RECONSTRUCTING DIFFUSION KURTOSIS TENSORS FROM SPARSE NOISY MEASUREMENTS

Yugang Liu[*◇] *Siming Wei*[†] *Quan Jiang*[‡] *Yizhou Yu*[◇]

∗ University of Electronic Science and Technology of China
† Zhejiang University ‡ Henry Ford Hospital
◇ University of Illinois at Urbana-Champaign

ABSTRACT

Diffusion kurtosis imaging (DKI) is a recent MRI based method that can quantify deviation from Gaussian behavior using a kurtosis tensor. DKI has potential value for the assessment of neurologic diseases. Existing techniques for diffusion kurtosis imaging typically need to capture hundreds of MRI images, which is not clinically feasible on human subjects. In this paper, we develop robust denoising and model fitting methods that make it possible to accurately reconstruct a kurtosis tensor from 75 or less noisy measurements. Our denoising method is based on subspace learning for multi-dimensional signals and our model fitting technique uses iterative reweighting to effectively discount the influences of outliers. The total data acquisition time thus drops significantly, making diffusion kurtosis imaging feasible for many clinical applications involving human subjects.

Index Terms— MRI, Kurtosis Tensors, Model Reconstruction, Optimization, Denoising

1. INTRODUCTION

Diffusion magnetic resonance imaging (MRI) provides a means to noninvasively probe the microstructure of biological tissues. Water diffusion in biological tissues attenuates the MRI signal. The amount of attenuation of diffusion-encoding gradient pulses along one specific direction depends on the probability density function of projected displacements of water molecules along that gradient direction [1]. Thus, by measuring the attenuation factors along multiple gradient directions, it becomes possible to reconstruct the full diffusion displacement probability density function for water molecules in biological tissues. Since microscale tissue structures, such as cellular compartments and membranes, determine the mobility of water molecules within, the reconstructed diffusion displacement probability density function can be conversely used to infer the tissue microstructure.

In tissues, such as brain gray matter, where the measured apparent diffusivity is largely independent of the orientation of the tissue, it is usually sufficient to characterize the diffusion characteristics with a scalar apparent diffusion coefficient (ADC). However, in anisotropic media, such as brain white matter, where the measured diffusivity is known to depend on the orientation of the tissue, no single ADC can characterize the orientation-dependent water mobility in these tissues. Because of this, a zero-mean trivariate Gaussian function was proposed to model the diffusion displacement probability density function. This led to diffusion tensor imaging (DTI), which reconstructs the full covariance matrix (tensor) of the Gaussian function from multiple measurements.

A diffusion tensor D is a second-order three-dimensional positive semidefinite symmetric tensor [2, 8]. Under a Cartesian coordinate system, it is represented by a real three-dimensional symmetric matrix, which has six independent elements $D = (D_{ij})$ with $D_{ij} = D_{ji}$ for $i, j = 1, 2, 3$. In this model, the MRI signal intensity $S_m(\mathbf{q})$ is expressed as follows.

$$\ln[S_m(\mathbf{q})] = \ln[S(0)] - bD_{app}(\mathbf{x}), \tag{1}$$

where $\mathbf{q} = b\mathbf{x}$, $\mathbf{x}(= [\mathrm{x}_1\ \mathrm{x}_2\ \mathrm{x}_3]^T)$ is a unit vector, $D_{app}(\mathbf{x})$ is the ADC value in the direction defined by $\mathbf{x}$

$$D_{app}(\mathbf{x}) = \sum_{i,j=1}^{3} D_{ij}\,\mathrm{x}_i\mathrm{x}_j, \tag{2}$$

and the parameter b is given by

$$b = (\gamma\delta g)^2 \left(\Delta - \frac{\delta}{3}\right), \tag{3}$$

where g is the gradient strength, γ is the proton gyromagnetic ratio, Δ is the separation time of the two diffusion gradients, δ is the duration of each gradient lobe. Combining (1) and (2), we have

$$\ln[S_m(\mathbf{q})] = \ln[S(0)] - b\sum_{i,j=1}^{3} D_{ij}\,\mathrm{x}_i\mathrm{x}_j. \tag{4}$$

