BLS-GSM method using steerable pyramids with $8$ orientations, including the parent neighborhood and a $3 \times 3$ spatial neighborhood as described in [34],
3) the ProbShrink estimator for multivariate data using undecimated wavelet transform [39] (with Daubechies wavelets of length $8$) and taking into account a $3 \times 3$ or no spatial neighborhood;
4) the Surevect estimator [22], extended to DTT (with Daubechies wavelets of length $8$);
5) the proposed estimator using a DTT where $\mathcal{V} = \{0.5, 1, 1.5, 2\}$, the ROV is represented in Fig. 6(b). The vector $\mathbf{q}_{j,\mathbf{m}}^{(b)}$ (resp. $\mathbf{q}_{j,\mathbf{m}}^{\mathrm{H}(b)}$) is such that it introduces a linear combination of the multichannel data in the primal (resp. dual) tree at the considered location and the 4 spatial nearest neighbors.

TABLE III

DENOISING RESULTS (AVERAGE VALUES COMPUTED OVER 3 CHANNELS) ON TUNIS IMAGE USING REDUNDANT TRANSFORMS (SEE TAB. I) WITH DAUBECHIES WAVELETS OF ORDER 4 (LENGTH 8).

| $\sigma^2$ | $\mathrm{SNR}_{\mathrm{init}}$ | Curvelets | BLS-GSM red + parent | Probshrink red ($3 \times 3$) | Probshrink red ($1 \times 1$) | Surevect DTT | Proposed DTT |
|---|---|---|---|---|---|---|---|
| 650.3 | 5.081 | 11.91 | 12.92 | 13.00 | 13.33 | 13.54 | **13.71** |
| 410.3 | 7.081 | 12.94 | 14.00 | 14.04 | 14.38 | 14.59 | **14.80** |
| 258.9 | 9.081 | 14.04 | 15.15 | 15.13 | 15.50 | 15.70 | **15.97** |
| 163.3 | 11.081 | 15.17 | 16.38 | 16.28 | 16.68 | 16.87 | **17.21** |
| 103.1 | 13.081 | 16.33 | 17.68 | 17.51 | 17.92 | 18.11 | **18.52** |
| 65.03 | 15.081 | 17.56 | 19.04 | 18.76 | 19.20 | 19.41 | **19.88** |

It is worth pointing out that the same noisy images as used in the non redundant case have been processed by the redundant transforms. As shown in Table III, curvelets do not seem really appropriate in this multichannel context in spite of their promising results in the monochannel one. ProbShrink and BLS-GSM methods are very efficient in the redundant case and ProbShrink shows its superiority when using an inter-component neighborhood. The methods using a DTT outperform the existing ones in all the cases. We point out that the DTT has a limited redundancy of a factor 2 compared with the other considered redundant decompositions. It can be noticed that our method provides better results than Surevect. The observed gain increases as the initial SNR increases and we obtain significant improvements with respect to critically decimated transforms of about $0.25$ dB. It is also interesting to note that the observed gain in terms of SNR leads to quite visible differences. In Fig. 4, cropped versions of the first channel of the Tunis image are displayed, for a low value of the initial SNR ($4.66$ dB). We can notice that the proposed

[5]We employ the `Curvelab 2.0` toolbox which can be downloaded from `http://www.curvelet.org`.

