Diffusion MRI Parametric Modelling

1 Parameter Estimation and Mapping

1.1 Weighted Linear Diffusion Tensor Model and Visualised Property Mappings

In this HARDI dataset, we use weighted linear diffusion tensor to model the brain slice 72, by squaring the measurement diagonal matrix. Three figures are extracted. Mean diffusivity map, Figure 1, blacker background and more accurate compared to unweighted. Fractional anisotropy map, Figure 2, depicts the regions such as the corpus callosum and web-like structures connecting the brain's peripheries. The third direction-encoded colour map, Figure 3, is calculated by normalising and scaling the largest eigenvector, which expresses the principal diffusion direction. We observe left-right is strongly marked in the central area which communicates between hemispheres, while laterally, fibres emphasise superior-inferior and front-back connections.

1.2 Ball-and-Stick Model and Optimiser

The next model fitted is Ball-and-Stick model, simulating isotropic diffusion and axon fibres. We optimise it with `fminunc` function in Matlab, tweaking `optimset`. `Display` helps identifying the stopping reason, whereas `MaxIter, MaxFunEvals, TolX, TolFun` help control the stopping criteria, e.g. increasing `ToX` from $1e^{-10}$ to $1e^{-6}$ stops the process earlier giving similar results. Using proposed starting point, the model reaches expected values. We observe significant deviations between estimated and actual measurements in Figure 4, which is expected from a squared residual norm (resnorm) of $2.87e^{7}$. Because if the standard deviation (std) of noise in each measurement is around 200, ideally resnorm is around $200^2 * 108 = 4.32e^6$.

1.3 Physically Constrain Ball-and-Stick Model with Transformation

Suspecting non-realistic parameters as the cause of the large resnorm, we transform parameters to constrained settings before calculating the signal in SSD (Sum of Square Differences). Once fitted, values are transformed again to real parameters. Original starting point is inversely transformed as input to optimiser. The constraints are positive S_0, d by squaring, f between 0 and 1 using sigmoid. θ and ϕ are not transformed due to periodicity. The resnorm and parameters are shown in Table 1. The resnorm decreased to $5.8720e^6$, closer to expected $4.32e^6$, because the parameters align better with realistic physical-world settings. The predictions in Figure 5 are closer to actual measurements.

1.4 Stochastic Starting Point Optimisation and Convergence Analysis

To avoid missing global minimum, we employ stochastic starting points. Using Gaussian noise at 20% we obtain 100 noisy starting points, rejecting non-realistic ones, with final distribution of points shown in Figure 6. Fitting and recording of results is performed for each point. Most (98%) of trials found same minimum $5.8720e^6$ indicating global minimum, while tiny portion falls in a local minimum of $1.5690e^7$ resnorm, as shown in Figure 7. The distribution in Figure 8 also agrees on a similar small range, as confirmed in Table 2 with small std. Trying some other voxels, Figure 9, and the resnorm results have same pattern in Figure 10. If probability of finding global minimum is P, then we are asking n in $1-(1-P)^n >= 0.95$, which can be written as $n >= ln(0.05)/ln(1-P)$. Our P seems to be 0.98 but let's suppose it's $P = 0.5$, 5 iterations are needed; thus, we use 10 iterations as a tolerance buffer.

1.5 Whole Slice Mapping and Parameter Visualisation

Using same procedure, i.e. constrain, stochastic starting point, and skipping no signal voxels, we create whole slice mapping. Resnorm map, Figure 11, shows sensible values between $5e^6$ and $1e^7$. Middle bundles of axons have higher error, suggesting difficulty in capturing complex diffusion, e.g. crossing fibres. S_0 map, Figure 12, has similar pattern as mean diffusivity map. Ball-and-Stick diffusivity map, Figure 13, shows granular brain structure, gyri and sulci, and gradual increase in diffusivity. Fractional anisotropy map, Figure 14, matches previous visualisation, showing proportion of axons even more precisely. Direction of fibre map in 2D, Figure 15, zoomed in Figure 16, shows smooth flows of axons, e.g. corpus callosum. Directionality near the scalp is unclear due to the complexity of the surface landscape.

