Image Denoising with 2-D Scale-Mixing Complex Wavelet Transforms

Norbert Remenyi, Orietta Nicolis, Guy Nason, and Brani Vidakovic

***Abstract*—This article introduces an image denoising procedure based on a two-dimensional scale-mixing complex-valued wavelet transform. Both the minimal (unitary) and redundant (maximum overlap) versions of the transform are used. The covariance structure of white noise in wavelet domain is established. Estimation is performed via empirical Bayesian techniques including versions that preserve the phase of the complex-valued wavelet coefficients and those that do not. The new procedure exhibits excellent quantitative and visual performance, which is demonstrated by simulation on standard test images.**

***Index Terms*—Image denoising, complex-valued wavelets, scale-mixing wavelet transform, empirical Bayes estimation, bivariate normal distribution, posterior mean.**

I. Introduction

WAVELET shrinkage methods that use complex-valued wavelets show excellent performance compared to real-valued wavelets. Typically, the observed wavelet coefficients are modeled by a statistical model where the true signal/image part corresponds to a location parameter. Under such an approach wavelet shrinkage becomes equivalent to a location estimation in the wavelet domain. Bayesian models are popular for wavelet coefficients due to the good behaviour of Bayes shrinkage rules, adaptivity via data-driven hyperparameter selection and signal sparsity being a strong driver for prior determination. A recent overview of Bayesian wavelet shrinkage can be found in [1], see also [2], [3] for general overviews of wavelets in statistics. Several papers considering Bayesian wavelet shrinkage with complex wavelets are available. For example, [4], [5], [6] and [7] focus on image denoising, in which the phase of the observed wavelet coefficients is preserved, but the modulus of the coefficients is thresholded or shrunk by a Bayes rule. The procedure introduced in [8] modifies both the phase and modulus of wavelet coefficients by a bivariate shrinkage rule.

We propose a Bayesian model in the domain of a complex scale-mixing discrete unitary compactly-supported wavelets that generalizes the method in [8] to 2-D signals. In estimating the signal the model is allowed to modify both phase and modulus. The choice of wavelet transform is motivated by the symmetry/antisymmetry of decomposing wavelets, which is possible only in the complex domain under condition of orthogonality (unitarity) and compact support. Symmetry is considered a desirable property of wavelets, especially when dealing with images, see, e.g., [9].

The 2-D discrete scale mixing wavelet transform can be computed by left- and right-multiplying the image by a wavelet matrix W and its Hermitian transpose $W^\dagger$, respectively. Mallat's algorithm is not used to perform this task, but it is implicit in the construction of matrix W.

We derive the covariance structure of white noise in the wavelet domain, induced by the 2-D discrete scale mixing complex wavelet transform, and show how the method of [8] can be extended to this transform. Since in some applications it might be important to preserve the estimated coefficients' phase we demonstrate how to use the proposed empirical Bayes method in a phase-preserving mode. We also show how to use the scale-mixing complex transform in a non-decimated manner, and defined the complex maximal overlap scale mixing 2-D complex wavelet transform (cMOSM-DWT). The MATLAB© implementation of the maximum overlap wavelet transform from [10] was modified to allow complex wavelet filters.

The rest of the paper is organized as follows. Section II defines the complex scale-mixing 2-D complex wavelet transform, and provides information on their basic properties and implementation. Section III introduces the Bayesian regression model, describes the noise covariance structure, and discusses the shrinkage rule. Section III also discusses a phase preserving version of the shrinkage as well how the model adapts to the maximal overlap transform. Section IV compares the proposed rule and its redundant modification with several recent and comparable proposals from the literature. Conclusions are presented in Section V followed by an Appendix that contains all technical derivations.

II. The 2-D Scale-Mixing Complex Wavelet Transform

The discrete complex wavelet transform (DCWT) can be considered a complex-valued extension to the standard discrete wavelet transform (DWT). The DCWT uses complex-valued filtering (analytic filter) for decomposing the real/complex signals into real and imaginary parts in the transform domain. Complex wavelet coefficients can be expressed via convolutions

$$c_{j-1,l} = \sum_k \overline{h}_{k-2l} c_{j,k}, \text{ and } d_{j-1,l} = \sum_k \overline{g}_{k-2l} c_{j,k}, \qquad (1)$$

N. Remenyi is working at Research Group, Sabre Holdings, Southlake, TX, 76092 USA, e-mail: nremenyi@gatech.edu

O. Nicolis is with Department of Statistics, University of Valparaiso, Chile, e-mail:orietta.nicolis@uv.cl

G.P. Nason is with the Statistics Group, School of Mathematics at University of Bristol, Bristol, BS8 1TW, UK, email: g.p.nason@bristol.ac.uk

B. Vidakovic is with the H. Milton Stewart School of Industrial and Systems Engineering at Georgia Institute of Technology, 755 Ferst Drive, Atlanta, GA 30332-0205 USA, e-mail: brani@gatech.edu

where $\overline{h}$ and $\overline{g}$ are complex conjugates of filters h and g. Conversely, the reconstruction is calculated as

$$c_{j,k} = \sum_l c_{j-1,l} h_{k-2l} + \sum_l d_{j-1,l} g_{k-2l}. \quad (2)$$

There are several versions of the 2-D wavelet transforms which correspond to different tessellations (or tilings), see [11]. For example, the complex wavelet atoms can be defined as,

$$\phi_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}) = 2^{(j_1+j_2)/2}\ \phi(2^{j_1}x - k_1, 2^{j_2}y - k_2) \quad (3)$$
$$\psi^u_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}) = 2^{(j_1+j_2)/2}\ \psi^u(2^{j_1}x - k_1, 2^{j_2}y - k_2), \quad (4)$$

where u is one of traditional wavelet directions h, v, and d, standing for horizontal $[\phi(x_1)\psi(x_2)]$, vertical $[\psi(x_1)\phi(x_2)]$ and diagonal $[\psi(x_1)\psi(x_2)]$, and $\boldsymbol{j} = (j_1, j_2) \in \mathbb{Z}^2$ and $\boldsymbol{k} = (k_1, k_2) \in \mathbb{Z}^2$.

The real-valued scale mixing transform was used extensively, some references are [11]–[15]. It is also known as "hyperbolic" [12], [15] and "rectangular" [14].

In the multiresolution context, any function $f \in \mathcal{L}_2(\mathbb{R}^2)$ can be represented as

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{\boldsymbol{k}} c_{(J_0,J_0),\boldsymbol{k}}\ \phi_{(J_0,J_0),\boldsymbol{k}}(\boldsymbol{x}) \\ &+ \sum_{j\ge J_0}\sum_{\boldsymbol{k}} d^h_{(J_0,j),\boldsymbol{k}}\ \psi^h_{(J_0,j),\boldsymbol{k}}(\boldsymbol{x}) \\ &+ \sum_{j\ge J_0}\sum_{\boldsymbol{k}} d^v_{(j,J_0),\boldsymbol{k}}\ \psi^v_{(j,J_0),\boldsymbol{k}}(\boldsymbol{x}) \\ &+ \sum_{j_1,j_2\ge J_0}\sum_{\boldsymbol{k}} d^d_{(j_1,j_2),\boldsymbol{k}}\ \psi^d_{(j_1,j_2),\boldsymbol{k}}(\boldsymbol{x}), \end{aligned} \quad (5)$$

which defines a *scale-mixing wavelet transform*. Notice that (j_1, j_2) in (3) and (4) can be indexed as $(j_1, j_1 + s)$ as well, where $s \in \mathbb{Z}$. Precise limits on the indices are given in (13) and (14).

The scale-mixing coefficients are defined as,

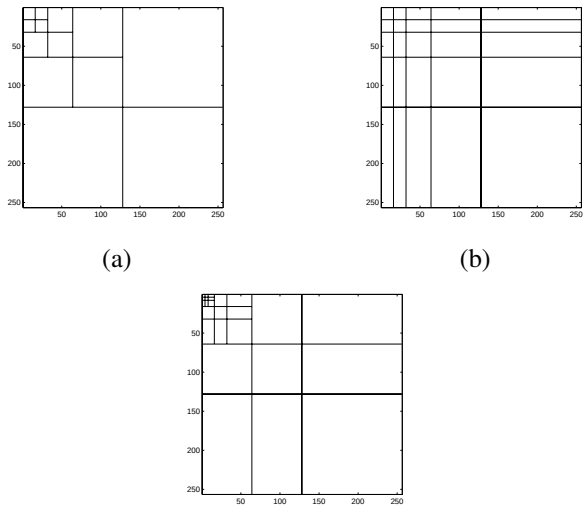
$$c_{(J_0,J_0),\boldsymbol{k}} = 2^{J_0} \cdot \int f(\boldsymbol{x})\ \overline{\phi}(2^{J_0}x_1 - k_1, 2^{J_0}x_2 - k_2)\ d\boldsymbol{x}, \quad (6)$$

