

A CURE for Noisy Magnetic Resonance Images: Chi-Square Unbiased Risk Estimation

Florian Luisier, *Member, IEEE*, Thierry Blu, *Fellow, IEEE*, and Patrick J. Wolfe, *Senior Member, IEEE*

Abstract—In this paper, we derive an unbiased expression for the expected mean-squared error associated with continuously differentiable estimators of the noncentrality parameter of a chi-square random variable. We then consider the task of denoising squared-magnitude magnetic resonance (MR) image data, which are well modeled as independent noncentral chi-square random variables on two degrees of freedom. We consider two broad classes of linearly parameterized shrinkage estimators that can be optimized using our risk estimate, one in the general context of undecimated filterbank transforms, and the other in the specific case of the unnormalized Haar wavelet transform. The resultant algorithms are computationally tractable and improve upon most state-of-the-art methods for both simulated and actual MR image data.

Index Terms—Chi-square distribution, filterbank transform, image denoising, magnetic resonance (MR) imaging, Rician noise, unbiased risk estimation.

I. INTRODUCTION

MAGNETIC resonance (MR) imaging is a fundamental *in vivo* medical imaging technique that provides high-contrast images of soft tissue without the use of ionizing radiation. The signal-to-noise ratio (SNR) of an acquired MR image is determined by numerous physical and structural factors, such as the static field strength, resolution, receiver bandwidth, and the number of signal averages collected at each encoding step [1], [2]. Current developments in MR imaging are focused primarily on lowering its inherently high scanning time, increasing the spatiotemporal resolution of the images themselves, and reducing the cost of the overall system. However, the pursuit of any of these objectives has a negative impact on the SNR of the acquired image. A postacquisition denoising step is therefore essential for visualization by clinicians and for meaningful computer-aided diagnoses [1].

Manuscript received June 14, 2011; revised February 22, 2012; accepted March 2, 2012. Date of publication April 5, 2012; date of current version July 18, 2012. This work was supported in part by the Swiss National Science Foundation Fellowship PBELP2-133245, the U.S. National Science Foundation under Grant DMS-0652743, and the General Research Fund CUHK410209 from the Hong Kong Research Grant Council. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jose M. Bioucas-Dias.

F. Luisier is with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: fluisier@seas.harvard.edu).

T. Blu is with the Department of Electronic Engineering, Chinese University of Hong Kong, Kowloon, Hong Kong (e-mail: thierry.blu@m4x.org).

P. J. Wolfe was with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA. He is now with University College London, London WC1E 7HB, U.K. (e-mail: wolfe@stat.harvard.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2012.2191565

In MR image acquisition, the data consist of discrete Fourier samples, usually referred to as k -space samples. These data are corrupted by random noise, due primarily to thermal fluctuations generated by a patient's body in the imager's receiver coils [1], [2]. This random degradation is well modeled by additive white Gaussian noise (AWGN) which independently corrupts the real and imaginary parts of the complex-valued k -space samples [3]. Assuming a Cartesian sampling pattern, the output image is straightforwardly obtained by computing the inverse discrete Fourier transform of the k -space samples, the resolution of the reconstructed image is then determined by the maximum k -space sampling frequency. Note that some recent works ([4] and references therein) aim to accelerate MR image acquisition time by undersampling the k -space. Non-Cartesian (e.g., spiral or random) sampling trajectories as well as nonlinear (usually sparsity-driven) reconstruction schemes are then considered. This axis of research is, however, outside the scope of this paper.

Following application of an inverse discrete Fourier transform, the resulting image may be considered as a complex-valued signal corrupted by independent and identically distributed samples of complex AWGN. In magnitude MR imaging, the image phase is disregarded and only the magnitude is considered for visualization and further analysis. Although the samples of the magnitude image remain statistically independent, they are no longer Gaussian, but rather are Rician distributed [5]. Contrary to the Gaussian case, this form of "noise" is *signal-dependent* in that both the mean and the variance of the magnitude samples depend on the underlying noise-free magnitude image. Consequently, generic denoising algorithms designed for AWGN reduction usually do not give satisfying results on Rician image data.

Denoising of magnitude MR images has thus gained much attention over the past several years. Two main strategies can be distinguished as follows: 1) the Rician data are treated directly, often in the image domain and 2) the denoising is applied to the squared magnitude MR image, which follows a (scaled) noncentral chi-square distribution on two degrees of freedom, whose noncentrality parameter is proportional to the underlying noise-free squared magnitude. An appealing aspect of this strategy is that it renders the bias due to the noise constant rather than signal dependent, thus enabling its straightforward removal. When reverting to the original image domain, a nonlinear mapping (e.g., a square root transformation) must be applied, which reintroduces a bias into the final image magnitude estimates. However, as indicated by our experiments below, this residual effect is far less pronounced than that

present in the Rician data, and thus denoising in the squared-image domain remains a widely used approach [6]–[12].

Following the first strategy previously mentioned, several Rician-based maximum likelihood estimators [13]–[15] have been proposed. In [15], this estimation is performed nonlocally among pixels having a similar neighborhood. A Bayesian maximum *a posteriori* estimator has been devised in [16], with the prior modeled in a nonparametric Markov random field framework. Very recently, Foi proposed a Rician-based variance stabilizing transform (VST) that makes the use of standard AWGN denoisers more effective [17]. Following the second strategy, a linear minimum mean-squared error (MMSE) filter applied to the squared magnitude image has been derived in [9] and [10]. In addition to these statistical model-based approaches, much work has also been devoted to the adaptation and enhancement of the nonlocal means filter originally developed by Buades *et al.* for AWGN reduction [18]. The core of this relatively simple, yet effective, denoising approach consists of a weighted averaging of similarly close (in spatial and photometric distance) pixels. Some of these adaptations operate directly on the Rician magnitude image data [19], [20], while others are applied to the squared magnitude image [8], [11], [12].

In addition to these image-domain approaches, several magnitude MR image denoising algorithms have also been developed for the wavelet domain. The sparsifying and decorrelating effects of wavelets and other related transforms typically result in the concentration of relevant image features into a few significant wavelet coefficients. Simple thresholding rules based on coefficient magnitude then provide an effective means of reducing the noise level while preserving sharp edges in the image. In the earliest uses of the wavelet transform for MR image denoising [21], [22], the Rician distribution of the data was not explicitly taken into account. Nowak subsequently proposed wavelet coefficient thresholding based on the observation that the empirical wavelet coefficients of the squared-magnitude data are unbiased estimators of the coefficients of the underlying squared-magnitude image, and that the residual scaling coefficients exhibit a signal-independent bias that is easily removable [6]. While the pointwise coefficient thresholding proposed in [6] is most natural in the context of an orthogonal discrete wavelet transform, Pižurica *et al.* subsequently developed a Bayesian wavelet thresholding algorithm applied in an undecimated wavelet representation [7]. Wavelet-based denoising algorithms that require the entirety of the *complex* MR image data [23]–[25] are typically applied separately to the real and imaginary components of the image.

In this paper, we develop a general result for chi-square unbiased risk estimation (CURE), which we then apply to the task of denoising squared-magnitude MR images. We provide two instances of effective transform-domain algorithms, each based on the concept of linear expansion of thresholds (LET) introduced in [26] and [27]. The first class of proposed algorithms consists of a pointwise continuously differentiable thresholding applied to the coefficients of an undecimated filterbank transform. The second class takes advantage of the conservation of the chi-square statistics across the lowpass channel of the unnormalized Haar wavelet transform. Owing

to the remarkable property of this orthogonal transform, it is possible to derive independent risk estimates in each wavelet subband, allowing a very fast denoising procedure. These estimates are then used to optimize the parameters of subband-dependent joint inter-/intra-scale LET.

This paper is organized as follows. We first derive, in Section II, an unbiased expression of the risk associated with estimators of the noncentrality parameter of a chi-square random variable having arbitrary (known) degrees of freedom. Then, in Section III, we apply this result to optimize pointwise estimators for undecimated filterbank transform coefficients, using LET. In Section IV, we give an expression for CURE directly in the unnormalized Haar wavelet domain, and propose a more sophisticated joint inter-/intrascale LET. We conclude in Section V with denoising experiments conducted on simulated and actual magnitude MR images, and evaluations of our methods relative to the current state of the art. Note that a subset of this paper (mainly part of Section III) has been presented at the 2011 IEEE International Conference on Image Processing [28].

II. CURE

Assume the observation of a vector $\mathbf{y} \in \mathbb{R}_+^N$ of N independent samples y_n , each randomly drawn from a noncentral chi-square distribution with (unknown) noncentrality parameter $x_n \geq 0$ and (known) common degrees of freedom $K > 0$. We use the vector notation $\mathbf{y} \sim \chi_K^2(\mathbf{x})$, recalling that (for integer K) the joint distribution $p(\mathbf{y}|\mathbf{x})$ can be seen as resulting from the addition of K independent vectors on \mathbb{R}_+^N whose coordinates are the *squares* of noncentered Gaussian random variables of unit variance. This observation model is statistically characterized by the data likelihood

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \prod_{n=1}^N p(y_n|x_n) \\ &= \prod_{n=1}^N \frac{1}{2} e^{-\frac{x_n+y_n}{2}} \left(\frac{y_n}{x_n}\right)^{\frac{K-2}{4}} I_{\frac{K}{2}-1}(\sqrt{x_n y_n}) \end{aligned} \quad (1)$$

where $I_\alpha(u) = \sum_{k \in \mathbb{N}} \{1/[k! \Gamma(k + \alpha + 1)]\} (u/2)^{2k + \alpha}$ is the α -order modified Bessel function of the first kind.

The chi-square distribution of (1) is most easily understood through its characteristic function, or Fourier transform

$$\hat{p}(\omega|\mathbf{x}) = \prod_{n=1}^N \frac{\exp\left(-\frac{j\omega_n x_n}{1 + 2j\omega_n}\right)}{(1 + 2j\omega_n)^{K/2}}. \quad (2)$$

For instance, by equating first- and second-order Taylor developments of this Fourier transform with the corresponding moments, it is straightforward to show that

$$\mathcal{E}\{\mathbf{y}\} = \mathbf{x} + K \cdot \mathbf{1}$$

and

$$\mathcal{E}\{\|\mathbf{y}\|^2\} = \|\mathbf{x}\|^2 + 2(K+2)\mathbf{1}^T \mathbf{x} + NK(K+2) \quad (3)$$

where $\mathcal{E}\{\cdot\}$ is the expectation operator.

