The Annals of Applied Statistics
2016, Vol. 10, No. 3, 1162–1165
DOI: 10.1214/16-AOAS937
Main article DOI: 10.1214/15-AOAS880

DISCUSSION OF "FIBER DIRECTION ESTIMATION IN DIFFUSION MRI"

BY JIAN KANG[1] AND LEXIN LI[2]

University of Michiga and University of California, Berkeley

We would like to congratulate the authors for their excellent and stimulating work. The proposed statistical methods address a series of important issues in the analysis of diffusion magnetic resonance imaging (dMRI) data. Their work will surely be the subject of much application and elaboration in the future.

Our discussion focuses mainly on three aspects: (1) the multiple tensor direction model, (2) the kernel smoothing method of multiple directions and (3) the fiber tracking algorithm. For each aspect, we begin with a brief summary of the contributions of the paper, then raise some questions and point out some potential alternatives and extensions.

1. Multiple tensor direction model. It is challenging to estimate multiple tensor matrices within a voxel using dMRI data, in that most existing multi-tensor models suffer the nonidentifiability issue. The multiple tensor direction model proposed in this paper focuses on a direct estimation of multiple directions within a voxel, which avoids the complications of estimating multiple tensor matrices. The proposed model is shown to be identifiable, is easy to interpret, and is to facilitate detection of crossing fiber tracts.

We raise two questions regarding this model. First, how does this model deal with label switching? According to model (2) of the paper, we write

$$\bar{S}(\mathbf{u};\boldsymbol{\gamma}) = S_0 \sum_{j=1}^{J} \tau_j \exp\{b\alpha_j(\mathbf{u}^{\mathrm{T}}\mathbf{m}_j)^2\},$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\mathrm{T}}, \boldsymbol{\gamma}_2^{\mathrm{T}}, \boldsymbol{\gamma}_3^{\mathrm{T}}, \ldots, \boldsymbol{\gamma}_J^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\gamma}_j = (\tau_j, \alpha_j, \mathbf{m}_j^{\mathrm{T}})^{\mathrm{T}}$, for $j = 1, \ldots, J$. Now consider $\tilde{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}_2^{\mathrm{T}}, \boldsymbol{\gamma}_1^{\mathrm{T}}, \boldsymbol{\gamma}_3^{\mathrm{T}}, \ldots, \boldsymbol{\gamma}_J^{\mathrm{T}})^{\mathrm{T}}$ and that $\boldsymbol{\gamma}_1 \neq \boldsymbol{\gamma}_2$. Then it is clear that $\boldsymbol{\gamma} \neq \tilde{\boldsymbol{\gamma}}$, but $\bar{S}(\mathbf{u};\boldsymbol{\gamma}) = \bar{S}(\mathbf{u};\tilde{\boldsymbol{\gamma}})$. We are curious how to get around this label switching issue to maintain model identifiability.

Second, the authors have developed an approximation to the log-likelihood on tessellations and used a grid search to obtain the maximum likelihood estimates

Received April 2016.
[1]Supported in part by NIH R01MH105561.
[2]Supported in part by NSF DMS-13-10319.
Key words and phrases. Diffusion tensor imaging, discussion.

TABLE 1

Average mean square errors for estimating $(\alpha_j, \tau_j, \mathbf{m}_j)$ when fixing σ and S_0 at different values for the three simulated cases. The results are based on 120 data replications

	J	α	τ	$\mathbf{m}$
		True parameter settings with $(\sigma, S_0) = (0.5, 10)$		
Case 1	$J = 1$	$\alpha_1 = 3.75$	$\tau_1 = 0.779$	$\mathbf{m}_1 = (1, 0, 0)^{\mathrm{T}}$
Case 2	$J = 1$	$\alpha_1 = 3.75$	$\tau_1 = 0.779$	$\mathbf{m}_1 = (0, 1, 0)^{\mathrm{T}}$
Case 3	$J = 2$	$\alpha_1 = 3.75$	$\tau_1 = 0.779$	$\mathbf{m}_1 = (1, 0, 0)^{\mathrm{T}}$
		$\alpha_2 = 3.75$	$\tau_2 = 0.779$	$\mathbf{m}_2 = (0, 1, 0)^{\mathrm{T}}$