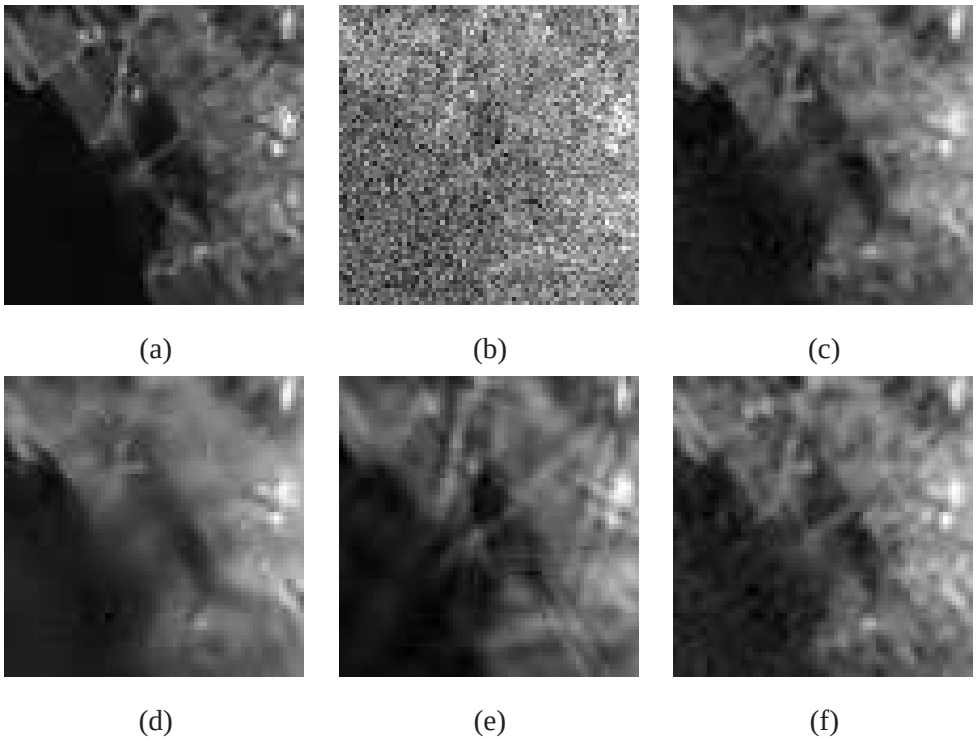


Fig. 4. Cropped versions of Tunis image (channel $b = 1$, initial SNR equal to $4.66$ dB) and (a) Original image, (b) Noisy image, (c) Denoised image using ProbShrink red. $(1 \times 1)$, (d) Denoised image using BLS-GSM red. + parent method, (e) Denoised image using curvelets and (f) Denoised image using our method (employing a DTT).

method (see Fig. 4-(f)) allows to better recover edges whereas the three others (see Fig. 4-(c,d,e)) result in more blurred images, where some of the original structures are missing. This is especially visible for the image denoised with the BLS-GSM estimator (see Fig. 4-(d)).

In the following, we focus on the method introduced in this paper and more specifically on the variations of its performance according to the parameter setup.

### *B. Pre-processing stage*

In order to improve the denoising performance in the multichannel context, additional linear procedures can be applied. Actually, different linear pre-processings of the components may be envisaged:

• The simplest idea consists in decorrelating the spectral components of the image to be estimated in order to process them separately. Knowing the noise covariance matrix $\mathbf{\Gamma^{(n)}}$, we can deduce the original data covariance matrix (assumed here to be spatially constant): $\mathbf{\Gamma^{(s)}} = \mathbf{\Gamma^{(r)}} - \mathbf{\Gamma^{(n)}}$, from the observed data covariance matrix $\mathbf{\Gamma^{(r)}}$. More precisely, by performing an eigendecomposition of $\mathbf{\Gamma^{(s)}}$, we seek for an orthogonal matrix $\mathbf{U^{(s)}}$ such that: $\mathbf{\Gamma^{(s)}} = \mathbf{U^{(s)}}\mathbf{D^{(s)}}(\mathbf{U}^{(s)})^\top$ where $\mathbf{D^{(s)}}$ is a diagonal matrix. Then,

the transformed multichannel image is $((\mathbf{U}^{(\mathbf{s})})^\top \mathbf{r}(\mathbf{k}))_{\mathbf{k}}$ and it is corrupted by a spatially white zero-mean Gaussian noise with covariance matrix $(\mathbf{U}^{(\mathbf{s})})^\top \mathbf{\Gamma}^{(\mathbf{n})} \mathbf{U}^{(\mathbf{s})}$. We then proceed to the nonlinear wavelet estimation of the decorrelated components as described in the previous sections.

