ESTIMATION OF RELAXATION TIME DISTRIBUTIONS IN MAGNETIC RESONANCE IMAGING

Kelvin J. Layton[†⋆] *Leigh A. Johnston*[†‡] *Peter M. Farrell*[†] *Bill Moran*[†] *Mark Morelande*[†]

[†]Department of Electrical and Electronic Engineering, The University of Melbourne
[⋆]NICTA Victorian Research Laboratory [‡]Florey Neuroscience Institutes (Parkville)

ABSTRACT

Recently there has been increasing interest in estimating the distribution of relaxation times contributing to a magnetic resonance signal. This paper shows that it is impractical to estimate the spread of such a distribution from typical measurements. Instead, a Bayesian estimator is developed for a discrete distribution, which is very robust to noise. Although the distribution spread is not modelled, the estimates capture the main features of the distribution such as the mode locations and often provide improved myelin water fraction estimates in simulation examples.

***Index Terms*—** MRI, T_2 relaxation, myelin water fraction, multiexponential analysis

1. INTRODUCTION

The signal decay in magnetic resonance imaging (MRI) has been shown to arise from a distribution of relaxation time constants, particularly for myelinated tissues [1]. Consequently, the accurate estimation of the distribution of T_2 relaxation time constants (or 'relaxation times') is of particular interest to study white matter diseases such as multiple sclerosis, which directly affects myelination. It is understood that the water within the myelin sheath has a relatively short T_2 ($\sim$10–20ms) compared to that from intracellular and extracellular water [2]. This property is used to quantify the amount of myelin within a voxel, known as the myelin water fraction (MWF), derived directly from the relaxation time distribution.

Previous attempts at estimating the distribution have involved fitting a large number of decaying exponentials using the nonnegative least squares (NNLS) algorithm [3]. The problem is inherently ill-conditioned and regularisation is needed to improve the sensitivity to noise. This regularisation introduces bias that favours less complex distributions and often the number of peaks is underestimated at clinically achievable signal-to-noise ratio (SNR) values.

Alternatively, a parametric model consisting of three discrete pools was proposed in [2] and subsequently tested in [1]. The estimator for this model was a Quasi-Newton fitting algorithm. However, in those simulations, the algorithm was initialised to the true values and poor performance was observed for initialisations outside 10% of the true values.

In this work, we analyse a parametric, yet continuous, model for the unknown distribution. The model consists of a mixture of inverse-gamma components, each with a weight, location and spread parameter. We show that estimation of such a model is impractical for clinical experiments and focus our attention on estimating a discrete distribution similar to [2]. To this end, we propose a robust Bayesian algorithm to overcome the problems associated with low SNR and algorithm initialisation. Finally, we compare the performances of our algorithm and NNLS for the purposes of myelin water fraction estimation.

The paper is organised as follows. In Section 2 we discuss the three models under consideration: non-parametric (NNLS), an inverse-gamma mixture and discrete relaxation times. In Section 3 we calculate the lower bound on estimation performance for the parametric models. Finally, in Sections 4 and 5 we present the estimation algorithm and corresponding simulations, respectively.

2. THEORY

The unknown distribution of relaxation times is observed through the amplitudes of the acquired echo signals. The measurements are described by the integration of the decaying signals from each contribution,

$$y(t_i) = \int f(\tau) e^{-t_i/\tau} d\tau + v_i \tag{1}$$

where $f(\cdot)$ is the distribution of relaxation times, t_i is the echo time and v_i is assumed to be an i.i.d. Gaussian random variable with known variance, σ^2. The problem of estimating a continuous distribution function from a finite number of observations is inherently ill-posed. In practice, we are forced to make some assumptions or approximations about the specific form of the distribution, f. In this work we consider three such approximations and examine the corresponding estimation performance.

