35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Merging squared-magnitude approaches to DWI denoising: An adaptive Wiener filter tuned to the anatomical contents of the image

Antonio Tristán-Vega[1], Véronique Brion[2], Gonzalo Vegas-Sánchez-Ferrero[1], Santiago Aja-Fernández[1]

***Abstract*— We present a new method for denoising of Diffusion Weighted Images (DWI) that shares several desirable features of state-of-the-art proposals: 1) it works with the squared-magnitude signal, allowing for a closed-form formulation as a Linear Minimum Mean Squared Error (LMMSE) estimator, a.k.a. Wiener filter; 2) it jointly accounts for the DWI channels altogether, being able to unveil anatomical structures that remain hidden in each separated channel; 3) it uses a Non-Local Means (NLM)-like scheme to discriminate voxels corresponding to different fiber bundles, being able to enhance the anatomical structures of the DWI. We report extensive experiments evidencing the new approach outperforms several related methods for all the range of input signal-to-noise ratios (SNR). An open-source C++ implementation of the algorithm is also provided for the sake of reproducibility.**

***Index Terms*— Diffusion MRI, denoising, NLM, LMMSE.**

I. Introduction

The noise pattern in Magnetic Resonance Imaging (MRI) has been modeled as Rician distributed, as long as the acquired image is the real envelope of a complex Gaussian [1]:

$$M = |(A_c + n_c) + j(A_s + n_s)|;\ A = |A_c + jA_s|, \qquad (1)$$

where A is the noise-free amplitude and n_c, n_s are Gaussians with zero mean and variance σ^2. Though MRI data show in general a large SNR, several methods for MRI denoising have been proposed including wavelet shrinkage [2], anisotropic diffusion [3], or maximum likelihood estimation [4], among others. When it comes to DWI, the SNR is typically much lower due to the strong attenuations introduced by the diffusion gradients, leading to a well-known bias in the quantitative parameters derived from Diffusion Tensor Images (DTI), such as the Mean Diffusivity (MD) or the Fractional Anisotropy (FA) [5]. Accordingly, denoising of DWI volumes has grabbed a great deal of attention in the recent literature. Filtering methods are usually based on the statistical properties of Rician noise, which involve signal-dependent moments that yield to non-linear and iterative processing [6]. For this reason, it is often advantageous to work with the squared envelope M^2, whose moments are trivially related to the signal A and the power of noise σ^2 [7]:

$$\begin{aligned} A^2 &= E\{M^2\} - 2\sigma^2; \\ A^4 &= E\{M^4\} - 8\sigma^2 E\{M^2\} + 8\sigma^4. \end{aligned} \qquad (2)$$

A number of denoising schemes for DWI have been designed with this philosophy [8], [9], aimed at the simultaneous fulfillment of three requirements: 1) An appropriate filter should account for the very low SNR inherent to DWI data; 2) It should combine the information present in all channels to reveal the overall structure of the white matter, and 3) It should preserve the fine details of small fiber bundles. As an example, the Wiener filter proposed in [10] uses the first two moments of M^2 to linearly estimate A^2:

$$\widehat{\mathbf{A}^2} = \langle \mathbf{M}^2 \rangle - 2\sigma^2 \cdot \mathbf{1} + \mathbf{C}_{\mathbf{A}^2\mathbf{M}^2} \mathbf{C}^{-1}_{\mathbf{M}^2\mathbf{M}^2} \left(\mathbf{M}^2 - \langle \mathbf{M}^2 \rangle \right), \qquad (3)$$

where the $N \times 1$ vectors $\mathbf{A}^2$ and $\mathbf{M}^2$ comprise the corresponding A^2 and M^2 from each DWI channel, $\mathbf{1}$ is an $N \times 1$ vector of all ones, and $\langle \cdot \rangle$ denotes the sample average in a neighborhood of the filtered voxel, which approximates the true expectation $E\{\cdot\}$. The covariance matrices $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2}$ and $\mathbf{C}_{\mathbf{M}^2\mathbf{M}^2}$, which are also estimated from sample averages, represent the correlations between the different DWI channels, i.e. the anatomy of the white matter hidden behind their joint information. At the same time, the linear term in M^2 accounts for the signal variability in the neighborhood, hence avoiding an excessive structural blurring. Indeed, eq. (3) may be seen as a general form for many DWI filtering techniques: in case $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2} \mathbf{C}^{-1}_{\mathbf{M}^2\mathbf{M}^2}$ is assumed diagonal, inter-channel dependencies are neglected and we get the channel-by-channel LMMSE filter in [8]. Going a step beyond, the linear correction may be neglected assuming $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2} = \mathbf{0}$, and we get the so-called Conventional Approach (CA) in [7]. However, this is too rough an assumption unless we have a homogeneous anatomy inside the whole neighborhood, and otherwise the image structure will become blurry. To palliate this artifact, the signal variability may be dramatically reduced by using the Unbiased NLM (UNLM) scheme, implying that only those voxels whose structure is similar enough are averaged in eq. (3) [9], [10].