• Instead of decorrelating the components, we may try to make them statistically independent or, at least, as independent as possible. A number of ICA (Independent Component Analysis) methods have been developed for this purpose in recent years [47]. In this case, a linear transform $\mathbf{V}^{(\mathbf{s})}$ (which is not necessarily orthogonal) is applied to the multichannel data.

The proposed estimator already includes an optimized linear combination of some of the components of the ROV. It is therefore expected to provide competitive results w.r.t. techniques involving some linear pre-processing. In order to make fair comparisons and evaluate the improvements resulting from the optimization of the linear part of the estimator, we provide simulations where the ROV is the same whatever the pre-processing is (we have chosen the same ROV as in the previous sections). In addition, when a decorrelation or an ICA is employed, the linear part of the estimator is chosen equal to the identity. We finally propose to compare these results with a simple linear MSE estimator based on a linear combination of coefficients from different channels.

Numerical results displayed in Table IV allow us to evaluate the proposed approach without optimization of the linear parameter vector, the same estimator combined with an ICA of the multichannel data (using the JADE algorithm [47]) or a pre-decorrelation stage and, finally our approach with an optimized linear part. From these results, it is clear that including some linear processing is useful for multichannel

TABLE IV

INFLUENCE OF DIFFERENT PRE-PROCESSINGS ON TUNIS IMAGE DENOISING ($\sigma^2 = 258.9$). SYMLETS OF LENGTH 16 ARE USED.

| Transform | Channel | $\text{SNR}_{\text{init}}$ | Without transf. | ICA | Decorrelation | MSE Lin. | Opt. lin. |
|---|---|---|---|---|---|---|---|
| DWT | $b=1$ | 8.664 | 13.84 | 14.66 | 15.15 | 15.18 | 15.75 |
| | $b=2$ | 9.653 | 14.39 | 15.03 | 15.36 | 15.28 | 15.89 |
| | $b=3$ | 8.926 | 15.15 | 13.85 | 15.11 | 15.26 | 15.84 |
| | Average | 9.081 | 14.46 | 14.51 | 15.21 | 15.24 | **15.83** |
| DTT | $b=1$ | 8.664 | 14.13 | 14.37 | 15.43 | 15.42 | 15.94 |
| | $b=2$ | 9.653 | 14.66 | 14.67 | 15.64 | 15.53 | 16.09 |
| | $b=3$ | 8.926 | 15.38 | 14.26 | 15.26 | 15.52 | 15.98 |
| | Average | 9.081 | 14.72 | 14.43 | 15.44 | 15.49 | **16.00** |

image denoising. The ICA only brings slight improvements, possibly due to the fact that the associated

transform is not orthogonal. Pre-decorrelating the data significantly increases the SNR, however the fully optimized version of our estimator remains the most effective method.

### *C. Influence of the neighborhoods*

The ROV can be defined as desired and plays a prominent role in the construction of our estimator. We study here the influence of different choices of the ROV:

1) ROV1 corresponds to an inter-component neighborhood. When a DWT is employed (see Fig. 5(a)), we have $\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [\left(r_{j,\mathbf{m}}^{(b')}(\mathbf{k})\right)_{b'}]^\top$, while for a DTT (see Fig. 6(a)), we use

$$\overline{\mathbf{r}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [\left(r_{j,\mathbf{m}}^{(b')}(\mathbf{k})\right)_{b'}, \left(r_{j,\mathbf{m}}^{\mathrm{H}(b')}(\mathbf{k})\right)_{b'}]^\top \quad \text{and} \quad \overline{\mathbf{u}}_{j,\mathbf{m}}^{(b)}(\mathbf{k}) = [\left(u_{j,\mathbf{m}}^{(b')}(\mathbf{k})\right)_{b'}]^\top \tag{53}$$