The first approximation is to grid the parameter space. In this case, a large but known sequence of relaxation times is defined, $\tau_1, \ldots, \tau_m$, covering a physically plausible range of times. This approximation is described by a distribution made up of m delta functions,

$$f(\tau) = \sum_{j=1}^{m} w_j \delta(\tau - \tau_j) \tag{2}$$

The corresponding signal model is obtained by combining (1) and (2),

$$y(t_i) = \sum_{j=1}^{m} w_j e^{-t_i/\tau_j} + v_i \tag{3}$$

This work is supported by NICTA Lifesciences, the Victorian Research Laboratory. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

We can construct a vector of measurements, $\boldsymbol{y} = [y(t_1), \ldots, y(t_n)]'$ and weights, $\boldsymbol{w} = [w_1, \ldots, w_m]'$ and rewrite (3) as a matrix equation,

$$\boldsymbol{y} = A\boldsymbol{w} + \boldsymbol{v} \tag{4}$$

where A is an $n \times m$ matrix with elements, $A_{i,j} = e^{-t_i/\tau_j}$. The problem of estimating the distribution is reduced to estimating the vector of weights, $\boldsymbol{w}$. Magnetic resonance experiments are generally expensive and time-consuming. This means the number of measurements is often significantly smaller than the number of weights we wish to estimate ($n < m$) and the system is underdetermined. Consequently, a least squares estimator that solves (4) will have poor noise performance. The solution proposed in [3] is to regularise the optimisation and solve,

$$\begin{array}{ll} \text{minimise} & \|\boldsymbol{y} - A\boldsymbol{w}\|_2^2 + \lambda\|C\boldsymbol{w}\|_2^2 \\ \text{subject to} & \boldsymbol{w} > 0 \end{array} \tag{5}$$

where C contains additional constraints (such as smoothness), weighted by the regularisation parameter, λ.

An alternative formulation considered in this paper is the parametric estimation of a mixture of distributions. We propose the distribution is made up of a small number of modes, each with an inverse-gamma distribution. The inverse-gamma distribution is particularly well-suited to our problem: it can approximate a wide range of distributions; it has positive support, which is suitable for relaxation times; and importantly it leads to a tractable integration in (1). The inverse-gamma mixture is given by

$$f(\tau) = \sum_{i=1}^{m} w_j \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} \tau^{-\alpha_j - 1} e^{-\beta_j/\tau} \tag{6}$$

The three parameters, w_j, α_j, β_j, characterise the weight, location and scale of the j^{th} mode, respectively. Substituting into the signal model (1) yields,

$$y(t_i) = \sum_{j=1}^{m} w_j \left(\frac{\beta_j}{t_i + \beta_j} \right)^{\alpha_j} + v(t_i) \tag{7}$$

Analogous to the non-parametric model, imposing this form on the distribution reduces the estimation problem to finding $3m$ parameters. In this way, this model can be seen as a parametric alternative to NNLS. An alternative parameterisation is to characterise each mode in the mixture by the mean and variance of the inverse-gamma distribution, given by

$$\nu_j = \frac{\beta_j}{\alpha_j - 1}; \qquad \rho_j^2 = \frac{\beta_j^2}{(\alpha_j - 1)^2(\alpha_j - 2)} \tag{8}$$

This parameterisation is useful to analyse the fundamental ability to estimate the location and spread of the relaxation times.

In the limit as $\rho_j^2 \to 0$ for fixed $\nu_j = \tau_j$, the distribution approaches a simple discrete model, consisting of a small number of weighted Dirac spikes at relaxation times, $\tau_j, j = 1, \ldots, m$. The corresponding signal model in (7) becomes

$$y(t_i) = \sum_{j=1}^{m} w_j e^{-t_i/\tau_j} + v_i \tag{9}$$

This model has been used previously for relaxation rate estimation [1]. It is fundamentally different from the non-parametric form in (3), since the times τ_j are unknown and must be estimated along with the weights.

