BAYESIAN FISHER INFORMATION CRITERION FOR SAMPLING OPTIMIZATION IN ASL-MRI

João Sanches[1,2], Inês Sousa[2] and Patrícia Figueiredo[2]

[1]Institute for Systems and Robotics
[2]Instituto Superior Técnico

ABSTRACT

Pulsed Arterial Spin Labeling (PASL) techniques potentially allow the absolute, non-invasive quantification of brain perfusion using Magnetic Resonance Imaging (MRI). This can be achieved by fitting a kinetic model to the data acquired at a number of inversion times (TI). Some model parameters such as the arterial transit time need to be estimated together with perfusion, while others are usually assumed to be known. The accuracy of the model estimation strongly depends on the distribution of the TI sampling points. Here, we propose a Bayesian framework for PASL perfusion estimation based on the Fisher information criterion, whereby the optimal sampling points can be determined taking into account the uncertainty of the model parameters as well as the amount of noise in the data. We show that the optimal sampling strategy for PASL depends on the a priori knowledge of the model parameters and this should therefore be taken into account.

***Index Terms*—** ASL, MRI, perfusion, Bayesian, Fisher.

1. INTRODUCTION

Perfusion describes the distribution of nutrients to the tissues by blood flow through the capillary bed and is defined as volume of blood per unit time and per unit volume of tissue. Arterial spin labeling (ASL) magnetic resonance imaging (MRI) techniques offer a non-invasive way of generating perfusion images that are potentially quantitative [1]. They consist on magnetically labeling the water molecules in the blood and then measuring the magnetization of the tissues after a certain time interval, the inversion time (TI). The magnetization difference ΔM as a function of TI in pulsed ASL (PASL) can be described by a standard kinetic model [2], illustrated in Fig.1 and defined in equation (1) where A is a constant and the vector $\Theta = \{f, \Delta t, \tau, r_{1b}, k\}$ contains the parameters perfusion, f, arterial transit time, Δt, bolus time width, τ, blood relaxation rate, r_{1b}, and a constant related with the tissue relaxation rate, k.

In principle, the magnetization collected at a single TI point is sufficient to obtain a perfusion estimate, provided that the values of the other model parameters are available or can be assumed. However, this is not always the case, particularly regarding the arterial transit time, which is often delayed in pathological conditions such as cerebrovascular disease. In these cases, it would be possible to estimate perfusion, as well as other unknown parameters, by fitting the PASL model to ΔM data collected at multiple TI points [3]. However, the intrinsically low signal to noise ratio (SNR) of ASL measurements usually requires substantial signal averaging, which could result in undesirably long scanning times. On the other hand, the accuracy of the estimated parameters strongly depends on the distribution of the TI sampling points. A judicious choice of the sampling points is therefore crucial in order to minimize scanning time, while optimizing estimation accuracy. Optimal sampling strategies have previously been designed based on the Fisher information matrix optimality criterion for the simultaneous estimation of perfusion, f, and the arterial transit time, Δt [4]. However, the uncertainly associated with the remaining model parameters was not take into account. Here, we propose a Bayesian framework for PASL perfusion estimation based on the *maximum a posteriori* (MAP) criterion, whereby the optimal sampling points can be determined taking into account the uncertainty of the model parameters as well as the amount of noise in the data.

2. PROBLEM FORMULATION

Let us consider the unknown function, $F(t, \Theta)$, where Θ is a vector of unknown parameters to be estimated from a set of N observations $\mathbf{y} = \{y_i\}$ taken at the N instants, $\mathbf{t} = \{t_i\}$. The goal is to choose the N optimal time points that maximize the accuracy of the Θ estimate by reducing the variance of the estimator.

Let us consider the following observation model

$$y_i = F(t_i, \Theta) + \eta \qquad (2)$$

where an *Additive White Gaussian Noise* (AWGN) model is adopted which means $p(\eta) \sim \mathcal{N}(0, \sigma_y^2)$. Therefore, $p(y_i|t_i, \Theta) = \mathcal{N}(F(t_i, \Theta), \sigma_y^2)$ where σ_y is the standard