$$\overline{\mathbf{r}}_{j,\mathbf{m}}^{\mathrm{H}(b)}(\mathbf{k}) = [\left(r_{j,\mathbf{m}}^{\mathrm{H}(b')}(\mathbf{k})\right)_{b'}, \left(r_{j,\mathbf{m}}^{(b')}(\mathbf{k})\right)_{b'}]^\top \qquad \overline{\mathbf{u}}_{j,\mathbf{m}}^{\mathrm{H}(b)}(\mathbf{k}) = [\left(u_{j,\mathbf{m}}^{\mathrm{H}(b')}(\mathbf{k})\right)_{b'}]^\top. \tag{54}$$

2) ROV2 corresponds to a combination of a spatial $3 \times 3$ and an inter-component neighborhood as considered in the previous sections and shown in Figs. 5(b) and 6(b).

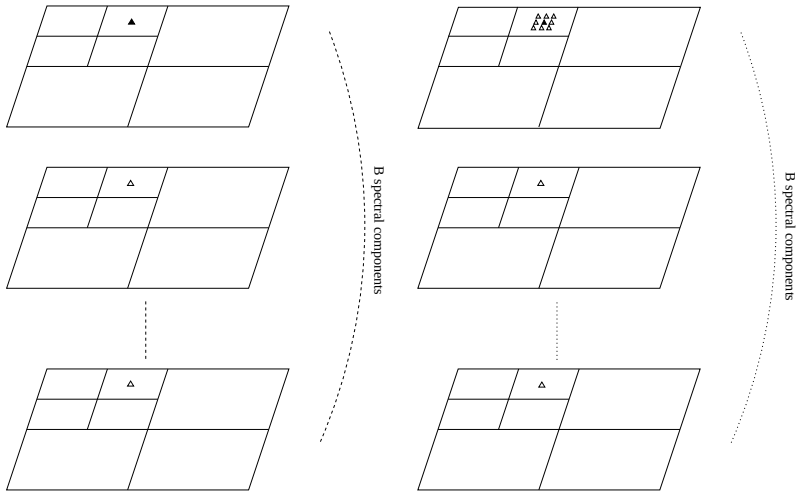


Fig. 5. Representation of the different considered ROVs in the DWT domain (the black triangle will be estimated taking into account the white ones); (a) ROV1 the purely inter-component one and (b) ROV2 combining inter-component and spatial dependencies.

The linear part of the estimator is defined as in Section V-A.

The corresponding results are given in Table V. In order to compare different possible wavelet choices, the results are provided both for symlets of length $16$ and a $4$-band filter bank given in [52] which is denoted by AC. These results can also be compared with the ones given in Section V-A where Daubechies