When designing an estimator $\hat{\mathbf{x}} = \mathbf{f}(\mathbf{y})$ of \mathbf{x} , a natural criterion is the value of its associated risk or expected MSE, defined here by

$$\begin{aligned} \mathcal{E}\{\text{MSE}\} &= \mathcal{E}\left\{\frac{1}{N}\|\mathbf{f}(\mathbf{y}) - \mathbf{x}\|^2\right\} \\ &= \frac{1}{N}\sum_{n=1}^N\left(\mathcal{E}\{f_n(\mathbf{y})^2\} - 2\mathcal{E}\{x_n f_n(\mathbf{y})\} + x_n^2\right). \end{aligned} \quad (4)$$

The expectation in (4) is with respect to the data \mathbf{y} . The vector \mathbf{x} of noncentrality parameters may either be considered as deterministic and unknown, or as random and independent of \mathbf{y} .

In practice, for any given realization of the data \mathbf{y} , the true MSE cannot be computed because \mathbf{x} is unknown. Yet, paralleling the general case of distributions in the exponential family [29], it is possible to establish a lemma that allows us to circumvent this issue, and *estimate* the risk without the knowledge of \mathbf{x} . The main technical requirement is that each component $f_n(\mathbf{y})$ of the estimator $\mathbf{f}: \mathbb{R}^N \rightarrow \mathbb{R}^N$ be continuously differentiable with respect to y_n , and so we introduce the following notation: $\partial\mathbf{f}(\mathbf{y}) = [(\partial/\partial y_n)f_n(\mathbf{y})]_{1 \leq n \leq N}$, $\partial^2\mathbf{f}(\mathbf{y}) = [(\partial^2/\partial y_n^2)f_n(\mathbf{y})]_{1 \leq n \leq N}$.

Lemma 1: Assume that the estimator $\mathbf{f}(\mathbf{y}) = [f_n(\mathbf{y})]_{1 \leq n \leq N}$ is such that each $f_n(\mathbf{y})$ is continuously differentiable with respect to y_n , with weakly differentiable partial derivatives $\partial f_n(\mathbf{y})/\partial y_n$ that do not increase “too quickly” for large values of \mathbf{y} , i.e., there exists a constant $s < 1/2$ such that for every $1 \leq n \leq N$, $\lim_{y_n \rightarrow +\infty} e^{-s y_n} |(\partial/\partial y_n)f_n(\mathbf{y})| = 0$. Then

$$\begin{aligned} \mathcal{E}\{\mathbf{x}^T \mathbf{f}(\mathbf{y})\} &= \mathcal{E}\{(\mathbf{y} - K \cdot \mathbf{1})^T \mathbf{f}(\mathbf{y})\} \\ &\quad - 4\mathcal{E}\left\{\left(\mathbf{y} - \frac{K}{2} \cdot \mathbf{1}\right)^T \partial\mathbf{f}(\mathbf{y})\right\} + 4\mathcal{E}\left\{\mathbf{y}^T \partial^2\mathbf{f}(\mathbf{y})\right\}. \end{aligned} \quad (5)$$

Proof: We first evaluate the scalar expectation $\mathcal{E}\{x_n f_n(\mathbf{y})\}$ appearing in (4), before summing up the contributions over n to get the expectation of the inner product $\mathbf{x}^T \mathbf{f}(\mathbf{y})$.

Let us consider the characteristic function given by (2). By differentiating the logarithm of this Fourier transform with respect to the variable ω_n , we find that

$$\frac{1}{\hat{p}(\omega|\mathbf{x})} \frac{\partial \hat{p}(\omega|\mathbf{x})}{\partial \omega_n} = -\frac{jK}{1+2j\omega_n} - \frac{jx_n}{(1+2j\omega_n)^2}.$$

After a rearrangement of the different terms involved, we get

$$x_n \hat{p}(\omega|\mathbf{x}) = j(1+2j\omega_n)^2 \frac{\partial \hat{p}(\omega|\mathbf{x})}{\partial \omega_n} - K(1+2j\omega_n) \hat{p}(\omega|\mathbf{x}).$$

Using $(1+2j\omega_n)^2 = 1+4j\omega_n+4(j\omega_n)^2$ and recalling that multiplication by $j\omega_n$ corresponds to differentiating with respect to y_n , while differentiation with respect to ω_n is equivalent to a multiplication with $-jy_n$, we can deduce that the probability density $p(\mathbf{y}|\mathbf{x})$ satisfies (in the sense of

distributions) the following linear differential equation:

$$\begin{aligned} x_n p(\mathbf{y}|\mathbf{x}) &= \left(1 + 4\frac{\partial}{\partial y_n} + 4\frac{\partial^2}{\partial y_n^2}\right)\{y_n p(\mathbf{y}|\mathbf{x})\} \\ &\quad - K\left(1 + 2\frac{\partial}{\partial y_n}\right)\{p(\mathbf{y}|\mathbf{x})\}. \end{aligned}$$

The expectation $\mathcal{E}\{x_n f_n(\mathbf{y})\}$ is simply the Euclidean inner product (with respect to Lebesgue measure) between $f_n(\mathbf{y})$ and $x_n p(\mathbf{y}|\mathbf{x})$. Continuous differentiability of $\mathbf{f}(\mathbf{y})$ coupled with weak differentiability of $\partial\mathbf{f}(\mathbf{y})$ implies that we may twice apply integration by parts to obtain

$$\begin{aligned} \mathcal{E}\{x_n f_n(\mathbf{y})\} &= \langle f_n(\mathbf{y}), x_n p(\mathbf{y}|\mathbf{x}) \rangle \\ &= \langle f_n(\mathbf{y}), \left(1 + 4\frac{\partial}{\partial y_n} + 4\frac{\partial^2}{\partial y_n^2}\right)\{y_n p(\mathbf{y}|\mathbf{x})\} \rangle \\ &\quad - K\langle f_n(\mathbf{y}), \left(1 + 2\frac{\partial}{\partial y_n}\right)\{p(\mathbf{y}|\mathbf{x})\} \rangle \\ &= \langle y_n \left(1 - 4\frac{\partial}{\partial y_n} + 4\frac{\partial^2}{\partial y_n^2}\right)\{f_n(\mathbf{y})\}, p(\mathbf{y}|\mathbf{x}) \rangle \\ &\quad - K\langle \left(1 - 2\frac{\partial}{\partial y_n}\right)\{f_n(\mathbf{y})\}, p(\mathbf{y}|\mathbf{x}) \rangle \\ &= \mathcal{E}\left\{\left(y_n - K\right)f_n(\mathbf{y}) - 4\left(y_n - \frac{K}{2}\right)\frac{\partial f_n(\mathbf{y})}{\partial y_n} \right. \\ &\quad \left. + 4y_n \frac{\partial^2 f_n(\mathbf{y})}{\partial y_n^2}\right\} \end{aligned}$$

plus additional integrated terms that do not depend on x_n , which vanish because of our (conservative) assumption on how $f_n(\mathbf{y})$ increases when $y_n \rightarrow +\infty$. (Asymptotic expansions of Bessel functions can be used to show that whenever $s < 1/2$, we have that $e^{s y_n} p(\mathbf{y}|\mathbf{x}) \rightarrow 0$ and $e^{s y_n} \partial p(\mathbf{y}|\mathbf{x})/\partial y_n \rightarrow 0$.) Finally, summing up over the index n yields (5). ■

We can then deduce a theorem that provides an unbiased estimate of the expected MSE given in (4). We term this random variable CURE.

Theorem 1: Let $\mathbf{y} \sim \chi_K^2(\mathbf{x})$ and assume that $\mathbf{f}(\mathbf{y})$ satisfies the regularity conditions of Lemma 1. Then, the random variable

$$\begin{aligned} \text{CURE} &= \frac{1}{N}\left(\|\mathbf{f}(\mathbf{y}) - (\mathbf{y} - K \cdot \mathbf{1})\|^2 - 4^T \left(\mathbf{y} - \frac{K}{2} \cdot \mathbf{1}\right)\right) \\ &\quad + \frac{8}{N}\left(\left(\mathbf{y} - \frac{K}{2} \cdot \mathbf{1}\right)^T \partial\mathbf{f}(\mathbf{y}) - \mathbf{y}^T \partial^2\mathbf{f}(\mathbf{y})\right) \end{aligned} \quad (6)$$

is an unbiased estimate of the risk, i.e., $\mathcal{E}\{\text{CURE}\} = \mathcal{E}\{\text{MSE}\}$.

Proof: As with other unbiased risk estimates, we express the MSE as a sum of three terms

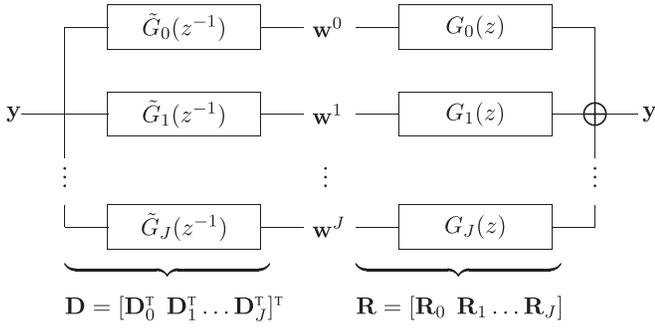
$$\|\mathbf{f}(\mathbf{y}) - \mathbf{x}\|^2 = \underbrace{\|\mathbf{f}(\mathbf{y})\|^2}_{\text{term 1}} - \underbrace{2\mathbf{x}^T \mathbf{f}(\mathbf{y})}_{\text{term 2}} + \underbrace{\|\mathbf{x}\|^2}_{\text{term 3}}$$

which we replace by a statistical equivalent *that does not depend on \mathbf{x} anymore*.

Term 1 needs no change, term 2 can be replaced according to (5) from Lemma 1, and term 3 can be reformulated using the noncentral chi-square moments of (3) to yield

$$\|\mathbf{x}\|^2 = \mathcal{E}\left\{\|\mathbf{y}\|^2\right\} - 2(K+2)\mathbf{1}^T \mathcal{E}\{\mathbf{y}\} + NK(K+2).$$

Putting everything together then leads directly to (6), and thus the theorem is proved. ■


 Fig. 1. Undecimated $(J + 1)$ -band analysis/synthesis filterbank.

In Sections III and IV, we propose specific examples of CURE-optimized transform-domain processing for estimating the unknown noncentrality parameter vector \mathbf{x} from data \mathbf{y} .

III. CURE-OPTIMIZED DENOISING IN UNDECIMATED FILTERBANK TRANSFORMS

In this section, we focus on processing within a $(J + 1)$ -band undecimated filterbank transform, as depicted in Fig. 1. This broad class of redundant representations notably includes the undecimated wavelet transform (UWT) and overlapping block discrete cosine transform.

A. Image-Domain CURE for Transform-Domain Processing

Nonlinear processing via undecimated filterbank transforms is a general denoising strategy long proven to be effective for reducing various types of noise degradations [27], [30]–[34]. It essentially boils down to performing a linear (and possibly redundant) analysis transformation of the data, which provides empirical coefficients which are then thresholded (possibly using a multivariate nonlinear function), the result of which is finally passed to a linear synthesis transformation. When treating signal-dependent noise, the entire denoising procedure is conveniently expressed in the generic form $\mathbf{f}(\mathbf{y}) = \mathbf{R}\boldsymbol{\theta}(\mathbf{D}\mathbf{y}, \bar{\mathbf{D}}\mathbf{y})$ [34].