In Section II we present an estimator for A^2 that combines the better noise removal ability of the LMMSE model with the better ability to preserve details of UNLM (see [10]). The statistical model in Section II-A inherits from eq. (3), but the sample moments are computed as non-homogeneous, NLM-like mixtures based on a novel structural DWI similarity measure introduced in Section II-B. Fine details are better preserved since the signal variability is accounted for through both $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2}$ and the shaped vicinities based on the NLM idea. Conversely, noise removal is also more efficient since all the channels are mixed together through $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2}$, and NLM weights allow using larger vicinities. These assertions are supported by the experiments in Section III, while the pros and cons of our proposal are discussed in Section IV.

[1]Laboratory of Image Processing, University of Valladolid (Spain). [2]NeuroSpin, CEA/DSV/I2BM, Gif-sur-Yvette, France. Work partially funded by grant numbers TEC2010-17982 from the Ministerio de Ciencia y Educación (Spain) and VA376A11-2, SAN103/VA40/1 from the Junta de Castilla y León (Spain). **Contact:** atriveg@lpi.tel.uva.es [www.lpi.tel.uva.es].

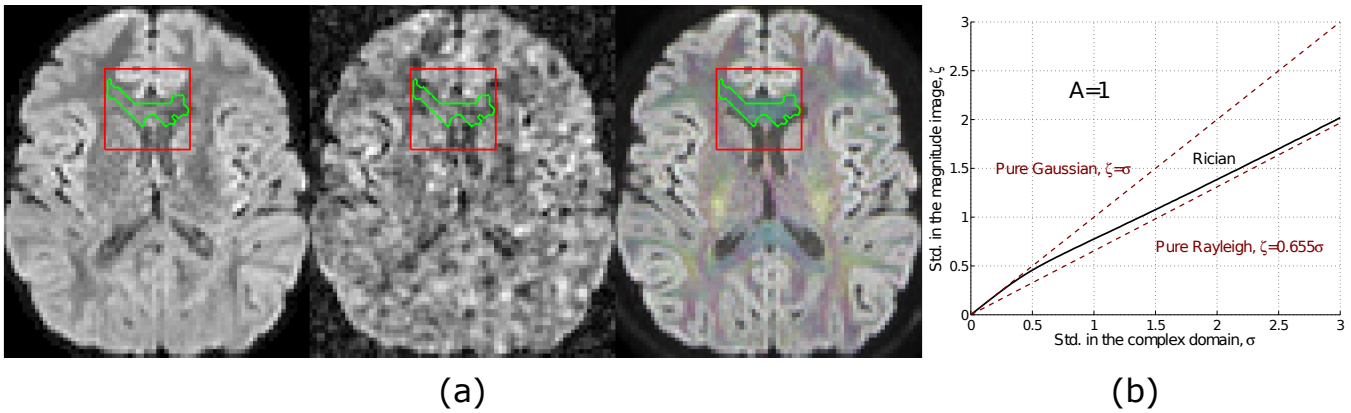

Fig. 1. **(a)** An axial slice of the *noise-free* phantom used in Section III (left), together with the noisy phantom (middle) and the RGB anatomical map obtained from the latter. An isotropic search window is centered at a voxel in the Corpus Callosum (red), and the voxels weighted with non-negligible factors highlighted in green. **(b)** Standard deviation of a Rician variable as a function of the standard deviation in the complex domain.

