NOISE ESTIMATION AND REMOVAL IN MR IMAGING: THE VARIANCE-STABILIZATION APPROACH

Alessandro Foi

Department of Signal Processing, Tampere University of Technology
P.O. Box 553, 33101, Tampere, Finland
web: http://www.cs.tut.fi/~foi   email: firstname.lastname@tut.fi

ABSTRACT

We develop optimal forward and inverse variance-stabilizing transformations for the Rice distribution, in order to approach the problem of magnetic resonance (MR) image filtering by means of standard denoising algorithms designed for homoskedastic observations.

Further, we present a stable and fast iterative procedure for robustly estimating the noise level from a single Rician-distributed image. At each iteration, the procedure exploits variance-stabilization composed with a homoskedastic variance-estimation algorithm.

Theoretical and experimental study demonstrates the success of our approach to Rician noise estimation and removal through variance stabilization. In particular, we show that the performance of current state-of-the-art algorithms specifically designed for Rician-distributed data can be matched by combining conventional algorithms designed for additive white Gaussian noise with optimal variance-stabilizing transformations.

1. INTRODUCTION

The magnitude of magnetic resonance (MR) images can be modeled by the Rice distribution. This distribution has two parameters: the unknown noise-free magnitude of the data, and the standard deviation of the additive noise that corrupts the real and imaginary parts of the data.

Estimation of the magnitude is a particularly challenging denoising problem because of two main reasons, namely heteroskedasticity and bias: first, the standard-deviation of the noise corrupting the magnitude depends also on the unknown magnitude itself; second, the expectation of the noisy magnitude differs from the unknown noise-free magnitude by a nonlinear function of the noise standard-deviation and of the noise-free magnitude. Special ad-hoc algorithms need to be designed for filtering MR images, in order to address both the heteroskedasticity and bias in the Rician-distributed data.

To enable the application to MR image filtering of conventional algorithms designed for homoskedastic observations (e.g., for data corrupted by additive white Gaussian noise), we develop optimal forward and inverse variance-stabilizing transformations for the Rice distribution (Section 3). The forward transformation makes the data accurately homoskedastic, and thus the noise removal can be accomplished by applying a homoskedastic denoising algorithm; the inverse transformation is designed to be applied on the denoised data and to return a maximum-likelihood (ML) estimate of the noise-free magnitude. To the best of our knowledge, this is the first approach of this kind to the MR image filtering problem.

A second fundamental contribution of this work consists in a stable and fast iterative procedure for robustly estimating the noise level from a single Rician-distributed image (Section 5). At each

iteration, the procedure exploits variance-stabilization composed with a homoskedastic variance-estimation algorithm. We require neither the presence of a dark uniform background, a preliminary segmentation of the data, nor a high signal-to-noise ratio.

Theoretical and experimental study demonstrates the success of our approach to Rician noise estimation and removal through variance stabilization. In particular, we show that by combining conventional algorithms designed for additive white Gaussian noise with optimal variance-stabilizing transformations, we can match the performance of current state-of-the-art algorithms specifically designed for Rician-distributed data.

2. PRELIMINARIES

In this section we review the main theoretical and technical elements necessary for the development of our contributions.

2.1 Rician-distributed data

Let $z \sim \mathcal{R}(v, \sigma)$ be the realization of a random variable with Rician probability density function (p.d.f.) with parameters $v \geq 0$ and $\sigma > 0$,

$$p(z|v, \sigma) = \frac{z}{\sigma^2} e^{-\frac{z^2 + v^2}{2\sigma^2}} I_0\left(\frac{zv}{\sigma^2}\right), \qquad z \geq 0, \tag{1}$$

where I_n denotes the modified Bessel function of order n, $I_n(x) = \sum_{m=0}^{\infty} \frac{(x/2)^{n+2m}}{m! \Gamma(n+m+1)}$ [16]. Equivalently, z can be obtained as

$$z = \sqrt{(c_r v + \sigma \eta_r)^2 + (c_i v + \sigma \eta_i)^2}, \tag{2}$$

where c_r and c_i are arbitrary constants such that $0 \leq c_r, c_i \leq 1 = c_r^2 + c_i^2$, and η_r and η_i are random variates independently distributed following the standard normal p.d.f., $\eta_r, \eta_i \sim \mathcal{N}(0, 1)$. The Rice distribution is used for modeling magnitude MR images or volumes, with an observation model of the form

$$z(x) \sim \mathcal{R}(v(x), \sigma), \quad x \in X, \tag{3}$$

where $X \subset \mathbb{Z}^d$ are the pixel ($d = 2$) or voxel ($d = 3$) coordinates, $v : X \to \mathbb{R}^+$ is the unknown original (noise-free) signal, and $z : X \to \mathbb{R}^+$ is the raw magnitude MR data.

2.2 $\mathcal{R}(\cdot, \sigma)$ as one-parameter family of distributions; scaling

The parameter σ is assumed as fixed and, until Section 4, we consider that its exact value is known a priori. Thus, z is treated as distributed according to a one-parameter family of Rician distributions, parametrized with respect to v. Figure 1 shows the distributions $\mathcal{R}(v, \sigma)$ for $v \in [0, 5]$ and a fixed $\sigma = 1$. Let us remark that assuming a particular value of σ (e.g., $\sigma = 1$) is not a fundamental restriction: from (2) one can easily see that if $z \sim \mathcal{R}(v, \sigma)$ then $\lambda z \sim \mathcal{R}(\lambda v, \lambda \sigma)$ for any $\lambda > 0$, which means that it suffices to carry out the analysis and computation of the transformations for, say, $\sigma = 1$, and then apply this result to all cases $\sigma > 0$ upon simple linear rescaling of data and parameters.
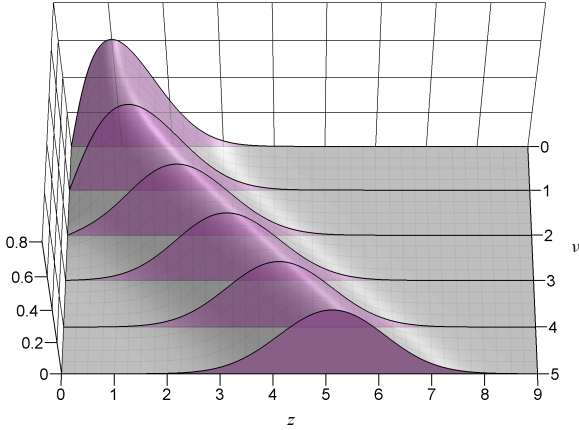
Figure 1: The one-parameter family of Rician p.d.f.'s $\mathcal{R}(\nu, 1)$ for $\nu \in [0, 5]$.