$$d^h_{(J_0,j),\boldsymbol{k}} = 2^{(J_0+j)/2} \cdot \int f(\boldsymbol{x})\ \overline{\psi^h}(2^{J_0}x_1 - k_1, 2^{j}x_2 - k_2)\ d\boldsymbol{x}, \quad (7)$$

$$d^v_{(j,J_0),\boldsymbol{k}} = 2^{(j+J_0)/2} \cdot \int f(\boldsymbol{x})\ \overline{\psi^v}(2^{j}x_1 - k_1, 2^{J_0}x_2 - k_2)\ d\boldsymbol{x}, \quad (8)$$

$$d^d_{(j_1,j_2),\boldsymbol{k}} = 2^{(j_1+j_2)/2} \cdot \int f(\boldsymbol{x})\ \overline{\psi^d}(2^{j_1}x_1 - k_1, 2^{j_2}x_2 - k_2)\ d\boldsymbol{x}, \quad (9)$$

where $\overline{\phi}$ and $\overline{\psi^u}$ are complex conjugates of ϕ and ψ^u. The complex scale-mixing detail coefficients are linked to the original image (2-D signal) through a matrix equation, as in the traditional 1- and 2-D cases. Let n be a power of two. Define W to be an $n \times n$ wavelet matrix composed by the complex scaling and wavelet filter coefficients h_k and g_k as on p.116 of [2], where the wavelet filter is given by $g_k = (-1)^k \overline{h}_{1+N-k}$ and N is a shift parameter which affects the location of the wavelet. Suppose that an $n \times n$ image (matrix) F is to be transformed into the wavelet domain.



(a) (b) (c)

Fig. 1. Tessellations for some 2-D wavelet transforms. (a) Traditional 2-D transform of depth 4; (b) Scale-mixing wavelet transform of depth 4; (c) General iterated scale-mixing transform of depth 2 per iteration.

The real-valued scale-mixing transform is given by $G = WFW'$ which was defined and utilized in [11]. Here we generalize this transform to the complex-valued case as

$$G = WFW^\dagger, \quad (10)$$

where W is a unitary matrix and $W^\dagger$ denotes its Hermitian transpose. Matrix G will be called the complex scale-mixing wavelet transform of matrix F. It represents a 2D implementation of (5) for signal $f(\mathbf{x})$ sampled in a form of matrix F.

The tessellation induced by transform in (10) is shown in Figure 1(b). A more general transform can be obtained as an iterative application of the transform in (10) with depth k, applied only on the "smooth part" of the previous iterative step. Figure 1(c) shows tessellation for an iterated transformation of depth two with three iterations. For more information, please refer to [11].

For transform (10) the balance of the total energy $E = \text{trace}(FF^\dagger)$ in the image F is preserved, since the unitary nature of W implies $E = \text{trace}(FF^\dagger) = \text{trace}(GG^\dagger)$, for $G = WFW^\dagger$. Our transform is versatile in that rectangular images (with dyadic dimensions) can be processed by properly choosing sizes of W and $W^\dagger$. Also, the complex wavelet generating the left-hand side matrix may be different from the wavelet generating the right-hand side matrix, as in $G = W_1 F W_2^\dagger$. The inverse transforms of these two modifications are obvious.

The scale-mixing 2-D wavelet transform is typically more compressive compared to the traditional 2-D wavelet transform, which is a desirable property when dimension reduction applications such as denoising or compression. Informally, in a depth k scale-mixing transform of an image $\left(1 - 2^{-k}\right)^2 \times 100\%$ of coefficients correspond to the atoms which are products of wavelet functions (possibly with different scales),

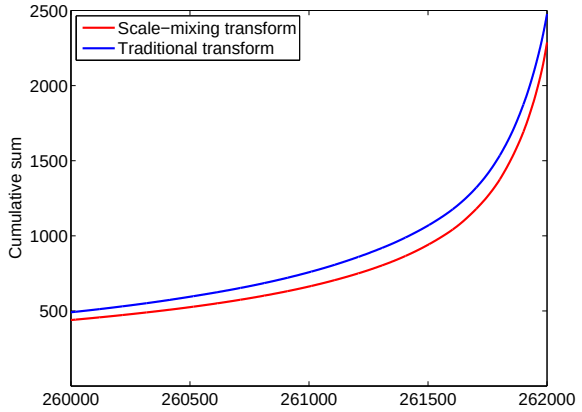


Fig. 2. Section of the Lorenz curve for scale-mixing and traditional 2-D wavelet transforms.

while for the traditional transform of the same depth k this proportion is $\left(1-2^{-2k}\right)/3 \times 100\%$. The rest of the coefficients correspond to the atoms containing at least one scaling function in the product. The first proportion is always larger (or equal when $k=1$) than the second. Informally, the squared wavelet coefficients that correspond to "differencing" atoms tend to be smaller compared to those of "averaging" and the Lorenz curve for squared wavelet coefficients in a scale-mixing transform falls typically below the Lorenz curve for the traditional transform of the same depth.

To illustrate this, we calculated the Lorenz curve for a noiseless test image "Lenna" of size 512×512 using the 2-D scale-mixing real wavelet transform and compared it to the 2-D periodic orthogonal wavelet transform available in WaveLab (http://www-stat.stanford.edu/~wavelab/). In particular, Daubechies 4 tap filter was used.

Figure 2 shows a section of the Lorenz curve plotting the cumulative sum of the ordered squared wavelet coefficients. We can see that the Lorentz curve of the scale-mixing wavelet transform is below the curve for the traditional wavelet transform, which translates to more efficient energy compression. This better energy compression results in superior denoising performance, which can be simply demonstrated by hard thresholding a noisy "Lenna" in the wavelet domain. In case of i.i.d normal noise with $\sigma = 50$, the resulting average mean squared error of a denoised image is approximately 10% better for the scale-mixing transform. This agrees with the results of [14], who shows that the scale-mixing wavelet transform allows to improve the compression rate (measured in the number of non-zero coefficient) compared to the traditional wavelet transform by factor of 2. Papers [13], [15] also refer to this transform's superiority due to the possibly anisotropic nature of the underlying images.

III. Complex Empirical Bayes Shrinkage

A. Statistical Model

As we indicated in the Introduction, wavelet shrinkage methods are formalized by a statistical model in the wavelet domain. Shrinkage is achieved by optimal location estimation under the proposed model.

For the application of 2-D image denoising, we start with the model

$$\boldsymbol{Y} = \boldsymbol{F} + \boldsymbol{e}, \tag{11}$$

where $\boldsymbol{Y}$ is an $n \times n$ matrix of real-valued noisy measurements of an unknown image $\boldsymbol{F}$ observed with noise matrix $\boldsymbol{e}$, where $\boldsymbol{e} \sim MN_{n,n}(\boldsymbol{0}, \sigma^2 I, I)$, MN denoting the matrix normal distribution. This noise model assumes that the components of $\boldsymbol{e}$ are independently distributed with $\mathrm{Cov}(\mathrm{Vec}(\boldsymbol{e})) = \sigma^2 I \otimes I = \sigma^2 I_{n^2 \times n^2}$ for noise variance parameter σ^2. The parameter σ^2 is assumed known, and in applications a robust plug-in estimator is used, as it is standard practice.

After applying the scale-mixing wavelet transform, the model in (11) becomes

$$\boldsymbol{D} = \boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{12}$$

where $\boldsymbol{D} = W\boldsymbol{Y}W^\dagger$, $\boldsymbol{\theta} = W\boldsymbol{F}W^\dagger$ and $\boldsymbol{\epsilon} = W\boldsymbol{e}W^\dagger$, and where W is the $n \times n$ unitary matrix from (10). The elements of matrix $\boldsymbol{D}$ are complex numbers $D_{\boldsymbol{j},\boldsymbol{k}}$, where

$$\boldsymbol{j} = (j_1, j_2),\ j_1, j_2 \ = \ J_0, \ldots, \log_2(n) - 1, \tag{13}$$

denotes the tessellation subband of the scale-mixing wavelet transform, and

$$\boldsymbol{k} = (k_1, k_2),\ (k_1, k_2) \in K_{j_1, j_2}, \tag{14}$$

where $K_{j_1,j_2} = \{0, \ldots, 2^{j_1} - 1\} \times \{0, \ldots, 2^{j_2} - 1\}$, is the location of complex wavelet coefficients within the subband. Due to the whitening property of wavelet transforms [16] many existing methods assume independence of the wavelet coefficients and model the wavelet coefficients one-by-one.

Here we consider term-by-term shrinkage, which is sometimes referred to as diagonal shrinkage. The model for an individual wavelet coefficient is

$$D_{\boldsymbol{j},\boldsymbol{k}} = \theta_{\boldsymbol{j},\boldsymbol{k}} + \epsilon_{\boldsymbol{j},\boldsymbol{k}},$$

where $\theta_{\boldsymbol{j},\boldsymbol{k}}$ needs to be estimated. We approach this problem from a Bayesian standpoint.