- 1) The circulant matrices $\mathbf{D} = [\mathbf{D}_0^T \ \mathbf{D}_1^T \ \dots \ \mathbf{D}_J^T]^T$ and $\mathbf{R} = [\mathbf{R}_0 \ \mathbf{R}_1 \ \dots \ \mathbf{R}_J]$ implement an arbitrary pair of analysis/synthesis undecimated filterbanks (Fig. 1) such that the perfect reconstruction condition $\mathbf{R}\mathbf{D} = \mathbf{I}_d$ is satisfied. Each component of the $N \times N$ circulant submatrix $\mathbf{D}_j = [d_{k,l}^j]_{1 \leq k,l \leq N}$ and $\mathbf{R}_j = [r_{k,l}^j]_{1 \leq k,l \leq N}$ is given by

$$\begin{cases} d_{k,l}^j = \sum_{n \in \mathbb{Z}} \tilde{g}_j[l - k + nN] \\ r_{k,l}^j = \sum_{n \in \mathbb{Z}} g_j[k - l + nN] \end{cases}, \text{ for } j = 0, \dots, J. \quad (7)$$

We assume that the considered analysis filters have unit norms, i.e., $\sum_n (d_{l,n}^j)^2 = 1$, for $l = 1, \dots, N$, $j = 0, \dots, J$. By convention, we also assume that, for $j = 1, \dots, J$, each \mathbf{D}_j implements a high-pass channel, i.e., $\forall l, \sum_n d_{l,n}^j = 0$. The complementary low-pass channel is then implemented by \mathbf{D}_0 with $\forall l, \sum_n d_{l,n}^0 = 2^{J/2}$.

Hence, denoting by $\mathbf{w}_j = \mathbf{D}_j \mathbf{y} = [w_l^j]_{1 \leq l \leq N}$ (resp. $\boldsymbol{\omega}_j = \mathbf{D}_j \mathbf{x} = [\omega_l^j]_{1 \leq l \leq N}$) each vector of noisy (resp. noise-free) transform coefficients, we have

$$\forall l = 1 \dots N, \begin{cases} \mathcal{E} \{w_l^j\} = \omega_l^j, \text{ for } j = 1 \dots J \\ \mathcal{E} \{w_l^0\} = \omega_l^0 + 2^{J/2} K. \end{cases} \quad (8)$$

While the noisy high-pass coefficients are unbiased estimates of their noise-free counterparts, the low-pass coefficients exhibit a constant bias $(2^{J/2} K)$ that must be removed.

- 2) The circulant matrix $\bar{\mathbf{D}} = [\bar{\mathbf{D}}_0^T \ \bar{\mathbf{D}}_1^T \ \dots \ \bar{\mathbf{D}}_J^T]^T$ implements a linear estimation of the variance $\bar{\mathbf{w}}_j = \bar{\mathbf{D}}_j \mathbf{y} = [\bar{w}_l^j]_{1 \leq l \leq N}$ of each transform coefficient w_l^j . The actual variance is given by

$$\begin{aligned} \text{var} \{w_l^j\} &= \sum_{n=1}^N (d_{l,n}^j)^2 \text{var} \{y_n\} \\ &\stackrel{(3)}{=} 4 \sum_{n=1}^N (d_{l,n}^j)^2 \left(x_n + \frac{K}{2} \right). \end{aligned} \quad (9)$$

Since $\mathcal{E} \{y_n\} = x_n + K$, the natural choice $\bar{\mathbf{D}}_j = [(d_{l,n}^j)^2]_{1 \leq l,n \leq N}$ achieves

$$\text{var} \{w_l^j\} = 4 \left(\mathcal{E} \{\bar{w}_l^j\} - \frac{K}{2} \right) \quad (10)$$

- 3) the vector function $\boldsymbol{\theta} : \mathbb{R}^L \times \mathbb{R}^L \rightarrow \mathbb{R}^L$, where $L = (J + 1)N$, can generally be arbitrary, from a simple pointwise thresholding rule to more sophisticated multivariate processing. In this section, we will however consider only subband-adaptive *pointwise* processing

$$\boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) = \left[\theta_l^j \left(w_l^j, \bar{w}_l^j \right) \right]_{0 \leq j \leq J, 1 \leq l \leq N}. \quad (11)$$

We further assume that the transform-domain pointwise processing of (11) is (at least) continuously differentiable, with piecewise-differentiable partial derivatives. Introducing the notation

$$\begin{aligned} \partial_1 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) &= \left[\frac{\partial \theta_l(\mathbf{w}, \bar{\mathbf{w}})}{\partial w_l} \right]_{1 \leq l \leq L} \\ \partial_2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) &= \left[\frac{\partial \theta_l(\mathbf{w}, \bar{\mathbf{w}})}{\partial \bar{w}_l} \right]_{1 \leq l \leq L} \\ \partial_{11}^2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) &= \left[\frac{\partial^2 \theta_l(\mathbf{w}, \bar{\mathbf{w}})}{\partial w_l^2} \right]_{1 \leq l \leq L} \\ \partial_{22}^2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) &= \left[\frac{\partial^2 \theta_l(\mathbf{w}, \bar{\mathbf{w}})}{\partial \bar{w}_l^2} \right]_{1 \leq l \leq L} \\ \partial_{12}^2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) &= \left[\frac{\partial^2 \theta_l(\mathbf{w}, \bar{\mathbf{w}})}{\partial w_l \partial \bar{w}_l} \right]_{1 \leq l \leq L} \end{aligned}$$

and denoting by “ \odot ” the Hadamard (elementwise) matrix product, we have the following.

Corollary 1: For pointwise processing of the form given by (11) and satisfying the requirements of Lemma 1, the risk

estimate of (6) takes the form

$$\begin{aligned}
\text{CURE} = & \frac{1}{N} \left(\|\mathbf{f}(\mathbf{y}) - (\mathbf{y} - K \cdot \mathbf{1})\|^2 - 4^T \left(\mathbf{y} - \frac{K}{2} \cdot \mathbf{1} \right) \right) \\
& + \frac{8}{N} \left(\mathbf{y} - \frac{K}{2} \cdot \mathbf{1} \right)^T \left((\mathbf{R} \odot \mathbf{D}^T) \partial_1 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) \right. \\
& \quad \left. + (\mathbf{R} \odot \bar{\mathbf{D}}^T) \partial_2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) \right) \\
& - \frac{8}{N} \mathbf{y}^T \left((\mathbf{R} \odot \mathbf{D}^T \odot \mathbf{D}^T) \partial_{11}^2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) \right. \\
& \quad \left. + (\mathbf{R} \odot \bar{\mathbf{D}}^T \odot \bar{\mathbf{D}}^T) \partial_{22}^2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}) \right) \\
& + \frac{16}{N} \mathbf{y}^T (\mathbf{R} \odot \mathbf{D}^T \odot \bar{\mathbf{D}}^T) \partial_{12}^2 \boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}}). \tag{12}
\end{aligned}$$

The proof of this result is straightforwardly obtained by developing the term $(\mathbf{y} - K/2 \cdot \mathbf{1})^T \partial \mathbf{f}(\mathbf{y}) - \mathbf{y}^T \partial^2 \mathbf{f}(\mathbf{y})$ from (6) for $\boldsymbol{\theta}(\mathbf{w}, \bar{\mathbf{w}})$ as defined in (11). A similar result for transform-domain denoising of mixed Poisson–Gaussian data is proved in [34].

For the remainder of this section, we drop the subband superscript j and the in-band location index l . We thus denote by w , \bar{w} , and ω any of the w_l^j , \bar{w}_l^j , and ω_l^j , for $j = 1, \dots, J$.

B. Choice of Thresholding Rule

We need to specify a particular shrinkage or thresholding rule to estimate each unknown high-pass coefficient ω from its noisy counterpart w . In the minimum-MSE sense, the optimal pointwise shrinkage factor is given by

$$\alpha^* = \arg \min_{\alpha} \mathcal{E} \left\{ (\alpha w - \omega)^2 \right\} \stackrel{(8),(10)}{=} 1 - \frac{4 \left(\mathcal{E} \{ \bar{w} \} - \frac{K}{2} \right)}{\mathcal{E} \{ w^2 \}}. \tag{13}$$

There are various possible implementations of the above formula to yield an effective shrinkage function. For the case of $K = 2$ degrees of freedom, Nowak proposed in [6] the function

$$\theta(w, \bar{w}) = \max \left(1 - \lambda \frac{4 \max(\bar{w} - 1, 1)}{w^2}, 0 \right) w \tag{14}$$

where the particular choice $\lambda = 3$ was motivated by a Gaussian prior on the noisy coefficients w . Our experiments have indicated that replacing $\max(\bar{w} - 1, 1)$ by \bar{w} gives slightly better MSE performance, and so, following the recent idea of LET [27], we propose the following shrinkage rule for arbitrary degrees of freedom:

$$\theta(w, \bar{w}; \mathbf{a}) = \sum_{i=1}^I a_i \max \left(1 - \lambda_i \frac{4 \bar{w}}{w^2}, 0 \right) w, \text{ with } I \ll N \tag{15}$$

$\theta_i(w, \bar{w})$

which can be seen as an optimized generalization of (14). To satisfy the requirements of Corollary 1, we implement a continuously differentiable approximation of the $\max(\cdot)$ function.

Empirically, we have observed $I = 2$ terms per subband to be the best choice in (15). The vector $\mathbf{a} \in \mathbb{R}^I$ of subband-adaptive parameters can be optimized in closed form via least

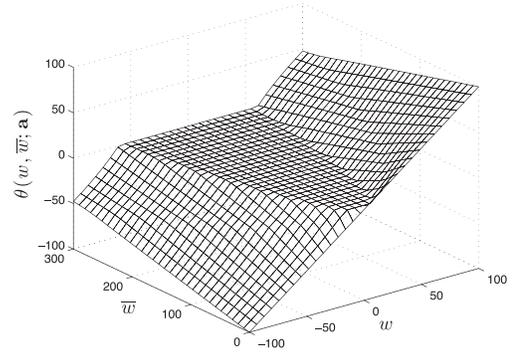


Fig. 2. Possible realization of the proposed thresholding rule of (15) ($I = 2$, $\lambda_1 = 3$, $\lambda_2 = 9$, $\mathbf{a} = [0.75 \ 0.25]^T$).

squares, while λ_1 and λ_2 can be optimized by minimizing the risk estimate of (12) directly. However, fixing $\lambda_1 = 3$ and $\lambda_2 = 9$ was observed to work well in all of our experiments, and leads to a much faster implementation. (We observed values close to (3, 9) to yield equivalent results ± 0.2 dB.) A potential realization of the proposed LET is displayed in Fig. 2.