2.3 Mean and variance of Rician data

The mean and variance of $z \sim \mathcal{R}(\nu, \sigma)$ are, respectively,

$$\mu = E\{z|\nu, \sigma\} = \sigma \sqrt{\frac{\pi}{2}} L\left(-\frac{\nu^2}{2\sigma^2}\right), \tag{4}$$

$$s^2 = \text{var}\{z|\nu, \sigma\} = 2\sigma^2 + \nu^2 - \frac{\pi\sigma^2}{2} L^2\left(-\frac{\nu^2}{2\sigma^2}\right), \tag{5}$$

where $L(x) = e^{x/2}\left[(1-x)I_0\left(-\frac{x}{2}\right) - xI_1\left(-\frac{x}{2}\right)\right]$. Equation (5) is illustrated, for the case $\sigma = 1$, in Figure 2. For large values of ν we have

$$E\{z|\nu, \sigma\} \approx \nu + \frac{\sigma^2}{2\nu}, \qquad \text{var}\{z|\nu, \sigma\} \approx \sigma^2 - \frac{\sigma^4}{2\nu^2}. \tag{6}$$

Two crucial issues follow from (4) and (5). Firstly, (5) implies that the noise variance is not uniform over the data. Secondly, the expectation (4) differs essentially from the parameter of interest, namely ν. The former problem is addressed by the (forward) variance-stabilizing transformation applied to the data before filtering, whereas the latter is addressed by the inverse transformation applied upon filtering, which is designed so to directly provide an estimate of ν out of the filtered transformed data.

2.4 Maximum-likelihood estimate $\nu_{\text{ML}}(z)$ of ν given z

If $\sigma = 1$, the maximum-likelihood (ML) estimate of ν given a single sample z, which we indicate by $\nu_{\text{ML}}(z)$, is the solution of the equation $\frac{z}{\nu}\frac{I_1(\nu z)}{I_0(\nu z)} = 1$ for $z > \sqrt{2}$ [15], and $\nu_{\text{ML}}(z) = 0$ for $z \leq \sqrt{2}$. This estimate will become useful for the construction of the optimized transformations in Section 3.2.

2.5 Stabilization of variance

The rationale behind applying a variance-stabilizing transformation is to remove the dependence of the noise variance from the value of the parameter ν and hence from the coordinate x. Without loss of generality, we look for transformations f such that the variance of the transformed data is stabilized to 1, i.e. $\text{var}\{f(z)|\nu, \sigma\} \approx 1$.

2.5.1 Scaling of the stabilizer for Rician data

With the same notation as in Section 2.2, if f is a stabilizer for $\mathcal{R}(\cdot, \sigma)$, then f_λ defined by

$$f_\lambda(w) = f(w/\lambda), \quad \forall w \in [0, +\infty), \tag{7}$$

is a stabilizer for $\mathcal{R}(\cdot, \lambda\sigma)$, because $\text{var}\{f(z)|\nu, \sigma\} = \text{var}\{f_\lambda(w)|\lambda\nu, \lambda\sigma\}$.

2.6 Three steps: stabilization, denoising, and inversion

Variance-stabilizing transformations are often exploited for the removal of signal-dependent noise through the following three-step procedure. First, the noise variance is stabilized by applying a variance-stabilizing transformation f to the data. This produces a signal in which the noise can be treated as additive with unitary variance. Second, the noise is removed using a conventional denoising algorithm for additive noise (e.g., additive white Gaussian noise). Third, an inverse transformation is applied to the denoised signal, obtaining the estimate of the signal of interest.

In what follows, we indicate the denoising algorithm by the operator Φ and the denoised signal before inversion by $D = \Phi(f(z))$. Denoising algorithms attempt to estimate the expectation, thus D can be treated as an approximation of $E\{f(z)|\nu, \sigma\}$.

2.7 Exact unbiased inverse

2.7.1 Estimation of $E\{z|\nu, \sigma\}$

Since f is necessarily a nonlinear mapping, we may have

$$E\{f(z)|\nu, \sigma\} \neq f(E\{z|\nu, \sigma\}), \tag{8}$$

and, thus,

$$f^{-1}(E\{f(z)|\nu, \sigma\}) \neq E\{z|\nu, \sigma\}, \tag{9}$$

which means that the inverse transformation applied after denoising (in the three-step procedure of Section 2.6) should not coincide with the algebraic inverse of f, as this would introduce bias in the estimation of $E\{z|\nu, \sigma\}$ from the observed z.

The problem of bias in variance-stabilized denoising is solved by the *exact unbiased inverse* [14],[7] that is defined by the mapping

$$\mathcal{I}_f : E\{f(z)|\nu, \sigma\} \longmapsto E\{z|\nu, \sigma\} = \mu. \tag{10}$$

Note that (10) assumes that the mapping $E\{z|\nu, \sigma\} \mapsto E\{f(z)|\nu, \sigma\}$ is invertible. In particular, we require this mapping to be strictly increasing, or, equivalently, that $E\{f(z)|\nu, \sigma\}$ is strictly increasing with ν. This condition supplants the traditional requirement of invertibility of f, which instead we may allow to be nonmonotone. The mappings $\nu \mapsto E\{z|\nu, \sigma\}$ and $\nu \mapsto E\{f(z)|\nu, \sigma\}$ are both necessarily smooth, because of the smoothness of $p(z|\nu, \sigma)$ with respect to ν. The exact unbiased inverse (10) is extended to values $D \notin \{E\{f(z)|\nu, \sigma\}, \nu \in [0, +\infty)\}$ as

$$\begin{aligned} \mathcal{I}_f(D) &= \inf_\nu E\{z|\nu, \sigma\} \quad \text{if } D \leq \inf_\nu E\{f(z)|\nu, \sigma\}, \\ \mathcal{I}_f(D) &= \sup_\nu E\{z|\nu, \sigma\} \quad \text{if } D \geq \sup_\nu E\{f(z)|\nu, \sigma\}. \end{aligned} \tag{11}$$