However, the complex structure of most tissues, consisting of various types of cells and membranes, can cause the diffusion displacement probability density function to deviate substantially from a Gaussian form. This deviation from Gaussian behavior can be quantified using a dimensionless metric called the excess kurtosis. A method, named diffusion kurtosis imaging (DKI), for estimating the excess kurtosis of water diffusion in vivo by means of pulsed-field-gradient MRI has been introduced in [5]. A diffusion kurtosis tensor W is a fourth-order three-dimensional fully symmetric array, which has fifteen independent elements $W = (W_{ijkl})$ with W_{ijkl} being invariant for any permutation of its indices $i, j, k, l = 1, 2, 3$. In this model, (1) can be expanded as follows:

$$\ln[S_m(\mathbf{q})] = \ln[S(0)] - bD_{app}(\mathbf{x}) + \frac{1}{6}b^2 D^2_{app}(\mathbf{x})K_{app}(\mathbf{x}), \tag{5}$$

where $K_{app}(\mathbf{x})$ is the apparent kurtosis coefficient (AKC) value in the direction $\mathbf{x}$,

$$D^2_{app}K_{app} = \left(\frac{1}{3}\sum_{1}^{3} D_{ii}\right)^2 \sum_{i,j,k,l=1}^{3} W_{ijkl}\,\mathrm{x}_i\mathrm{x}_j\mathrm{x}_k\mathrm{x}_l. \tag{6}$$

Combining (5), (2) and (6), we have

$$\begin{aligned}\ln[S_m(\mathbf{q})] \;=\; & \ln[S(0)] - bD_{app} + \\ & \frac{1}{6}b^2\left(\frac{1}{3}\sum_{1}^{3} D_{ii}\right)^2 \sum_{i,j,k,l=1}^{3} W_{ijkl}\,\mathrm{x}_i\mathrm{x}_j\mathrm{x}_k\mathrm{x}_l.\end{aligned} \tag{7}$$

The non-Gaussian behavior of water molecules may contain useful information related to tissue structure and patho-physiology. Hence, the diffusion kurtosis imaging (DKI) has important biological and clinical significance. Sharp differences between the diffusion kurtosis in white and gray matters have been reported in [5]. DKI is believed to have potential value for the assessment of neurologic diseases, such as multiple sclerosis and epilepsy, with associated white matter abnormalities. In addition, DKI may be useful for investigating abnormalities in tissues with isotropic structures, such as brain gray matter, where techniques like DTI are less applicable.

Existing techniques for diffusion kurtosis imaging typically need to capture hundreds of MRI images with distinct combinations of b values and gradient directions [5, 6], which takes a long time to finish and is not clinically feasible on human subjects as they need to stay still during the entire data acquisition stage. In this paper, we develop robust denoising and model fitting methods that makes it possible to reconstruct a kurtosis tensor from 75 or less noisy measurements. Our denoising method is based on subspace learning for multi-dimensional signals and our model fitting technique uses iterative reweighting to effectively discount the influences of outliers. The total data acquisition time thus drops significantly, making diffusion kurtosis imaging feasible for many clinical applications involving human subjects.

2. ROBUST MODEL FITTING

From the above section, we know a diffusion tensor D has 6 unknowns and a diffusional kurtosis tensor W has 15 unknowns. To reconstruct the kurtosis tensor from multiple measurements, we have to simultaneously estimate the diffusion tensor because both of them are involved in (5) and (7). Therefore, we need to simultaneously estimate 21 unknowns during kurtosis tensor reconstruction. Suppose there are $n+1$ ($n \geq 21$) real noisy measurements associated with distinct $\mathbf{q}$ vectors, $\{S(0), S_r(\mathbf{q}_1), \cdots, S_r(\mathbf{q}_n)\}$, where $\mathbf{q}_i = b_i \mathbf{x}^i$. We cast kurtosis tensor reconstruction as a least-squares minimization of the differences between real and predicted measurements,