II. The Joint Anisotropic LMMSE Filter for DWI

A. Linear filtering model

According to the previous discussion, we compute $\widehat{\mathbf{A}^2}$ from $\mathbf{M}^2$ as the linear estimator with minimum squared error:

$$\widehat{\mathbf{A}^2} = \langle \mathbf{M}^2 \rangle_{\mathcal{N}} - 2\sigma^2 \cdot \mathbf{1} + \mathbf{C}_{\mathbf{A}^2\mathbf{M}^2} \mathbf{C}^{-1}_{\mathbf{M}^2\mathbf{M}^2} \left(\mathbf{M}^2 - \langle \mathbf{M}^2 \rangle_{\mathcal{N}} \right), \quad (4)$$

where the sample moments $\langle \cdot \rangle_{\mathcal{N}}$ are now computed in an anisotropic vicinity $\mathcal{N}$ as discussed in Section II-B. $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2}$ and $\mathbf{C}_{\mathbf{M}^2\mathbf{M}^2}$ may be represented in different ways: 1) The estimation proposed in [6] is assumption-free. 2) In [8], it is assumed that no correlations exist between the channels. 3) As opposed, we assume the channels are completely correlated, so the covariance matrices become [10]:

$$\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2} = \varsigma \langle \mathbf{A}^2 \rangle_{\mathcal{N}} \langle \mathbf{A}^2 \rangle^T_{\mathcal{N}}; \quad (5)$$

$$\mathbf{C}_{\mathbf{M}^2\mathbf{M}^2} = \varsigma \langle \mathbf{A}^2 \rangle_{\mathcal{N}} \langle \mathbf{A}^2 \rangle^T_{\mathcal{N}} + 4\sigma^2 \mathrm{diag}(\langle \mathbf{A}^2 \rangle_{\mathcal{N}}) + 4\sigma^4 I_N, \quad (6)$$

where $\varsigma = \frac{1}{N} \sum_{i=1}^{N} (\langle A_i^4 \rangle_{\mathcal{N}} - \langle A_i^2 \rangle^2_{\mathcal{N}}) / \langle A_i^2 \rangle^2_{\mathcal{N}}$. This assumption properly accounts for the joint information in the DWI, and grants also an efficient computation and inversion of $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2}$. In practice, we split the estimates $\widehat{\mathbf{A}^2}$ in those channels corresponding to either baselines or actual DWIs.

B. From anatomical contents to adaptive neighborhoods

With the same philosophy behind the NLM, we define the moments of $I(\mathbf{x}_i)$ at $\mathbf{x}_i$ inside a shaped vicinity $\mathcal{N}$ as:

$$\langle I(\mathbf{x}_i) \rangle_{\mathcal{N}} = \frac{1}{Z} \sum_{l \in \mathcal{N}} \exp\left(-\frac{d(\mathbf{x}_i, \mathbf{x}_l)}{h \cdot \alpha^2} \right) I(\mathbf{x}_l), \quad (7)$$

where $d(\mathbf{x}_i, \mathbf{x}_l)$ is a *distance* between the voxels $\mathbf{x}_i$ and $\mathbf{x}_l$ that measures their structural similarity, α^2 is the expected value of such *distance* for similar voxels, and h is a free parameter. The normalization constant Z assures the mean value of I is preserved. The definition of $d(\cdot,\cdot)$ may also vary, but it should fulfill certain conditions:

- As shown in Fig. 1(a), an isolated DWI channel cannot reveal the white matter structure, even when the SNR is very high. For this reason, distances based on the actual anisotropy of the white matter like [11] are preferable to those based on the raw DWI channels, like [9], [10].
- $E\{d(\cdot,\cdot)\}$ must be statistically characterized to compute α^2 [12], so non-linear channel combinations like that in [11] are less desirable than linear mixtures.

To meet these requirements, let us define three unit, mutually orthogonal vectors (we use the canonical basis), $\{\mathbf{u}_R, \mathbf{u}_G, \mathbf{u}_B\}$, and let us assume the unit gradient used to acquire the i-th DWI is $\mathbf{g}_i$. A RGB image is formed by combining the DWI channels M_i depending on their orientations; for the R channel (and conversely for the G and B channels):

$$R(\mathbf{x}) = \sum_{i=1}^{N} w_{R_i} \cdot M_i(\mathbf{x}), \;\; w_{R_i} = \left(\sum_{i=1}^{N} |\mathbf{u}_R^T \mathbf{g}_i| \right)^{-1} |\mathbf{u}_R^T \mathbf{g}_i|. \quad (8)$$