2.7.2 Estimation of ν

The exact unbiased inverse $\mathcal{I}_f$ can be composed with the mapping $E\{z|\nu, \sigma\} \mapsto \nu$, thus obtaining the *exact unbiased inverse* $\mathcal{V}_f$ *for the estimation of* ν:

$$\mathcal{V}_f : E\{f(z)|\nu, \sigma\} \longmapsto \nu. \tag{12}$$

$$\begin{aligned} \mathcal{V}_f(D) &= 0 \qquad \text{if } D \leq \inf_\nu E\{f(z)|\nu, \sigma\}, \\ \mathcal{V}_f(D) &= +\infty \quad \text{if } D \geq \sup_\nu E\{f(z)|\nu, \sigma\}. \end{aligned} \tag{13}$$

2.7.3 Maximum-likelihood interpretation of $\mathcal{I}_f$ and $\mathcal{V}_f$

Under the rather generic assumption that $D - E\{f(z)|\nu, \sigma\}$ is distributed according to a unimodal distribution with mode at 0, it can be easily shown (with a proof and motivation analogous to that in [14]) that $\mathcal{I}_f(D)$ and $\mathcal{V}_f(D)$ are maximum-likelihood estimates of $E\{z|\nu, \sigma\}$ and ν, respectively. We refer the reader to [14] for further details about this form of inversion.

## 3. VARIANCE-STABILIZING TRANSFORMATIONS FOR RICIAN DATA

3.1 Asymptotics for large ν

Starting from (4) and (5), we can express the variance s^2 as a function of the mean μ (so-called *variance function*) as

$$s^2(\mu) = \text{var}\{z|\mu, \sigma\} = \sigma^2\left(1 - \frac{\sigma^2}{2\mu^2} + O\left(\frac{1}{\mu^4}\right)\right). \tag{14}$$

Table 1: The terms included in the optimization functional (16).

Accuracy of stabilization	$F_{\text{stabil}}(f) = \int_0^{\nu_{\max}} (\text{std}\{f(z)|\nu,1\} - 1)^2 \, d\nu$
Smoothness of f	$F_{\text{smooth}}(f) = \int_0^{z_{\max}} (f''(z))^2 \, dz$
Asymptotic	$F_{\text{asympt}}(f) = \int_0^{z_{\max}} \frac{1}{(z_{\max} - z + \epsilon)^4} \left(f(z) - f_{\text{asympt}}(z) \right)^2 \, dz$
Closeness of $\mathcal{V}_f(f(z))$ to $\nu_{\text{ML}}(z)$	$F_{\text{inverse}}(f) = \int_0^{z_{\max}} \left(\mathcal{V}_f(f(z)) - \nu_{\text{ML}}(z) \right)^2 \, dz$
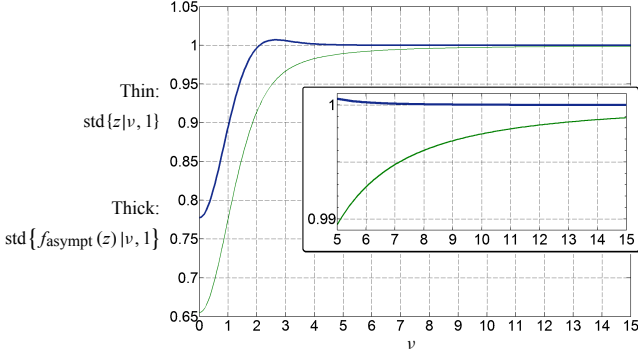
Figure 2: The standard deviation of the stabilized Rician data $\text{std}\{f_{\text{asympt}}(z)|\nu,\sigma\}$ (thick line) versus the standard deviation of the nonstabilized data $\text{std}\{z|\nu,\sigma\}$, for fixed $\sigma = 1$ plotted as function of ν.

This can be used for computing the asymptotic stabilizer for large values of ν (i.e. for large values of μ) as the indefinite integral [6]

$$\int^z \text{var}\{z|\mu,\sigma\}^{-\frac{1}{2}} \, d\mu = \int^z \frac{1}{s(\mu)} d\mu.$$

After neglecting the "big O" term in (14), by solving $\frac{1}{\sigma} \int^z \frac{\mu}{\sqrt{\mu^2 - \frac{\sigma^2}{2}}} d\mu$, we arrive to the primitive

$$f_{\text{asympt}}(z) = \sqrt{\frac{z^2}{\sigma^2} - \frac{1}{2}} + a, \tag{15}$$

where $a \in \mathbb{R}$ is an arbitrary constant and $z \geq \sigma/\sqrt{2}$; for completeness, we can define $f_{\text{asympt}}(z) = a$ for all $z < \sigma/\sqrt{2}$. Figure 2 illustrates the stabilization provided by (15); as can be seen in this figure, the stabilization improves rather quickly with ν. However, the stabilization remains poor for smaller values of ν, as it is natural to expect from a design based on asymptotics.

3.2 Optimization of the stabilizer for finite ν

To achieve good stabilization also for small values of ν, we resort to the numerical direct optimization procedure [8, 9]. Without loss of generality, let $\sigma = 1$. We consider a finite parameter range $\Theta = [0, \nu_{\max}]$. Because of the exponential decay of the p.d.f. (1), only a finite range of values $Z = [0, z_{\max}]$ is numerically relevant for Θ; values $z > z_{\max}$ are treated by defining $f(z) = f_{\text{asympt}}(z)$ (15), with the constant $a = f(z_{\max}) - \sqrt{z_{\max}^2 - \frac{1}{2}}$.