$$\min_{D_{ij}, W_{ijkl}} \sum_{i=1}^{n} \left(\ln[S_r(\mathbf{q}_i)] - \ln[S_m(\mathbf{q}_i)]\right)^2, \tag{8}$$

where $S_m(\mathbf{q}_i)$ is the model-based prediction defined in (5). We solve this nonlinear minimization in two stages. First, we remove the nonlinear term and minimize

$$\sum_{i=1}^{n} (\ln[S_r(\mathbf{q}_i)] - (\ln[S(0)] - b_i D_{app}(\mathbf{x}^i))^2$$

as a linear least-squares problem. This linear minimization provides an initial estimation of the diffusion tensor only. Second, solve the full nonlinear problem in (8) using the Levenberg-Marquardt method [7]. The diffusion tensor is initialized using linear least-squares solution and the unknowns in the diffusion kurtosis tensor are initialized to zero.

Since we aim to use sparse measurements only and they are further contaminated with noise and outliers, the above nonlinear method usually does not give rise to robust results. Therefore, we further integrate nonlinear optimization with a robust statistical technique, iteratively re-weighted least squares (IRLS) [9], to reduce the influences of noise and outliers. (8) thus becomes the following weighted least-squares minimization.

$$\min_{D_{ij}, W_{ijkl}} \sum_{i=1}^{n} w(\mathbf{q}_i) \left(\ln[S_r(\mathbf{q}_i)] - \ln[S_m(\mathbf{q}_i)]\right)^2 \tag{9}$$

where $w(\mathbf{q}_i)$ is the weighting term for the measurement associated with $\mathbf{q}_i$. We define one *meta-iteration* as the process of running the Levenberg-Marquardt method on the weighted least-squares problem in (9) for a reasonable number of iterations (say 10) while keeping the weights fixed. At the end of each meta-iteration, we update the weights in (9) according to the final residual error of each measurement in the current meta-iteration. A larger residual error indicates a larger likelihood for the measurement to be an outlier, and therefore it should receive a smaller weight in the next meta-iteration to reduce its influence on the estimated kurtosis tensor. We use a Gaussian kernel to define the weighting term as follows.

$$w(\mathbf{q}_i) = \exp\left(-\left(\ln[S_r(\mathbf{q}_i)] - \ln[S_m(\mathbf{q}_i)]\right)^2 / 2\sigma_e^2\right), \tag{10}$$

where σ_e is the standard deviation of the residual errors of all measurements. We typically use only 2 to 3 meta-iterations in our experiments and the last meta-iteration is run until convergence.

Note that the above model fitting algorithm does not explicitly enforce the constraint that the diffusion tensor is a positive semidefinite matrix even though it does return such a matrix most of the time. When this algorithm does not return a matrix satisfying this constraint, we switch to a more expensive procedure based on convex optimization. Let $\mathbf{d}$ be the 9-element vector by flattening the 3x3 diffusion tensor D, and $\mathbf{k}$ be the 15-element vector assembled from the fifteen independent elements of the kurtosis tensor W. The weighted least-squares minimization in (9) can be written in a vector form as follows.

$$\min_{\mathbf{d},\mathbf{k}} \sum_{i=1}^{n} w(\mathbf{q}_i) \left(b_i + \mathbf{a}_i^T \mathbf{d} - \mathbf{f}_i^T \mathbf{k}\right)^2 \tag{11}$$

where $b_i = S_r(\mathbf{q}_i) - S(0)$, $\mathbf{a}_i$ and $\mathbf{f}_i$ are coefficient vectors for $\mathbf{d}$ and $\mathbf{k}$, respectively. $\mathbf{a}_i$ and $\mathbf{f}_i$ are derived from (7) and include the cross-terms, $\mathrm{x}_i\mathrm{x}_j$ and $\mathrm{x}_i\mathrm{x}_j\mathrm{x}_k\mathrm{x}_l$, respectively. They are specific to $\mathbf{q}_i$. If we further stack $\mathbf{a}_i$'s and $\mathbf{f}_i$'s into matrices and b_i's into a vector, (11) can be further rewritten in the following matrix-vector form,