	Average mean square errors (MSE $\times 10^{-3}$)					
	α	τ	$\mathbf{m}$	α	τ	$\mathbf{m}$
	Fix $(\sigma, S_0) = (0.5, 10)$					
Case 1	20.2 (27.7)	0.1 (0.2)	11.6 (21.3)			
Case 2	19.1 (25.6)	0.1 (0.2)	12.5 (23.4)			
Case 3	308.9 (413.1)	0.2 (0.2)	0.8 (0.6)			
	Fix $(\sigma, S_0) = (0.75, 10)$			Fix $(\sigma, S_0) = (0.25, 10)$		
Case 1	65.8 (64.1)	0.2 (0.2)	0.9 (1.4)	36.1 (41.7)	0.2 (0.2)	86.3 (53.7)
Case 2	73.2 (87.9)	0.2 (0.3)	1.0 (2.1)	38.2 (42.1)	0.2 (0.2)	78.9 (53.5)
Case 3	545.4 (637.3)	0.2 (0.3)	0.9 (0.7)	345.1 (425.3)	0.2 (0.3)	0.8 (0.6)
	Fix $(\sigma, S_0) = (0.5, 5)$			Fix $(\sigma, S_0) = (0.5, 15)$		
Case 1	2731.0 (241.1)	48.9 (0.0)	0.5 (0.5)	23.1 (30.6)	67.2 (4.2)	48.6 (44.2)
Case 2	2731.2 (241.4)	48.9 (0.0)	0.5 (0.4)	20.0 (28.9)	67.4 (4.3)	49.0 (45.3)
Case 3	717.1 (488.6)	136.3 (6.4)	1.1 (1.0)	345.3 (428.0)	19.6 (1.5)	0.9 (0.5)
	Fix $(\sigma, S_0) = (0.75, 5)$			Fix $(\sigma, S_0) = (0.25, 15)$		
Case 1	2342.5 (226.2)	48.9 (0.0)	0.5 (0.4)	41.1 (43.5)	69.4 (4.5)	97.7 (74.5)
Case 2	2351.3 (236.1)	48.9 (0.0)	0.5 (0.4)	39.4 (47.5)	69.1 (3.7)	103.2 (75.1)
Case 3	522.2 (480.5)	134.9 (6.8)	1.2 (1.0)	329.4 (523.4)	19.6 (1.6)	0.8 (0.5)

for $\{(\alpha_j, \mathbf{m}_j)^{\mathrm{T}}\}_{j=1}^{J}$. The resulting optimization can be done in a parallel fashion, but depends on two unknown parameters S_0 and σ. The authors have suggested to use an independent data set to estimate these two parameters. We are curious how sensitive the proposed method is to the estimates of S_0 and σ, and if one can directly estimate S_0 and σ using the same set of data. Toward that end, we performed a simple simulation using the code provided by the authors. We simulated data from the proposed model with $\sigma = 0.5$, $S_0 = 10$, $\alpha_j = 3.75$ and $\tau_j = 0.779$ for all $j = 1, \ldots, J$. We considered three cases, where cases 1 and 2 are single-direction in that $J = 1$ and $\mathbf{m}$ equals $(1, 0, 0)^{\mathrm{T}}$ and $(0, 1, 0)^{\mathrm{T}}$, respectively, whereas case 3 is multi-direction in that $J = 2$, $\mathbf{m}_1 = (1, 0, 0)^{\mathrm{T}}$ and $\mathbf{m}_2 = (0, 1, 0)^{\mathrm{T}}$. In the estimation procedure, we assumed J is known, and fixed σ and S_0 at different values, such that $\sigma = 0.25, 0.5, 0.75$ and $S_0 = 5, 10, 15$. Table 1 summarizes the

resulting estimates of $(\alpha, \tau, \mathbf{m})$ based on 120 data replications. We see from the table that the average mean square errors for estimating α and τ both increase when σ and S_0 are not set at the true values. It is especially so when the signal-to-noise ratio, S_0/σ, is set at a small value, corresponding to rows 2 to 4 of the first column of the table. On the other hand, the error for estimating $\mathbf{m}$ decreases for the single direction cases. Does this imply the feasibility of estimating σ and S_0 using the same set of data directly? In addition, it is worth noting that, for the multi-direction case, the estimates of directions are stable to the changes of σ and S_0, consistently producing small errors, which shows robustness of the proposed method for multiple directions.