The direct optimization [8, 9] operates by progressively modifying the stabilizer f over Z with the goal of minimizing a cost functional $F(f) : f \mapsto \mathbb{R}$. This functional F comprises a main term F_{stabil}, which incorporates the accuracy of stabilization over Θ, and a few penalty terms, which enable the additional properties needed for the practical use of stabilizer. The various terms are given in Table 1, and the overall functional F is obtained by summing these terms as

$$F(f) = F_{\text{stabil}}(f) + \lambda_{\text{smooth}} \cdot F_{\text{smooth}}(f) + \\ + \lambda_{\text{asympt}} \cdot F_{\text{asympt}}(f) + \lambda_{\text{inverse}} \cdot F_{\text{inverse}}(f), \tag{16}$$

where $\lambda_{\text{smooth}}, \lambda_{\text{inverse}}, \lambda_{\text{asympt}} \geq 0$ are penalty parameters. In order to guarantee the existence of the exact unbiased inverses $\mathcal{I}_f$ (10) and $\mathcal{V}_f$ (12), the optimization is constrained to functions for which the mapping $E\{z|\nu,1\} \mapsto E\{f(z)|\nu,1\}$ is strictly increasing. Without loss of generality, $f(0) = 0$.

Let us explain the role of the various penalty terms in (16) (see also Table 1). Firstly, F_{smooth} regularizes the stabilizer, penalizing oscillating solutions. While F_{stabil} appreciates only the stabilization for $\nu \in \Theta$, the largest values $z \in Z$ influence more the stabilization for $\nu > \nu_{\max}$ than that for $\nu \in \Theta$; F_{asympt} forces f to approach f_{asympt} as z approaches $z_{\max}$, thus preserving the good stabilization provided by f_{asympt} for $\nu > \nu_{\max}$. The epsilon $\epsilon = 2 \times 10^{-16}$ in the denominator is used solely for ensuring numerical well posedness. The role of F_{inverse} is more subtle as it operates directly over the exact unbiased inverse $\mathcal{V}_f$ of f; specifically, if there are singularities that cannot be filtered by Φ, i.e. $D = \Phi(f(z)) \approx f(z) \neq E\{f(z)|\nu,1\}$, then $\mathcal{V}_f(D)$ can be quite different from $\nu = \mathcal{V}_f(E\{f(z)|\nu,1\})$; F_{inverse} ensures that in such a case $\mathcal{V}_f(D)$ will not be far from $\nu_{\text{ML}}(z)$, which, in maximum-likelihood sense, is the best estimate that we can have out of an individual z sample alone.

Different penalty parameters correspond to a different optimized stabilizer $\arg\min_f F(f)$. In this paper, we use the two optimized stabilizers shown in Figure 3, which for brevity are referred to as the stabilizers "A" and "B".

3.3 Exact unbiased inverse

The exact unbiased inverse transformations $\mathcal{I}_f$ and $\mathcal{V}_f$ for the optimized f can be computed by evaluating numerically $E\{f(z)|\nu,1\}$ with $\nu \in [0, \nu_{\max}]$. For $D > \max_{\nu \in [0,\nu_{\max}]} E\{f(z)|\nu,1\} = E\{f(z)|\nu_{\max},1\}$, we can consider its asymptotic expression for large values of D. In particular, it can be shown [1] that the difference between the exact unbiased inverse $\mathcal{I}_f$ and the algebraic inverse f^{-1} is asymptotically zero. Thus, from (15), we have for large D

$$\mathcal{I}_f(D) \approx f_{\text{asympt}}^{-1}(D) = \sigma\sqrt{(D-a)^2 + \frac{1}{2}}. \tag{17}$$

Further, by leveraging asymptotic expansions of (4) we obtain

$$\mathcal{V}_f(D) = \mathcal{I}_f(D)\left(1 - \frac{\sigma^2}{2\mathcal{I}_f^2(D)} + \dots\right) \approx \frac{\sigma(D-a)^2}{\sqrt{(D-a)^2 + \frac{1}{2}}}.$$

For values of D outside of the range of $E\{f(z)|\nu,\sigma\}$, $\nu \in \mathbb{R}^+$, the definitions (11) and (13) are always valid. Of course, $\inf_\nu E\{z|\nu,\sigma\} = E\{z|0,\sigma\}$, $\sup_\nu E\{z|\nu,\sigma\} = +\infty$, and, because of (15), $\sup_\nu E\{f(z)|\nu,\sigma\} = +\infty$. Further, since the mapping $E\{f(z)|\nu,1\}$ is strictly increasing with ν, we also have $\inf_\nu E\{f(z)|\nu,\sigma\} = E\{f(z)|0,\sigma\}$. Note that the infima are attained. Therefore, the range $\{E\{f(z)|\nu,\sigma\}, \nu \geq 0\} = [E\{f(z)|0,\sigma\}, +\infty)$, and for $D \leq E\{f(z)|0,\sigma\}$ we have $\mathcal{I}_f(D) = E\{z|0,\sigma\}$ and $\mathcal{V}_f(D) = 0$.

Exact unbiased inverses $\mathcal{V}_f(D)$ corresponding to the stabilizers in Figure 3 are plotted within the same figure.