$$\min_{\mathbf{d},\mathbf{k}} \|\mathbf{b} + \mathbf{A}\mathbf{d} - \mathbf{F}\mathbf{k}\|_2^2, \tag{12}$$

where $\mathbf{b}$ is a $n \times 1$ vector, $\mathbf{A}$ is a $n \times 9$ matrix, and $\mathbf{F}$ is a $n \times 15$ matrix. The i-th element of $\mathbf{b}$ is $\sqrt{w(\mathbf{q}_i)}b_i$; The i-th row of $\mathbf{A}$ and $\mathbf{F}$ are $\sqrt{w(\mathbf{q}_i)}\mathbf{a}_i^T$ and $\sqrt{w(\mathbf{q}_i)}\mathbf{f}_i^T$, respectively.

If $\mathbf{d}$ is known, (12) becomes a linear least-squares problem with $\mathbf{k}$ as the unknown. The optimal solution in this case is

$$\mathbf{k}^* = \mathbf{F}^+(\mathbf{b} + \mathbf{A}\mathbf{d}), \tag{13}$$

where $\mathbf{F}^+$ is the Moose-Penrose pseudo-inverse of $\mathbf{F}$. If we substitute this formulation for optimal $\mathbf{k}$ into (12) and further add the constraint that the diffusion tensor D formed from the vector $\mathbf{d}$ should be symmetric and positive semidefinite, we have the following convex optimization,

$$\begin{aligned} \min_{\mathbf{d}} \quad & \left\|(\mathbf{I} - \mathbf{F}\mathbf{F}^+)(\mathbf{b} + \mathbf{A}\mathbf{d})\right\|_2^2, \\ \text{s.t.} \quad & D(\mathbf{d}) \succeq \mathbf{0}. \end{aligned} \tag{14}$$

This convex optimization sets $\mathbf{d}$ as the only unknown and can be conveniently solved using the SeDuMi toolbox [10] in MATLAB.

Given the solution of (14), we extract the eigenvectors of the diffusion tensor $D(\mathbf{d})$ and use them as the orthogonal axes of a local frame. We further transform all $\mathbf{q}$ vectors to this local frame and solve the nonlinear problem in (9) again within this local frame using a revised Levenberg-Marquardt method. In this local frame, we only

need to optimize the three diagonal elements of the diffusion tensor D because all of its off-diagonal elements should be zero. We only need to guarantee that the three diagonal elements are all nonnegative to enforce the original constraint that D is positive semidefinite. Our revised Levenberg-Marquardt method can easily guarantee such nonnegativity during every iteration of the optimization. Our experiments have confirmed that solving the kurtosis tensor using this procedure can achieve much more accurate results than directly using the pseudo-inverse in (13), where $\mathbf{d}$ is replaced with the solution for the convex optimization in (14).

Note that a complete dataset for diffusion tensor imaging or diffusion kurtosis imaging is typically three dimensional and consists of a large number of pixels. Each pixel contains multiple measurements with the aforementioned $\mathbf{q}$ vectors, $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$. Thus, the aforementioned model fitting is repeated for every pixel.

3. SUBSPACE BASED DENOISING

The robust model fitting technique discussed in the previous section can resist the influences of noise and outliers to a certain extent. Nonetheless, when the noise level is very high, it may fail to accurately reconstruct a kurtosis tensor model. In this section, we develop a preprocessing step that performs noise removal before model fitting. Note that we are actually dealing with multi-dimensional signals if we stack the multiple measurements at the same pixel together as a vector, $\mathbf{s} = [S(0)\ S(\mathbf{q}_1)\ \cdots\ S(\mathbf{q}_n)]^T$. Since all components in this vector are derived from the same underlying biological structure, they are highly correlated. On the other hand, since each component is measured separately, the noise in different components is much less correlated. Instead of denoising each component separately, we chose to denoise all components simultaneously. Because of the correlation present among different components, the idea is to learn a lower-dimensional subspace that the noise free multi-dimensional signal belongs [3]. The noisy signal can then be projected to this subspace to have its noise part removed.