1.6 Exploring Transformation and Informed Starting Point

We explore sidedly the transformation of parameters. In particular, switching to an exponential transformation instead of squaring for positivity and adding constraints to angles. The former provides one-to-one, monotonically increase, and could fit the exponential nature of signals. The latter, while theoretically unconstrained, mapping into limited period may save computational time. Results in Table 3 show that using angle transformation often leads to higher resnorm, fitting angles to 0 even with noisy starting point. Positive transformation pairs converge into the same global minimum, despite different unconstrained parameters.

Informed starting point may speed up the process. We use parameters calculated in Q1.1.1. The ideal mapping found is using S_0 directly, mean diffusivity as d, fractional anisotropy as f, the normalised eigenvalues using $\arccos(z)$ and $\arctan 2(y, x)$ to estimate θ and ϕ. Global minimum is ensured with preliminary trials. We record the computing time to map whole slice in Table 4, where transformation of angles is excluded. We observe that informed starting point has variable effects depending on the transformation used. It is quicker when using squaring, but slows down when using exponential. More surprisingly, using exponential with non-informed starting point boosted the speed of convergence from thousands of seconds to quickest $397.42s$. Future exercise leverages this result.

2 Uncertainty Estimation

2.1 Classical Bootstrapping to Estimate Uncertainty

Now we want to estimate the uncertainty of parameters, starting with classical bootstrapping. We sample measurements with replacement and fit the model to new data, using stochastic starting point and constraint settings. The process is repeated 1000 times to obtain a distribution of parameters and their resnorms, in Figure 17. Table 5 shows the 2σ and 95% range of parameters and resnorm.

We observe that resnorm seems normally distributed with little skewness, but significantly higher than previous predictions, deteriorating from $5.8720e^6$ to $9e^7$ levels, indicating the importance of experimental design as disruptions lead to poorer performance. The distribution of S_0 is slightly skewed to the right, as 95% mass is on the right side of the 2σ range. The values are smaller too, ranging from $4.2579e^3$ to $2.5e^3$ levels. d is normally distributed too with less skewness, the previous estimation falls within the range. f is unimodal with 759/1000 of samples >= 0.99, meaning most predictions estimate majority of this voxel as axons. Originally, the design predicted a more balanced distribution with 35% axons. θ is interesting in having a bimodal and symmetric distribution around 0, which likely reflects the symmetry of the polar angle in the horizontal plane. ϕ also shows a narrower normal distribution with little skewness, with periodic outliers expected from angles. In summary, the bootstrap predictions yield higher resnorm and distinct parameter distributions.

Strikingly, when tried with other voxels at different positions of the brain, they all show similar patterns, as shown in Figures 18 - 20. Not only the does the bootstrap model yield higher resnorm, but it also predicts similar model settings across voxels.

2.2 MCMC Metropolis-Hastings to Estimate Uncertainty

We also attempt posterior estimation with Monte Carlo Markov Chain (MCMC), using Metropolis-Hastings (MH) algorithm. The best set of previously estimated parameters serves as the starting point. The perturbation was initially chosen heuristically as two times the range of the parameters, but it led to low acceptance ratio as shown in Figure 21. After trials, a refined perturbation covariance matrix is chosen to be $[0.0015, 0.0122, 0.0060, 0.0004, 0.0010]$ in unconstrained settings, achieving the desired 20-50% acceptance ratio (see Figure 22).

The sampling procedure is as follows. For each iteration, a new set of unconstrained parameters is proposed by adding Gaussian noise with the perturbation covariance. Then we calculate the log posterior, which, assuming flat priors, is proportional to the log likelihood. The latter assuming a Gaussian distribution is given by $-0.5 * SSD/\hat{\sigma}^2$ where $\hat{\sigma}^2$ is the variance of residuals, calculated with estimated and actual measurements. The log evidence cancels out in the ratio, no need to calculate. We accept the new parameters with probability $\alpha = \exp(logPosterior_{proposed} - logPosterior_{current})$. An adaptive mechanism is added to perturbation covariance, checking acceptance rate every 100 iterations, and increase by 10% if below 20% rate, decrease by 10% if above 50% rate. Finally, we postprocess the 10000 samples with 2000 burn-in and 4 thinning, and convert back to constrained parameters. Resnorm is also calculated for each sample. The best sample is recorded in Table 6.

