1. INITIAL REMARKS

This interesting and timely article attempts an important goal: to formalize the inference about diffusion tensors from diffusion weighted images in a single subject in the presence of measurement and artifact noise. The article’s main contributions are:

- A heteroscedastic linear model to account for noise in diffusion-weighted MRI data along with theoretical support for the use of a (one-step) weighted least squares algorithm to solve it.
- Asymptotic distributions of the estimated eigenstructure of the diffusion tensor under degenerate and nondegenerate cases, in addition to pseudolikelihood ratio tests for classifying each tensor into one of those cases.

As the authors explain, inferences about the diffusion tensor in a single subject are usually based on quantities derived from the tensor, the most common being scalar functions of the eigenvalues such as fractional anisotropy (FA) and trace, and the principal diffusion direction (PDD), the eigenvector corresponding to the largest eigenvalue. Whereas standard statistics are often used to analyze the scalar quantities, formal modeling of the PDD is not usually seen. Perhaps this is because statistical methods for unit vectors in three-dimensional space are not as widely known in the general scientific community, even though they have been studied extensively in the field of directional statistics (Mardia and Jupp 2005).

Especially because tractography algorithms are based on the PDD, it is important to have a characterization of the uncertainty in that vector as a result of noise. The authors provide this in an asymptotic sense as the number of measurements gets large. When the true tensor is oblate (i.e., the two largest eigenvalues are equal) or isotropic (i.e., all three eigenvalues are equal), the PDD is not defined, making the uncertainty infinite. Algorithms often deal with this problem by thresholding a function of the eigenvalues such as FA or CL (Westin et al. 2002), the idea being that if the tensor is not sufficiently anisotropic, then the PDD is not to be trusted. Instead of numerically trying appropriate thresholds, the authors mathematically derive the uncertainty and provide formal tests to classify whether the tensor is isotropic, oblate, prolate, or fully anisotropic.

A difficulty with the article is that the mathematical results obscure how the obtained quantities actually can be computed and used in practice. More specifically, theorem 1 proves the asymptotic normality of $\hat{\theta}$ and provides a way to estimate the asymptotic covariance, whereas the results of section 2.2 give the asymptotic distribution of the eigenstructure of the tensor as a function of the asymptotic covariances $\Sigma_U, \Sigma_{U_{11}},$ and $\Sigma_{U_{22}}$. However, the results do not explain how these parameters depend on the underlying true diffusion and noise parameters. Clearly, the asymptotic covariances depend on the acquisition scheme $z_1, \ldots, z_n$. The authors have preferred to give general results and to not commit themselves to particular acquisitions.

Admittedly, general analytical expressions are difficult to obtain, so a solution offered by the authors is to compute the asymptotic covariances by simulation, which is what they implement in section 3. The particular acquisition used by the authors in their simulation consists of $m = 5$ baseline images with $b = 0$ and $n - m = 25$ directions of diffusion gradients arranged uniformly in three-dimensional space with $b \neq 0$. In this comment, I hope to provide some insight into the asymptotic covariance parameters, based on an asymptotic version of this acquisition scheme. Specifically, I derive analytic forms of the asymptotic covariances of $\hat{\theta}$ and $U, U_{11},$ and $U_{22}$. It turns out that these provide a surprising answer to the question of what is the optimal ratio between the number of measurements at $b = 0$ versus the number of measurements at $b \neq 0$.

Armin Schwartzman is Assistant Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115 (E-mail: armins@hsph.harvard.edu).
2. ACQUISITION

As the authors mention in section 5, their results can be used to design acquisition schemes. The key conditions for theorem 1 that are affected by the acquisition scheme are (C2) and (C5), so it makes sense to ask when these hold. First, note that if \( b_1 = b \) constant for all \( i = 1, \ldots, n \), then \( A_0 \) is not invertible for any \( n \) and has an eigenvalue 0 with eigenvector \((b, 1, 0, 0, 1, 0, 1)T /\sqrt{3} + b^2\). For this reason, it is common in DWI to make some measurements, say \( m \) of them, at \( b = 0 \), and the remaining \( n - m \) measurements at a constant value \( b > 0 \). In their simulation, the authors use \( m = 5 \) and \( n - m = 25 \). A question that may be asked is what is a good proportion between the two.