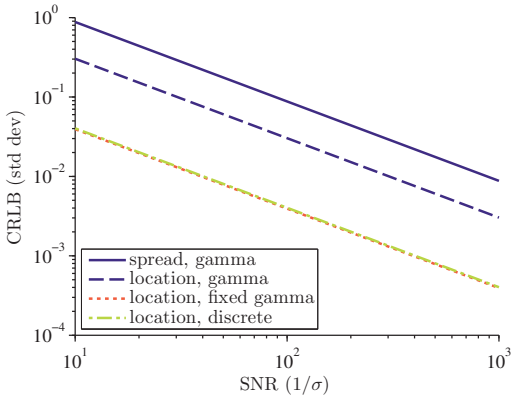

Fig. 1. The estimation bounds for the spread and location parameters of different parametric distributions.

3. ESTIMATION BOUNDS

The minimum variance of an unbiased estimator is given by the Cramer Rao Lower Bound (CRLB). This bound is often used as a benchmark to compare estimators, but an equally valid use is to analyse the uncertainty in an estimation problem. It reveals our ability to estimate the parameters for a given experimental setup. Of particularly interest in this work, is the effect of the distribution spread on parameter estimation.

Although we calculate the Fisher Information Matrix (FIM) for the signal model with the initial parameterisation (in terms of α_j and β_j), it is much more instructive to consider the estimator's performance in terms of the location and spread of relaxation times. Thus we use the relationships in (8) to transform the matrix according to

$$I(\nu, \rho) = J' I(\alpha, \beta) J \tag{10}$$

where J is the Jacobian matrix of the mapping associated with (8).

We calculate the CRLB for the parameters of three different distribution models. Firstly the inverse-gamma mixture parameters, w_j, ν_j and ρ_j, representing the weight, location and spread, respectively. We also consider a discrete distribution with parameters for the weights, w_j, and locations, τ_j. Finally for comparison we consider a gamma mixture with the spread parameter fixed and known leaving only the weight, w_j, and location, ν_j, to be estimated. We evaluate the CRLB for biologically realistic distributions, consisting of two modes: a slow mode with $\nu_1 = \tau_1 = 100\text{ms}$, $\rho_1 = 10\text{ms}$ and a fast mode with $\nu_2 = \tau_2 = 20\text{ms}$, $\rho_2 = 10\text{ms}$. These features are common for white matter tissue in the cortex. The experimental setup is typical of a MR sequence optimised for relaxation time estimation, consisting of 32 echoes with echo times equally spaced between 10ms and 506ms.

Fig. 1 displays the CRLB for the parameters of the first mode of the mixture distribution under different model assumptions. The plot demonstrates that the spread parameter is exceedingly difficult to estimate. For an SNR of 20, the spread of the slow mode can only be estimated with a standard deviation of $\sim$500ms, five times greater than the value of the location parameter. These results also highlight that estimating the location is much harder when the spread is unknown.

To achieve useful estimates of the weight, location and spread parameters of the distribution, the SNR would need to be $\sim$3000, far beyond that achievable on a clinical MR system. Alternatively we

would need to collect in the order of 10^5 echoes, which would bloat the acquisition time unsatisfactorily. Similar plots for the second mode or different parameter values yield the same conclusions, we cannot estimate the spread of the relaxation time distribution.

The reason for such poor estimation performance is that a large number of distributions will produce very similar measurements. For reasonable noise levels, the difference is indistinguishable and we cannot pick the correct spread model. This result is not specific to an inverse-gamma distribution and holds for other parametric distributions that incorporate a spread.

Since we have no means to estimate the spread parameter, we must assume a known spread or adopt a discrete distribution model that doesn't model it. We will see that the discrete distribution is suitable for estimating the main contributions to the signal.

4. ESTIMATION ALGORITHM

Surprisingly, the simple model in (9) does not lend itself well to a simple estimation algorithm. The nonlinear relationship between the parameters and the signal creates a poorly behaved cost function, with local minima and large regions in parameter space where the cost function is essentially flat. These features will be problematic for a naïve gradient-based optimisation algorithm.