Figure 23 shows the results in a balanced Gaussian distribution, with close 2σ and percentile ranges. This time, most of resnorm is around $6e^6$, thus much better than the classic bootstrap model. The distributions of other scalars are also more balanced and natural, with S_0 around $4.2e^3$, d around $1e^{-3}$, f around 0.35, θ around 1, and ϕ around -2.6. This model provides more confident estimates within a certain range. Overall, the MCMC model does a closer prediction to previous best parameters. Using another voxel (100, 125) as an example, we confirm the acceptance rate in Figure 24, with a distinct model captured for this voxel in Figure 25. Located in the temporal cortical area, where gray matter is expected to dominate over white matter, the estimates effectively show lower diffusivity d, tiny fraction f and minimal directionality along the axons.

2.3 Laplace Approximation and Parametric Bootstrap

Two other methods are used to estimate uncertainty, Laplace approximation and parametric bootstrap. In Laplace Approximation, we sample from the multi-variate Gaussian distribution, with best estimate as mean and Σ covariance matrix. Σ accounting for noise is calculated as $2\hat{\sigma}^2 * inv(Hessian)$, where the $Hessian$ is obtained from the `fminunc` function. 1000 samples are taken and the distribution is shown in Figure 26. The results are mostly sensible, except for d which has high variance, and therefore reject sampling is used when d is negative. This causes instability in resnorm. Other scalars have ranges close to each other, indicating a balanced Gaussian distribution, as expected.

Parametric bootstrap is achieved by adding noise to signal created by best estimate. The σ for the Gaussian noise chosen is the square root of SSD in unbiased form. We fit parameters to the new signal and repeat this 1000 times. Figure 27 shows its distribution. Overall decent predictions, but a few parameter predictions led to a rather high ResNorm, lacking some robustness.

Comparing intervals we notice that in general Laplace approximation is confident in its estimation, with narrowest S_0 and f, but high variance in d. So it is not quite as robust as MCMC. Classical bootstrap is the least confident of all with deviating results, and parametric bootstrap is in between with narrowest angle estimation.

3 Model Selection

3.1 Ball-and-Stick Model and ISBI 2015 Dataset

Turning to modelling voxels in ISBI 2015 dataset. We continue with the same suggested starting point, noise perturbation, constrained settings, and run 100 stochastic starting points. Observing Figure 28, we almost always hit the global minimum with tolerance at $4e-13$ level. Table 7 records best parameters and resnorm. Given that resnorm is around $1.5106e^1$ and 0.04 std noise would ideally match to 5.7792, it's about half magnitude away. We conclude this Ball-and-Stick model is acceptable but could be better. A visualisation of estimated vs actual measurements is shown in Figure 29.

3.2 Diffusion Tensor and Zeppelin-and-Stick Family Models

We further extend the range of models to three more models. Firstly non-linear diffusion tensor model, we estimate our starting point using the linear fitting as in Q1.1. We constrain S_0 with positivity and ensure diffusion tensor is positive definite. Table 8 records results and Figure 30 visualises estimation. Then, Zeppelin-and-Stick (ZS) model, characterised by two eigenvalues, we incorporate $lambda_2$. Same approach in modelling, we constrain and set starting point. However, the first starting points wasn't ideal as only small portion reaches global minimum, Figure 31. After refining, informed starting point reaches global minimum almost in all trials, Figure 32. Table 9 records results and Figure 33 visualises estimation. Thirdly, for the variant Zeppelin-and-Stick-with-Tortuosity (ZST) model, no extra parameter is needed as $lambda_2$ is derived from other parameters. Using the same approach, we ensure minimum resnorm distribution as Figure 34 shows. Table 10 records results and Figure 35 visualises estimation. Notice that f predicted is slightly higher than in previous models.