Let \( r \) be a random unit vector on the sphere and define \( x = (r_1^2, 2r_1r_2, r_1r_3, r_2^2, 2r_2r_3, r_3^2)^T \). Let \( r_j \) be iid samples from the distribution of \( r \) and, without loss of generality, assume that \( b_1 = 0 \) for \( i = 1, \ldots, m \) and \( b_1 = b \) (constant) for \( i = m + 1, \ldots, n \). Suppose both \( m, n \to \infty \) but \( m/n \to \gamma \), where, by definition, \( 0 \leq \gamma \leq 1 \). The strong law of large numbers applies because the sphere is bounded. Thus \( A_{0/n} \to A_{0} = E(zz^T) \), where \( z \) is a random vector that takes the value \((1, 0, 0, 0, 0, 0, 0)^T \) with probability \( \gamma \) and \((1, -bx^T)^T \) with probability \( 1 - \gamma \). Here, \( A_{0} \) can be written as

\[
\frac{1}{n}A_n \to A_{0} = \begin{pmatrix}
1 & -b(1 - \gamma)E(x)T \\
-b(1 - \gamma)E(x) & (1 - \gamma)b^2E(xx^T)
\end{pmatrix}.
\]

(1)

When \( \gamma = 0 \), \( A_{0} \) has a zero eigenvalue corresponding to the same eigenvector \((b, 1, 0, 0, 0, 0, 0)^T /\sqrt{3} + b^2 \) as \( A_{0/n} \). When \( \gamma = 1 \), \( A_{0} \) is obviously noninvertible. Therefore, a necessary condition for \( A_{0} \) to be invertible is that \( 0 < \gamma < 1 \). That this is also a sufficient condition is proved next when \( r \) is uniform on the sphere. The principle is somehow clearer when proved for general dimension \( p \) (in DWI, \( p = 3 \)). In addition, presentation is easier if the elements of \( x \) are reordered as was done by Salvador et al. (2005), as follows.

**Proposition 1.** Let \( r = (r_1, \ldots, r_p)^T \) be a random unit vector uniformly distributed on the \((p - 1)\)-dimensional sphere, and let \( x \) be the \((p + 1)/2\)-dimensional vector \( x = (r_1^2, \ldots, r_p^2, 2r_1r_2, \ldots, 2r_{p-1}r_p)^T \) formed by taking the \( p \) squared diagonals of \( r \) and then twice all \((p - 1)/2\) pairs of off-diagonals of \( r \). Then the matrix \( A_{0} \), defined by the right side of (1) is positive definite if and only if \( 0 < \gamma < 1 \).

**Proof.** By symmetry, all of the odd moments of \( r \) are 0. The relevant even moments are \( E(r_{i}^2) = 1/p \), \( E(r_i^4) = 3/(p(p + 2)) \), and \( E(r_i^4r_j^2) = 1/(p(p + 2)) \) for \( i \neq j \) (Mardia and Jupp 2005, p. 186). Therefore, \( E(x^T) = (I_p^T/p, 0^T) \), where \( I_p^T \) denotes a vector of 1’s of length \( p \) and \( 0^T \) fills in up to length \( p(p + 1)/2 \). In addition,

\[
E(xx^T) = \frac{1}{p(p + 2)} \begin{pmatrix}
2I_p & 1_p^T \\
0 & 4I_{p(p-1)/2}
\end{pmatrix},
\]

(2)

with inverse

\[
(E(xx^T)^{-1} = \frac{p(p + 2)}{4} \begin{pmatrix}
2I_p - 1_p^T T/(p + 2) \\
0 & I_{p(p-1)/2}
\end{pmatrix},
\]

(3)

and, interestingly, \( E(x^T)(E(xx^T)^{-1} = (I_p^T, 0^T) \), so that \( E(x^T)(E(xx^T)^{-1}E(x) = 1 \). Replacing in (1) and using the formula for the determinant of a partitioned matrix (Schott 2005, p. 250), we get \(|A_0| = (1 - \gamma)b^2E(xx^T)| \), which is positive if and only if \( 0 < \gamma < 1 \).

Once \( 0 < \gamma < 1 \) has been established, an asymptotically equivalent acquisition scheme is to sample the iid \( z_i \)'s as

\[
z = \begin{pmatrix}
1, \theta^T \\
(1, -bx^T)^T
\end{pmatrix}
\]

(4)

with probability \( \gamma \) and \( 1 - \gamma \), where \( B_{\gamma} = S_0^2 A_{\gamma} \) when \( \lambda = 0 \), where \( A_{\gamma} \) is as given by (1).