We adopt a Bayesian framework that leads to a numerically robust algorithm and allows us to incorporate prior information about the biological tissue. The cornerstone equation is Bayes rule,

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \frac{\ell(\boldsymbol{x}|\boldsymbol{y})\pi_0(\boldsymbol{x})}{\pi(\boldsymbol{y})} \tag{11}$$

where $\boldsymbol{x}$ is a vector of unknown parameter. The numerical challenge of computing (11) for a relatively wide prior, π_0, and narrow likelihood, ℓ, is overcome using a technique known as progressive correction [4]. Although we do not use a Monte Carlo approximation, we apply the same principle of 'flattening' the likelihood and iteratively correct our estimate of the posterior. We define a schedule of s corrections, $\gamma_1, \ldots, \gamma_s$, with the intermediate posterior at the j^{th} correction step given by

$$\pi_j(\boldsymbol{x}|\boldsymbol{y}) = \tfrac{1}{\eta_j}\ell^{\gamma_j}(\boldsymbol{x})\pi_{j-1}(\boldsymbol{x}|\boldsymbol{y}) \tag{12}$$

where η_j is a normalising constant. When $\sum_j \gamma_j = 1$, the final posterior π_s is the required one defined in (11). To make the computation in (12) tractable, we use a linearised approximation of the likelihood and a Gaussian prior with mean, $\boldsymbol{\mu}_0$, and covariance Γ_0. In this case, we do not compute the exact posterior, but rather a Gaussian approximation to it. The approximate likelihood is obtained by linearising the nonlinear function, $\boldsymbol{h}(\cdot)$, which describes the signal model in (9),

$$\hat{\ell}^{\gamma_j}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{h}(\boldsymbol{\mu}_{j-1}) + J(\boldsymbol{\mu}_{j-1})(\boldsymbol{x} - \boldsymbol{\mu}_{j-1}); \tfrac{1}{\gamma_j}\Sigma); \tag{13}$$

where the J is the Jacobian of the function $\boldsymbol{h}(\cdot)$. Analogous to the Kalman filter, the posterior at each step is a Gaussian with mean and variance given by,

$$\boldsymbol{\mu}_j = \boldsymbol{\mu}_{j-1} + K_j(\boldsymbol{y} - \boldsymbol{h}(\boldsymbol{\mu}_{j-1})) \tag{14a}$$

$$\Gamma_j = (I - K_j J)\Gamma_{j-1} \tag{14b}$$

where $K_j = \Gamma_{j-1}J'(J\Gamma_{j-1}J' + \frac{1}{\gamma_j}\Sigma)^{-1}$.

Initially, this approximation will be poor but the likelihood will be wide due do severe flattening. As the algorithm progresses, the approximate posterior approaches the true posterior and the linearisation becomes more accurate. This process is very similar to simulated annealing in optimisation and we will see that the resulting algorithm is very robust to local minima, particularly at low SNR. The final algorithm is described in Algorithm 1

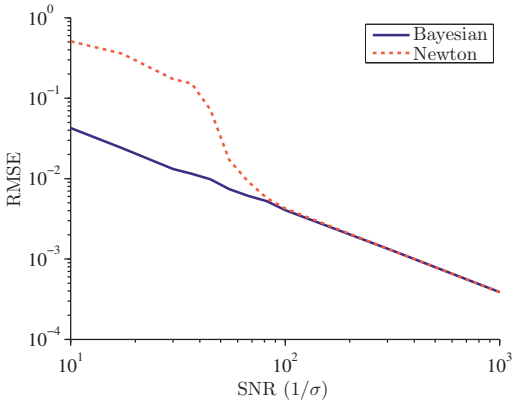

Fig. 2. The RMSE at various SNR values for two estimators: a gradient-based optimisation and the proposed Bayesian estimator.