$$\Delta M(t, \Theta) = \begin{cases} 0 & \text{if } t < \Delta t \\ \frac{Af}{k} e^{-r_{1b}t} \left(e^{k(t-\Delta t)} - 1 \right) & \text{if } \Delta t \leq t < \Delta t + \tau \\ \frac{Af}{k} e^{-r_{1b}t} \left(e^{k(t-\Delta t)} - e^{k(t-\Delta t - \tau)} \right) & \text{if } t \geq \Delta t + \tau \end{cases} \tag{1}$$
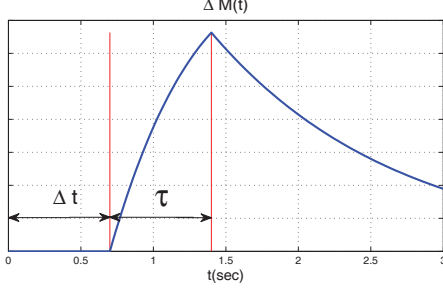
Fig. 1. PASL signal ΔM as a function of the inversion time TI, according to a standard kinetic model.

deviation of the noise. The estimation of the vector Θ by using the *maximum a posteriori* (MAP) [5] criterion may be formulated as the following optimization task:

$$\hat{\Theta} = \arg \min_{\Theta} E(\mathbf{y}, \mathbf{t}, \Theta) \tag{3}$$

where the energy function to be minimized is

$$E(\mathbf{y}, \mathbf{t}, \theta) = -\log \left[p(\mathbf{y}|\mathbf{t}, \Theta) p(\Theta) \right] \tag{4}$$

The distribution function $p(\mathbf{y}|\mathbf{t}, \Theta)$ models the acquisition process and $p(\Theta)$ incorporates the a priori knowledge about the parameters to be estimated. Assuming statistical independence of the observations $p(\mathbf{y}|\mathbf{t}, \Theta) = \prod_{i=1}^{N} p(y_i|t_i, \Theta)$. Here, the P elements of the vector $\Theta = \{\theta_1, \theta_2, ..., \theta_P\}$ are assumed independent and Gaussian distributed with distribution $\theta_i \sim N(\theta_{0i}, \sigma_i^2)$. The values $\{\theta_{0i}, \sigma_i\}$ reflect the a priori knowledge about the model parameters to be estimated obtained, e.g., from experimental data. In particular, the standard deviation σ_i accounts for the uncertainty about θ_i. Therefore, the distribution of Θ is a multivariate Gaussian distribution, $p(\Theta) = \mathcal{N}(\Theta_0, \mathbf{C})$, where $\mathbf{C} = diag(\{\sigma_1^2, \sigma_2^2, ... \sigma_P^2\})$ is a diagonal covariance matrix.

The energy function (4) may be written as follows:

$$E(\mathbf{y}, \mathbf{t}, \Theta) = \underbrace{\frac{1}{2\sigma_y^2} \sum_{i=1}^{N} (F(t_i, \Theta) - y_i)^2}_{\text{Data fidelity term}}$$
$$+ \underbrace{\frac{1}{2} \sum_{i=1}^{P} \frac{(\theta_i - \theta_{0i})^2}{\sigma_i^2}}_{\text{Prior term}}. \tag{5}$$

The estimation of the model parameters Θ may be obtained by computing the stationary point of $E(\mathbf{y}, \mathbf{t}, \Theta)$ with respect to Θ,

$$\nabla_\Theta E(\mathbf{y}, \mathbf{t}, \Theta) = 0, \tag{6}$$

which is equivalent to the following set of equations:

$$\frac{\partial E(\mathbf{y}, \mathbf{t}, \Theta)}{\partial \theta_k} = \sum_{i=1}^{N} \left[(F(t_i, \Theta) - y_i) \frac{\partial F(t_i, \Theta)}{\partial \theta_k} \right]$$
$$+ \frac{\sigma_y^2}{\sigma_k^2} (\theta_k - \theta_{0k}) = 0 \tag{7}$$

where $1 \leq k \leq P$.

3. OPTIMAL SAMPLING STRATEGY

The N optimal sampling time points $\mathbf{t} = \{t_i\}$ to estimate Θ depend on the function $F(t, \Theta)$ and on the distribution of the parameters, $p(\Theta)$. They are optimal, according the minimum variance of the estimator for a given vector of parameters Θ, if the determinant of the *Fisher Information* matrix [6], is maximum:

$$\mathbf{t}^*(\Theta) = \arg \max_{\mathbf{t}} \mathcal{J}(\mathbf{t}, \Theta) \tag{8}$$

where $\mathcal{J}(\mathbf{t}, \Theta) = |H_{kr}(\Theta)|$.