2. Kernel smoothing of multiple directions. An overarching challenge in biomedical imaging analysis is how to borrow information and strength across neighboring voxels or even distant regions. By taking into account the complex spatial dependency structure of imaging data, it is to increase the power of signal detection and to improve the efficiency of parameter estimation. The same challenge arises in modeling spatially distributed tensor directions, and it can be even more difficult when the multiple tensor directions appear in the same locations. This paper is among the first to address the challenge by utilizing a combination of kernel smoothing and clustering of tensor directions. The associated theoretical properties are also established, which is a timely contribution to the field of nonparametric statistics in the non-Euclidean space.

For the proposed kernel smoothing, we have a few questions. First, the theoretical justification was based upon a simplification that treats the *estimated* diffusion directions as if they were *observed*. Such a simplification may not well capture the uncertainty introduced in the estimation process. We are curious if the same properties can be established by incorporating the variation of the estimated diffusion directions. Second, the theoretical justification has been developed only for smoothing along a single fiber tract. It is unclear whether it can be extended to multiple fibers in a straightforward fashion. Third, cross-validation (CV) has been employed in this paper to determine the kernel bandwidth. It is noteworthy that, for the multi-fiber scenario, CV could be computationally intensive. Is there any fast and simple alternative, such as the *solve-the-equation-plug-in* method [Jones, Marron and Sheather (1996), Raykar and Duraiswami (2006)] for the kernel density estimation, that can be developed in this context?

In addition to kernel smoothing, a potential alternative is to apply the smoothing spline techniques to the diffusion direction data. The key step would be to construct a set of spatially varying basis function taking values on the tensor direction space, and to develop a modeling strategy for multiple tensors occurring in the same location. A marked point process in tensor direction space can be used in that it can assign a random probability measure on the tensor shapes and the number of tensors. It is interesting to further explore this direction and compare the two smoothing techniques.

3. Fiber tracking. It is known that most existing deterministic fiber tracking methods are sensitive to the tensor estimation, and they typically assume one single direction in each voxel. By contrast, the proposed fiber tracking algorithm is more robust and practically useful for handling multiple directions within one voxel.

The proposed algorithm hinges on a number of tuning parameters, for instance, the threshold for separation angles and the number of neighboring voxels to search. It is important to understand how sensitive the fiber tracking estimation is to those parameters. From an application perspective, any rule of thumb for specifying those parameters under different scenarios would be practically useful. It would also be informative to compare the proposed tracking algorithm to the Bayesian method of Friman, Farnebäck and Westin (2006), which naturally incorporates the uncertainty to construct the fiber tracts.

Furthermore, a Bayesian hierarchical model for fiber tracking offers a potentially useful alternative. Such a model would consist of three levels of hierarchy. At the first level, a nonparametric prior model, such as a Gaussian process on a one-dimensional manifold, needs to be developed for multiple fiber tracts over a three-dimensional space. At the second level, given the fiber tracts, a generative model for the voxel-specific multiple tensor directions needs to be constructed. At the third level, the sampling distribution of the dMRI data is to be constructed. The multi-tensor direction model proposed in this paper can be employed for this purpose. This Bayesian hierarchical model is useful in that it provides a direct statistical inferential capability on the fiber tracts based on the observed dMRI data under a unified modeling framework. On the other hand, the complexity of the hierarchical model poses numerous challenges, including theoretical investigation of large support of the prior model and posterior consistency on the fiber tract estimation, as well as practical consideration, for instance, development of an appropriate posterior computational algorithm. Future research is warranted for this line of work.

REFERENCES

FRIMAN, O., FARNEBÄCK, G. and WESTIN, C.-F. (2006). A Bayesian approach for stochastic white matter tractography. *Medical Imaging*, *IEEE Transactions on* **25** 965–978.

JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91** 401–407. MR1394097

RAYKAR, V. C. and DURAISWAMI, R. (2006). Fast optimal bandwidth selection for kernel density estimation. In *Proceedings of the Sixth SIAM International Conference on Data Mining* 524–528. SIAM, Philadelphia, PA. MR2337970

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF MICHIGAN
1415 WASHINGTON HEIGHTS, 3651 SPH I
ANN ARBOR, MICHIGAN 48109-2029
USA
E-MAIL: jiankang@umich.edu

DIVISION OF BIOSTATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
344B LI KA SHING CENTER, MC 3370
BERKELEY, CALIFORNIA 94720-3370
USA
E-MAIL: lexinli@berkeley.edu