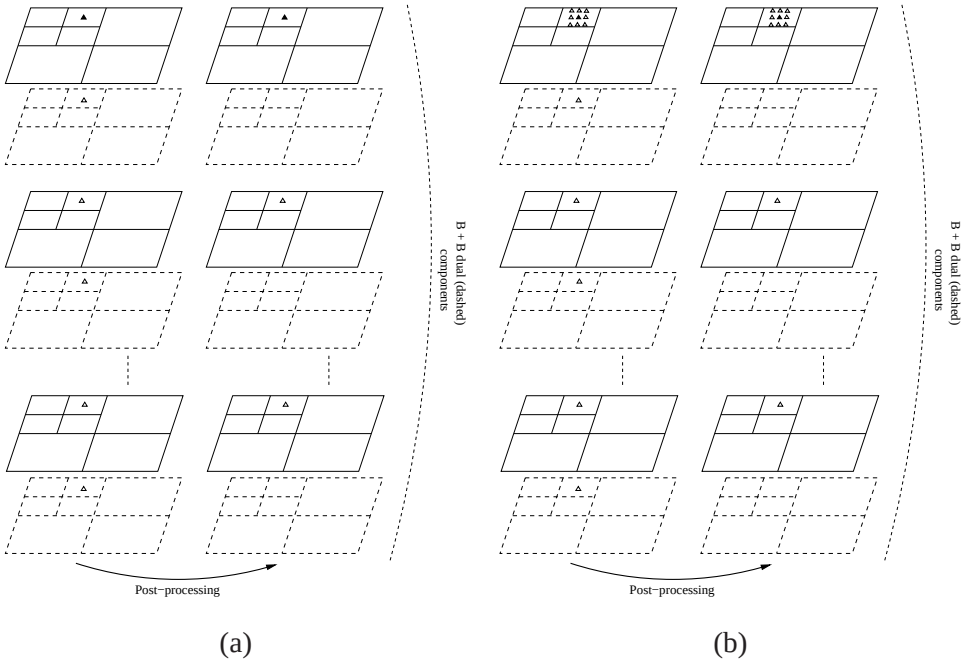


Fig. 6. Representation of the different considered ROVs in the DTT domain, with and without post-processing stage (the black triangle will be estimated taking into account the white ones); (a) ROV1 the purely inter-component one and (b) ROV2 combining inter-component and spatial dependencies.

TABLE V

INFLUENCE OF THE NEIGHBORHOOD IN TUNIS IMAGE DENOISING (AVERAGE VALUES COMPUTED OVER 3 CHANNELS ARE PROVIDED AND $\sigma^2 = 258.9$) USING SYMLETS (LENGTH 16) (TOP) AND AC FILTER BANK (LENGTH 16) (BOTTOM).

| Transform | $SNR_{init}$ | ROV1 | ROV2 | Transform | $SNR_{init}$ | ROV1 | ROV2 |
|---|---|---|---|---|---|---|---|
| DWT (symlets) | 9.081 | 15.42 | **15.83** | DWT (AC) | 9.081 | 15.49 | **15.76** |
| DTT (symlets) | 9.081 | 15.77 | **16.00** | DTT (AC) | 9.081 | 15.88 | **16.01** |

filters of length $8$ are used.

Concerning the neighborhood influence, we note that taking into account spatial dependence leads to a significant improvement w.r.t. inter-component dependence.

Concerning the wavelet choice, it appears that the 4-band AC wavelets yield slightly better results than the dyadic symlets choosing ROV1 and equivalent results choosing ROV2. Both outperform Daubechies wavelets wathever the ROV chosen.

### *D. Various noise levels*

In this section, we consider that the image channels are corrupted at different noise levels. Thus, the noise is spatially white, zero-mean, Gaussian with covariance matrix $\mathbf{\Gamma}_2^{(\mathbf{n})} = \text{Diag}(\sigma_1^2, \ldots, \sigma_B^2)$. The

TABLE VI

DENOISING RESULTS ON TUNIS IMAGE CONSIDERING $\mathbf{\Gamma}_2^{(n)}$ AND USING SYMLETS (LENGTH 16).

| Channel | $\sigma_b^2$ | $\text{SNR}_{\text{init}}$ | Surevect DWT | Proposed DWT | Surevect DTT | Proposed DTT |
|---|---|---|---|---|---|---|
| $b = 1$ | 25.89 | 18.66 | 20.58 | 21.16 | 20.85 | 21.24 |
| $b = 2$ | 258.9 | 9.653 | 18.53 | 18.61 | 18.75 | 18.82 |
| $b = 3$ | 491.9 | 6.138 | 14.20 | 14.55 | 14.51 | 14.69 |
| Average | | 11.49 | 17.76 | 18.11 | 18.04 | **18.25** |

resulting numerical results are displayed in Table VI with the corresponding noise levels, when our estimator is used with ROV2. Noticeable differences can be observed when comparing Surevect with our method both considering DWT and DTT transforms.