B. Wavelet Coefficients' Error Structure

Several papers considering Bayesian wavelet shrinkage with complex wavelets are available. Among those [4], [5], and [17] focus on image denoising by preserving the phase of the observed wavelet coefficients but thresholding or shrinking the modulus by the Bayes rule. We will build on the complex empirical Bayes (CEB) procedure proposed by [8], which modifies both the phase and modulus of wavelet coefficients by a bivariate shrinkage rule. The method in [8] was implemented in the context of nonparametric regression (1-D signal denoising), here we make extension to the 2-D case.

In model (11) an i.i.d. normal noise model $\boldsymbol{e} \sim MN_{n,n}(\boldsymbol{0}, \sigma^2 I, I)$ is assumed, however, in the scale-mixing complex wavelet domain the real and imaginary parts of the transformed noise $\boldsymbol{\epsilon} = W\boldsymbol{e}W^\dagger$ become correlated. We show that

$$\mathrm{Cov}\left\{\mathrm{Re}(\mathrm{vec}(\boldsymbol{\epsilon})), \mathrm{Re}(\mathrm{vec}(\boldsymbol{\epsilon}))\right\} = \frac{\sigma^2}{2}\left[I_{n^2 \times n^2} + \mathrm{Re}\left\{WW' \otimes \overline{WW'}\right\}\right],$$

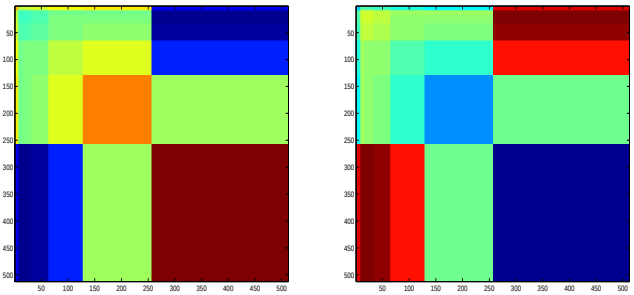
$$\begin{aligned}&\mathrm{Cov}\left\{\mathrm{Re}(\mathrm{vec}(\boldsymbol{\epsilon})),\mathrm{Im}(\mathrm{vec}(\boldsymbol{\epsilon}))\right\} = \\ &\qquad -\frac{\sigma^2}{2}\left[\mathrm{Im}\left\{WW'\otimes\overline{WW'}\right\}\right], \\ &\mathrm{Cov}\left\{\mathrm{Im}(\mathrm{vec}(\boldsymbol{\epsilon})),\mathrm{Im}(\mathrm{vec}(\boldsymbol{\epsilon}))\right\} = \\ &\qquad \frac{\sigma^2}{2}\left[I_{n^2\times n^2}-\mathrm{Re}\left\{WW'\otimes\overline{WW'}\right\}\right]. \quad (15)\end{aligned}$$

In the above expressions W' and $\overline{W}$ denote the transpose and the conjugate of W, respectively. Note that $W^\dagger = \overline{W'}$. For the detailed derivations, see Appendix. Note that calculating the Kronecker products in the above expressions can be prohibitive for large matrices. However, for term-by-term shrinkage only the diagonal elements of the above covariance matrices are needed. We consider elements $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Re}(D_{\boldsymbol{m},\boldsymbol{n}})\right\}$, $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{m},\boldsymbol{n}})\right\}$, $\mathrm{Cov}\left\{\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{m},\boldsymbol{n}})\right\}$, only for $\boldsymbol{j}=\boldsymbol{m}$ and $\boldsymbol{k}=\boldsymbol{n}$ and neglect the other elements of the covariance matrices because the relationships among neighboring coefficients are not accounted for. Computation of the above diagonal elements is feasible by using the identity

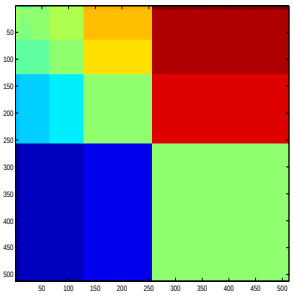
$$\mathrm{diag}(A\otimes B) = \mathrm{vec}(\mathrm{diag}(B)\cdot\mathrm{diag}(A)'),$$

where A and B are square matrices and $\mathrm{diag}(\cdot)$ are the diagonal elements of a square matrix.

Figure 3 shows the covariance structure of the complex wavelet coefficients for $\sigma = 10$, $J_0 = 3$ and $n = 512$. The matrices shown are $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$, $\mathrm{Cov}\left\{\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$, and $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$. We can see that $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$ and $\mathrm{Cov}\left\{\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$ have a symmetric structure, while $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$ is antisymmetric.



(a) $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$ (b) $\mathrm{Cov}\left\{\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$



(c) $\mathrm{Cov}\left\{\mathrm{Re}(D_{\boldsymbol{j},\boldsymbol{k}}),\mathrm{Im}(D_{\boldsymbol{j},\boldsymbol{k}})\right\}$

Fig. 3. Covariance structure of the noise for $\sigma = 10$, $J_0 = 3$ and $n = 512$.

C. Wavelet Coefficient Prior and Posterior

By representing the complex-valued wavelet coefficients as bivariate real-valued random variables, the likelihood model for the observed wavelet coefficients becomes $D_{\boldsymbol{j},\boldsymbol{k}}|\theta_{\boldsymbol{j},\boldsymbol{k}} \sim N_2(\theta_{\boldsymbol{j},\boldsymbol{k}},\Sigma_{\boldsymbol{j}})$, where $\Sigma_{\boldsymbol{j}}$ is determined by (15) for each subband or dyadic level $\boldsymbol{j} = (j_1, j_2)$ in the 2-D wavelet decomposition. Following [8], a natural choice for the prior on $\theta_{\boldsymbol{j},\boldsymbol{k}}$ is a bivariate mixture of the form

$$\begin{aligned}\theta_{\boldsymbol{j},\boldsymbol{k}} &\sim (1-\gamma_{\boldsymbol{j}})\delta_0 + \gamma_{\boldsymbol{j}} N_2(\mathbf{0}, V_{\boldsymbol{j}}),\\ \gamma_{\boldsymbol{j}} &\sim Ber(p_{\boldsymbol{j}}),\end{aligned}$$

where δ_0 is the point mass at $(0,0)^T$ commonly considered in the wavelet shrinkage literature. This prior is the bivariate extension of the prior considered by [18], related to [19] and [20]. Conjugacy of the normal distribution results in the posterior distribution

$$\theta_{\boldsymbol{j},\boldsymbol{k}}|D_{\boldsymbol{j},\boldsymbol{k}} \sim (1-\tilde{p}_{\boldsymbol{j},\boldsymbol{k}})\delta_0 + \tilde{p}_{\boldsymbol{j},\boldsymbol{k}} N_2(\mu_{\boldsymbol{j},\boldsymbol{k}}, \tilde{V}_{\boldsymbol{j}}),$$

where

$$\begin{aligned}&\tilde{p}_{\boldsymbol{j},\boldsymbol{k}} = \\ &\qquad \frac{p_{\boldsymbol{j}} f(D_{\boldsymbol{j},\boldsymbol{k}}|\gamma_{\boldsymbol{j}}=1)}{p_{\boldsymbol{j}} f(D_{\boldsymbol{j},\boldsymbol{k}}|\gamma_{\boldsymbol{j}}=1) + (1-p_{\boldsymbol{j}}) f(D_{\boldsymbol{j},\boldsymbol{k}}|\gamma_{\boldsymbol{j}}=0)},\\ &f(D_{\boldsymbol{j},\boldsymbol{k}}|\gamma_{\boldsymbol{j}}=1) = \\ &\qquad (2\pi|V_{\boldsymbol{j}}+\Sigma_{\boldsymbol{j}}|)^{-1/2}\exp\left\{-D'_{\boldsymbol{j},\boldsymbol{k}}(V_{\boldsymbol{j}}+\Sigma_{\boldsymbol{j}})^{-1}D_{\boldsymbol{j},\boldsymbol{k}}/2\right\},\\ &f(D_{\boldsymbol{j},\boldsymbol{k}}|\gamma_{\boldsymbol{j}}=0) = \\ &\qquad (\pi|\Sigma_{\boldsymbol{j}}|)^{-1/2}\exp\left\{-D'_{\boldsymbol{j},\boldsymbol{k}}\Sigma_{\boldsymbol{j}}^{-1}D_{\boldsymbol{j},\boldsymbol{k}}/2\right\},\\ &\tilde{V}_{\boldsymbol{j}} = \left(V_{\boldsymbol{j}}^{-1}+\Sigma_{\boldsymbol{j}}^{-1}\right)^{-1} \text{ and } \mu_{\boldsymbol{j},\boldsymbol{k}} = \tilde{V}_{\boldsymbol{j}}\Sigma_{\boldsymbol{j}}^{-1}D_{\boldsymbol{j},\boldsymbol{k}}.\end{aligned}$$

The posterior mean of $\theta_{\boldsymbol{j},\boldsymbol{k}}$ becomes

$$\mathbb{E}(\theta_{\boldsymbol{j},\boldsymbol{k}}) = \tilde{p}_{\boldsymbol{j},\boldsymbol{k}}\mu_{\boldsymbol{j},\boldsymbol{k}}. \quad (16)$$

Similarly to [8], we estimate the prior parameters $p_{\boldsymbol{j}}$ and $V_{\boldsymbol{j}}$ by the data-driven empirical Bayes approach maximizing the logarithm of the marginal likelihood. The marginal log-likelihood is

$$\sum_{\boldsymbol{j}}\sum_{\boldsymbol{k}}\log\left\{p_{\boldsymbol{j}} f(D_{\boldsymbol{j},\boldsymbol{k}}|\gamma_{\boldsymbol{j}}=1) + (1-p_{\boldsymbol{j}}) f(D_{\boldsymbol{j},\boldsymbol{k}}|\gamma_{\boldsymbol{j}}=0)\right\},$$

which can be maximized separately for each dyadic rectangle $\boldsymbol{j}$ to select the hyperparameters $p_{\boldsymbol{j}}$ and $V_{\boldsymbol{j}}$.