The elements of the *Fisher Information* matrix are defined as follows:

$$H_{kr} = \mathcal{E}_{\mathbf{Y}} \left[\frac{\partial^2 \log p(Y, \Theta)}{\partial \theta_k \partial \theta_r} \right], \tag{9}$$

where $\mathcal{E}_{\mathbf{Y}}()$ is the expectation with respect to the multivariate random variable $\mathbf{Y}$. These elements, according with (5), are:

$$H_{kr}(\Theta) = \frac{1}{\sigma_y^2} \sum_{i=1}^{N} \left[\frac{\partial F(t_i, \Theta)}{\partial \theta_k} \frac{\partial F(t_i, \Theta)}{\partial \theta_r} \right] + \frac{1}{\sigma_k^2} \delta_{k,r}. \tag{10}$$

The analytical solution of (8) is usually difficult mainly due to the complexity and non-continuity of the derivatives of $F(t, \Theta)$. Here, the set of optimum time points is determined on an incremental basis, whereby each time point is computed at a time, as a function of the previously computed time points, and added to these.

Let us consider the cost function $\mathcal{J}_n(t, \mathbf{t}_{n-1}, \Theta)$ where $\mathbf{t}_{n-1} = \{t_1, t_2, ..., t_{n-1}\}$ are the first $n-1$ optimum time points estimated up to the $(n-1)^{th}$ iteration. The nth optimum time point, t_n, is obtained by solving the following optimization problem:

$$t_n(\Theta) = \arg \max_{t} \mathcal{J}_n(t, \mathbf{t}_{n-1}, \Theta) \tag{11}$$

where $\mathcal{J}_n(t, \mathbf{t}_{n-1}, \Theta)$ is the determinant of the matrix with elements

$$H_{k,r}(t, \mathbf{t}_{n-1}, \Theta) = H_{k,r}(\mathbf{t}_{n-1}, \mathbf{t}_{n-2}, \Theta)$$
$$+ \frac{1}{\sigma_y^2} \frac{\partial F(t, \Theta)}{\partial \theta_k} \frac{\partial F(t, \Theta)}{\partial \theta_r} \quad (12)$$

where

$$H_{k,r}(\mathbf{t}_{n-1}, \mathbf{t}_{n-2}, \Theta) = \frac{1}{\sigma_y^2} \sum_{i=1}^{n-1} \left[\frac{\partial F(t_i, \Theta)}{\partial \theta_k} \frac{\partial F(t_i, \Theta)}{\partial \theta_r} \right]$$
$$+ \frac{1}{\sigma_k^2} \delta_{k,r} \quad (13)$$

was incrementally estimated in the previous $n - 1$ steps.

Equations (11), (12) and (13) describe an incremental procedure to compute the N optimum time points where in each iteration a 1D cost function is maximized.

The set of points obtained from (8), or equivalently from the incremental approach (11), depends on the value of the parameter vector Θ which is not known but for which there is a prior knowledge incorporated in $p(\Theta)$.

Here, the following strategy is proposed for the identification of the set of optimal sampling time points for ASL model estimation:

1. Sample the prior distribution $p(\Theta)$ to obtain a vector of parameters Θ_i.

2. Use equation (8) to obtain a collection of M optimal time points, $\mathbf{T}_i$, for the parameter value Θ_i.

3. Add the resulting time point distribution, $\mathbf{T}_i$, to a running histogram of optimal time points, $h(t, \mathbf{T})$.

4. Repeat steps 1) to 3) to cover the full distribution $p(\Theta)$.

5. Compute the cumulative curve from the final histogram to extract the required N optimal points by partitioning the area under the final histogram in N interval with the same area .

In general terms, the cumulative curves, e.g. Fig.2 and Fig.3, give the optimal density of the N desired sampling points that should be used to maximize $\mathcal{J}(\mathbf{t}, \Theta)$. In fact, as shown in these figures, if N samples are decided to be acquired, its optimal distribution is not uniform, as expected, but following a different distribution.