4. NOISE-LEVEL MISMATCH

Up to now, we assumed that the parameter σ was known exactly. However, in practice, there can be some mismatch between the true value of σ and the value $\hat{\sigma} = \lambda\sigma$ assumed in place of σ: it is then
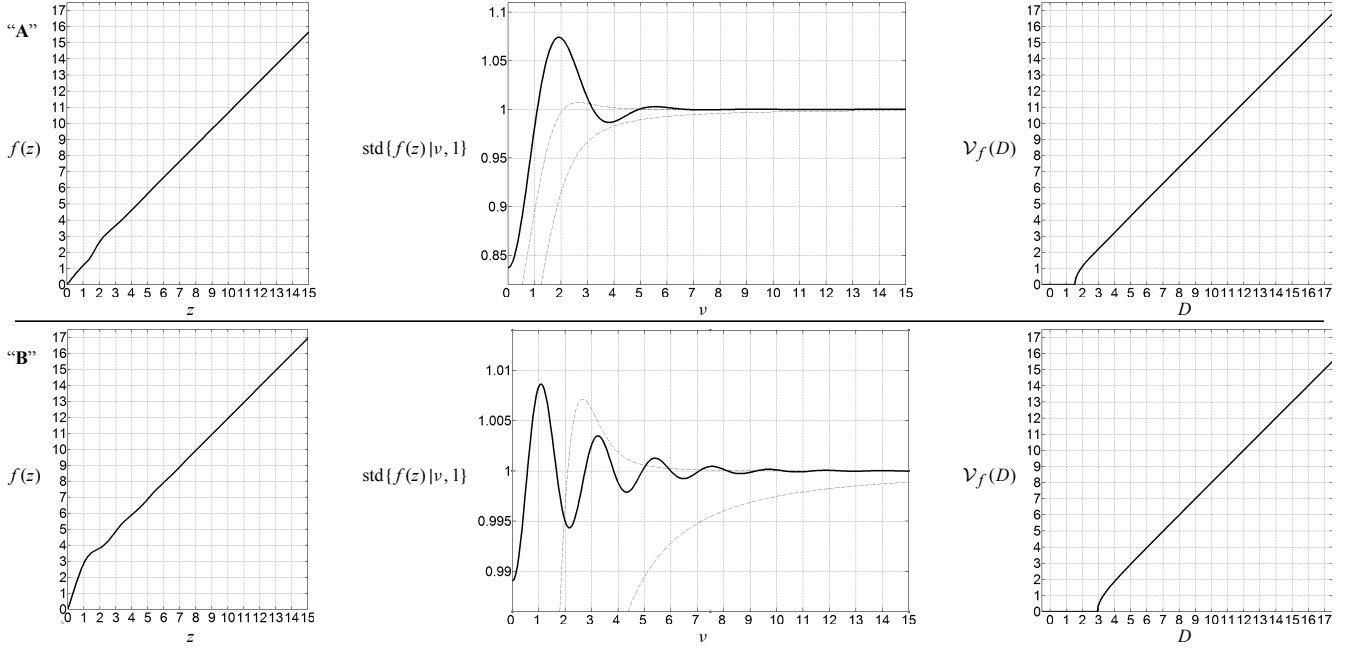
Figure 3: Optimized stabilizers obtained with penalty parameters (top) $\lambda_{\text{asympt}} = 1$, $\lambda_{\text{smooth}} = 10^{-2}$, $\lambda_{\text{inverse}} = 10^{-\frac{1}{2}}$, and (bottom) $\lambda_{\text{asympt}} = 1$, $\lambda_{\text{smooth}} = 10^{-4}$, $\lambda_{\text{inverse}} = 0$: (from left to right) f, $\text{std}\{f(z)|\nu,\sigma\}$, and the exact unbiased inverse $\mathcal{V}_f$, for $\sigma = 1$. The thin dashed curves in the center plots represent $\text{std}\{f_{\text{asympt}}(z)|\nu,1\}$ and $\text{std}\{z|\nu,1\}$.
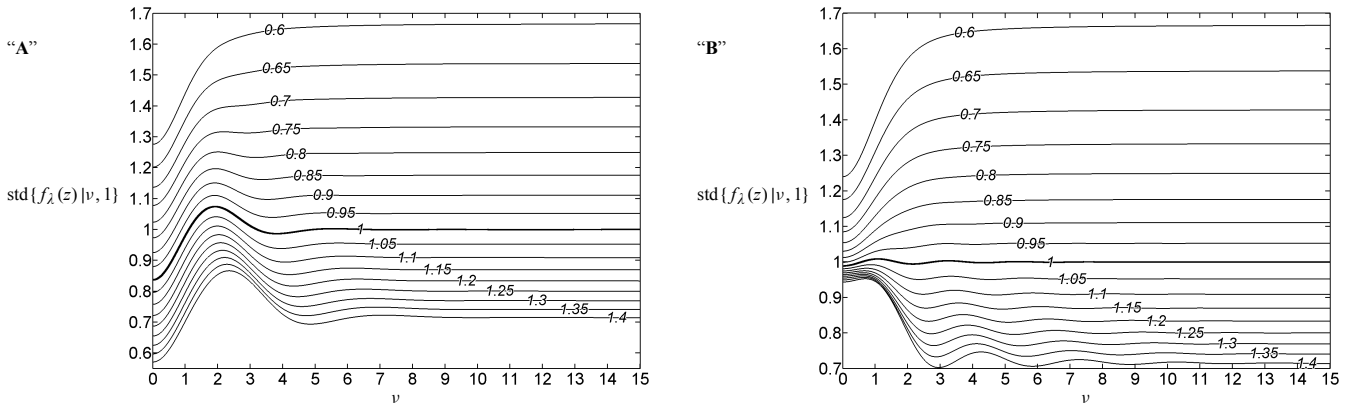

Figure 4: Noise-level mismatch. Standard deviation of the transformed data $\text{std}\{f_\lambda(z)|\nu,1\}$, for different values of λ, as indicated by the italic numbers superimposed on the curves. The plots correspond to the stabilizers f of Figure 3.

natural to ask ourselves how well can $f_{\hat{\sigma}}$ (7) be used for stabilizing $\mathcal{R}(\cdot,\sigma)$, if f is the stabilizer for $\mathcal{R}(\cdot,1)$.

Figure 4 shows the standard deviations $\text{std}\{f_\lambda(z)|\nu,1\} = \text{std}\{f_{\lambda\sigma}(z)|\sigma\nu,\sigma\}$ of the transformed data for a wide range of values of λ, where f is the stabilizer "A" or "B" from Figure 3. Values of λ smaller or larger than 1 correspond, respectively, to underestimation or overestimation of the standard-deviation parameter σ in the definition of the distribution of z (1)-(3).