After obtaining the hyperparameters, we estimate each $\theta_{\boldsymbol{j},\boldsymbol{k}}$ by the posterior mean $\hat{\theta}_{\boldsymbol{j},\boldsymbol{k}} = \tilde{p}_{\boldsymbol{j},\boldsymbol{k}}\mu_{\boldsymbol{j},\boldsymbol{k}}$ and return to the time domain by $\hat{\boldsymbol{Y}} = W^\dagger\hat{\boldsymbol{\theta}}W$, which provides a denoised estimate of the original image. The denoising procedure described above will be called Complex Scale-Mixing Empirical Bayes (*cSM-EB*) shrinkage.

D. Phase-preserving Shrinkage

The bivariate shrinkage procedure *cSM-EB* proposed above changes the phase of the observed wavelet coefficients. Some authors, for example [5] and [17], argue that keeping the phase information unchanged is important for image denoising applications. It is possible to estimate $V_{\boldsymbol{j}}$ by a constrained empirical Bayes procedure so that the phase of the estimated coefficients

remains unchanged. The estimator of $\theta_{\boldsymbol{j},\boldsymbol{k}}$ for this case is $\hat{\theta}_{\boldsymbol{j},\boldsymbol{k}} = \tilde{p}_{\boldsymbol{j},\boldsymbol{k}} A_{\boldsymbol{j}} D_{\boldsymbol{j},\boldsymbol{k}}$, where $A_{\boldsymbol{j}} = \left(V_{\boldsymbol{j}}^{-1} + \Sigma_{\boldsymbol{j}}^{-1}\right)^{-1} \Sigma_{\boldsymbol{j}}^{-1}$. If $A_{\boldsymbol{j}} = \lambda_{\boldsymbol{j}} I_2$, then we preserve the phase of $D_{\boldsymbol{j},\boldsymbol{k}}$ in the shrinkage procedure and the shrinkage still remains nonlinear through $\tilde{p}_{\boldsymbol{j},\boldsymbol{k}}$. Note that $\Sigma_{\boldsymbol{j}}$ is known and $V_{\boldsymbol{j}}$ is estimated by the empirical Bayes procedure. It can be shown that if $V_{\boldsymbol{j}}$ is of the form

$$V_{\boldsymbol{j}} = v_{\boldsymbol{j}} \begin{pmatrix} 1 & \Sigma_{12}^{\boldsymbol{j}}/\Sigma_{11}^{\boldsymbol{j}} \\ \Sigma_{12}^{\boldsymbol{j}}/\Sigma_{11}^{\boldsymbol{j}} & \Sigma_{22}^{\boldsymbol{j}}/\Sigma_{11}^{\boldsymbol{j}} \end{pmatrix},$$

then $A_{\boldsymbol{j}}$ becomes a diagonal matrix of the form $A_{\boldsymbol{j}} = v_{\boldsymbol{j}}(\Sigma_{11}^{\boldsymbol{j}} + v_{\boldsymbol{j}})^{-1} I_2$. In fact, it is straightforward to show that only $V_{\boldsymbol{j}}$ that are scalar multiples of $\Sigma^{\boldsymbol{j}}$, or, equivalently, $A_{\boldsymbol{j}}$ that are scalar multiples of identity lead to phase preservation. This means that we restrict the structure of $V_{\boldsymbol{j}}$ as a function of a single unknown hyperparameter $v_{\boldsymbol{j}}$, which is subsequently estimated by the empirical Bayes procedure. Hyperparameter $v_{\boldsymbol{j}}$ can be thought to be a function of the known elements of $\Sigma_{\boldsymbol{j}}$, which can give different interpretations to the shrinkage rule. For example setting $v_{\boldsymbol{j}} = w_{\boldsymbol{j}} \Sigma_{11}^{\boldsymbol{j}} (\Sigma_{11}^{\boldsymbol{j}} + \Sigma_{22}^{\boldsymbol{j}})^{-1}$ results in $A_{\boldsymbol{j}} = w_{\boldsymbol{j}} (\Sigma_{11}^{\boldsymbol{j}} + \Sigma_{22}^{\boldsymbol{j}} + w_{\boldsymbol{j}})^{-1} I_2$. Although the above modification keeps the wavelet phase information unchanged, in practice it does not provide superior results to the *cSM-EB* procedure: the average mean squared error and visual appearance are very similar. Thus, we conclude that keeping the phase information of the image does not provide measurable benefits over the original procedure. An alternative way of preserving the phase is to consider simple multiwavelet-style hard thresholding of the complex wavelet coefficients, but simulations showed poor quantitative and visual results.

E. Complex maximum overlap wavelet transform

The complex scale mixing discrete wavelet transform can be operated in a non-decimated, or translation invariant manner, see, for example, [21] or [22].

Unlike the minimal (unitary) transform discussed above, the transforms defined under these names are highly redundant containing the convolutions of all shifts of the original signal due to omitting the decimation step in the algorithm. The redundancy is beneficial providing superior denoising performance and the transform can also be defined for arbitrary length input signals.

This wavelet transform was called maximal overlap DWT (MODWT) by [10], and we will follow their terminology here. The transform produces detail/wavelet coefficient vectors $W_{J-1}, W_{J-2}, \ldots, W_{J_0}$ and scaling coefficient vector V_{J_0} all of length n, the length of the original signal. Note that $J = \log_2(n)$. Therefore the total length of the transformed signal is $(J - J_0 + 1)n$ compared to n for the traditional orthogonal wavelet transform.

In [10] authors defined the transform of signal X such that $\|X\|^2 = \sum_{j=J_0}^{J-1} \|W_j\|^2 + \|V_{J_0}\|^2$ thus preserving the energy of the input signal. If we represent the wavelet transform with a rectangular matrix $W_{(J-J_0+1)n \times n}$, $W'W = I_n$ holds, however $WW' \neq I_{(J-J_0+1)n}$. Using a complex wavelet filter we define the complex maximal overlap DWT (cMODWT). Property $W^\dagger W = I_n$ allows us to define the complex maximal overlap scale mixing 2-D wavelet transform (cMOSM-DWT) by $G_{(J-J_0+1)n \times (J-J_0+1)n} = W F_{n \times n} W^\dagger$, for input image $F_{n \times n}$. Similarly as before, the inverse transform is defined as $F_{n \times n} = W^\dagger G_{(J-J_0+1)n \times (J-J_0+1)n} W$, and the total energy $E = \text{trace}(FF^\dagger)$ in the image F is preserved. Note that W is not unitary, but since $W^\dagger W = I_n$ we have $E = \text{trace}(FF^\dagger) = \text{trace}(GG^\dagger)$. Similarly to (15) we can show that in case of the cMOSM-DWT

$$\begin{aligned}
&\text{Cov}\left\{\text{Re}(\text{vec}(\boldsymbol{\epsilon})), \text{Re}(\text{vec}(\boldsymbol{\epsilon}))\right\} = \\
&\qquad \frac{\sigma^2}{2}\left[\text{Re}\{(\overline{W}W') \otimes (W\overline{W'})\} + \right. \\
&\qquad \left. \text{Re}\{(WW') \otimes (\overline{WW'})\}\right], \\
&\text{Cov}\left\{\text{Re}(\text{vec}(\boldsymbol{\epsilon})), \text{Im}(\text{vec}(\boldsymbol{\epsilon}))\right\} = \\
&\qquad -\frac{\sigma^2}{2}\left[\text{Im}\{(\overline{W}W') \otimes (W\overline{W'})\} + \right. \\
&\qquad \left. \text{Im}\{(WW') \otimes (\overline{WW'})\}\right], \\
&\text{Cov}\left\{\text{Im}(\text{vec}(\boldsymbol{\epsilon})), \text{Im}(\text{vec}(\boldsymbol{\epsilon}))\right\} = \\
&\qquad \frac{\sigma^2}{2}\left[\text{Re}\{(\overline{W}W') \otimes (W\overline{W'})\} - \right. \\
&\qquad \left. \text{Re}\{(WW') \otimes (\overline{WW'})\}\right]. \qquad (17)
\end{aligned}$$

For the detailed derivations, see Appendix. Note that the expressions of (15) are the special case of (17), because for the standard unitary W, $W\overline{W'} = \overline{W}W' = I_{n \times n}$.