4. EXPERIMENTAL RESULTS

Sets of optimal TI sampling points were obtained for the estimation of parameters f and Δt, using different levels of data noise, σ_y, as well as different levels of the parameter uncertainty, σ_f and $\sigma_{\Delta t}$. Physiologically plausible parameter distributions were considered, according to values in the literature [4]:
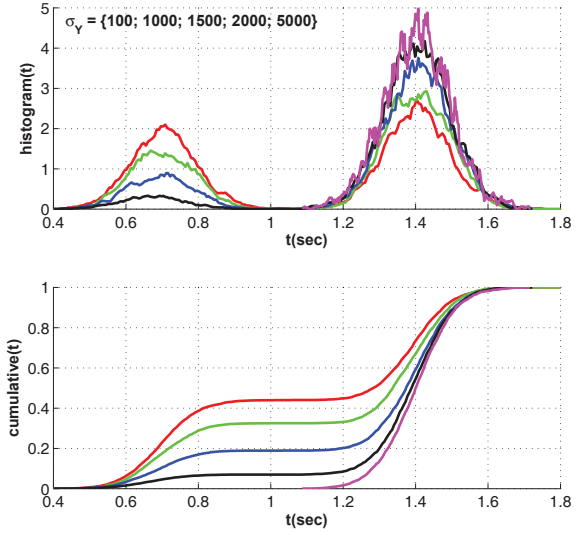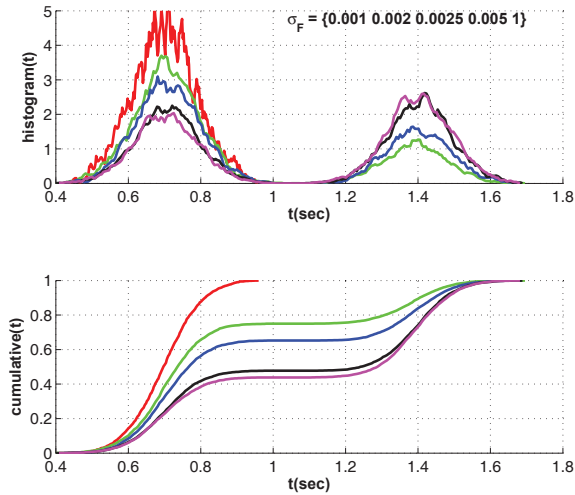

Fig. 2. Histograms and cumulative curves obtained by optimization using different values for the amount of noise corrupting the data $\sigma_Y = \{100, 1000, 1500, 2000, 5000\}$ (red, green, blue, black, magenta).

- $p(f) = \mathcal{N}(0.012, 0.05^2) s^{-1}$

- $p(\Delta t) = \mathcal{N}(0.7, 0.3^2) s$

- $p(\tau) = \mathcal{N}(0.7, 0.1^2) s$

- $p(k) = \mathcal{N}(-0.16, 0.01^2)$

- $p(r1_b) = \mathcal{N}(0.63, 0.05^2) s^{-1}$

In order to investigate the effects of different amounts of noise corrupting the data on the optimal distribution of the sampling time points, the following noise levels were considered, $\sigma_Y = \{100, 1000, 1500, 2000, 5000\}$, while keeping the parameter uncertainty levels constant, at $\sigma_f = 0.05 s^{-1}$ and $\sigma_{\Delta t} = 0.3 s$. The resulting histograms and corresponding cumulative curves are shown in Fig. 2. It can be observed that the optimal sampling points are, in general, distributed around the values $\Delta t_0 = 0.7 s$ and $\Delta t_0 + \tau_0 = 1.4 s$. For moderate noise levels, there is only a small advantage of sampling around $\Delta t_0 + \tau_0 = 1.4 s$. However, as noise levels increase, the advantage of sampling around $\Delta t_0 + \tau_0 = 1.4 s$ becomes greater. This behaviour can be understood in terms of the fact that the curve assumes its greatest value at this time point, which therefore becomes the sampling point of choice when noise levels increase.

In order to investigate the effects of different amounts of uncertainty on the prior knowledge of the model parameter f on the optimal distribution of the sampling time points, the following uncertainty levels were considered $\sigma_f = \{0.001, 0.002, 0.0025, 0.005, 1\} s^{-1}$, while keeping
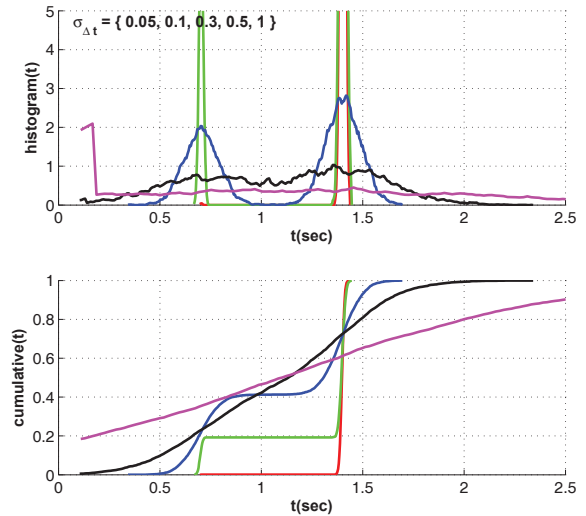
Fig. 3. Histograms and cumulative curves obtained by optimization using parameter uncertainties $\sigma_f = \{0.001, 0.002, 0.0025, 0.005, 1\}s^{-1}$ (red, green, blue, black, magenta).