From (7) and (15), follows that f_λ is asymptotically affine for large z, with derivative approaching $\frac{1}{\lambda}$. Therefore, as can be seen from the plots in Figure 4, we have

$$\text{std}\{f_{\lambda\sigma}(z)|\sigma\nu,\sigma\} = \text{std}\{f_\lambda(z)|\nu,1\} \xrightarrow[\nu\to+\infty]{} \frac{1}{\lambda}. \qquad (18)$$

5. NOISE-LEVEL ESTIMATION

Exploiting the results from the previous sections, we devise a general iterative scheme based on variance stabilization aimed at esti-

mating the value of the σ parameter in (1)-(2) from a single realization z.

Let $\mathfrak{E}$ denote an estimator of the standard deviation σ of the homoskedastic noise corrupting a signal. Popular examples for estimating σ of additive white Gaussian noise (AWGN) in natural images are the median or mean absolute deviation of the high-pass filtered signal:

$$\mathfrak{E}_{\text{MedianAD}}\{z\} = \text{med}\{|H\{z\}|\}/\Phi^{-1}(3/4), \qquad (19)$$

$$\mathfrak{E}_{\text{MeanAD}}\{z\} = \text{mean}\{|H\{z\}|\}\sqrt{\pi/2}, \qquad (20)$$

where $H\{z\} = z \circledast w_{\text{hi}}$, and w_{hi} is a high-pass convolutional kernel having zero mean and unit L^2-norm,

$$\int w_{\text{hi}} = 0, \qquad \int |w_{\text{hi}}|^2 = 1,$$

such as, e.g., a wavelet function.

5.1 Iterative scheme for estimating σ

The proposed scheme is expressed by the following recursive system:

$$\begin{cases} \hat{\sigma}_1 = \mathfrak{E}\{z\}, \\ \hat{\sigma}_{k+1} = \mathfrak{E}\{f_{\hat{\sigma}_k}(z)\}\hat{\sigma}_k, \quad k \geq 1. \end{cases} \quad (21)$$

The idea of this recursion originates from (18). The estimate $\hat{\sigma}_k$ is used to define a variance-stabilizing transformation for z. If the estimated value $\hat{\sigma}_k$ is correct, then the transformation $f_{\hat{\sigma}_k}$ successfully stabilizes the data and when $\mathfrak{E}$ is applied to the stabilized data it should return a value $\mathfrak{E}\{f_{\hat{\sigma}_k}(z)\}$ close to 1. If the estimated value $\hat{\sigma}_k$ is not correct (e.g., an under-estimate of σ), then the stabilization is not accurate, being roughly the inverse of the misestimation ratio (18), $\mathfrak{E}\{f_{\hat{\sigma}_k}(z)\} \approx \frac{\sigma}{\hat{\sigma}_k}$. Hence, we correct the current estimate $\hat{\sigma}_k$ by multiplying it by $\mathfrak{E}\{f_{\hat{\sigma}_k}(z)\}$. Observe that if $\mathfrak{E}\{f_{\hat{\sigma}}(z)\} = 1$ for some value $\hat{\sigma}$, then this $\hat{\sigma}$ is a fixed point for (21) and we want the sequence $\hat{\sigma}_k$ to converge to such $\hat{\sigma}$. The system (21) is initialized by the estimator $\mathfrak{E}$ applied on the non-stabilized data z.

In practice, we can guarantee converge with exponential rate to an accurate and stable estimate $\hat{\sigma}$ of the true value σ by application of the contraction mapping theorem. Due to space limitation, here below we provide only a sketch of the proof under a simplified modelling.

5.2 Convergence of the iterative scheme for estimating σ

The study of convergence is based on a perturbation analysis of the univariate mapping $\mathfrak{M}_z : \hat{\sigma}_k \mapsto \hat{\sigma}_{k+1}$ defined by

$$\mathfrak{M}_z(\hat{\sigma}) = \mathfrak{E}\{f_{\hat{\sigma}}(z)\}\hat{\sigma},$$

modeling $\mathfrak{E}\{f_{\hat{\sigma}}(z)\} \approx \text{mean}_{x \in X}\{\text{std}\{f_{\hat{\sigma}}(z)|v(x), \sigma\}\}$. To give a rough and very preliminary intuition of our study, let us consider the case where $\mathfrak{E}\{f_{\sigma}(z)\} = 1$. This case would arise under the ideal hypotheses that the stabilization is exact, i.e. $\text{std}\{f_{\sigma}(z)|v, \sigma\} \equiv 1$, and that the estimator $\mathfrak{E}$ successfully returns this value, i.e. $\mathfrak{E}\{f_{\sigma}(z)\} = \text{std}\{f_{\sigma}(z)|v, \sigma\} \equiv 1$. In this case, σ is a fixed point for the mapping $\mathfrak{M}_z$, i.e. $\sigma = \mathfrak{M}_z(\sigma)$, and any of the following conditions may be leveraged to ensure convergence of the sequence $\hat{\sigma}_k$ (21) to σ:

1. $\mathfrak{M}_z$ is a contraction, i.e. there exists $\gamma \in [0, 1)$ such that $|\mathfrak{M}_z(\sigma') - \mathfrak{M}_z(\sigma'')| \leq \gamma |\sigma' - \sigma''|$ for any pair σ', σ''.

2. $\mathfrak{M}_z$ is continuous and $|\mathfrak{M}_z(\hat{\sigma}) - \sigma| < |\hat{\sigma} - \sigma|$ for $\hat{\sigma} \neq \sigma$.