Algorithm 1: Estimation Algorithm

for $j = 1 \ldots s$ **do**
  Calculate: $J(\boldsymbol{\mu}_{j-1})$ and K_j
  Calculate $\boldsymbol{\mu}_j$ and Γ_j according to (14)
Return estimate: $\hat{\boldsymbol{x}} = \boldsymbol{\mu}_s$

The effectiveness of the algorithm is demonstrated by calculating the empirical MSE of the estimates. A discrete version of the distribution in Section 3 was simulated. That is, 32 measurements were generated from a two mode distribution with $w_1 = 0.7$, $w_2 = 0.3$, $\tau_1 = 100\text{ms}$ and $\tau_2 = 20\text{ms}$. Independent Gaussian noise was added and the estimation was performed using Algorithm 1 and a gradient-based optimisation algorithm for comparison. The procedure was repeated for 500 independent noise realisations for each of the 12 SNR values to generate a plot of the root-mean-square error (RMSE). Fig. 2 illustrates the sharp degradation in the performance of the gradient-based optimisation algorithm for low SNR values. In this case, the problematic cost function makes finding the global minimum difficult. Conversely, the progressive correction implemented in the Bayesian algorithm overcomes this limitation and performs well at all SNR values.

5. SIMULATIONS

The algorithm developed above assumes a known number of modes in the distribution mixture. We relax this assumption by performing model selection using Akaike's information criteria [5]. This selects the number of modes that best describes the data, with an appropriate penalty for increasing model complexity. Technically, model selection should be performed using the maximum likelihood estimate (MLE) of each candidate model; instead we use the Bayesian algorithm above. Under certain non-restrictive conditions, the Bayesian estimate asymptotically converges to the MLE [6]. This justifies the use of the proposed algorithm, and the robustness it provides.

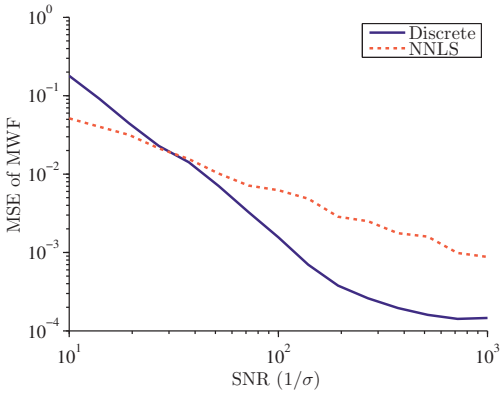

Fig. 3. The MSE of the myelin water fraction using different estimation algorithms.

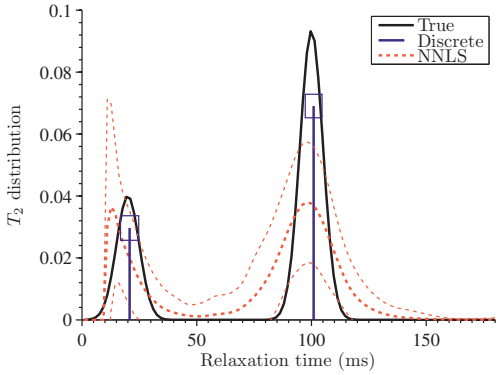

Fig. 4. Estimated T_2 distributions at an SNR of 100, under different model assumptions. The mean distribution and standard deviation from 200 trials are indicated by thick and thin lines, respectively. The discrete model weights are scaled by 0.1 for display.

Simulations were performed to analyse the effect of the discrete distribution assumption on myelin water fraction estimation. For comparison, regularised NNLS was performed with an optimal λ chosen according to the "least squares-based constraint" [7]. This iterative strategy selects λ such that the χ^2 misfit from the regularised solution is 1% larger than the χ^2 for the unregularised solution. The myelin water fraction was defined as the integral of the estimated distribution over the range $0-50\text{ms}$, divided by the total mass of the distribution. The true distribution was the inverse-gamma form in (6), with a slow component ($w_1 = 0.7$, $\nu_1 = 100\text{ms}$, $\rho_1 = 5\text{ms}$) and a fast component ($w_2 = 0.3$, $\nu_2 = 20\text{ms}$, $\rho_2 = 5\text{ms}$), typical of a voxel located in white matter.