### *E. Increased number of channels*

A strong advantage of the proposed method is that, unlike many multicomponent approaches limited to RGB (3 components) images, it may process any kind of multichannel images, whatever the number of channels is. We consider here the 6 channel Trento image. We apply the Surevect estimator (both using DWT and DTT), the BLS-GSM estimator (taking into account the parent coefficient), and our estimator using ROV2. From the results provided in Table VII, we see that, while the number of channels is

TABLE VII

RESULTS OBTAINED APPLYING DIFFERENT ESTIMATORS ON TRENTO IMAGE ($\sigma^2 = 258.9$).

| Channel | $\text{SNR}_{\text{init}}$ | Surevect DWT | Proposed DWT | BLS-GSM red + parent | Surevect DTT | Proposed DTT |
|---|---|---|---|---|---|---|
| $b = 1$ | -2.907 | 8.661 | 8.945 | 8.311 | 8.984 | 9.251 |
| $b = 2$ | -6.878 | 8.375 | 8.427 | 6.536 | 8.805 | 8.839 |
| $b = 3$ | -3.836 | 8.288 | 8.443 | 7.341 | 8.647 | 8.761 |
| $b = 4$ | 2.428 | 9.525 | 9.799 | 9.836 | 9.901 | 10.01 |
| $b = 5$ | 4.765 | 11.18 | 11.52 | 11.38 | 11.61 | 11.77 |
| $b = 6$ | -1.560 | 9.545 | 9.685 | 8.167 | 9.945 | 10.00 |
| Average | -1.331 | 9.262 | 9.470 | 8.596 | 9.649 | **9.770** |

increased, our method still outperforms the other ones especially when a DTT is used. With the increase of the number of channels, the reduced redundancy of the DTT becomes another attractive feature of the

proposed approach.

## VI. Conclusion

In this paper, we have proposed a nonlinear Stein based estimator for wavelet denoising of multichannel data. Due to its flexible form, the considered estimator generalizes many existing methods, in particular block-based ones. Although the proposed approach has been applied to satellite images, it could also be used in any multivariate signal denoising problem. Besides, the estimator has been used in conjunction with real dual-tree wavelet transforms but complex ones or other frame decompositions could be envisaged as well. In the context of frame representations, it should however be noticed that the proposed estimator minimizes the risk over the frame coefficients and not on the reconstructed signal, which may be suboptimal [21], [53]. Another question that should be investigated in future work is the ability of the proposed framework to exploit inter-scale dependencies in addition to spatial and inter-component ones, as considered in [21] for the mono-channel case. In order to obtain an interscale denoising method, an appropriate ROV should be defined and the interscale statistics of the noise should be available.

## References

[1] M. C. Abrams and S. C. Cain, "Sampling, radiometry and image reconstruction for polar and geostationary meteorological remote sensing systems," in *Proc. of SPIE, Image Reconstruction from Incomplete Data II*, P. J. Bones, M. A. Fiddy, and R. P. Millane, Eds., vol. 4792, Dec. 2002, pp. 207–215.

[2] B. R. Corner, R. M. Narajanan, and S. E. Reichenbach, "Noise estimation in remote sensing imagery using data masking," *Int. J. Remote Sensing*, vol. 24, no. 4, pp. 689–702, 2003.

[3] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: CBMS-NSF, SIAM Lecture Series, 1992.

[4] S. Mallat, *A wavelet tour of signal processing*. San Diego, CA, USA: Academic Press, 1998.

[5] D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *Appl. and Comp. Harm. Analysis*, vol. 1, no. 1, pp. 100–115, Dec. 1993.

[6] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Aug. 1994.

[7] ——, "Adapting to unknown smoothness via wavelet shrinkage," *J. American Statist. Ass.*, vol. 90, pp. 1200–1224, Dec. 1995.

[8] G. P. Nason, "Wavelet shrinkage using cross-validation," *J. Royal Statistical Society B*, vol. 58, pp. 463–479, 1996.