These conditions are met within a neighborhood of σ provided that $\mathfrak{M}_z$ is smooth and $\left|\frac{\partial}{\partial\hat{\sigma}}\mathfrak{M}_z(\hat{\sigma})\right|_{\hat{\sigma}=\sigma}| < 1$. Simple derivations yield $\frac{\partial}{\partial\hat{\sigma}}\mathfrak{M}_z(\hat{\sigma})\big|_{\hat{\sigma}=\sigma} = \sigma\frac{\partial}{\partial\hat{\sigma}}\mathfrak{E}\{f_{\hat{\sigma}}(z)\}\big|_{\hat{\sigma}=\sigma} + 1$. If we substitute $\mathfrak{E}\{f_{\hat{\sigma}}(z)\}$ with $\text{mean}_{x \in X}\{\text{std}\{f_{\hat{\sigma}}(z)|v(x), \sigma\}\}$, the bound on $\frac{\partial}{\partial\hat{\sigma}}\mathfrak{M}_z(\hat{\sigma})\big|_{\hat{\sigma}=\sigma}$ can be replaced by $\left|\frac{\partial}{\partial\hat{\sigma}}\text{std}\{f_{\hat{\sigma}}(z)|v, \sigma\}\big|_{\hat{\sigma}=\sigma} + \frac{1}{\sigma}\right| < \frac{1}{\sigma}$, $v \geq 0$. Upon scaling, this coincides to $\left|\frac{\partial}{\partial\lambda}\text{std}\{f_\lambda(z)|\frac{v}{\sigma}, 1\}\big|_{\lambda=1} + 1\right| < 1$, $v \geq 0$. Thus, under the above ideal hypotheses, the iterative scheme (21) can successfully be used to estimate σ provided that

$$-2 < \frac{\partial}{\partial\lambda}\text{std}\{f_\lambda(z)|v, 1\}\Big|_{\lambda=1} < 0, \quad v \geq 0. \quad (22)$$

Figure 5 gives an illustration of the $\frac{\partial}{\partial\lambda}\text{std}\{f_\lambda(z)|v, 1\}\big|_{\lambda=1}$ for the stabilizers "A" and "B"; it can be seen that the inequality (22) indeed holds for both transformations.

6. EXPERIMENTS

This section is structured in three parts. First, we consider the denoising problem assuming exact knowledge of σ; second, we focus on the problem of estimating the value of σ from a given image; third, we present denoising experiments where the stabilization is made with an estimate $\hat{\sigma}$ of σ, in order to illustrate the robustness
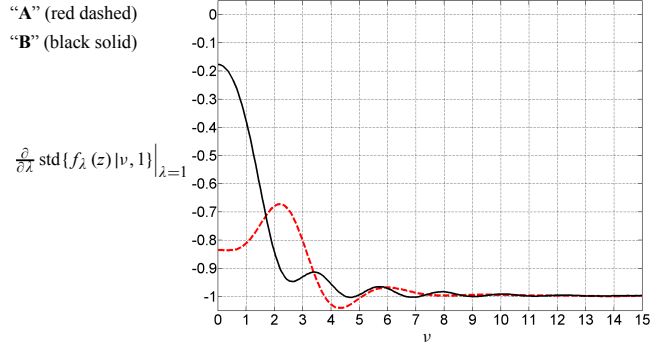

Figure 5: Partial derivative of the standard deviation of the stabilized Rician data $\text{std}\{f(z)|v, \sigma\}$ for fixed $\sigma = 1$ with respect to the σ value asssumed by the stabilizer, plotted as function of v.

with respect to stabilization based on an imprecise value of σ. As test data, we use the T1 phantom of size $181\times217\times181$ from the BrainWeb dataset [2] corrupted with different levels of Rician noise.

6.1 Denoising (exact σ)

To validate our stabilization approach, we compare the denoising performance of a state-of-the-art filter specifically designed for Rician noise removal against its counterpart designed for Gaussian observations. The latter filter is applied between our forward variance stabilizing transformation "A" and the corresponding exact unbiased inverse. In particular, as filters we use the optimized blockwise volumetric NLmeans algorithm with wavelet mixing (OB-NLM3D-WM) [4], in its Gaussian [3] and Rician [17] versions. Here the stabilizer is scaled according to the exact noise level σ. For the sake of memory requirements when using the OB-NLM3D-WM algorithms [12], only the $181\times217\times51$ middle portion of the phantom was processed. The PSNR results are given in rows 2 and 4 of Table 2. There are only marginal differences between the results obtained with the Rician version of the algorithm and those obtained using the Gaussian version inserted in the proposed variance-stabilization framework.

6.2 Noise estimation

To validate our iterative scheme for estimating σ, we compare the state-of-the-art robust estimator [5, 13] for Rician data, which is based on the $\mathfrak{E}_{\text{MedianAD}}$ (19) and exploits both automatic segmentation and iterative SNR correction [11] in order to achieve unbiased estimation for the Rice distribution, against its direct counterpart for additive white Gaussian noise (obtained by disabling the iterative correction [11]), which we use at every iteration of the scheme (21) in place of the operator $\mathfrak{E}$. For the variance stabilization, we use the optimized transformation "B". About three to five iterations of (21) are usually sufficient to reach numerical convergence of $\hat{\sigma}_k$ with a relative precision (stopping rule) of $\frac{|\hat{\sigma}_k - \hat{\sigma}_{k-1}|}{\hat{\sigma}_k} < 10^{-4}$.

Figure 6 shows the average relative error $1 - \frac{\hat{\sigma}}{\sigma}$ over 10 independent replications where the estimated values $\hat{\sigma}$ are obtained with our approach ("VST + Gaussian MAD") and by the method [5] ("Rician MAD"). As can be seen in the plots, there is no essential difference in performance between the two methods, which confirms that our recursive technique based on variance stabilization is successful in enabling the accurate estimation of $\hat{\sigma}$ from Rician observations using estimators of the standard deviation designed for additive Gaussian noise.