3. ASYMPTOTIC DISTRIBUTION OF EIGENVALUES

To better understand the results of the authors’ section 2.2, explicit expressions for \( \Sigma_0 \) and \( \Sigma_U \) can be obtained as follows. Assume that the acquisition scheme where the \( z_i \)'s are iid as in (4). For simplicity, suppose the measurement errors are homoscedastic. Appealing to the strong law of large numbers as before, we have

\[
\frac{1}{n}B_n = \frac{1}{n} \sum_{i=1}^{n} z_i z_i^T \exp(2z_i^T \theta) \to B_{\theta} = E(zz^T \exp(2z^T \theta)),
\]

(5)

and similarly, as stated in theorem 1c, \((1/n)G_n \) and \((1/n)F_n \) converge to the same limit \( F_{\theta} \) by Slutsky’s theorem, the authors’ equation (5) becomes \( \sqrt{n}(\hat{\theta}(k) - \theta_0) \to \mathcal{N}(0, \Sigma_0) \), where \( \Sigma_0 = B_{\theta}^{-1}F_{\theta}B_{\theta}^{-1} = \sigma^2 B_{\theta}^{-1} \). Because \( \theta = (\Sigma_0, B_{\theta}^T) \) and \( B = \text{vecs}(D) \), assumption (C9) holds under the conditions of theorem 1, and \( \Sigma_D \) is equal to the 6 \times 6 lower diagonal block of \( \Sigma_0 \).

When \( D \) is isotropic, two simplifications occur. First, \( \Gamma = I_3 \), \( \Lambda = \lambda I_3 \), and the authors’ equation (8) reads \( U_n = \sqrt{n}(D - D) \to U \). Thus \( U = \Sigma_D = D \). Second, according to the acquisition scheme (4), \( z_i^T \theta = \log S_0 \) with probability \( \gamma \) and \( z_i^T \theta = \log b - bx^T \) with probability \( 1 - \gamma \); thus

\[
B_{\theta} = S_0^2 \begin{pmatrix}
\gamma + (1 - \gamma)e^{-2b\lambda} & -(1 - \gamma)e^{-2b\lambda}bE(x^T) \\
-(1 - \gamma)e^{-2b\lambda}bE(x) & (1 - \gamma)e^{-2b\lambda}b^2E(xx^T)
\end{pmatrix}
\]

(6)

Note that \( B_{\theta} = S_0^2 A_{\gamma} \) when \( \lambda = 0 \), where \( A_{\gamma} \) is as given by (1).
Recall that $\Sigma_U$ is the lower $6 \times 6$ diagonal block of $\Sigma_\theta = \sigma^2 B_\gamma^{-1}$. Applying the Schur inversion formula for block matrices (Schott 2005, p. 247) to (6) and deleting the first row and first column gives

$$\Sigma_U = \frac{\sigma^2}{S_0^2 b^2} \left( (E(xx^T))^{-1} \begin{pmatrix} 2 I_p - C_p(\gamma) & I_p I_p^T \\ 0 & I_p(p-1)/2 \end{pmatrix} \right).$$

Expression (7) is valid for any distribution of $r$ on the sphere and demonstrates the dependence on the signal-to-noise ratio and the advantage of using high values of $b$ in the acquisition. When $r$ is uniform, expressions (2) and (3) reduce (7) explicitly to

$$\Sigma_U = \frac{p(p+2)\sigma^2 e^{2b\gamma}}{4S_0^2 b^2(1-\gamma)} \begin{pmatrix} 2 I_p - C_p(\gamma) & I_p I_p^T & 0 \\ 0 & I_p(p-1)/2 \end{pmatrix},$$

with $C_p(\gamma) = 2/(p+2) - e^{-2b\gamma}(1-\gamma)/\gamma$. In particular for DWI ($p = 3$), we get

$$\Sigma_U = \frac{15\sigma^2 e^{2b\gamma}}{4S_0^2 b^2(1-\gamma)} \begin{pmatrix} 2 I_3 - C_3(\gamma) & I_3 I_3^T & 0 \\ 0 & I_3(3-1)/2 \end{pmatrix},$$

with $C_3(\gamma) = 2/5 - e^{-2b\gamma}(1-\gamma)/\gamma$.

From this calculation, we learn that when $D$ is isotropic, the diagonal entries of the estimation residuals $U_n = \sqrt{n}(D - \bar{\lambda} I_1) \rightarrow U$ [the covariance of which is indicated by the upper $3 \times 3$ diagonal block in (8)] are asymptotically uncorrelated with the off-diagonal entries (the covariance of which is indicated by the lower $3 \times 3$ diagonal block). The correlation between the diagonal entries of $U$ themselves is symmetric with respect to those entries, as expected from the isotropy. In particular, the diagonal entries of $U$ become uncorrelated ($C_3(\gamma) = 0$) when the fraction $\gamma$ is chosen to be equal to

$$\gamma_0 = \frac{1}{1 + 2 e^{2b\gamma}/(p+2)},$$

are positively correlated when $\gamma < \gamma_0$ and negatively correlated if $\gamma > \gamma_0$. Moreover, when $\gamma = \gamma_0$, the diagonal entries of $U$ have exactly twice the variance as the off-diagonal entries, and $U$ has the distribution known in random matrix theory as the Gaussian orthogonal ensemble (Mehta 1991). This distribution has been proposed for modeling variability of diffusion tensors (Schwartzman 2006). Using $b\lambda = .7$ as done by the authors gives $\gamma_0 = .381$.