[9] Y. Wang, "Function estimation via wavelet shrinkage for long-memory data," *The Annals of Statistics*, vol. 24, pp. 466–484, 1996.

[10] N. Weyrich and G. T. Warhola, "Wavelet shrinkage and generalized cross validation for image denoising," *IEEE Trans. on Image Proc.*, vol. 7, no. 1, pp. 82–90, Jan. 1998.

[11] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. on Inform. Theory*, vol. 45, no. 7, pp. 2225–2238, Nov. 1999.

[12] K. Tang, J. Astola, and Y. Neuvo, "Nonlinear multivariate image filtering techniques," *IEEE Trans. on Image Proc.*, vol. 4, no. 6, pp. 788–798, Jun. 1995.

[13] A. K. Fletcher, V. Goyal, and K. Ramchandran, "On multivariate estimation by thresholding," in *Proc. Int. Conf. on Image Processing*, vol. 1, Barcelona, Spain, Sep. 14-17 2003, pp. 61–64.

[14] A. K. Fletcher and K. Ramchandran, "Estimation error bounds for denoising by sparse approximation," in *Proc. Int. Conf. on Image Processing*, vol. 1, Barcelona, Spain, Sep. 4-17 2003, pp. 113–116.

[15] P. Scheunders, "Wavelet thresholding of multivalued images," *IEEE Trans. on Image Proc.*, vol. 13, no. 4, pp. 475–483, Apr. 2004.

[16] A. Benazza-Benyahia and J.-C. Pesquet, "Wavelet-based multispectral image denoising with Bernoulli-Gaussian models," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, Grado, Italy, Jun. 8-11 2003.

[17] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for bayesian wavelet based image denoising," *IEEE Trans. on Image Proc.*, vol. 11, no. 5, pp. 545–557, May 2002.

[18] P. Scheunders and J. Driesen, "Least-squares interband denoising of color and multispectral images," in *Proc. Int. Conf. on Image Processing*, vol. 2, Singapore, Oct. 24-27 2004, pp. 985–988.

[19] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.

[20] J.-C. Pesquet and D. Leporini, "A new wavelet estimator for image denoising," in *IEE Sixth International Conference on Image Processing and its Applications*, vol. 1, Dublin, Ireland, Jul. 14-17 1997, pp. 249–253.

[21] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Inter-scale orthonormal wavelet thresholding," *IEEE Trans. on Image Proc.*, vol. 16, no. 3, pp. 593–606, Mar. 2007.

[22] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. on Image Proc.*, vol. 14, no. 11, pp. 1814–1830, Nov. 2005.

[23] T. T. Cai and B. W. Silverman, "Incorporating information on neighboring coefficients into wavelet estimation," *Sankhya*, vol. 63, Series B., pp. 127–148, 2001.

[24] F. Abramovich, P. Besbeas, and T. Sapatinas, "Empirical Bayes approach to block wavelet function estimation," *Computational Statistics & Data Analysis*, vol. 39, no. 4, pp. 435–451, Jun. 2002.

[25] C. Chaux, A. Benazza-Benyahia, and J.-C. Pesquet, "A block-thresholding method for multispectral image denoising," in *Proc. SPIE, Ann. Meeting, Wavelets X*, vol. 5914, San Diego, CA, USA, Aug. 2005, pp. 1–H1,1–H13.

[26] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov. 2005.

[27] C. Chaux, L. Duval, and J.-C. Pesquet, "Image analysis using a dual-tree $M$-band wavelet transform," *IEEE Trans. on Image Proc.*, vol. 15, no. 8, pp. 2397–2412, Aug. 2006.

[28] A. Antoniadis, D. Leporini, and J.-C. Pesquet, "Wavelet thresholding for some classes of non-Gaussian noise," *Statistica Neerlandica*, vol. 56, no. 4, pp. 434–453, Dec. 2002.

[29] H. Maître, Ed., *Le traitement d'images*. Paris, France: Hermès Science Publications, 2003.