C. Implementation

The overall subband-adaptive transform-domain estimator is thus, for $\mathcal{I} = \{0, \dots, J\} \times \{1, 2\}$

$$\begin{aligned}
\mathbf{f}(\mathbf{y}) = & \sum_{i \in \mathcal{I}} a_i \mathbf{f}_i(\mathbf{y}) = a_1^0 \underbrace{(\mathbf{R}_0 \mathbf{w}_0 - K)}_{\text{Lowpass bias removed}} \\
& + \sum_{j=1}^J \sum_{i=1}^2 a_i^j \mathbf{R}_j \boldsymbol{\theta}_i^j(\mathbf{w}_j, \bar{\mathbf{w}}_j)
\end{aligned}$$

with the CURE-optimized parameters $\mathbf{a} = [a_i]_{i \in \mathcal{I}}$ the solution to $\mathbf{M} \mathbf{a} = \mathbf{c}$, where

$$\begin{cases} \mathbf{c} = \left[(\mathbf{y} - K \cdot \mathbf{1})^T \mathbf{f}_i(\mathbf{y}) - 4 \left(\left(\mathbf{y} - \frac{K}{2} \cdot \mathbf{1} \right)^T \partial \mathbf{f}_i(\mathbf{y}) \right. \right. \\ \quad \left. \left. - \mathbf{y}^T \partial^2 \mathbf{f}_i(\mathbf{y}) \right) \right]_{i \in \mathcal{I}}, \\ \mathbf{M} = [\mathbf{f}_i(\mathbf{y})^T \mathbf{f}_j(\mathbf{y})]_{i, j \in \mathcal{I}}. \end{cases} \tag{16}$$

IV. CURE-OPTIMIZED DENOISING VIA UNNORMALIZED HAAR WAVELET TRANSFORM

In the previous section, we have considered the general case of an undecimated filterbank transform and derived the corresponding *image-domain* MSE estimate. Owing to the intractability of the noncentral chi-square distribution after an arbitrary (even orthogonal) transformation, an explicit *transform-domain* risk estimate is generally unobtainable. Remarkably, in the particular case of the *unnormalized* Haar wavelet transform, the derivation of such an explicit subband-dependent MSE estimate is possible. Its construction is presented in this section.

A. Unnormalized Haar Wavelet-Domain CURE

The 1-D unnormalized Haar discrete wavelet transform consists of a critically sampled two-channel filterbank

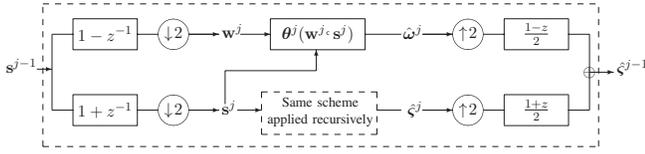


Fig. 3. Signal-dependent noise reduction in the unnormalized Haar discrete wavelet transform.

(see Fig. 3.) On the analysis side, the low-pass (resp. high-pass) channel is implemented by the unnormalized Haar scaling (resp. wavelet) filter whose z -transform is $\tilde{H}(z^{-1}) = 1 + z^{-1}$ (resp. $\tilde{G}(z^{-1}) = 1 - z^{-1}$). To achieve a perfect reconstruction, the synthesis side is implemented by the low-pass and high-pass filters $H(z) = (1 + z)/2$ and $G(z) = (1 - z)/2$. At a given scale j , the unnormalized Haar scaling and wavelet coefficients of the observed data $\mathbf{y} = \mathbf{s}^0$ are given by

$$s_n^j = s_{2n}^{j-1} + s_{2n-1}^{j-1}, \quad w_n^j = s_{2n}^{j-1} - s_{2n-1}^{j-1}. \quad (17)$$

Similarly, the unnormalized Haar scaling and wavelet coefficients of the noncentrality parameter vector of interest $\mathbf{x} = \boldsymbol{\zeta}^0$ are given by

$$\varsigma_n^j = \varsigma_{2n}^{j-1} + \varsigma_{2n-1}^{j-1}, \quad \omega_n^j = \varsigma_{2n}^{j-1} - \varsigma_{2n-1}^{j-1}. \quad (18)$$

Since the sum of independent noncentral chi-square random variables is a noncentral chi-square random variable whose noncentrality parameter and number of degrees of freedom are the summed noncentrality parameters and number of degrees of freedom [35], the empirical scaling coefficients follow a noncentral chi-square distribution, i.e., $\mathbf{s}^j \sim \chi_{K_j}^2(\boldsymbol{\zeta}^j)$, where $K_j = 2^j K$. Moreover, since the squared low-pass filter coefficients are the low-pass filter coefficients themselves, the scaling coefficients can be used as estimates of the variance of the (same-scale) wavelet coefficients. In the notation of Section III-A, this means that $\bar{\mathbf{w}}^j = \mathbf{s}^j$ for $j = 1, \dots, J$.

Denoting by N_j the number of samples at a given scale j and assuming a subband-adaptive processing $\boldsymbol{\theta}^j: \mathbb{R}^{N_j} \times \mathbb{R}^{N_j} \rightarrow \mathbb{R}^{N_j}$, the MSE in each high-pass subband j is given by $\text{MSE}_j = (1/N_j) \|\boldsymbol{\theta}^j(\mathbf{w}^j, \mathbf{s}^j) - \boldsymbol{\omega}^j\|^2$, and we have the following theorem.

Theorem 2: Let $\boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) = \boldsymbol{\theta}^j(\mathbf{w}^j, \mathbf{s}^j)$ be an estimator of the unnormalized Haar wavelet coefficients $\boldsymbol{\omega} = \boldsymbol{\omega}^j$ of \mathbf{x} at scale j , satisfying the conditions of Lemma 1. Then the random variable

$$\begin{aligned} \text{CURE}_j &= \frac{1}{N_j} \left(\|\boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) - \mathbf{w}\|^2 - 4^{\text{T}} \left(\mathbf{s} - \frac{K_j}{2} \cdot \mathbf{1} \right) \right) + \frac{8}{N_j} \\ &\quad \times \left(\left(\mathbf{s} - \frac{K_j}{2} \cdot \mathbf{1} \right)^{\text{T}} \partial_1 \boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) + \mathbf{w}^{\text{T}} \partial_2 \boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) \right) - \frac{8}{N_j} \\ &\quad \times \left(\mathbf{w}^{\text{T}} \left(\partial_{11}^2 \boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) + \partial_{22}^2 \boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) \right) + 2\mathbf{s}^{\text{T}} \partial_{12}^2 \boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) \right) \end{aligned} \quad (19)$$

is an unbiased estimate of the risk for subband j , i.e., $\mathcal{E}\{\text{CURE}_j\} = \mathcal{E}\{\text{MSE}_j\}$.

Proof: We consider the case $j = 1$, so that we may use $K = K_j/2$, $\mathbf{y} = \mathbf{s}^{j-1}$, and $\mathbf{x} = \boldsymbol{\zeta}^{j-1}$ to ease notation. We first

develop the squared error between $\boldsymbol{\omega}$ and its estimate $\boldsymbol{\theta}(\mathbf{w}, \mathbf{s})$

$$\begin{aligned} \mathcal{E} \left\{ \|\boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) - \boldsymbol{\omega}\|^2 \right\} &= \mathcal{E} \left\{ \|\boldsymbol{\theta}(\mathbf{w}, \mathbf{s})\|^2 \right\} - 2 \underbrace{\mathcal{E} \left\{ \boldsymbol{\omega}^{\text{T}} \boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) \right\}}_{\text{(I)}} \\ &\quad + \underbrace{\|\boldsymbol{\omega}\|^2}_{\text{(II)}}. \end{aligned} \quad (20)$$

We can then evaluate the two expressions (I, II) that involve the unknown $\boldsymbol{\omega}$.

1) Computation of $\mathcal{E} \left\{ \boldsymbol{\omega}^{\text{T}} \boldsymbol{\theta}(\mathbf{w}, \mathbf{s}) \right\} = \sum_{n=1}^{N_j} \mathcal{E} \left\{ \omega_n \theta_n \right\}(\mathbf{w}, \mathbf{s})$. We can successively write

$$\begin{aligned} \mathcal{E} \left\{ \omega_n \theta_n(\mathbf{w}, \mathbf{s}) \right\} &\stackrel{(18)}{=} \mathcal{E} \left\{ x_{2n} \theta_n(\mathbf{w}, \mathbf{s}) \right\} - \mathcal{E} \left\{ x_{2n-1} \theta_n(\mathbf{w}, \mathbf{s}) \right\} \\ &\stackrel{(5);(17)}{=} \mathcal{E} \left\{ w_n \left(\theta_n(\mathbf{w}, \mathbf{s}) + 4 \left(\partial_{11}^2 \theta_n(\mathbf{w}, \mathbf{s}) \right. \right. \right. \\ &\quad \left. \left. \left. + \partial_{22}^2 \theta_n(\mathbf{w}, \mathbf{s}) - \partial_2 \theta_n(\mathbf{w}, \mathbf{s}) \right) \right) \right\} \\ &\quad - 4 \mathcal{E} \left\{ (s_n - K) \partial_1 \theta_n(\mathbf{w}, \mathbf{s}) \right. \\ &\quad \left. - 2s_n \partial_{12}^2 \theta_n(\mathbf{w}, \mathbf{s}) \right\}. \end{aligned} \quad (21)$$

2) Computation of $\|\boldsymbol{\omega}\|^2 = \sum_{n=1}^{N_j} \omega_n^2$. We can successively write

$$\omega_n^2 = \mathcal{E} \left\{ \omega_n w_n \right\} \stackrel{(21)}{=} \mathcal{E} \left\{ w_n^2 - 4(s_n - K) \right\}. \quad (22)$$

Inserting (21) and (22) into (20) yields the desired equality, for $j > 1$, the proof is similar. ■

Subband superscript j will be omitted below, as we consider any of the J wavelet subbands.

B. CUREshrink

A natural choice of subband-adaptive estimator in orthogonal wavelet representations is *soft thresholding*, introduced by Weaver *et al.* [21] and theoretically justified by Donoho [36]. In contrast to the AWGN scenario, a signal-dependent threshold is required here. As in the case of Poisson noise removal [37], [38], we wish to adapt the original ‘‘uniform’’ soft thresholding as

$$\theta_n(\mathbf{w}, \mathbf{s}; a) = \text{sign}(w_n) \max(|w_n| - a\sqrt{s_n}, 0). \quad (23)$$

In *SUREshrink* and *PUREshrink* [37]–[39] for Gaussian (resp. Poisson) noise reduction, a is set to the value that minimizes the corresponding unbiased risk estimate. Similarly, we may select a to yield the minimum CURE value according to (19) on the basis of observed data \mathbf{y} , resulting in a *CUREshrink* denoising procedure. To comply with the requirements of Theorem 2, we use a continuously differentiable approximation to soft thresholding. Fig. 4 shows the empirical accuracy of CURE as a practical criterion for choosing the best value of a , we have also observed a pointwise LET approach, as in (15), to yield comparable denoising results.