Because the interest is usually in estimating $D$ rather than $\delta_0$, the fraction $\gamma$ could be chosen to minimize the variance in the estimation of $\hat{\lambda}$ as opposed to that of the vector $\theta$. Using $\hat{\lambda} = \text{tr}(D)/3$ as the estimate of $\lambda$ (as the authors suggest at the end of sec. 2.2), we get from (8) that the asymptotic variance of $\sqrt{n}(\hat{\lambda} - \lambda)$ is proportional to

$$f_p(\gamma) = \frac{2 - C_p(\gamma)}{1 - \gamma} = \frac{2(p+1)}{(p+2)(1-\gamma)} + \frac{e^{-2b\gamma}}{\gamma},$$

which is minimized at

$$\gamma^* = \frac{1}{1 + \sqrt{2} e^{2b\gamma}(p+1)/(p+2)}.$$
Again, using $b\lambda = .7$ as done by the authors gives $\gamma^* = .282$ (Fig. 3). This value is smaller than the optimal $\gamma$ obtained in the previous section, but the minimum is not as sharp.

When $D$ is not fully isotropic, the calculations are complicated by the fact that they depend on the particular orientation in space of the anisotropic axes. However, it can be argued that if $r$ is sampled uniformly, then the estimation variance should not depend on those axes. Then, without loss of generality, one may assume that $\Gamma = I_3$. Then again, we have that $\Sigma_U = \Sigma_D$, and the required submatrices $\Sigma_{U_{11}}$ and $\Sigma_{U_{22}}$ can be taken as the appropriate submatrices of (8). Note that because of the symmetry, we get that in fact, $\Sigma_{U_{11}} = \Sigma_{U_{22}}$. Similarly, the variance $\sigma_{ij}$ in the authors’ corollary 2 is equal to any one of the upper three diagonal entries of (8) and thus is also proportional to (9). Interestingly, the minimization problem in terms of $\gamma$ turns out to be the same as in the isotropic case.

The distribution of the eigenvector matrices $C$, $C_{11}$, and $C_{22}$ in theorems 3 and 4 and corollaries 1 and 2 are more difficult to assess directly from the written densities. However, intuition dictates that $C$ should be uniformly distributed on the orthogonal group $O(3)$, whereas $C_{11}$ and $C_{22}$ should be uniformly distributed on the orthogonal group $O(2)$ in the oblate and prolate cases. The authors did not write this observation explicitly, but their figures 1 and 2 support this conjecture. Figures 2(b) and 2(f) make this clear for the oblate and prolate case. Similarly, the angle histogram plots in figure 1 indicate a uniform distribution on the sphere, because the uniform density in spherical coordinates is $\sin(\phi)/(4\pi)$.

4. SUMMARY AND FINAL REMARKS

Using the authors’ results, I have suggested a stochastic acquisition scheme [eq. (4)] and derived from it analytical forms for the covariances $\Sigma_\theta$ and $\Sigma_\gamma$. These formulas provide insight into the covariance between the entries of the estimated diffusion tensor $D$ and validate a model of variability for diffusion tensors proposed by Schwartzman (2006).

In addition, using these formulas, I have calculated the optimal fraction of measurements $\gamma^*$ at $b = 0$ in two different ways. When the objective is to optimize the rate of convergence of the estimate $\hat{\theta}(k)$, the optimal fraction is $4/9$. When the objective is to minimize the asymptotic variance of the estimate of the eigenvalues of $D$, the optimal fraction is $.28$, although this criterion is less sensitive to departures of $\gamma$ from the optimum. Both values are higher than one would intuitively guess, because measurements at $b = 0$ do not provide information about $D$. This might be necessary to compensate for the singularity of the design when those measurements are not present.

As a final remark, one issue that the authors overlook is the multiple-testing problem. This issue presents itself in two ways. First, it is present in the hierarchical testing scheme proposed in section 2.3. Classification of a diffusion tensor requires evaluation of one, two, or three tests in a certain order, so $p$ values need to be adjusted accordingly. Second, a large-scale multiple testing problem is present when the classification scheme is applied to the tens of thousands of voxels in the brain. Because of this, the marginal $p$ values reported in Section 4 and Figure 3 are much lower than they would be were multiple testing corrections applied.

ADDITIONAL REFERENCES