We use Independent Component Analysis (ICA) to learn the aforementioned subspace. ICA is performed on the collection of noisy $(n+1)$-dimensional vectors from all pixels or a subset of pixels, such as the pixels in a 2D slice of the original 3D dataset. The result is $n+1$ independent components (ICs) each of which is a $(n+1)$-dimensional vector. In contrast to Principal Component Analysis (PCA), ICA renders the output signals as statistically independent as possible by evaluating higher-order statistics. We use the popular FastICA algorithm introduced in [4], which performs ICA by maximizing the non-Gaussianity of the signal components.

To extract a desired signal subspace, we need to first sort the independent components in decreasing order of importance and then choose a subspace from this ordered list. Let $\mathbf{U} = [\mathbf{u}_0\ \mathbf{u}_1\ \cdots\ \mathbf{u}_n]$ be the $(n+1) \times (n+1)$ matrix of the complete set of independent components, whose linear combination can losslessly represent the $(n+1)$-dimensional signal at each pixel. Suppose the signal at a pixel is expressed as $\mathbf{s} = \sum_{i=0}^{n} c_i \mathbf{u}_i$. Then the vector of coefficients $\mathbf{c} = [c_0\ c_1\ \cdots\ c_n]^T$ can be obtained by solving the linear system,

$$\mathbf{U}\mathbf{c} = \mathbf{s}. \tag{15}$$

We solve such projection coefficients for signals from all pixels and compute the variance of the coefficients corresponding to each independent component. An independent component is considered important if its projection coefficients have a large variance. Finally, we sort the independent components in decreasing order of importance.

A common criterion for choosing a subspace is based on minimum description length (MDL). However, as mentioned in [3], the MDL estimator tends to significantly underestimate the number of independent components corresponding to noise. In practice, we simply visualize the variances of the projection coefficients for all components in the aforementioned order and visually identify a locally maximum gap in variance between two adjacent components. All components after that gap are classified as noise components and discarded. The components before the gap form the desired subspace. Suppose there are r remaining components. This subspace is represented by the truncated matrix, $\mathbf{U}^{(r)} = [\mathbf{u}'_1\ \mathbf{u}'_2\ \cdots\ \mathbf{u}'_r]$.

To extract the noise free part of a signal, we need to project the original noisy signal to this subspace. Again, the projection of a signal is a linear combination of these r components, $\hat{\mathbf{s}} = \sum_{i=1}^{r} c'_i \mathbf{u}'_i$. The vector of coefficients $\mathbf{c}' = [c'_1\ c'_2\ \cdots\ c'_r]^T$ can be obtained by minimizing $\|\mathbf{s} - \hat{\mathbf{s}}\|^2$, which is again a linear least-squares problem, whose solution is $\mathbf{c}' = {\mathbf{U}^{(r)}}^{+}\mathbf{s}$, where ${\mathbf{U}^{(r)}}^{+} = \left({\mathbf{U}^{(r)}}^T \mathbf{U}^{(r)}\right)^{-1} {\mathbf{U}^{(r)}}^T$ is the Moose-Penrose pseudo-inverse of $\mathbf{U}^{(r)}$.