C. Joint Inter-/Intra-Scale CURE-LET

To decrease the usual ringing artifacts inherent to orthogonal transform-domain thresholding, more sophisticated denoising functions must be considered. In particular, the integration of inter-scale dependencies between wavelet coefficients

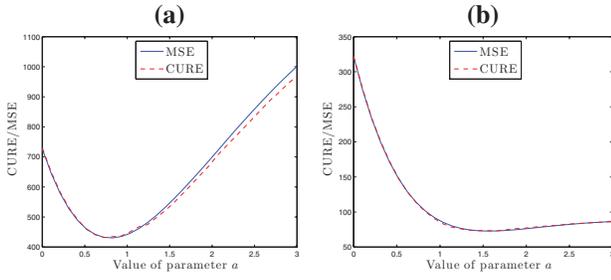


Fig. 4. Minimum CURE versus MMSE threshold selection for the CUREshrink adapted soft thresholding of (23). (a) $N = 128 \times 128$ samples, $\mathbf{s} \sim \chi^2_8(\mathbf{s})$, input SNR = 15 dB. (b) $N = 256 \times 256$, $\mathbf{s} \sim \chi^2_{16}(\mathbf{s})$, input SNR = 10 dB.

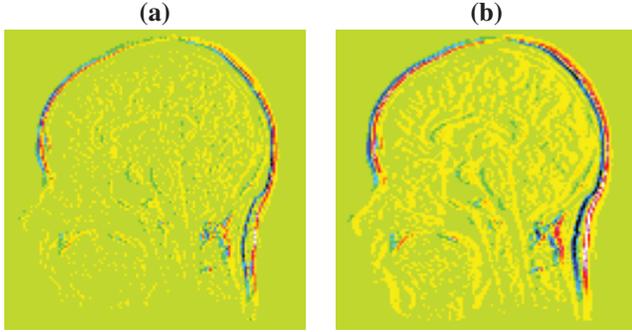


Fig. 5. (a) Child and (b) its group-delay compensated parent in a particular subband at the first scale of a 2-D unnormalized Haar wavelet transform. In this 16-color map, the background yellowish value is zero, with the most significant coefficients appearing either in blue (negative) or in red (positive).

(the so-called “parent–child” relationship) has already been shown to significantly increase the denoising quality in both AWGN reduction [26], [40]–[42] and Poisson intensity estimation [38].

To this end, Fig. 5 shows an example of a group-delay-compensated parent \mathbf{p} and its child \mathbf{w} for a particular subband at the first scale of a 2-D unnormalized Haar wavelet transform. For the unnormalized Haar wavelet transform, the group-delay compensated parent $\mathbf{p} = [p_n]_{1 \leq n \leq N}$ is simply given by $p_n = s_{n+1} - s_{n-1}$ [38]. From Fig. 5, we observe that both the signs and the locations of the significant (i.e., large-magnitude) coefficients persist across scale. To take advantage of this persistence, we propose the following LET approach, inspired by (15):

$$\begin{aligned} \theta_n(\mathbf{w}, \mathbf{s}; \mathbf{a}) &= \sum_{k=1}^2 a_k \max\left(1 - \lambda_k \frac{4\gamma_n(\mathbf{s})}{\gamma_n^2(\mathbf{w})}, 0\right) w_n \\ &+ \sum_{k=1}^2 a_{k+2} \max\left(1 - \lambda_k \frac{4\gamma_n(\mathbf{s})}{\gamma_n^2(\mathbf{p})}, 0\right) w_n \\ &+ \sum_{k=1}^2 a_{k+4} \max\left(1 - \lambda_k \frac{4\gamma_n(\mathbf{s})}{\gamma_n^2(\mathbf{w})}, 0\right) p_n \\ &+ \sum_{k=1}^2 a_{k+6} \max\left(1 - \lambda_k \frac{4\gamma_n(\mathbf{s})}{\gamma_n^2(\mathbf{p})}, 0\right) p_n. \end{aligned} \quad (24)$$

Here the function $\gamma_n(\mathbf{u}) = 1/\sqrt{2\pi} \sum_k |u_k| e^{-(n-k)^2/2}$ implements a normalized Gaussian smoothing of the magnitude

of its argument. This local filtering accounts for similarities between neighboring wavelet, scaling, and parent coefficients.

The proposed denoising function of (24) thus integrates both the inter- and intra-scale dependencies that naturally arise in the Haar wavelet transform. It involves a set \mathbf{a} of eight parameters that can be optimized via least squares, as well as parameters λ_1 and λ_2 that can be fixed in advance without noticeable loss in denoising quality, we use $\lambda_1 = 1$ and $\lambda_2 = 9$ in all experiments below. Considering the processing of all wavelet coefficients in a given subband j , (24) reads as $\boldsymbol{\theta}(\mathbf{w}, \mathbf{s}; \mathbf{a}) = \sum_{k=1}^8 a_k \boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s})$. The optimal (in the minimum CURE sense) set of linear parameters is then the solution of the linear system of equation $\mathbf{M}\mathbf{a} = \mathbf{c}$, where

$$\begin{cases} \mathbf{c} = \left[\mathbf{w}^T \boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s}) - 4 \left(\mathbf{s} - \frac{K_j}{2} \cdot \mathbf{1} \right)^T \partial_1 \boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s}) \right. \\ \quad \left. + 8\mathbf{s}^T \partial_{12}^2 \boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s}) + 4\mathbf{w}^T \right. \\ \quad \left. \times \left(\partial_{11}^2 \boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s}) + \partial_{22}^2 \boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s}) - \partial_2 \boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s}) \right) \right]_{1 \leq k \leq 8} \\ \mathbf{M} = [\boldsymbol{\theta}_k(\mathbf{w}, \mathbf{s})^T \boldsymbol{\theta}_l(\mathbf{w}, \mathbf{s})]_{1 \leq k, l \leq 8}. \end{cases}$$

Finally, the bias is removed from the low-pass residual subband at scale J as $\hat{\mathbf{z}}^J = \mathbf{s}^J - 2^J K \cdot \mathbf{1}$.

Although all the results presented in Sections II–IV have been developed for 1-D signals, they can be straightforwardly applied to d -dimensional signals as well, when considering separable transforms. Note that in the d -dimensional unnormalized Haar wavelet transform, the number of degrees of freedom of the noncentral- χ^2 -distributed low-pass subband at scale j is given by $K_j = 2^{d \cdot j} K$.

V. APPLICATION TO MAGNITUDE MR IMAGE DENOISING

In magnitude MR imaging, the observed image consists of the magnitudes $|m_n|$ of N independent complex measurements m_n , where

$$\begin{cases} \Re\{m_n\} \sim \mathcal{N}(\Re\{\mu_n\}, \sigma^2) \\ \Im\{m_n\} \sim \mathcal{N}(\Im\{\mu_n\}, \sigma^2). \end{cases} \quad (25)$$

Our objective is to estimate the original (unknown) magnitudes $|\mu_n| = \sqrt{\Re\{\mu_n\}^2 + \Im\{\mu_n\}^2}$ from their noisy observations $|m_n|$. If we define two N -dimensional vectors

$$\begin{cases} \mathbf{x} = [|\mu_n|^2/\sigma^2]_{1 \leq n \leq N} \in \mathbb{R}_+^N \\ \mathbf{y} = [|m_n|^2/\sigma^2]_{1 \leq n \leq N} \in \mathbb{R}_+^N \end{cases} \quad (26)$$

then the data likelihood for \mathbf{y} is the product of N independent noncentral χ^2 distributions with $K = 2$ degrees of freedom and noncentrality parameter x_n , i.e., (1) with $K = 2$.

We then denoise the magnitude MR image \mathbf{m} according to the following steps.

- 1) If necessary, estimate the noise variance σ^2 using known techniques.
- 2) Apply the nonlinear mapping defined in (26) to the magnitude image \mathbf{m} to obtain the normalized squared-magnitude image \mathbf{y} .
- 3) Apply a CURE-optimized denoising algorithm to obtain an estimate $\hat{\mathbf{x}} = \mathbf{f}(\mathbf{y})$ of \mathbf{x} .
- 4) Produce the final estimate $\hat{\boldsymbol{\mu}}$ of the unknown magnitude MR image $\boldsymbol{\mu}$, by inverting the nonlinear mapping specified in (26).

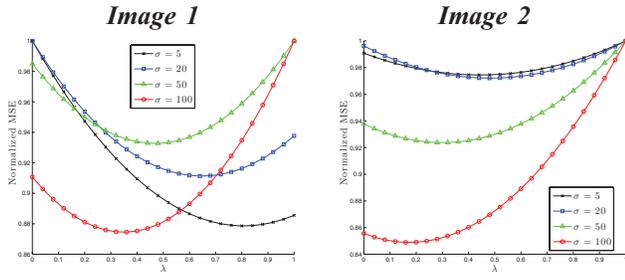


Fig. 6. Sensitivity to the choice of λ in (27). The ordinate shows the normalized MSE relative to the reference case $\lambda = 0$.

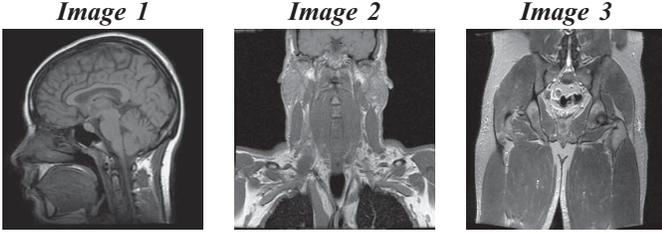


Fig. 7. Test set of high-quality 256×256 magnitude MR images used in the experiments of Section V.