Denoising via empirical Bayes shrinkage is defined in the same fashion for the cMOSM-DWT as in case of the unitary transform, but it is computationally more expensive because of the added redundancy. The proposed denoising method based on the maximal overlap transform will be called (*cMOSM-EB*) method.

IV. Simulations and comparison

In this section we discuss the performance of the proposed estimators (*cSM-EB*) and (*cMOSM-EB*) and compare it to some other established image denoising methods. In our simulations three standard test images (`Lenna`, `Peppers`, and `Barbara`) of size 512×512 were considered. We added i.i.d. normal noise with standard deviation σ to the test images. In the simulation study we assumed σ to be known. In practical applications, however, the noise variance can be estimated using a robust median estimator from the finest scale wavelet coefficients, which is commonly performed in signal [23] and image denoising [24]. Three different noise levels were considered, $\sigma = 10$, $\sigma = 25$ and $\sigma = 50$. These noise levels correspond to peak signal-to-noise ratio (PSNR) of 28.13, 20.17, and 14.15, where the PSNR is defined as

$$\text{PSNR} = 10 \log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right), \qquad (18)$$

where MAX_I is the maximum pixel value of the image (255 in our case), and MSE is the mean squared error, which is equal to σ^2 if no image denoising is applied.

The coarsest decomposition level was set at $J_0 = 3$. Reconstruction of the image was evaluated by the average PNSR and the average Structural SIMilarity (SSIM) index, averaging over $M = 50$ simulation runs. The MSE in (18) for each simulation run k was calculated calculated as

$$\text{MSE}_k = \frac{\|\hat{\boldsymbol{Y}}_k - \boldsymbol{Y}\|_F^2}{n^2},$$

where $\boldsymbol{Y}$ is the $n \times n$ true (noiseless) image, and $\hat{\boldsymbol{Y}}_k$ is the estimated image from the kth simulation run. The SSIM index for each simulation run was calculated using an available MATLAB package by [25].

The performance of our methods is compared to the Hidden Markov Model (*HMM*) of [26], to the Complex Hidden Markov Tree (*cHMT*) model of [27], to the Block-Matching and 3D filtering (*BM3D*) algorithm of [28], and to the method by [24]. The method by [24], denoted as *Seles*, uses a redundant 2-D wavelet transform on a symmetric extension of the image. In order to compare the performance with methods using the orthogonal wavelet transform, we modified the implementation of [24] to use traditional 2-D periodic orthogonal wavelets transform and denote it as *SelesPO*.

For methods *HMM* and *cHMT* the pixel values were normalized to [0,1] as suggested by the authors. Note that method *cHMT* was based on the 2-D dual-tree complex wavelet transform, which provides superior performance over the regular complex wavelet transforms.

The results are summarized in Table I. The proposed methods exhibit excellent performance comparable to most of the state-of-the-art methods. The *BM3D* method provides superior denoising results, but it is based on a fairly complicated multistep procedure consisting of block grouping, 3D transform, collaborative hard-thresholding, and Wiener filtering. We emphasize that the proposed method achieves comparable results by using a simpler and more straightforward transform taking advantage of a powerful statistical shrinkage model.

Visual results are presented in Figure 4 for `Lenna` and in Figure 5 for `Peppers` with noise level $\sigma = 25$ using methods *cSM-EB*, *HMM*, *cMOSM-EB*, *cHMT*, and *BM3D*. We also presented the right eye of the denoised `Lenna` to illuminate the visual performance of the methods. It can be seen that the proposed methods provide good visual performance, comparable to some of the state-of-the-art methods.

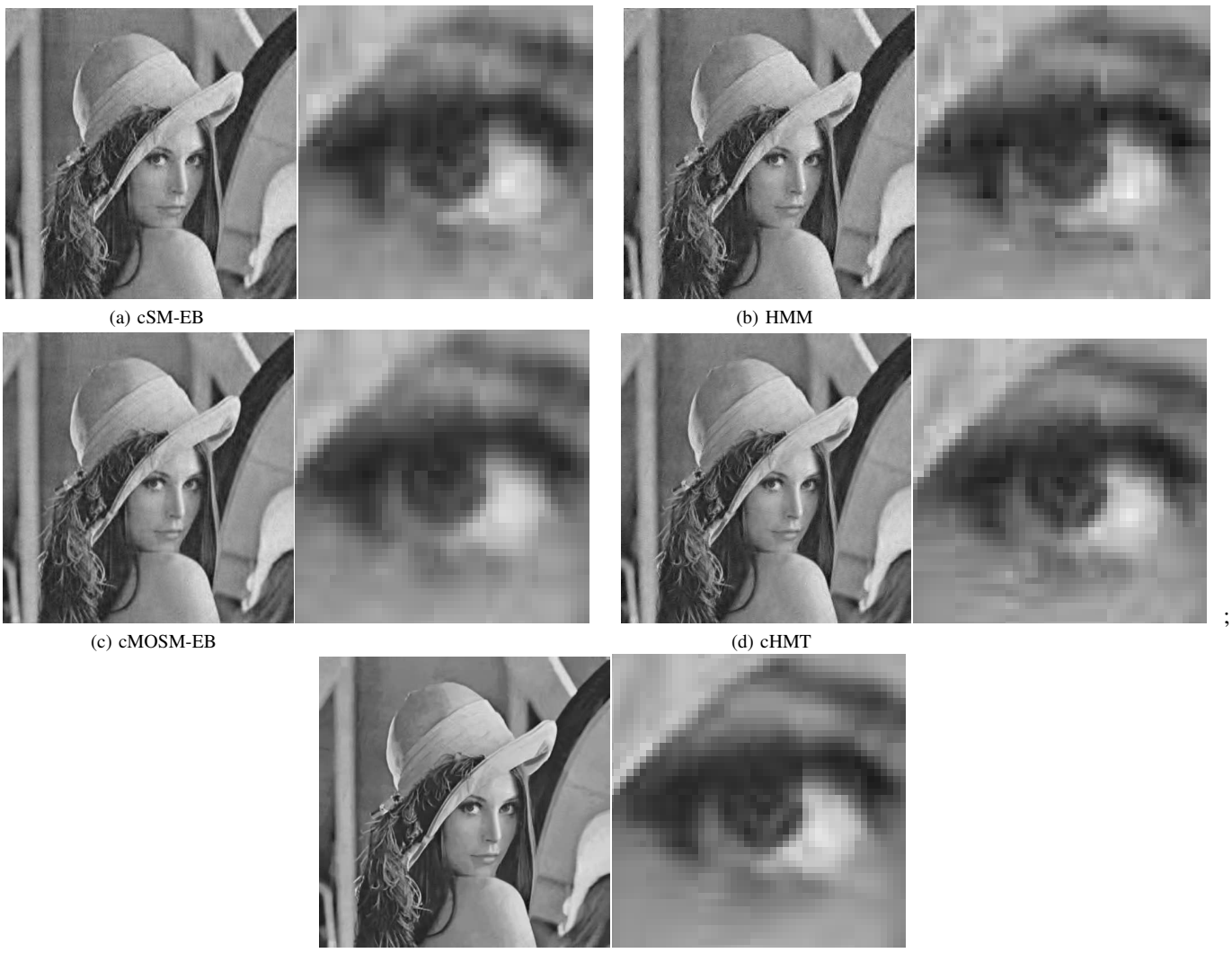
TABLE I
AVERAGE PNSR AND SSIM (IN BRACKETS) OF THE PROPOSED *cSM-EB* AND *cMOSM-EB* PROCEDURES COMPARED TO ESTABLISHED METHODS

Picture	Method	$\sigma = 10$	$\sigma = 25$	$\sigma = 50$
Lenna	cSM-EB	34.06	29.91	27.02
		(0.9523)	(0.8751)	(0.7788)
	SelesPO	33.91	29.55	26.47
		(0.9526)	(0.8795)	(0.7846)
	HMM	33.80	29.23	26.09
		(0.9518)	(0.8661)	(0.7410)
	cMOSM-EB	35.07	30.80	27.67
		(0.9638)	(0.9040)	(0.8182)
	Seles	34.31	30.15	27.12
		(0.9547)	(0.8880)	(0.8024)
	cHMT	35.05	30.47	26.36
		(0.9619)	(0.9046)	(0.7940)
	BM3D	35.89	32.04	28.98
		(0.9688)	(0.9255)	(0.8646)
Peppers	cSM-EB	33.60	29.82	26.97
		(0.9491)	(0.8787)	(0.7887)
	SelesPO	33.44	29.39	26.15
		(0.9495)	(0.8827)	(0.7968)
	HMM	33.58	28.69	25.62
		(0.9531)	(0.8795)	(0.7567)
	cMOSM-EB	34.46	30.67	27.68
		(0.9614)	(0.9077)	(0.8295)
	Seles	33.65	29.82	26.76
		(0.9493)	(0.8849)	(0.8045)
	cHMT	34.51	30.40	26.10
		(0.9605)	(0.9112)	(0.8028)
	BM3D	35.07	31.92	29.03
		(0.9768)	(0.9399)	(0.8712)
Barbara	cSM-EB	32.09	27.00	23.90
		(0.9491)	(0.8787)	(0.7887)
	SelesPO	31.95	26.61	23.45
		(0.9584)	(0.8689)	(0.7426)
	HMM	31.67	26.42	22.96
		(0.9515)	(0.8518)	(0.7076)
	cMOSM-EB	33.31	27.95	24.38
		(0.9681)	(0.8945)	(0.7781)
	Seles	32.37	27.13	23.89
		(0.9601)	(0.8805)	(0.7621)
	cHMT	32.98	27.78	24.04
		(0.9642)	(0.8904)	(0.7794)
	BM3D	34.95	30.66	27.21
		(0.9768)	(0.9399)	(0.8712)