Contrarily to [11], this anatomical map in Fig. 1(a) robustly characterizes both the anisotropy and the directionality of the white matter, even when computed from low SNR data. The variance of each RGB channel is easy to compute, since it is a weighted sum of N independent Rician variables. Though the variance ζ^2 of a Rician variable M is not trivially related to σ^2 in eq. (1), see Fig. 1(b), for a reasonable SNR ($A/\sigma > 2$ in the figure) we may set $\zeta \simeq \sigma$. From [12], and assuming the RGB channels are roughy uncorrelated, we write in eq. (7):

$$\frac{d(\mathbf{x}_i, \mathbf{x}_l)}{\alpha^2} = \frac{1}{\sigma^2} \frac{d_R(\mathbf{x}_i, \mathbf{x}_l) + d_G(\mathbf{x}_i, \mathbf{x}_l) + d_B(\mathbf{x}_i, \mathbf{x}_l)}{\sum_{i=1}^{N} w_{R_i}^2 + \sum_{i=1}^{N} w_{G_i}^2 + \sum_{i=1}^{N} w_{B_i}^2}, \quad (9)$$

where $d_R(\cdot,\cdot)$, $d_G(\cdot,\cdot)$, and $d_B(\cdot,\cdot)$ are computed as follows.

C. Computation of the structural similarity in the RGB map

In the original NLM, the $d(\mathbf{x}_i, \mathbf{x}_l)$ are computed as the Euclidean distances between the vectors of voxels taken from vicinities centered at $\mathbf{x}_i$ and $\mathbf{x}_l$. We prefer the method in [12], which is both computationally efficient and robust to noise: once the mean value $\overline{R}(\mathbf{x})$ and derivatives $R_k(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}_k} R(\mathbf{x})$ (conversely for G and B) of each channel are computed using the least squares method proposed in [12], we define:

$$d_R(\mathbf{x}_i, \mathbf{x}_l) = \frac{\left| \overline{R}(\mathbf{x}_i) - \overline{R}(\mathbf{x}_l) \right|^2}{h/h_{\mathrm{eff}}} + \sum_{k=1}^{3} \delta_k \frac{\left| R_k(\mathbf{x}_i) - R_k(\mathbf{x}_l) \right|^2}{h/h_{\mathrm{eff}}}, \quad (10)$$

where the constant δ_k is related to the discretization of the derivatives computation, and the term h/h_{eff} accounts for the effective reduction of the power of noise in the mean value and gradients of the image, see [12] for details.