Note that the last step of the above procedure has a direct impact on the final image domain MSE. Ideally, we would like to use the nonlinear mapping \mathcal{T} that minimizes $\mathcal{E} \{ \|\mathcal{T}(\hat{\mathbf{x}}) - \boldsymbol{\mu}\|^2 \}$. Yet, such a nontrivial MMSE design would lead us beyond the scope of this paper. Instead, we propose to use the algebraic inverse of (26), i.e., $\hat{\boldsymbol{\mu}} = \sigma [\sqrt{\hat{x}_n}]_{1 \leq n \leq N}$. In the literature, there are two main approaches to deal with potentially negative estimated values \hat{x}_n : either taking their absolute value as in [7], or setting them to zero as in [8]. The first strategy usually yields a lower variance, but a higher bias than the second. Since we are looking for the best bias–variance tradeoff (i.e., the lowest MSE), we propose a convex combination of both approaches as

$$\hat{\boldsymbol{\mu}} = \sigma \left[\lambda \sqrt{|f_n(\mathbf{y})|} + (1 - \lambda) \sqrt{\max(f_n(\mathbf{y}), 0)} \right]_{1 \leq n \leq N} \quad (27)$$

where $\lambda \in [0, 1]$. In Fig. 6, we evaluate the MSE-sensitivity of the final estimate $\hat{\boldsymbol{\mu}}$ with respect to the value of λ . As observed, the MMSE choice of λ depends on the noise level, as well as on the image content. Since $\lambda = 0$ usually provides a lower bias, but a higher variance than $\lambda = 1$, the MMSE choice of λ generally lies in between these two extreme values. In all our experiments, we have thus used $\lambda = 0.5$.

We evaluate denoising performance objectively using three full-reference image quality metrics.

- 1) The standard peak SNR (PSNR), defined as $\text{PSNR} = 10 \log_{10}(N \|\boldsymbol{\mu}\|_{\infty}^2) / (\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2)$.
- 2) A contrast-invariant PSNR, defined as

$$\text{CIPSNR} = 10 \log_{10} \frac{N \|\boldsymbol{\mu}\|_{\infty}^2}{\|(a^* \hat{\boldsymbol{\mu}} + b^*) - \boldsymbol{\mu}\|^2}$$

where the affine parameters a^*, b^* are given by

$$(a^*, b^*) = \arg \min_{a, b} \|(a \hat{\boldsymbol{\mu}} + b) - \boldsymbol{\mu}\|^2$$

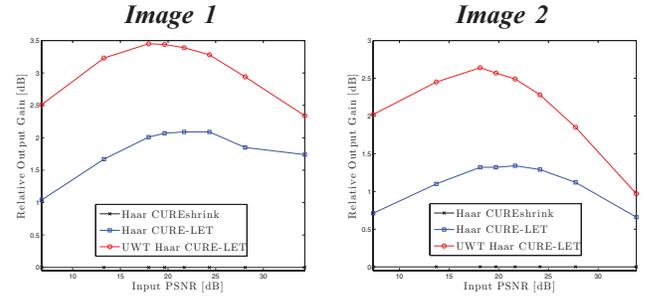


Fig. 8. PSNR comparisons among pointwise undecimated Haar CURE-LET (Section III-B) and joint intra-/inter-scale unnormalized Haar CURE-LET (Section IV-C), shown relative to unnormalized Haar CUREshrink (Section IV-B).

$$\Leftrightarrow \begin{cases} a^* = \frac{N \boldsymbol{\mu}^T \hat{\boldsymbol{\mu}} - \mathbf{1}^T \boldsymbol{\mu} \mathbf{1}^T \hat{\boldsymbol{\mu}}}{N \hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}} - (\mathbf{1}^T \hat{\boldsymbol{\mu}})^2} \\ b^* = \frac{\mathbf{1}^T \boldsymbol{\mu} \hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^T \hat{\boldsymbol{\mu}} \mathbf{1}^T \hat{\boldsymbol{\mu}}}{N \hat{\boldsymbol{\mu}}^T \hat{\boldsymbol{\mu}} - (\mathbf{1}^T \hat{\boldsymbol{\mu}})^2}. \end{cases}$$

- 3) The mean of the structural similarity index map (SSIM), which is a popular visual quality metric introduced in [43] (see <https://ece.uwaterloo.ca/z70wang/research/ssim/>).

In all simulated experiments, we have assumed that the variance σ^2 of the complex Gaussian noise is known. In practice, a reliable estimate can be obtained in signal-free or constant regions of the image by moment-matching [6], [10] or maximum likelihood techniques [13], [15], [44]. When no background is available, more sophisticated approaches can be considered [45].

To simulate various input noise levels, several values for σ have been selected in the range $\sigma \in [5, 100]$. The set of high-quality magnitude MR test images used is shown in Fig. 7, and may be obtained from <http://bigwww.epfl.ch/luisier/MRIdenoising/TestImages.zip>.

A. Comparisons Between Variants of the Proposed Approach

Before comparing our approach with some state-of-the-art MR image denoising methods, we first evaluate the performance of the various variants of our approach. In Fig. 8, we compare the results obtained using the pointwise CURE-LET thresholding of (15) applied in the undecimated Haar wavelet transform domain, and the joint intra-/inter-scale LET denoising function of (24) applied to the unnormalized Haar wavelet transform coefficients. These are shown relative to a baseline provided by the unnormalized Haar CUREshrink approach of (23). As expected, the joint intra-/inter-scale denoising function of (24) outperforms the simple soft-thresholding of (23) by 1–3 dB. Moreover, pointwise thresholding applied in a shift-invariant setting outperforms (by 0.5–1.5 dB) a more sophisticated thresholding in a shift-variant one.

Note that the shift-invariance of the unnormalized Haar wavelet transform can be increased by applying the so-called cycle-spinning technique [30]. In Fig. 9, we show the PSNR improvements brought by averaging the results of several cycle-spins (CS). As observed, 16 CS allow a near match

TABLE I
PSNR AND CIPSNR COMPARISONS

σ	PSNR results						CIPSNR results					
	5	10	20	30	50	100	5	10	20	30	50	100
Image	<i>Image 1 (256 × 256)</i>											
Input	34.35	28.09	21.69	17.97	13.28	6.74	34.52	28.58	22.72	19.69	16.85	15.16
[6]	37.95	33.46	29.09	26.51	23.20	18.35	38.00	33.57	29.41	27.05	24.02	19.46
[7]	33.13	30.47	28.56	25.13	18.56	10.48	33.55	30.89	28.64	26.29	22.89	18.72
[10]	37.29	32.86	28.10	25.22	21.69	17.33	37.30	32.91	28.33	25.57	22.17	17.91
[8]	39.15	34.98	30.64	28.14	24.82	19.11	39.15	34.99	30.64	28.15	24.84	19.28
[17] + [27]	39.44	34.88	30.34	27.75	24.37	18.84	39.44	34.89	30.35	27.77	24.44	19.15
[17] + [46]	40.24	35.95	31.59	29.19	26.02	21.03	40.24	35.98	31.67	29.26	26.10	21.33
CURE-LET Haar/CS = 16	39.00	34.84	30.71	28.15	24.87	20.32	39.00	34.85	30.82	28.40	25.30	20.75
CURE-LET UWT Haar	38.32	34.48	30.61	28.28	25.24	20.81	38.33	34.49	30.64	28.36	25.44	21.13
Image	<i>Image 2 (256 × 256)</i>											
Input	34.00	27.77	21.48	17.84	13.32	7.07	34.33	28.31	22.24	18.88	15.44	13.06
[6]	35.88	31.72	27.64	25.21	22.10	17.84	36.01	31.88	28.00	25.76	22.90	18.83
[7]	29.35	27.62	25.89	24.73	19.45	11.65	30.25	28.49	26.35	24.98	22.02	18.13
[10]	35.84	31.24	26.58	24.07	21.00	17.47	35.86	31.38	26.88	24.58	21.65	18.17
[8]	36.11	32.33	28.80	26.60	23.48	19.21	36.14	32.34	28.81	26.61	23.54	19.29
[17] + [27]	36.85	32.82	28.85	26.51	23.64	19.42	36.85	32.83	28.86	26.54	23.71	19.61
[17] + [46]	37.15	33.17	29.37	27.23	24.45	20.26	37.16	33.18	29.38	27.23	24.46	20.32
CURE-LET Haar/CS = 16	36.55	32.54	28.72	26.47	23.57	19.67	36.58	32.59	28.86	26.73	24.03	20.29
CURE-LET UWT Haar	36.42	32.71	29.08	26.89	24.02	20.08	36.48	32.75	29.13	26.98	24.22	20.52
Image	<i>Image 3 (256 × 256)</i>											
Input	33.84	27.73	21.64	18.13	13.71	7.36	34.07	28.01	22.13	19.06	16.07	14.07
[6]	35.75	31.50	27.36	24.99	22.07	18.08	35.94	31.77	27.79	25.51	22.64	18.53
[7]	28.20	26.32	24.63	24.15	20.98	12.31	28.96	26.93	25.07	24.46	22.34	18.25
[10]	35.86	31.19	26.70	24.14	21.06	17.40	35.93	31.35	26.93	24.41	21.34	17.68
[8]	35.92	31.96	28.34	26.31	23.37	18.77	35.94	31.98	28.37	26.36	23.46	18.81
[17] + [27]	36.44	32.38	28.66	26.58	23.83	19.46	36.45	32.40	28.68	26.61	23.89	19.58
[17] + [46]	36.84	32.84	29.11	27.05	24.41	19.93	36.86	32.85	29.12	27.07	24.43	19.99
CURE-LET Haar/CS = 16	36.36	32.17	28.23	25.98	23.19	19.41	36.42	32.27	28.44	26.27	23.55	19.67
CURE-LET UWT Haar	36.21	32.25	28.59	26.50	23.81	20.04	36.25	32.29	28.66	26.62	24.01	20.31

PSNRs and CIPSNRs have been averaged over 10 noise realizations.

in performance to the shift-invariant transform. The cycle-spinning technique has the further advantage of being easily implementable in parallel.

B. Comparisons to State-of-the-Art MR Image Denoising Methods

As benchmarks for evaluating our CURE-LET approach, we have retained five state-of-the-art MR image denoising techniques: two wavelet-based algorithms [6], [7] (code at <http://telin.ugent.be/sanja/>), a spatially adaptive linear MMSE filter [10], an unbiased nonlocal means filter specifically designed for MR data [8] (code at <http://personales.upv.es/jmanjon/denoising/nlm2d.htm>), and the recent approach based on Rician-adapted VST [17] followed by an AWGN denoiser. We have considered two AWGN denoisers: the SURE-LET algorithm described in [27] (code at

bigwww.epfl.ch/luisier/SURE-LET.zip), which has a similar complexity as the proposed CURE-LET approach, and the state-of-the-art BM3-D [46] (code at <http://www.cs.tut.fi/foi/GCF-BM3-D/>), which is a sophisticated two-pass denoising algorithm. For each of these methods, we have used the tuning parameters suggested in their respective publications and software, except for the linear MMSE filter, where we have hand-optimized (in the MMSE sense) the size of the filter support.

We have considered two CURE-LET variants: 16 CS of the joint intra-/inter-scale thresholding of (24) applied in the unnormalized Haar wavelet transform and the pointwise thresholding of (15) applied in the undecimated Haar wavelet transform. The corresponding MATLAB code can be downloaded at <http://fluisier.webnode.com/software/>.