V. CONCLUSION AND FURTHER WORK

This paper introduced a discrete complex scale mixing 2-D wavelet transform, and maximal overlap version, and applied it in the context of image denoising. For both transforms we derived the covariance structure of wavelet-transformed white noise. Our shrinkage procedures, cSM-EB and cMOSM-EB, are based on empirical Bayes and utilize non-zero covariances between real and imaginary parts of the wavelet coefficients. We discussed the possibility of phase-preserving shrinkage in this framework. Overall, the methods we propose are conceptually simple, practically efficient, and exhibit often superior denoising performance when contrasted with comparable wavelet-based techniques.

Our methods currently involve diagonal shrinkage and better performance still could undoubtedly be achieved by using more general shrinkage methods that exploit structure across and between scales, such as block-shrinkage or tree-based shrinkage methods. Another generalization would be to go beyond Gaussianity and explore different sparse models and priors, as well as to assess comparative performance and compressibility of various multiscale transforms involving real



(a) cSM-EB (b) HMM

(c) cMOSM-EB (d) cHMT ;

(e) BM3D

Fig. 4. Denoised `Lenna` ($\sigma = 25$) and right eye using various methods.

and complex wavelets.

In the spirit of reproducible research a suite of MATLAB demo files for implementing cSM-EB and cMOSM-EB shrinkage can be found at http://gtwavelet.bme.gatech.edu/cSM-EB2.zip

APPENDIX
COVARIANCE STRUCTURE OF IID NOISE IN WAVELET DOMAIN (15), (17)

Let $Y \sim MN_{n,n}(M, C, D)$ denote an $n \times n$ normal random matrix Y with $E(Y) = M$ and $\mathrm{Cov}\{\mathrm{vec}(Y)\} = D \otimes C$, where C and D represents the column and row covariance structures, respectively. Let A and B be $n \times n$ complex-valued matrices. We will use the fact that $Z = BYA'$ has (complex) covariance structure

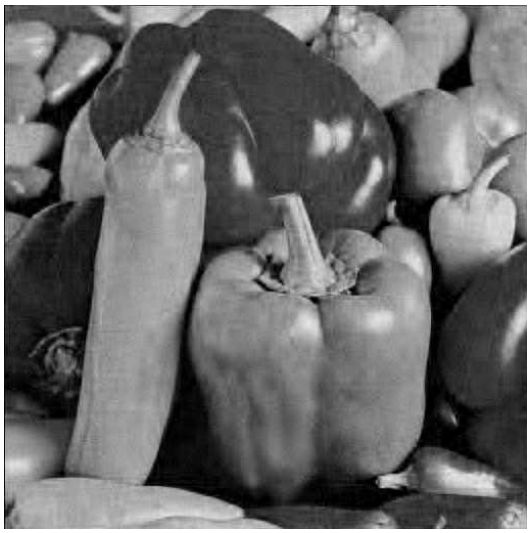
$$\begin{aligned}\mathrm{Cov}\{\mathrm{vec}(Z)\} &= \mathrm{Cov}\{\mathrm{vec}(BYA')\} \\ &= \mathrm{Cov}\{(A \otimes B)\mathrm{vec}(Y)\} \\ &= (A \otimes B)\mathrm{Cov}\{\mathrm{vec}(Y)\}(A \otimes B)^\dagger \\ &= (A \otimes B)(D \otimes C)(A^\dagger \otimes B^\dagger) \\ &= (ADA^\dagger) \otimes (BCB^\dagger).\end{aligned}$$

Here we use identities $\mathrm{vec}(BYA') = (A \otimes B)\mathrm{vec}(Y)$, $\overline{AB} = \overline{A}\,\overline{B}$, $(A \otimes B)' = A' \otimes B'$ and $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$. Note that for a complex matrix A we denote $\overline{A}$ as the conjugate of A, A' as the transpose of A and $\overline{A'} = A^\dagger$ as the Hermitian transpose of A. Also $\mathrm{vec}(Y)$ stands for vectorization, which stacks the columns of matrix Y into one column-vector.

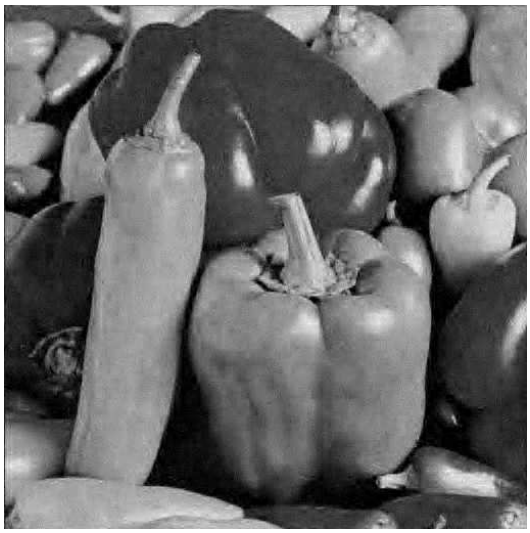
Following the model in (11), $\boldsymbol{e}$ is an $n \times n$ random matrix with i.i.d. $N(0, \sigma^2)$ elements, therefore $\boldsymbol{e} \sim MN_{n,n}(0, \sigma^2 I, I)$. After applying the 2-D scale-mixing wavelet transform, the noise in the wavelet domain becomes $\boldsymbol{\epsilon} = W\boldsymbol{e}W^\dagger$. Let $\mathrm{Re}(\boldsymbol{\epsilon}) = (\boldsymbol{\epsilon} + \overline{\boldsymbol{\epsilon}})/2$ and $\mathrm{Im}(\boldsymbol{\epsilon}) = (\boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}})/2i$. The covariance structure of $\boldsymbol{\epsilon}$ becomes

$$\begin{aligned}&\mathrm{Cov}\left[\mathrm{Re}\{\mathrm{vec}(\boldsymbol{\epsilon})\}, \mathrm{Re}\{\mathrm{vec}(\boldsymbol{\epsilon})\}\right] = \\ &\quad \mathrm{Cov}\left[\{\mathrm{vec}(\boldsymbol{\epsilon}) + \mathrm{vec}(\overline{\boldsymbol{\epsilon}})\}/2, \{\mathrm{vec}(\boldsymbol{\epsilon}) + \mathrm{vec}(\overline{\boldsymbol{\epsilon}})\}/2\right] = \\ &\quad [\mathrm{Cov}\{\mathrm{vec}(\boldsymbol{\epsilon}), \mathrm{vec}(\boldsymbol{\epsilon})\} + \mathrm{Cov}\{\mathrm{vec}(\overline{\boldsymbol{\epsilon}}), \mathrm{vec}(\boldsymbol{\epsilon})\} + \\ &\quad + \mathrm{Cov}\{\mathrm{vec}(\boldsymbol{\epsilon}), \mathrm{vec}(\overline{\boldsymbol{\epsilon}})\} + \mathrm{Cov}\{\mathrm{vec}(\overline{\boldsymbol{\epsilon}}), \mathrm{vec}(\overline{\boldsymbol{\epsilon}})\}]/4,\end{aligned}$$

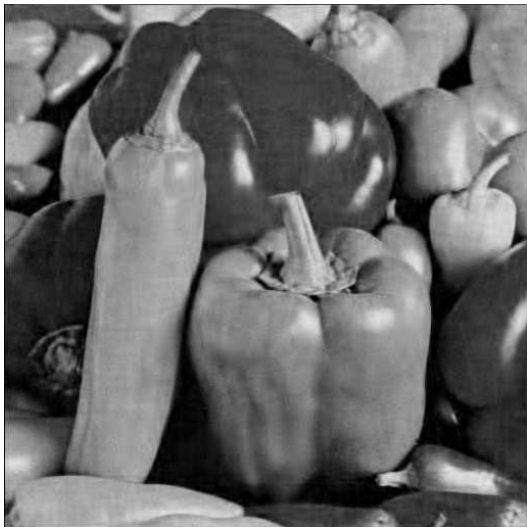
$$\mathrm{Cov}\left[\mathrm{Re}\{\mathrm{vec}(\boldsymbol{\epsilon})\}, \mathrm{Im}\{\mathrm{vec}(\boldsymbol{\epsilon})\}\right] =$$