Looking at the best resnorm of models in Table 11, ZS is the lowest hence best model by resnorm, indicating the effectiveness of additional parameter. The second place is Ball-and-Stick, followed by ZST. Diffusion tensor performed the worst.

3.3 Information Criteria and Model Ranking

In order to account for the trade-off of the number of parameters, we use AIC and BIC to assess their performance. AIC is calculated with log likelihood assuming Gaussian noise and estimated std, thus number of parameters N becomes $N + 1$ for the std. Then use $2N + Klog(SSD/K)$ where K is number of measurements. BIC is calculated in a similar fashion with $Nlog(K) + 2Klog(SSD/K)$. These are calculated for each of the four models and the results were recorded in Table 12. We observe consistent rankings across ICs, having ZS as the best model, meaning best balance between model complexity and goodness of fit. Adding parameters to better explain the data could be desirable at this stage.

3.4 Adding Ball-and-Two-Stick as Candidate

As previous findings suggest, we add a more complex model Ball-and-Two-Stick (BTS) where sticks in two directions (f_2, θ_2, ϕ_2) are added, to relax the parallel constraint. We use the same settings as before, and after informed starting point, the distribution of resnorm is shown in Figure 36. Table 13 records results and Figure 37 visualises estimation. We end up having $1.1736e^1$ as ResNorm, suggesting better fit. Notice how f is split into three parts, indicating the importance of allowing multiple directions. Reviewing the ICs of five models in Table 14, BTS is the best model by both ICs, followed by ZS. This proves the aforementioned point that adding parameters to better explain the data is desirable.

3.5 Consistency among Voxels and Selection by Cross-Validation

To look at the consistency of findings, we run all five models on six provided voxels and record their respective ICs. The results are recorded in Table 15. The best model varies among voxels, preferring ZS in voxels 3, 5, 6, and BTS in voxels 1, 2, 4. Diffusion Tensor is consistently the worst. Model families are consistent as well, e.g. if ZS is the best model for a voxel, ZST usually surpasses Ball-and-Stick too, and vice versa. Note that a missing measurement in voxel 5 at 1338 is replaced by the mean of the voxel signal.

Contrary to IC approach, we also test selection by cross validation, using resnorm as the metric. We split the data into 6 folds, iteratively leaving one for validation and the rest as training. In Table 16, we record the resnorms, average and number of folds with minimum resnorm. The ideal resnorm is around $9.63e^{-1}$ for 602 measurements at 0.04 std, so our results are sensible. BTS model achieves lowest average resnorm across folds, and by voting it is also the model with the most best-folds, positioning it as the best model by cross validation. ZS comes second, with these two outperforming all others. Bayesian model selection could be another option to compare posterior of models. But it's not easy to set priors and making a decision when models don't significantly surpass one another.

3.6 Whole Slice Model Selection Mapping and Akaike Weights

Going back to the HARDI dataset, we use Ball-and-Stick and ZS models to fit the whole slice. We record their resnorm and calculate the corrected AIC, due to smaller measurement size. The correction term added to AICc is $2N(N + 1)/(K - N - 1)$. The map created with AICc is shown in Figure 38, representing the best model using colour. We observe that Ball-and-Stick dominates central ventricle areas, whereas ZS form border region in what is probably subcortical white matter and surface of cortex with gyri and sulci. The left and right sides are overall consistent, despite imperfect symmetry in ventricles and subcortical white matter. These locations have anatomical relevance. For instance, Ball-and-Stick's isotropic diffusion better describes CSF and gray matter in clustered regions, whereas ZS's anisotropic diffusion better describes axons and wired structures.

Akaike weights can provide a smoother transition compared to binary decisions, using $exp((AIC_min - AIC_i)/2)$. We visualise this as values for red and blue intensity in Figure 39. Transition areas between two models give us a notion of not only the preference but also the proportion. For instance, the caudate putamen lateral to ventricles strongly resonates with Ball-and-Stick, probably due to its condensed gray matter cells. In contrast, some regions in subcortical white matter are consistent with Zeppelin structure, indicating the large proportion of white matter and guided anisotropic. By using Akaike weights, we observe a more gradual transition between models, which is more realistic.