In Table I, the PSNR and CIPSNR results of the various methods are displayed. As observed, the Rician VST + BM3-

TABLE II
 MEAN SSIM COMPARISONS

σ	5	20	30	50	100	5	20	30	50	100	5	20	30	50	100
Image	Image 1 (256 × 256)					Image 2 (256 × 256)					Image 3 (256 × 256)				
Input	0.87	0.40	0.26	0.14	0.04	0.88	0.46	0.33	0.19	0.06	0.88	0.47	0.34	0.20	0.06
[6]	0.96	0.79	0.68	0.52	0.32	0.93	0.76	0.65	0.49	0.29	0.93	0.74	0.64	0.50	0.32
[7]	0.89	0.78	0.65	0.47	0.27	0.78	0.68	0.65	0.46	0.28	0.80	0.69	0.64	0.53	0.33
[10]	0.97	0.78	0.67	0.51	0.26	0.93	0.68	0.56	0.41	0.23	0.93	0.70	0.59	0.44	0.26
[8]	0.97	0.78	0.68	0.52	0.28	0.93	0.75	0.67	0.53	0.31	0.92	0.77	0.69	0.56	0.31
[17] + [27]	0.97	0.78	0.66	0.48	0.21	0.94	0.76	0.66	0.53	0.30	0.94	0.78	0.70	0.59	0.32
[17] + [46]	0.98	0.80	0.71	0.58	0.38	0.94	0.75	0.68	0.57	0.39	0.94	0.78	0.72	0.62	0.43
CURE-LET Haar/CS = 16	0.97	0.85	0.76	0.63	0.43	0.94	0.80	0.72	0.59	0.39	0.94	0.78	0.70	0.58	0.40
CURE-LET UWT Haar	0.97	0.84	0.76	0.64	0.45	0.94	0.81	0.74	0.62	0.41	0.94	0.79	0.72	0.61	0.43

Output SSIMs have been averaged over 10 noise realizations.

 TABLE III
 COMPUTATION TIMES IN [s]

Image size	[6]	[7]	[10]	[8]	[17]+[27]	[17]+[46]	CURE-LET Haar/CS = 1	CURE-LET Haar/CS = 16	CURE-LET UWT Haar
256 × 256	0.2	0.5	0.1	33.8	0.3	1.6	0.2	2.8	0.9
512 × 512	0.8	2.4	0.5	145.7	1.9	6.9	0.6	8.2	6.6

Computation times have been averaged over 10 runs, [8], [10] do not use precompiled MEX files.

D generally obtains the best results (+0.5 dB compared to our CURE-LET.) This mostly comes from the high-quality of BM3-D. Indeed, our CURE-LET approach usually outperforms the Rician VST (+0.3 dB) when the latter is followed by the similar complexity SURE-LET denoiser. The CURE-LET denoising gains are about +1.3 dB relative to the method of [6], +3.6 dB compared to [7], +2.2 dB compared to [10], and +0.5 dB compared to [8]. In Table II, the mean SSIMs of the various methods are reported. As observed, the proposed CURE-LET approach generally obtains the highest SSIM scores. Overall, our method thus compares favorably to most state-of-the-art approaches, especially under very noisy conditions, in which case the signal-dependent nature of the noise is more pronounced. Note that further denoising gains are likely to be obtained by considering more sophisticated thresholding rules, image-adaptive overcomplete dictionaries (e.g., [47]), or a linear expansion of nonlocal means, as proposed in [48]. A more evolved CURE-optimized algorithm could then reach the state-of-the-art performance achieved by Rician VST + BM3-D.

Fig. 10 presents a visual comparison of the various MR image denoising algorithms. As observed, the CURE-LET denoising result offer a good balance between noise suppression and fine-structure preservation. In particular, the traditional artifacts resulting from wavelet thresholding [see Fig. 10(c) and (d)] are greatly reduced thanks to the image-content-based optimization of the numerous LET parameters. This subjective observation is confirmed by the higher SSIM score obtained by the proposed denoising approach.

In Table III, we report the computation time of the various algorithms considered. All have been executed on MATLAB

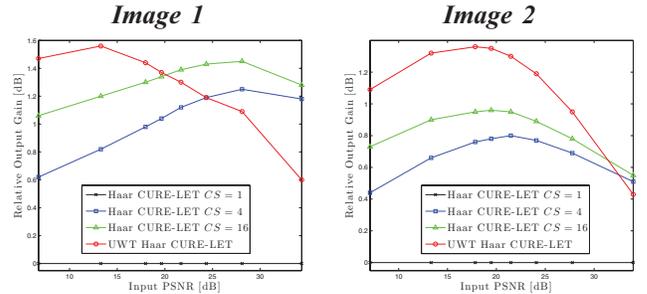


Fig. 9. PSNR improvements brought by cycle-spinning the unnormalized Haar wavelet transform.

R2011a running under Mac OS X equipped with a 2.66 GHz Intel Core 2 Duo processor. As observed, the proposed CURE-LET algorithms take 0.2–3 s to denoise a 256 × 256 image. When applied within an undecimated filterbank transform, most of the computational load is dedicated to the independent reconstructions of the processed subbands and their corresponding first- and second-order derivatives [see (16)].

C. Denoising of a Magnitude MR Knee Image

We have also applied our CURE-LET denoising approach to an actual magnitude MR image of the knee. This 512 × 512 16-bit raw image has been acquired on a Siemens 1.5-T Magnetom Sonata MR system, following a sagittal T2-weighted protocol. The standard deviation of the complex Gaussian noise has been estimated from a signal-free region \mathcal{S} of the squared data, as $\hat{\sigma} = \sqrt{(1/2) \sum_{n \in \mathcal{S}} |m_n|^2}$, and subsequently treated as known.

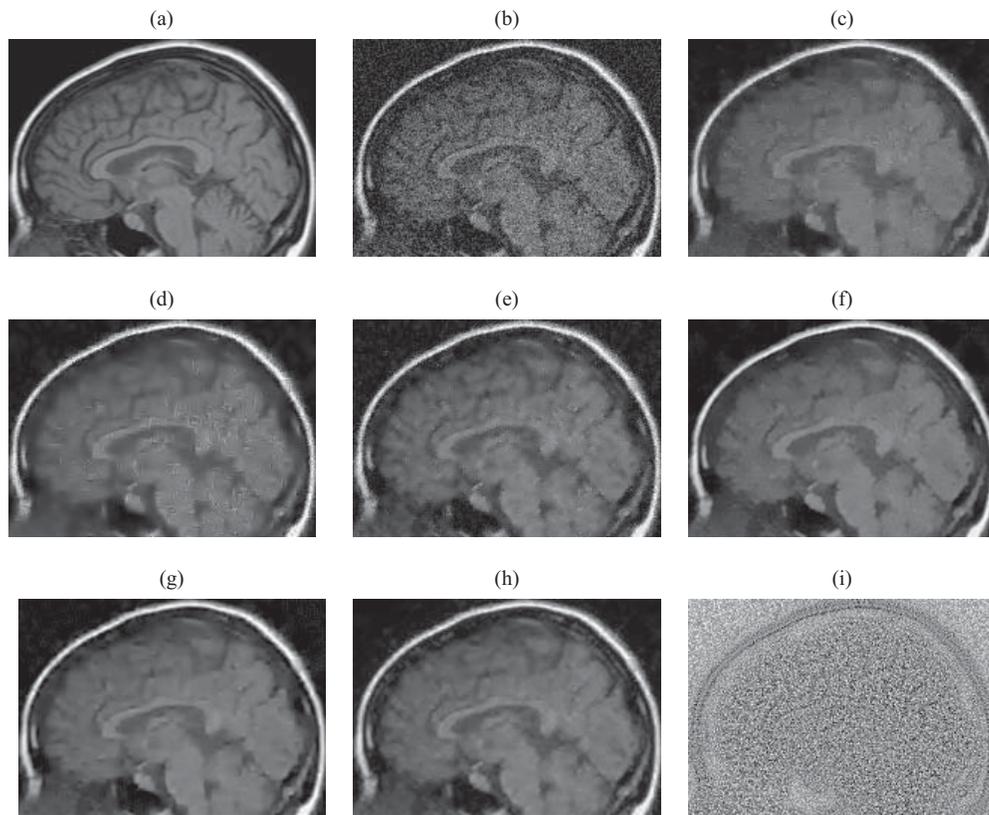


Fig. 10. Visual Comparisons. (a) Zoom in image 1. (b) Noisy version: $SSIM = 0.319$. (c) Denoised by [6]: $SSIM = 0.734$. (d) Denoised by [7]: $SSIM = 0.726$. (e) Denoised by [10]: $SSIM = 0.653$. (f) Denoised by [8]: $SSIM = 0.719$. (g) Denoised by [17] + [27]: $SSIM = 0.721$. (h) Denoised by Haar CURE-LET (CS = 16): $SSIM = 0.800$. (i) Residual between (b) and (h).

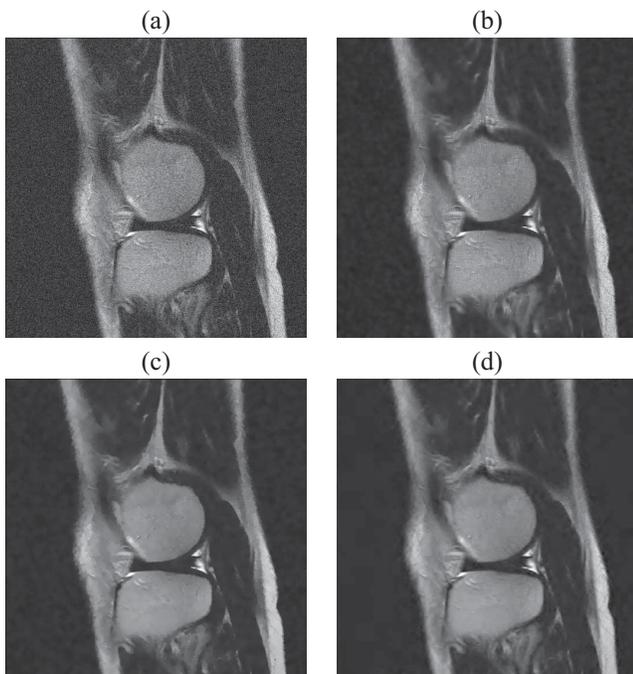


Fig. 11. Denoising of a magnitude MR image of the knee. (a) Raw 16-bit data: $\hat{\sigma} = 38$. (b) Denoised via [7]. (c) Via [8]. (d) Via Haar CURE-LET (CS = 16).

Fig. 11 shows a comparison of denoising results. As observed, in the CURE-LET approach of Fig. 11(d) the

noise is efficiently attenuated and the contrast is significantly improved, owing to a significant reduction of the signal-dependent bias introduced by the noise.

VI. CONCLUSION

In this paper, we have derived an unbiased risk estimation procedure, and applied it to the problem of magnitude MR image denoising, where the squared value of each pixel comprises an independent noncentral chi-square variate on two degrees of freedom. Our approach can be used to optimize the parameters of essentially any continuously differentiable estimator for this class of problems, and here we have focused our attention on transform-domain algorithms in particular.