(a) cSM-EB



(b) HMM



(c) cMOSM-EB

(d) cHMT

(e) BM3D

Fig. 5. Denoised `Peppers` ($\sigma = 25$) using various methods.

$$\begin{aligned}&\text{Cov}\left[\{\text{vec}(\boldsymbol{\epsilon})+\text{vec}(\overline{\boldsymbol{\epsilon}})\}/2,\{\text{vec}(\boldsymbol{\epsilon})-\text{vec}(\overline{\boldsymbol{\epsilon}})\}/2i\right]=\\&\left[\text{Cov}\left\{\text{vec}(\boldsymbol{\epsilon}),\text{vec}(\boldsymbol{\epsilon})\right\}+\text{Cov}\left\{\text{vec}(\overline{\boldsymbol{\epsilon}}),\text{vec}(\boldsymbol{\epsilon})\right\}-\right.\\&\left.-\text{Cov}\left\{\text{vec}(\boldsymbol{\epsilon}),\text{vec}(\overline{\boldsymbol{\epsilon}})\right\}-\text{Cov}\left\{\text{vec}(\overline{\boldsymbol{\epsilon}}),\text{vec}(\overline{\boldsymbol{\epsilon}})\right\}\right]/4\bar{i},\end{aligned}$$

$$\begin{aligned}&\text{Cov}\left[\text{Im}\{\text{vec}(\boldsymbol{\epsilon})\},\text{Im}\{\text{vec}(\boldsymbol{\epsilon})\}\right]=\\&\text{Cov}\left[\{\text{vec}(\boldsymbol{\epsilon})-\text{vec}(\overline{\boldsymbol{\epsilon}})\}/2i,\{\text{vec}(\boldsymbol{\epsilon})-\text{vec}(\overline{\boldsymbol{\epsilon}})\}/2i\right]=\\&\left[\text{Cov}\left\{\text{vec}(\boldsymbol{\epsilon}),\text{vec}(\boldsymbol{\epsilon})\right\}-\text{Cov}\left\{\text{vec}(\overline{\boldsymbol{\epsilon}}),\text{vec}(\boldsymbol{\epsilon})\right\}-\right.\\&\left.-\text{Cov}\left\{\text{vec}(\boldsymbol{\epsilon}),\text{vec}(\overline{\boldsymbol{\epsilon}})\right\}+\text{Cov}\left\{\text{vec}(\overline{\boldsymbol{\epsilon}}),\text{vec}(\overline{\boldsymbol{\epsilon}})\right\}\right]/4i\bar{i}.\end{aligned}$$

Further we have that

$$\begin{aligned}&\text{Cov}\left\{\text{vec}(\boldsymbol{\epsilon}),\text{vec}(\boldsymbol{\epsilon})\right\}=\\&\quad\text{Cov}\left\{\text{vec}(W\boldsymbol{e}W^{\dagger}),\text{vec}(W\boldsymbol{e}W^{\dagger})\right\}=\\&\quad\text{Cov}\left\{(\overline{W}\otimes W)\text{vec}(\boldsymbol{e}),(\overline{W}\otimes W)\text{vec}(\boldsymbol{e})\right\}=\\&\quad(\overline{W}\otimes W)\text{Cov}\left\{\text{vec}(\boldsymbol{e})\right\}(\overline{W}\otimes W)^{\dagger}=\\&\quad(\overline{W}\otimes W)(\sigma^2I\otimes I)(W'\otimes\overline{W'})=\\&\quad\sigma^2(\overline{W}W')\otimes(W\overline{W'}).\end{aligned}$$

Similarly

$$\begin{aligned}&\text{Cov}\left\{\text{vec}(\overline{\boldsymbol{\epsilon}}),\text{vec}(\boldsymbol{\epsilon})\right\}=\\&\quad\text{Cov}\left\{\text{vec}(\overline{W}\boldsymbol{e}W'),\text{vec}(W\boldsymbol{e}W^{\dagger})\right\}=\\&\quad\text{Cov}\left\{(W\otimes\overline{W})\text{vec}(\boldsymbol{e}),(\overline{W}\otimes W)\text{vec}(\boldsymbol{e})\right\}=\\&\quad(W\otimes\overline{W})\text{Cov}\left\{\text{vec}(\boldsymbol{e})\right\}(\overline{W}\otimes W)^{\dagger}=\\&\quad(W\otimes\overline{W})(\sigma^2I\otimes I)(W'\otimes\overline{W'})=\\&\quad\sigma^2(WW')\otimes(\overline{WW'}),\end{aligned}$$

and

$$\begin{aligned}&\text{Cov}\left\{\text{vec}(\boldsymbol{\epsilon}),\text{vec}(\overline{\boldsymbol{\epsilon}})\right\}=\\&\quad\text{Cov}\left\{\text{vec}(W\boldsymbol{e}W^{\dagger}),\text{vec}(\overline{W}\boldsymbol{e}W')\right\}=\\&\quad\text{Cov}\left\{(\overline{W}\otimes W)\text{vec}(\boldsymbol{e}),(W\otimes\overline{W})\text{vec}(\boldsymbol{e})\right\}=\\&\quad(\overline{W}\otimes W)\text{Cov}\left\{\text{vec}(\boldsymbol{e})\right\}(W\otimes\overline{W})^{\dagger}=\\&\quad(\overline{W}\otimes W)(\sigma^2I\otimes I)(\overline{W'}\otimes W')=\\&\quad\sigma^2(\overline{WW'})\otimes(WW').\end{aligned}$$

And finally we have that

$$\begin{aligned}&\text{Cov}\left\{\text{vec}(\overline{\boldsymbol{\epsilon}}),\text{vec}(\overline{\boldsymbol{\epsilon}})\right\}=\\&\quad\text{Cov}\left\{\text{vec}(\overline{W}\boldsymbol{e}W'),\text{vec}(\overline{W}\boldsymbol{e}W')\right\}=\\&\quad\text{Cov}\left\{(W\otimes\overline{W})\text{vec}(\boldsymbol{e}),(W\otimes\overline{W})\text{vec}(\boldsymbol{e})\right\}=\\&\quad(W\otimes\overline{W})\text{Cov}\left\{\text{vec}(\boldsymbol{e})\right\}(W\otimes\overline{W})^{\dagger}=\\&\quad(W\otimes\overline{W})(\sigma^2I\otimes I)(\overline{W'}\otimes W')=\\&\quad\sigma^2(W\overline{W'})\otimes(\overline{W}W').\end{aligned}$$

Therefore

$$\begin{aligned}&\text{Cov}\left\{\text{Re}(\text{vec}(\boldsymbol{\epsilon})),\text{Re}(\text{vec}(\boldsymbol{\epsilon}))\right\}=\\&\quad\frac{\sigma^2}{4}\left[(\overline{W}W')\otimes(W\overline{W'})+(WW')\otimes(\overline{WW'})+\right.\\&\quad\left.+(\overline{WW'})\otimes(WW')+(W\overline{W'})\otimes(\overline{W}W')\right]=\\&\quad\frac{\sigma^2}{2}\left[\text{Re}\{(\overline{W}W')\otimes(W\overline{W'})\}+\right.\\&\quad\left.+\text{Re}\{(WW')\otimes(\overline{WW'})\}\right],\end{aligned}$$

$$\begin{aligned}&\text{Cov}\left\{\text{Re}(\text{vec}(\boldsymbol{\epsilon})),\text{Im}(\text{vec}(\boldsymbol{\epsilon}))\right\}=\\&\quad\frac{\sigma^2}{4\bar{i}}\left[(\overline{W}W')\otimes(W\overline{W'})+(WW')\otimes(\overline{WW'})-\right.\\&\quad\left.-(\overline{WW'})\otimes(WW')-(W\overline{W'})\otimes(\overline{W}W')\right]=\\&\quad-\frac{\sigma^2}{4i}\left[2i\text{Im}\{(\overline{W}W')\otimes(W\overline{W'})\}+\right.\\&\quad\left.+2i\text{Im}\{(WW')\otimes(\overline{WW'})\}\right]=\\&\quad-\frac{\sigma^2}{2}\left[\text{Im}\{(\overline{W}W')\otimes(W\overline{W'})\}+\right.\\&\quad\left.+\text{Im}\{(WW')\otimes(\overline{WW'})\}\right],\end{aligned}$$

$$\begin{aligned}&\text{Cov}\left\{\text{Im}(\text{vec}(\boldsymbol{\epsilon})),\text{Im}(\text{vec}(\boldsymbol{\epsilon}))\right\}=\\&\quad\frac{\sigma^2}{4i\bar{i}}\left[(\overline{W}W')\otimes(W\overline{W'})-(WW')\otimes(\overline{WW'})-\right.\\&\quad\left.-(\overline{WW'})\otimes(WW')+(W\overline{W'})\otimes(\overline{W}W')\right]=\\&\quad\frac{\sigma^2}{2}\left[\text{Re}\{(\overline{W}W')\otimes(W\overline{W'})\}-\right.\\&\quad\left.-\text{Re}\{(WW')\otimes(\overline{WW'})\}\right].\end{aligned}$$