Fig. 3 demonstrates that for a wide range of SNR values, the myelin water fraction estimated using a simple discrete model is superior to that obtained from an optimally regularised NNLS algorithm. At very low SNR, regularised NNLS is able to trade a large amount of variance for a small amount of bias, whereas the estimated discrete distribution may have its smallest mode larger than 50ms, resulting in a MWF estimate of zero and a large MSE.

Fig. 4 illustrates the estimation results for 200 trials at an SNR of 100 (the noise standard deviation is 1% of the signal amplitude). The figure highlights the inability of NNLS to accurately estimate the true spread, as predicted by our analysis. The regularisation introduces a bias into the shape of the estimated distributions. Particularly relevant is the location of the peak of the fast component, which is significantly underestimated. Conversely, the simple discrete model provides relatively unbiased estimates for the locations and weights of the distribution components. Indeed, at this SNR, for the purposes of myelin water fraction estimation, Fig. 3 indicates the estimates obtained from the discrete model are more accurate than those from the biased NNLS distribution.

6. DISCUSSION

This work considered a parametric alternative to the commonly used NNLS algorithm for the estimation of relaxation time distributions. The analysis demonstrated that the measurements have a weak dependence on the spread of the T_2 components and consequently estimating such spread is impractical. Although NNLS attempts to estimate a continuous distribution, the spread is more dependent on the regularisation parameter than the true distribution. Additionally, we have illustrated that this regularisation adds a bias to the estimator, which is unsuitable for quantitative analysis such as myelin water fraction estimation.

We adopt a simple model consisting of discrete relaxation times. The model has been previously proposed together with a gradient based MLE estimator, but this did not demonstrate robust performance at clinically achievable SNR values. The Bayesian algorithm overcomes the problems associated with the MLE cost function and the corresponding estimator's sensitivity to noise and initial estimate. This means the estimator is well poised for practical applications. The discrete model, although naïve to distribution spread, proved to provide accurate estimates of the location and weights of the distribution modes. The corresponding myelin water fraction estimates were more accurate for a broad range of SNR values, justifying further research into parametric models for distribution estimation.

7. REFERENCES

[1] T. Andrews, J. L. Lancaster, S. J. Dodd, C. Contreras-Sesvold, and P. T. Fox, "Testing the three-pool white matter model adapted for use with T2 relaxometry," *Magnetic Resonance in Medicine*, vol. 54, no. 2, pp. 449–454, 2005.

[2] J. L. Lancaster, T. Andrews, L. J. Hardies, S. Dodd, and P. T. Fox, "Three-pool model of white matter," *Journal of Magnetic Resonance Imaging*, vol. 17, no. 1, pp. 1–10, 2003.

[3] K. P. Whittall and A. L. MacKay, "Quantitative interpretation of NMR relaxation data," *Journal of Magnetic Resonance (1969)*, vol. 84, no. 1, pp. 134 – 152, 1989.

[4] N. Oudjane and C. Musso, "Progressive correction for regularized particle filters," in *Proceedings of the Third International Conference on Information Fusion*, 2002, vol. 2.

[5] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[6] M. T. Chao, "The asymptotic behavior of Bayes' estimators," *The Annals of Mathematical Statistics*, vol. 41, no. 2, pp. 601–608, 1970.

[7] S. J. Graham, P. L. Stanchev, and M. J. Bronskill, "Criteria for analysis of multicomponent tissue T2 relaxation data," *Magnetic Resonance in Medicine*, vol. 35, no. 3, pp. 370–378, 1996.