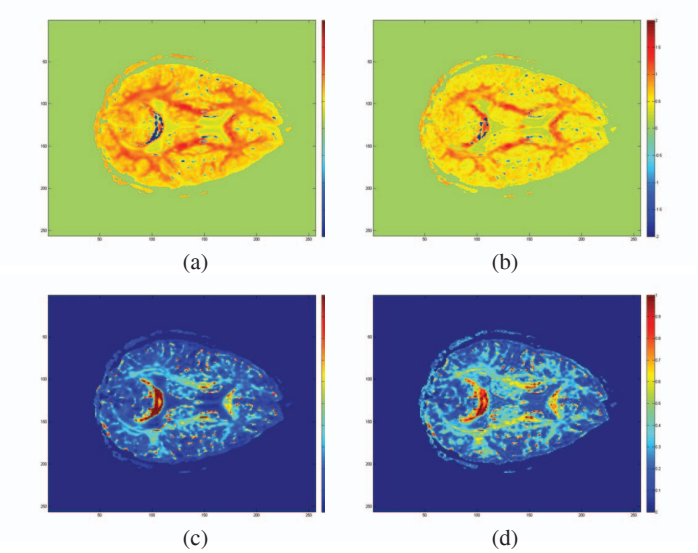
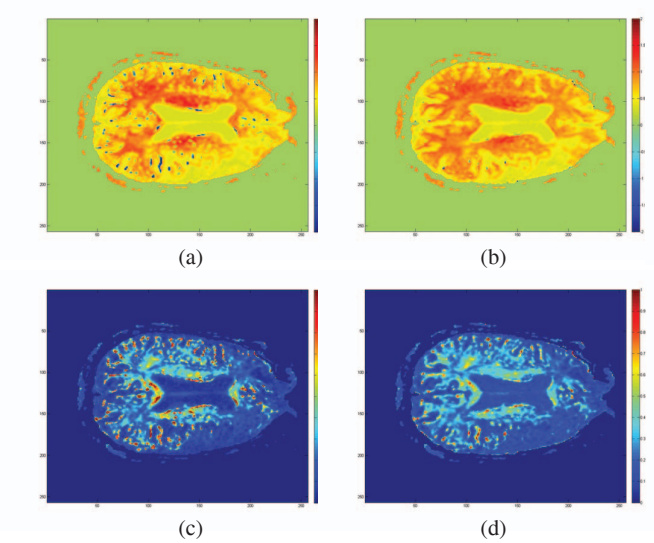


Fig. 1. Comparison of model fitting results with and without iterative reweighting. (a) mean kurtosis with reweighting; (b) mean kurtosis without reweighting; (c) standard deviation of kurtosis with reweighting; (d) standard deviation of kurtosis without reweighting.

4. EXPERIMENTAL RESULTS

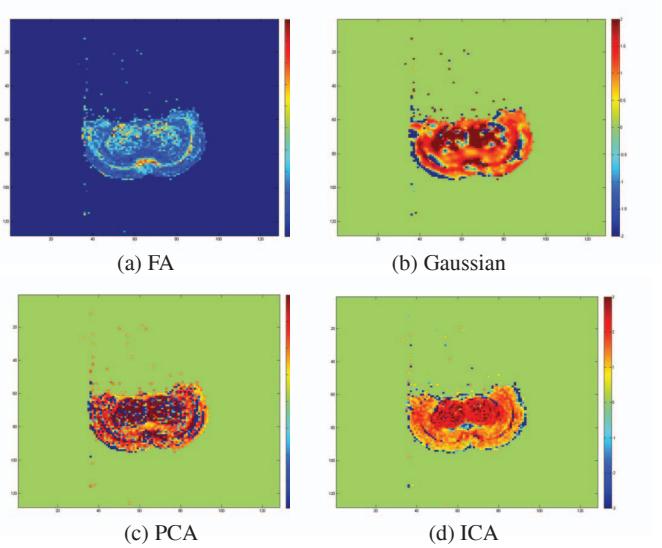
We have performed various experiments to validate our algorithms. The measurements were collected with five levels of b values, which are 500, 1000, 1500, 2000, and 2500, respectively. At each distinct b value, 10 to 15 gradient directions were sampled over the unit sphere, resulting in 58 to 75 noisy measurements per image pixel.

In our experiments, we estimate scalar quantities, including mean and standard deviation of kurtosis, because clinical applications typically adopt such scalar quantities instead of the originally reconstructed tensorial data. Mean kurtosis is defined to be the integral of the apparent kurtosis coefficient in (6) over the unit sphere [8]. Standard deviation of kurtosis is defined to be the integral of the squared difference between the apparent kurtosis coefficient and the mean kurtosis over the unit sphere. Intuitively, mean kurtosis has relatively large values in regions with a relatively high fiber density, which is true not only in the white matter but also in regions of the



(a) (b) (c) (d)

Fig. 2. Comparison of model fitting results with and without ICA-based subspace denoising. (a) mean kurtosis without denoising; (b) mean kurtosis with denoising; (c) standard deviation of kurtosis without denoising; (d) standard deviation of kurtosis with denoising.