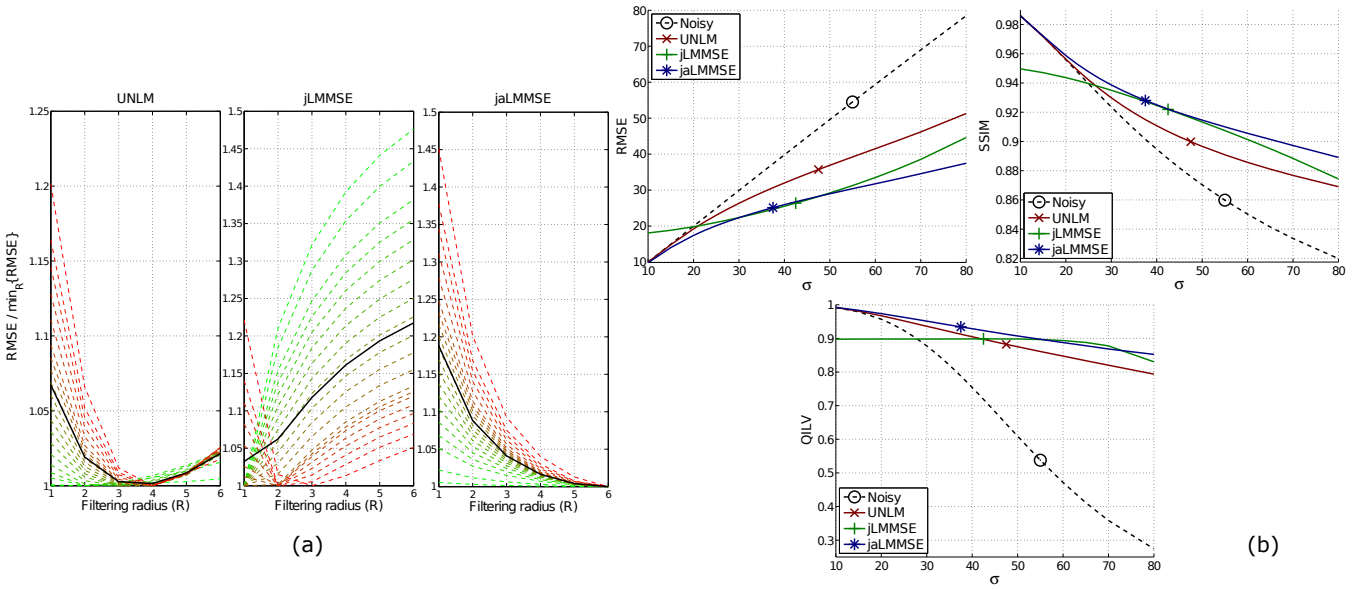

Fig. 2. **(a)** Normalized RMSE in the DWIs (inside an FA mask) as a function of the isotropic filtering radius used for each algorithm. Dashed lines correspond to each power of noise considered, from very high (green) to very low (red) SNR. Solid lines represent mean values for all powers of noise. **(b)** Quality indices (inside the FA mask) as a function of the input power of noise for each algorithm (with the optimum filtering radius inferred from **(a)**).

III. Experiments and Results

In this Section we compare three different filters for DWI: 1) In case we assume $\mathbf{C}_{\mathbf{A}^2\mathbf{M}^2} = 0$ in eq. (4), we get an **UNLM** filter; we have used $h = 1.2$ in all cases after [10], [12]. 2) If we set $h = \infty$ with the whole eq. (4), we get the isotropic joint LMMSE (**jLMMSE**) in [10]. 3) The model in eq. (4), with a value of $h = 2.0$ (empirically fixed) is our new proposal, so-called joint anisotropic LMMSE (**jaLMMSE**). We use a test bed inspired in the *noise-free* phantom in [13]: starting from a DWI data set (60 gradients, 1 baseline, matrix: $128 \times 128 \times 66$, isotropic resolution $2 \times 2 \times 2$ mm^3) with a dual b-value of 800 s/mm^2 and 1000 s/mm^2, we filter each volume with the UNLM technique described in [10]. The logarithm of the signal at each voxel is fitted in the basis of Spherical Harmonics up to order 4, and a synthetic logarithmic signal at $b = 900$ s/mm^2 is averaged from both original b-values. This signal is sampled at 15 evenly spaced gradient directions, and used to compute the final *clean* DWI shown in Fig. 1(a). The noisy M are obtained from A as in eq. (1) for different σ. To assess the quality of each filtered image, we compute three indices between them and the phantom: the Root Mean Squared Error (RMSE), the Structural SIMilarity (SSIM, [14]), and the Quality Index from Local Variance (QILV, [15]). While the former two sum up how well noise is removed, the latter is more sensitive to over-blurring artifacts.

In the first experiment we aim at finding the optimal size of the vicinities $\mathcal{N}$ to be used by each algorithm, as depicted in Fig. 2(a). For each input σ, we compute the output RMSE versus the filtering radius (isotropic, as the resolution of the phantom itself). This curve is normalized by the minimum RMSE achievable in each case. While jLMMSE requires very small radii (because it uses isotropic neighborhoods), the other methods are more robust to large vicinities (since they are adaptive to the anatomy). In particular, jaLMMSE may use arbitrarily large radii since it preserves the structures by two means: the computation of moments in shaped neighborhoods, and the LMMSE linear correction in eq. (4).

For the optimal filtering radii inferred from Fig. 2(a) for each algorithm, we show in Fig. 2(b) the output qualities obtained as a function of the input σ. The RMSE and SSIM values demonstrate that jaLMMSE is a more efficient noise removal engine than the other two approaches for all the range of input SNR. As reported in [10], UNLM is preferable to jLMMSE for small σ: in this case, the improvement in the SNR does not compensate for the blurring introduced by the isotropic jLMMSE computations. For larger (and more realistic) values of σ, however, jLMMSE represents a better trade-off between SNR improvement and blurring, and it outperforms UNLM. If we look at the QILV index, indeed, the response of jLMMSE is quite flat, meaning this index is governed by the blurring the isotropic vicinities produce in the output. Our jaLMMSE, however, exhibits the nice capability of UNLM to preserve fine details for high SNR (small σ), and it better removes the noise for low SNR (high σ) than any other method. With regard to the QILV measure, the LMMSE correction of jaLMMSE proves itself useful to better preserve fine details in the output over the pure UNLM.