These general expressions describe the covariance structure for the maximum overlap version of the wavelet transform. Note, however, that for the orthogonal wavelet transform $W\overline{W'}=\overline{W}W'=I_{n\times n}$, therefore the above expressions simplify to

$$\begin{aligned}&\text{Cov}\left\{\text{Re}(\text{vec}(\boldsymbol{\epsilon})),\text{Re}(\text{vec}(\boldsymbol{\epsilon}))\right\}=\\&\qquad\frac{\sigma^2}{2}\left[I_{n^2\times n^2}+\text{Re}\left\{WW'\otimes\overline{WW'}\right\}\right],\\&\text{Cov}\left\{\text{Re}(\text{vec}(\boldsymbol{\epsilon})),\text{Im}(\text{vec}(\boldsymbol{\epsilon}))\right\}=\\&\qquad-\frac{\sigma^2}{2}\left[\text{Im}\left\{WW'\otimes\overline{WW'}\right\}\right],\\&\text{Cov}\left\{\text{Im}(\text{vec}(\boldsymbol{\epsilon})),\text{Im}(\text{vec}(\boldsymbol{\epsilon}))\right\}=\\&\qquad\frac{\sigma^2}{2}\left[I_{n^2\times n^2}-\text{Re}\left\{WW'\otimes\overline{WW'}\right\}\right].\end{aligned}$$

Also note that using the expressions above, the wavelet domain covariance structure of a more complicated, general (not white) noise structure $\boldsymbol{e}\sim N_{n,n}(0,C,D)$ can also be derived. Above we only considered the white noise case, therefore $D\otimes C=\sigma^2I\otimes I$.

ACKNOWLEDGMENT

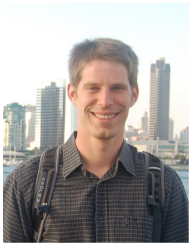
Work of Orietta Nicolis was partially funded by the national Chilean project Fondecyt ID 1131147. Atlanta Clinical and Translational Science Institute (ACTSI) provided partial support for work of Brani Vidakovic. Guy Nason was supported by EPSRC grants EP/K020951 and EP/I01697X from The Energy Programme, a RCUK cross-council initiative led by EPSRC and contributed to by ESRC, NERC, BBSRC and STFC.

REFERENCES

[1] N. Remenyi and B. Vidakovic, "Bayesian wavelet shrinkage strategies: A review," in *Multiscale Signal Analysis and Modeling*, X. Shen and A. I. Zayed, Eds. Springer-Verlag, New York, 2013, pp. 317–346.

[2] B. Vidakovic, *Statistical Modeling by Wavelets*, ser. Wiley series in probability and mathematical statistics: Applied probability and statistics. John Wiley & Sons, Inc., New York, 1999.

[3] G. P. Nason, *Wavelet Methods in Statistics with R*, First ed. Springer, 2008.
[4] J.-M. Lina and B. Macgibbon, "Non-linear shrinkage estimation with complex Daubechies wavelets," in *Proceedings of SPIE, Wavelet Applications in Signal and Image Processing V*, ser. 3169, 1997, pp. 67–79.
[5] J.-M. Lina, "Image processing with complex Daubechies wavelets," *Journal of Mathematical Imaging and Vision*, vol. 7, pp. 211–223, 1997.
[6] J.-M. Lina, P. Turcotte, and B. Goulard, "Complex dyadic multiresolution analyses," in *Advances in Imaging and Electron Physics*, P. W. Hawkes, Ed. Elsevier, 1999, vol. 109, pp. 163 – 197.
[7] S. Barber and G. P. Nason, "Denoising real images using complex-valued wavelets," in *Stochastic Geometry, Biological Structure and Images*, R. Aykroyd, K. Mardia, and M. Langdon, Eds. Department of Statistics, University of Leeds, 2003, pp. 91–95.
[8] ——, "Real nonparametric regression using complex wavelets," *Journal of the Royal Statistical Society, Series B*, vol. 66, pp. 927–939, 2004.
[9] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, pp. 205–220, 1992.
[10] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000.
[11] P. Ramírez-Cobo, K. S. Lee, A. Molini, A. Porporato, G. Katul, and B. Vidakovic, "A wavelet-based spectral method for extracting self-similarity measures in time-varying two-dimensional rainfall maps," *Journal of Time Series Analysis*, vol. 32, no. 4, pp. 351–363, 2011.
[12] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov, "Hyperbolic wavelet approximation," *Constructive Approximation*, vol. 14, no. 1, pp. 1–26, 1998.
[13] M. H. Neumann and R. von Sachs, "Wavelet thresholding in anisotropic function classes and application to adaptive estimation," *Annals of Statistics*, vol. 25, no. 1, pp. 38–76, 1997.
[14] V. Zavadsky, "Image approximation by rectangular wavelet transform," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 2, pp. 129–138, 2007.
[15] S. G. Roux, M. Clausel, B. Vedel, S. Jaffard, and P. Abry, "Self-similar anisotropic texture analysis: The hyperbolic wavelet transform contribution," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4353–4363, 2013.
[16] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, 1992.
[17] D. Clonda, J.-M. Lina, and B. Goulard, "Complex Daubechies wavelets: properties and statistical image modelling," *Signal Processing*, vol. 84, pp. 1–23, 2004.
[18] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *Journal of the Royal Statistical Society, Series B*, vol. 60, pp. 725–749, 1998.
[19] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *Journal of the American Statistical Association*, vol. 92, pp. 1413–1421, 1997.
[20] M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika*, vol. 85, no. 2, pp. 391–401, 1998.
[21] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, ser. Lecture Notes in Statistics, A. Antoniadis and G. Oppenheim, Eds., vol. 103. Springer-Verlag, New York, 1995, pp. 125–150.
[22] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, ser. Lecture Notes in Statistics, A. Antoniadis and G. Oppenheim, Eds., vol. 103. Springer-Verlag, New York, 1995, pp. 281–300.
[23] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
[24] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744–2756, 2002.
[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
[26] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.
[27] J. K. Romberg, H. Choi, R. G. Baraniuk, and N. Kingsbury, "A hidden markov tree model for the complex wavelet transform," Department of Electrical and Computer Engineering, Rice University, Tech. Rep., 2002.
[28] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.



Norbert Remenyi is a statistician at Sabre Research, an operations research group of Sabre Holdings. He received his Ph.D. degree in Statistics (Industrial and Systems Engineering) at Georgia Institute of Technology in 2012. His research focus is on Bayesian modeling in the wavelet domain with applications to nonparametric regression, image denoising, and estimation of partially linear models. He was an ORISE research fellow with the Centers for Disease Control and Prevention (CDC) in 2012. Previously Remenyi earned a master's degree in Engineering Management at Budapest University of Technology and Economics (Hungary) as well as a master's degree in Experimental Statistics at New Mexico State University.

Orietta Nicolis received her Ph.D. degree in Applied Statistics in 1999 from the University of Padova, Italy. She was an Assistant Professor of Statistics at the University of Bergamo, Italy, Department of Information Technology and Mathematical Methods. Since 2012 she is an Associate Professor at the University of Valparaiso, Chile. In addition to teaching responsibilities at the university, she also leads several scientific research projects. Dr. Nicolis's research interests include the study of multiscale methods, image processing, spatial point processes, and spatio-temporal models with applications to medical images and environmental data.

Guy Nason is Professor of Statistics in the School of Mathematics at the University of Bristol, U.K. His research focuses on multiscale methods in statistics, regression, time series and variance stabilization. Nason received his bachelor's degree and PhD at the University of Bath in 1988 and 1992 respectively, and obtained a distinction in the Diploma in Mathematical Statistics at the University of Cambridge in 1989. He is a Fellow of the Royal Statistical Society, UK and member of the Institute of Mathematical Statistics, the International Statistical Institute, the International Association for Statistical Computing and an invited Fellow of the Institute of Mathematics and its Applications.

Brani Vidakovic is a Professor of Statistics at the School of Industrial and Systems Engineering and Department Biomedical Engineering at Georgia Institute of Technology and Emory University. His research focuses on biostatistics, Bayesian methodology, and statistical modeling in multiscale domains with applications in bioengineering, medicine and environmental sciences. Vidakovic received his bachelor and masters degrees in mathematics from University of Belgrade (Serbia) in 1978, and 1981, respectively, and Ph. D. in statistics at Purdue University (Indiana, USA) in 1992. He was an Assistant and Associate Professor at Institute of Statistics and Decision Sciences at Duke University form 1992 to 2000 prior to joining Georgia Institute of Technology. Vidakovic is a Fellow of American Statistical Association and elected member of International Statistical Institute.