[30] P. Steffen, P. N. Heller, R. A. Gopinath, and C. S. Burrus, "Theory of regular $M$-band wavelet bases," *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, pp. 3497–3511, Dec. 1993.

[31] P. Hall, S. Penev, G. Kerkyacharian, and D. Picard, "Numerical performance of block thresholded wavelet estimators," *Statist. Comput.*, vol. 7, no. 2, pp. 115–124, 1997.

[32] P. Hall, G. Kerkyacharian, and D. Picard, "Block threshold rules for curve estimation using kernel and wavelet methods," *Ann. Statist.*, vol. 26, no. 3, pp. 922–942, 1998.

[33] ——, "On the minimax optimality of block thresholded wavelet estimators," *Statist. Sinica*, vol. 9, pp. 33–49, 1999.

[34] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. on Image Proc.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[35] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Proc.*, vol. 41, pp. 3445–3462, Dec. 1993.

[36] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. on Image Proc.*, vol. 10, no. 7, pp. 1056–1068, Jul. 2001.

[37] L. Şendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. on Signal Proc.*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.

[38] A. Benazza-Benyahia and J.-C. Pesquet, "An interscale multivariate MAP estimation of multispectral images," in *Proc. Eur. Sig. and Image Proc. Conference*, Wien, Austria, Sep. 6-10 2004.

[39] A. Pižurica and W. Philips, "Estimating probability of presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. on Image Proc.*, vol. 15, no. 3, pp. 654–665, Mar. 2006.

[40] R. Coifman and D. Donoho, *Translation-invariant de-noising*, ser. Lecture Notes in Statistics. New York, NY, USA: Springer Verlag, 1995, vol. 103, pp. 125–150.

[41] G. P. Nason and B. W. Silverman, *The stationary wavelet transform and some statistical applications*, ser. Lecture Notes in Statistics. New York, NY, USA: Springer Verlag, 1995, vol. 103, pp. 281–300.

[42] J.-C. Pesquet, H. Krim, and H. Carfatan, "Time-invariant orthogonal wavelet representations," *IEEE Trans. on Signal Proc.*, vol. 44, no. 8, pp. 1964–1970, Aug. 1996.

[43] H.-Y. Gao, "Wavelet shrinkage denoising using the non-negative garrote," *J. Comput. Graph. Statist.*, vol. 7, no. 4, pp. 469–488, 1998.

[44] L. Breiman, "Better subset regression using the nonnegative garrote," *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.

[45] L. Şendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *Signal Processing Letters*, vol. 9, no. 12, pp. 438–441, Dec. 2002.

[46] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. on Mult. Model. Simul.*, vol. 4, pp. 1168–1200, Nov. 2005.

[47] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[48] X. Guyon, *Champs aléatoires sur un réseau, modélisations, statistiques et applications*. Paris: Masson, 1993.

[49] C. Chaux, L. Duval, and J.-C. Pesquet, "Étude du bruit dans une analyse $M$-bandes en arbre dual," in *Proc. GRETSI*, Louvain, Belgique, Sep. 2005, pp. 229–232.

[50] C. Chaux, J.-C. Pesquet, and L. Duval, "Noise covariance properties in dual-tree wavelet decompositions," *IEEE Trans. on Inform. Theory*, vol. 53, no. 12, pp. 4680–4700, Dec. 2007.

[51] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM J. on Mult. Model. Simul.*, vol. 5, no. 3, pp. 861–899, Mar. 2006.

[52] O. Alkin and H. Caglar, "Design of efficient $M$-band coders with linear-phase and perfect-reconstruction properties," *IEEE Trans. on Signal Proc.*, vol. 43, no. 7, pp. 1579–1590, Jul. 1995.

[53] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant bases," in *Proc. Int. Conf. on Image Processing*, vol. III, San Antonio, TX, USA, Sep. 16-19 2007, pp. 113–116.