Finally, we have computed corresponding DTI volumes from one of the images used to generate the phantom ($b = 1000$ s/mm^2), and from its filtered versions. We have estimated FA histograms over partial segmentations of the corpus callosum (cc) and the cingulum (cg), see Fig. 3: While jLMMSE seems to negatively bias the FA due to the increased partial volume effect casued by the isotropic computations, jaLMMSE shows approximately the same FA values as the noisy image (as expected for this SNR, see [5]). On the contrary, UNLM overestimates the FA, as suggested by the displacement of the histogram peak at the cc.

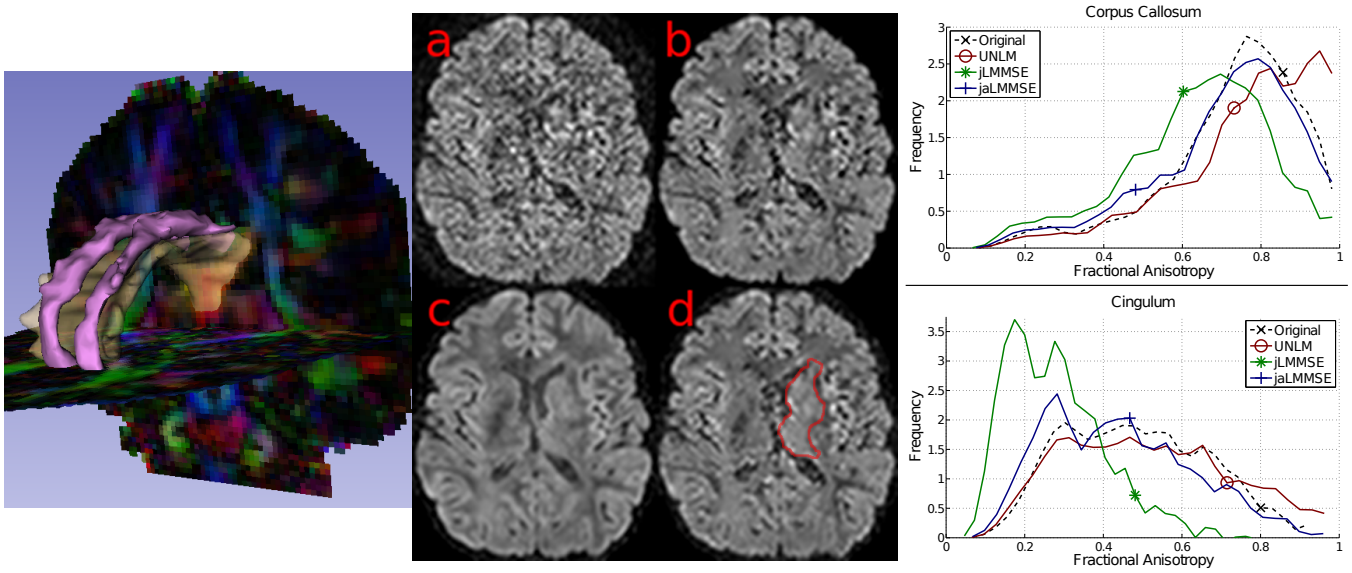

Fig. 3. The corpus callosum and cingulum have been partially segmented (**left**) in a DWI data set with $b = 1000$ s/mm^2 (**middle**; a) original, b) UNLM-filtered, c) jLMMSE-filtered, d) jaLMMSE-filtered -compare the more effective noise removal of jaLMMSE over UNLM in the highlighted section-). A DTI volume is obtained in each case using standard Least Squares, and histograms of the corresponding FAs computed for each segmented tract (**right**).

IV. Discussion and conclusions

The joint anisotropic LMMSE filter for DWI provides a nice trade-off between the dramatic improvement of the SNR and the preservation of anatomical details. Specifically, it is able to *clean* the DWI channels without introducing a systematic bias in the FA. As compared to the joint LMMSE in [10], it avoids mixing together different fiber bundles, or even the background and the fiber bundles, maintaining the anisotropy of the white matter. Compared to the pure UNLM, it combines the information from all the DWI channels producing a more effective noise removal. At the same time, the LMMSE correction may keep visible fine details in small tracts that might be otherwise blurred out by UNLM, due to the raw estimation of the signal as its second order moment. Its computational load is still fairly reasonable, thanks to the efficient computation of the NLM weights based on distances in the features space [12]. A complete C++ implementation may be downloaded from `http://www.nitrc.org/projects/jalmmse_dwi`.