(a) FA (b) Gaussian (c) PCA (d) ICA

Fig. 3. A comparison of model fitting results among various denoising methods. (a) Fractional anisotropy; (b) mean kurtosis with Gaussian denoising performed on each data channel separately; (c) mean kurtosis with PCA-based subspace denoising; (d) mean kurtosis with ICA-based subspace denoising.

gray matter filled with tiny fibers. Standard deviation of kurtosis, on the other hand, computes the degree of consistency among different diffusion directions. It has relatively large values in regions filled with crossing fibers.

Fig. 1 shows kurtosis fitting results for a normal subject. The results obtained with iterative reweighting better reveal details and are more consistent between the mean and standard deviation of kurtosis. On the other hand, the results obtained without iterative reweighting overestimate the standard deviation of kurtosis because it considers the noise as part of the standard deviation. They also appear to be washed out and show less spatial details.

Fig. 2 shows kurtosis fitting results for a stroke patient. It is obvious from the images that one side of the brain has much less neural fibers. A comparison is shown between results obtained with and without ICA-based subspace denoising. It is easy to verify that the results obtained with ICA-based denoising are cleaner with much less noisy dots.

Fig. 3 shows denoising results for an extremely noisy dataset collected for a pair of mouse brains. We compared our ICA-based subspace denoising with two other methods. One is Gaussian denoising applied to individual data channels. The other is PCA-based subspace denoising. Our denoising method succeeded on this low-quality dataset while the other two methods failed. The fractional anisotropy of the original dataset is also shown as a reference image.

5. CONCLUSIONS

We have developed robust denoising and model fitting methods that make it possible to accurately reconstruct a kurtosis tensor from 60 or less noisy measurements. Our denoising method is based on subspace learning for multi-dimensional signals and our model fitting technique uses iterative reweighting to effectively discount the influences of outliers. The total data acquisition time thus drops significantly, making diffusion kurtosis imaging feasible for many clinical applications involving human subjects.

Acknowledgments This work was partially supported by National Science Foundation (IIS 09-14631) and China Scholarship Council.

6. REFERENCES

[1] D.C. Alexander. Multiple-fiber reconstruction algorithms for diffusion mri. *Ann. N.Y. Acad. Sci.*, 1064:113–133, 2005.

[2] D.L. Bihan, J.-F. Mangin, C. Poupon, C.A. Clark, S. Pappata, N. Molko, and H. Chabriat. Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, 13:534–546, 2001.

[3] P. Gruber, K. Stadlthanner, M. Böhm, F.J. Theis, E.W. Lang, A.M. Tomé, A.R. Teixeira, C.G. Puntonet, and J.J. Gorriz Saéz. Denoising using local projective subspace methods. *Neurocomputing*, 69:1485–1501, 2006.

[4] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13:411–430, 2000.

[5] J.H. Jensen, J.A. Helpern, A. Ramani, H. Lu, and K. Kaczynski. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine*, 53:1432–1440, 2005.

[6] M. Lazar, J.H. Jensen, L. Xuan, and J.A. Helpern. Estimation of the orientation distribution function from diffusional kurtosis imaging. *Magnetic Resonance in Medicine*, 60:774–781, 2008.

[7] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes in C*. Cambridge Univ. Press, New York, 1988.

[8] L. Qi, D. Han, and E.X. Wu. Principal invariants and inherent parameters of diffusion kurtosis tensors. *Journal of Mathematial Analysis and Applications*, 349:165–180, 2009.

[9] C.J. Stone. *A Course in Probability and Statistics*. Duxbury Press, 1996.

[10] J.F. Sturm. Using sedumi 1.02, a matlab toolbox for optimizaton over symmetric cones. *Optimization Methods & Software*, 11(2):625–653, 1999.