In this vein, we first developed a pointwise LET estimator applied to the coefficients of an arbitrary undecimated filter-bank transform. We then considered the specific case of the unnormalized Haar wavelet transform, which is a multiscale orthogonal transform allowing the derivation of *subband-dependent* CURE denoising strategies. We also introduced a subband-adaptive joint inter-/intra-scale LET that outperforms a simpler estimator similar to soft thresholding.

We then applied our proposed CURE-optimized algorithms to test images artificially degraded by noise, and observed them to compare favorably with most state-of-the-art techniques, both quantitatively and qualitatively. Finally, we showed an example of denoising results obtained on an actual magnitude MR image, in order to show the practi-

cal efficacy of our approach to MR image denoising via CURE.

ACKNOWLEDGMENT

The authors would like to thank W.-Y. I. Tseng for providing the MRI data, and A. Pižurica, J. Manjón, K. Dabov, and A. Foi for making their respective software implementations available online.

REFERENCES

- [1] G. Wright, "Magnetic resonance imaging," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 56–66, Jan. 1997.
- [2] C. L. Epstein, *Introduction to the Mathematics of Medical Imaging*, 2nd ed. Philadelphia, PA: SIAM, 2008.
- [3] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Med. Phys.*, vol. 12, no. 2, pp. 232–233, 1985.
- [4] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.
- [5] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Mag. Reson. Med.*, vol. 34, no. 6, pp. 910–914, 1995.
- [6] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Trans. Image Process.*, vol. 8, no. 10, pp. 1408–1419, Oct. 1999.
- [7] A. Pižurica, W. Philips, I. Lemahieu, and M. Achero, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 323–331, Mar. 2003.
- [8] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, "MRI denoising using non local means," *Med. Image Anal.*, vol. 12, no. 2, pp. 514–523, 2008.
- [9] S. Aja-Fernández, M. Niethammer, M. Kubicki, M. E. Shenton, and C.-F. Westin, "Restoration of DWI data using a Rician LMMSE estimator," *IEEE Trans. Med. Imag.*, vol. 27, no. 10, pp. 1389–1403, Oct. 2008.
- [10] S. Aja-Fernández, C. Alberola-Lopez, and C.-F. Westin, "Noise and signal estimation in magnitude MRI and Rician distributed images: A LMMSE approach," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1383–1398, Aug. 2008.
- [11] H. Liu, C. Yang, N. Pan, E. Song, and R. Greens, "Denoising 3D MR images by the enhanced non-local means filter for Rician noise," *Mag. Reson. Imag.*, vol. 28, no. 10, pp. 1485–1496, 2010.
- [12] T. Thaipanich and C.-C. J. Kuo, "An adaptive nonlocal means scheme for medical image denoising," *Proc. SPIE*, vol. 7623, pp. 76230-1–76230-12, Feb. 2010.
- [13] J. Sijbers, A. J. den Dekker, P. Scheunders, and D. V. Dyck, "Maximum-likelihood estimation of Rician distribution parameters," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 357–361, Jun. 1998.
- [14] J. Sijbers and A. J. den Dekker, "Maximum-likelihood estimation of signal amplitude and noise variance from MR data," *Mag. Reson. Med.*, vol. 51, no. 3, pp. 586–594, 2004.
- [15] L. He and I. Greenshields, "A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images," *IEEE Trans. Med. Imag.*, vol. 28, no. 2, pp. 165–172, Feb. 2009.
- [16] S. P. Awate and R. T. Whitaker, "Feature-preserving MRI denoising: A nonparametric empirical Bayes approach," *IEEE Trans. Med. Imag.*, vol. 26, no. 9, pp. 1242–1255, Sep. 2007.
- [17] A. Foi, "Noise estimation and removal in MR imaging: The variance-stabilization approach," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Mar.–Apr. 2011, pp. 1809–1814.
- [18] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, Jul. 2005.
- [19] P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise non local means denoising filter for 3D magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 425–441, Apr. 2008.
- [20] Y. Gal, A. J. H. Mehnert, A. P. Bradley, K. McMahon, D. Kennedy, and S. Crozier, "Denoising of dynamic contrast-enhanced MR images using dynamic nonlocal means," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 302–310, Feb. 2010.
- [21] J. B. Weaver, Y. Xu, D. M. Healy, and L. D. Cromwell, "Filtering noise from images with wavelet transforms," *Magn. Reson. Med.*, vol. 21, no. 2, pp. 288–295, 1991.
- [22] Y. Xu, J. Weaver, D. Healy, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Trans. Image Process.*, vol. 3, no. 6, pp. 747–758, Nov. 1994.
- [23] J. C. Wood and K. M. Johnson, "Wavelet packet denoising of magnetic resonance images: Importance of Rician noise at low SNR," *Magn. Reson. Med.*, vol. 41, no. 3, pp. 631–635, 1999.
- [24] S. Zaroubi and G. Goelman, "Complex denoising of MR data via wavelet analysis: Application for functional MRI," *Magn. Reson. Imag.*, vol. 18, no. 1, pp. 59–68, 2000.
- [25] P. Bao and L. Zhang, "Noise reduction for magnetic resonance images via adaptive multiscale products thresholding," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1089–1099, Sep. 2003.
- [26] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 593–606, Mar. 2007.
- [27] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778–2786, Nov. 2007.
- [28] F. Luisier and P. J. Wolfe, "Chi-square unbiased risk estimate for denoising magnitude MR images," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 1593–1596.
- [29] L. D. Brown, *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Hayward, CA: Institute of Mathematical Statistics, 1986.
- [30] R. R. Coifman and D. L. Donoho, "Translation invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics*. New York: Springer-Verlag, 1995, pp. 125–150.
- [31] G. Nason and B. W. Silverman, *The Stationary Wavelet Transform and Some Statistical Applications*. New York: Springer-Verlag, 1995.
- [32] A. Gyaourova, C. Kamath, and I. K. Fodor, "Undecimated wavelet transforms for image de-noising," Lawrence Livermore National Laboratory, Livermore, CA, Tech. Rep., 2002.
- [33] J.-L. Starck, J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 297–309, Feb. 2007.
- [34] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, Mar. 2011.
- [35] P. B. Patnaik, "The non-central χ^2 - and f -distribution and their applications," *Biometrika*, vol. 36, nos. 1–2, pp. 202–232, 1949.
- [36] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [37] K. Hirakawa, F. Baqai, and P. J. Wolfe, "Wavelet-based Poisson rate estimation using the Skellam distributions," *Proc. SPIE*, vol. 7246, pp. 1–12, Feb. 2009.
- [38] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of Poisson-corrupted images," *Signal Process.*, vol. 90, no. 2, pp. 415–427, 2010.
- [39] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.
- [40] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.
- [41] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, Nov. 2002.
- [42] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1138–1351, Nov. 2003.
- [43] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [44] J. Sijbers, D. Poot, A. J. Dekker, and W. Pintjens, "Automatic estimation of the noise variance from the histogram of a magnetic resonance images," *Phys. Med. Biol.*, vol. 52, no. 5, pp. 1335–1348, 2007.
- [45] J. Rajan, D. Poot, J. Juntu, and J. Sijbers, "Noise measurement from magnitude MRI using local estimates of variance and skewness," *Phys. Med. Biol.*, vol. 55, no. 16, pp. 441–449, 2010.
- [46] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

- [47] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representations," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [48] D. Ville and M. Kocher, "Non-local means with dimensionality reduction and SURE-based parameter selection," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2683–2690, Sep. 2011.



Florian Luisier (S'05–M'10) was born in Bagnes, Switzerland, in 1981. He received the Masters degree in microengineering and the Ph.D. degree in computer, communication, and information sciences from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2005 and 2010, respectively.

He was with the Biomedical Imaging Group, EPFL, from 2005 to March 2010. He is now a Post-Doctoral Fellow with the School of Engineering and Applied Sciences, Harvard University, Cambridge,

MA. His current research interests include statistical signal/image/video processing, image restoration, biomedical imaging, multiresolution representation, estimation and detection theory, and inverse problems.

Dr. Luisier was a recipient of the Young Author Best Paper Award in 2009 from the IEEE Signal Processing Society for a paper co-authored with Prof. T. Blu and Prof. M. Unser.



Thierry Blu (M'96–SM'06–F'12) was born in Orléans, France, in 1964. He received the Diplôme d'Ingénieur degree from École Polytechnique, Palaiseau, France, in 1986, and Télécom Paris (ENST), Paris, France, in 1988, and the Ph.D. degree in electrical engineering from ENST in 1996, for a study on iterated rational filterbanks, applied to wideband audio coding.

He was with the Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne, Switzerland, from 1998 and 2007. He is now a Profes-

sor with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong. His current research interests include multiwavelets,

multiresolution analysis, multirate filterbanks, interpolation, approximation and sampling theory, sparse sampling, image denoising, psychoacoustics, biomedical imaging, optics, and wave propagation.

Prof. Blu was the recipient of two Best Paper Awards by the IEEE Signal Processing Society in 2003 and 2006. He is a co-author of a paper that received the Young Author Best Paper Award in 2009 from the same society. He was an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2002 to 2006, the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2010, and *Elsevier Signal Processing* from 2008 to 2011. He has been on the Board of the *Eurasip Journal on Image and Video Processing* since 2010.



Patrick J. Wolfe (S'96–M'03–SM'08) received the undergraduate degree in electrical engineering and music from the University of Illinois at Urbana-Champaign, Urbana, and the Ph.D. degree from the University of Cambridge, Cambridge, U.K.

He is Professor of statistics and Honorary Professor of computer science with University College London, London, U.K. He was a U.S. National Science Foundation Graduate Research Fellow with the University of Cambridge. From 2001 to 2004, he held a Fellowship and College Lectureship of

engineering and computer science with University of Cambridge. He was an Assistant Professor from 2004 to 2008 and an Associate Professor from 2008 to 2011 with Harvard University, Cambridge, MA, and a Faculty Member of the Harvard–MIT Speech and Hearing Biosciences and Technology Program. His current research interests include modeling and inference for graphs and networks, statistical imaging and image processing, time-series and time–frequency analysis, audio signal processing, and acoustic modeling.

Prof. Wolfe was a recipient of the Presidential Early Career Award for Scientists and Engineers in 2008. He received honors from the IEEE, the Acoustical Society of America, the Royal Statistical Society, and the International Society for Bayesian Analysis. He currently serves as an Elected Member of the Image, Video, and Multidimensional Signal Processing Technical Committee of the IEEE Signal Processing Society, on the Editorial Board of its *Signal Processing Magazine*, and as a member of the Defense Advanced Research Projects Agency Information Science and Technology Board.