The current scheme may also be further developed in several ways. First, we have considered a stationary Rician model for the noise pattern (i.e., a constant value of σ for the entire Field of View), which may not be a realistic assumption with parallel acquisitions or Echo-Planar imaging. Yet, Rician statistics may be easily generalized to a non-central Chi model to tackle certain multi-coil reconstruction procedures. Finally, the entire evaluation of our proposal has been intended for DTI-like data sets. With High Angular Resolution volumes, which are usually acquired with larger b-values, the optimal parameters inferred from the experiments may be no longer appropriate. Indeed, an adaptive value of h fitted to the SNR and the level of detail at each voxel is an interesting improvement, together with alternative designs of the projection weights to compute the RGB anatomical map.

References

[1] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Res. Med.*, vol. 34, pp. 910–914, 1995.

[2] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile Wavelet domain noise filtration technique for medical imaging," *IEEE Trans. Med. Imag.*, vol. 22, pp. 323–331, 2003.

[3] G. Gerig, O. Kübler, R. Kikinis, and F. Jolesz, "Nonlinear anisotropic filtering of MRI," *IEEE Trans. Med. Imag.*, vol. 11, pp. 221–232, 1992.

[4] J. Sijbers and A. J. den Dekker, "Maximum likelihood estimation of signal amplitude and noise variance form MR data," *Magn. Res. Imaging*, vol. 51, pp. 586–594, 2004.

[5] D. K. Jones and P. J. Basser, "Squashing peanuts and smashing pumpkins: how noise distorts diffusion weighted MR data," *Magn. Res. Med.*, vol. 52, pp. 979–993, 2004.

[6] M. Martín-Fernández, E. Muñoz-Moreno, L. Cammoun, J.-P. Thiran, C.-F. Westin, and C. Alberola, "Sequential anisotropic multichannel Wiener filtering with Rician bias correction applied to 3D regularization of DWI data," *Med. Imag. Anal.*, vol. 13, pp. 19–35, 2009.

[7] G. McGibney and M. Smith, "Unbiased signal-to-noise ratio measure for MR images," *Med. Phys.*, vol. 20, pp. 1077–1078, 1993.

[8] 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, pp. 1389–1403, 2008.

[9] N. Wiest-Daesslé, S. Prima, P. Coupé, S. Morrissey, and C. Barillot, "Non-Local Means variants for denoising of diffusion-weighted and diffusion tensor MRI," in *Procs. of MICCAI*, ser. LNCS, vol. 4792, 2007, pp. 344–351.

[10] A. Tristán-Vega and S. Aja-Fernández, "DWI filtering using joint information for DTI and HARDI," *Med. Imag. Anal.*, vol. 14, pp. 205–218, 2010.

[11] V. Brion, O. Riff, I. Kezele, M. Descoteaux, D. Le Bihan, J.-F. Mangin, C. Poupon, and F. Poupon, "Real-time Rician noise correction applied to real-time HARDI and HYDI," in *Procs. of ISMRM*, 2011, p. 1930.

[12] A. Tristán-Vega, V. García-Pérez, S. Aja-Fernández, and C.-F. Westin, "Efficient and robust nonlocal means denoising of MR data based on salient features matching," *Comput. Meth. Prog. Biomed.*, vol. 105, pp. 131–144, 2012.

[13] A. Tristán-Vega and S. Aja-Fernández, "Design and construction of a realistic DWI phantom for filtering performance assessment," in *Procs. of MICCAI*, ser. LNCS, vol. 5761, 2009, pp. 951–958.

[14] Z. Wang, A.-C. Bovik, H.-R. Sheikh, and E.-P. Simoncelli, "Image quality assessment: form error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, pp. 600–612, 2004.

[15] S. Aja-Fernández, R. San-José-Estépar, C. Alberola-López, and C.-F. Westin, "Image quality assessment based on local variance," in *Procs. of IEEE EMBC*, New York, USA, 2005, pp. 4815–4818.