6.3 Denoising with estimated σ

Here we repeat the experiments of Section 6.1 with the only modification that the value of the noise level σ assumed by the transformations in not the exact one but the one estimated by the proposed estimation algorithm, as detailed in Section 6.2. The PSNR results

σ (%)	1	3	5	7	9	11	13	15
Noisy observations z	40.04	30.52	26.11	23.21	21.05	19.32	17.89	16.66
Rician OB-NLM3D-WM with exact σ	43.14	38.14	35.18	33.11	31.51	30.19	29.04	28.02
Rician OB-NLM3D-WM with estimated $\hat{\sigma}$	43.15	38.14	35.18	33.11	31.51	30.18	29.04	28.01
VST + Gaussian OB-NLM3D-WM with exact σ	43.21	38.25	35.25	33.14	31.52	30.19	29.03	27.99
VST + Gaussian OB-NLM3D-WM with estimated $\hat{\sigma}$	43.22	38.25	35.24	33.14	31.52	30.15	29.01	27.97

Table 2: Denoising of Rician-distributed volumetric phantom data. PSNR (dB) results obtained using the optimized blockwise volumetric NLmeans denoising algorithm with wavelet mixing in its version specifically designed for Rician data (Rician OB-NLM3D-WM) versus those obtained using the proposed variance-stabilization framework (VST) combined with the standard version of the denoising algorithm, which is designed for Gaussian data (Gaussian OB-NLM3D-WM).
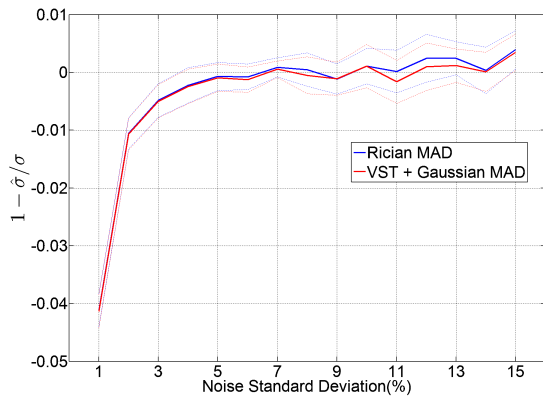

Figure 6: Noise-level estimation performance as $1 - \frac{\hat{\sigma}}{\sigma}$. The thin dashed lines show the unit deviation from the mean result (thick line) over 10 independent replications.

are given in rows 3 and 5 of Table 2. There is basically no significant difference between these results and those obtained using the exact value of σ (rows 2 and 4 of the table).

7. CONCLUSIONS

We developed optimized variance-stabilizing transformations for the Rician distribution, as well as the corresponding exact unbiased inverse transformations. This makes possible the successful application of denoising algorithms designed for filtering data corrupted by AWGN for the more challenging problem of MR image filtering. We have also verified the stability of the variance stabilization with respect to misestimation of the noise-level parameter. From this analysis we derived an algorithm for estimating the noise level from a single Rician-distributed image which is based on conventional noise standard-deviation estimators for AWGN.

The developed framework delivers state-of-the-art results in both estimation and removal of Rician noise using simpler algorithms designed for AWGN instead of ad-hoc algorithms specifically designed for Rician-distributed data.

An open-source Matlab implementation of the proposed framework is provided at [10].

Ongoing research is aimed at generalizing these results to generic exponential distribution families.

8. ACKNOWLEDGMENT

We would like to thank Eric Diaz (University of California, San Diego) for drawing the author's attention to the problem of MR filtering, and José Manjón and Pierrick Coupé for making their software available to the public.

REFERENCES

[1] Bar-Lev, S.K., and P. Enis, "On the construction of classes of variance stabilizing transformations", *Statistics & Probability Letters*, vol. 10, no. 2, pp. 95–100, July 1990.

[2] BrainWeb: Simulated Brain Database, available online at http://www.bic.mni.mcgill.ca/brainweb

[3] Coupé, P., P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An Optimized Blockwise NonLocal Means Denoising Filter for 3-D Magnetic Resonance Images", *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 425–441, 2008.

[4] Coupé, P., S. Prima, P. Hellier, C. Kervrann, C. Barillot, "3D Wavelet Sub-Bands Mixing for Image Denoising", *Int. Journal of Biomedical Imaging*, 2008.

[5] Coupé P., J. Manjón, E. Gedamu, D. Arnold, M. Robles, and D. L. Collins, "Robust Rician noise estimation for MR images", *Medical Image Analysis*, vol. 14, no. 4, pp. 483–493, Aug. 2010.

[6] Curtiss, J.H., "On transformations used in the analysis of variance", *The Annals of Mathematical Statistics*, vol. 14, no. 2, pp. 107–122, June 1943.

[7] Foi, A., "Clipped noisy images: heteroskedastic modeling and practical denoising", *Signal Processing*, vol. 89, no. 12, pp. 2609–2629, Dec. 2009. doi:10.1016/j.sigpro.2009.04.035

[8] Foi, A., "Optimization of variance-stabilizing transformations", preprint available at http://www.cs.tut.fi/~foi/.

[9] Foi, A., "Direct optimization of nonparametric variance-stabilizing transformations", presented at *8èmes Rencontres de Statistiques Mathématiques, CIRM*, Luminy, Dec. 2008.

[10] Foi, A., RiceOptVST − Matlab software for Rician noise estimation and removal, available online at http://www.cs.tut.fi/~foi/RiceOptVST/, 2011.

[11] Koay, C.G., and P.J. Basser, "Analytically exact correction scheme for signal extraction from noisy magnitude MR signals", *Journal of Magnetic Resonance*, vol. 179, no. 2, pp. 317–322, April 2006.

[12] Manjón, J., Matlab software for Non-Local Means denoising, version 1.0, February 2010, available at http://personales.upv.es/jmanjon/denoising/arnlm.html

[13] Manjón, J., Matlab software for robust Rician noise estimation, version 1.0, March 2010, available at http://personales.upv.es/jmanjon/analysis/noise.htm

[14] Mäkitalo, M., and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising", *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99–109, Jan. 2011.

[15] Meyer, P.L., "The Maximum Likelihood Estimate of the Non-Centrality Parameter of a Non-Central χ^2 Variate", *J. Amer. Statis. Assoc.*, vol. 62, no. 320, pp. 1258–1264, Dec. 1967.

[16] Watson G.N., *Treatise on the Theory of Bessel Functions*, second edition, Cambridge University Press, 1944.

[17] Wiest-Daesslé, N., S. Prima, P. Coupé, S.P. Morrissey, C. Barillot, "Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI", *Proc. 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2008*, pp. 171–179, New York, Sep. 2008.