Fig. 4. Histograms and cumulative curves obtained by optimization using parameter uncertainties $\sigma_{\Delta t} = \{0.05, 0.1, 0.3, 0.5, 1\}s$ (red, green, blue, black, magenta).

Δt uncertainty levels constant, at $\sigma_{\Delta t} = 0.3s$, and the noise level constant, at $\sigma_Y = 500$. The resulting histograms and corresponding cumulative curves are shown in Fig. 3. It can be observed that, as the uncertainty on the value of f decreases, the distribution of the optimal sampling points moves towards the point $\Delta t_0 = 0.7s$ and away from the point $\Delta t_0 + \tau_0 = 1.4s$. This behaviour can be understood in terms of the fact that better knowledge of the parameter f will concentrate estimation efforts on the other unknown parameter, Δt.

In order to investigate the effects of different amounts of uncertainty on the prior knowledge of the model parameter Δt on the optimal distribution of the sampling time points, the following uncertainty levels were considered, $\sigma_{\Delta t} = \{0.05, 0.1, 0.3, 0.5, 1\}s$, while keeping f uncertainty levels constant, at $\sigma_f = 0.05s^{-1}$, and the noise level constant, at $\sigma_Y = 500$. The resulting histograms and corresponding cumulative curves are shown in Fig. 4. In this case, similarly to what was observed as a function of the uncertainty of f, better knowledge of Δt moves the optimal sampling points away from $\Delta t_0 = 0.7s$ and towards $\Delta t_0 + \tau_0 = 1.4s$.

5. CONCLUSIONS

A fast and computationally efficient method was implemented to obtain the optimal set of TI sampling points for the estimation of ASL-MRI perfusion model parameters. This proposed method is developed within a Bayesian framework, such that the choice of the optimal sampling points is based on the

Fisher information matrix optimality criterion, but further incorporating prior knowledge about the uncertainty of the model parameters to be estimated. We show that the optimal distributions of the TI sampling points strongly depend on the uncertainty of the model parameters. Our results therefore suggest that a Bayesian approach should be considered when optimizing a multiple-TI ASL perfusion experiment, so that the a priori knowledge of model parameters may be taken into account. Future work will consist on validating the optimal distributions obtained here through Monte Carlo experiments on artificial data, as well as through empirical evidence obtained from real ASL data.

6. REFERENCES

[1] E T Petersen, I Zimine, Y-C L Ho, and X Golay, "Non-invasive measurement of perfusion: a critical review of arterial spin labelling techniques," *Br J Radiol*, vol. 79, no. 944, pp. 688–701, 2006.

[2] Richard B. Buxton, Lawrence R. Frank, Eric C. Wong, Bettina Siewert, Steven Warach, and Robert R. Edelman, "A general kinetic model for quantitative perfusion imaging with arterial spin labeling," *Magnetic Resonance in Medicine*, vol. 40, no. 3, pp. 383–396, 1998.

[3] P. M. Figueiredo, S. Clare, and P. Jezzard, "Quantitative perfusion measurements using pulsed arterial spin labeling: effects of large region-of-interest analysis.," *Journal of Magnetic Resonance Imaging*, vol. 21, no. 6, pp. 676–682, June 2005.

[4] J. Xie, D. Gallichan, R. N. Gunn, and P. Jezzard, "Optimal design of pulsed arterial spin labeling mri experiments.," *Magnetic Resonance in Medicine*, vol. 59, no. 4, pp. 826–834, April 2008.

[5] Todd K. Moon and Wynn C. Stirling, *Mathematical methods and algorithms for signal processing*, Prentice-Hall, 2000.

[6] Steven M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*, Prentice Hall PTR, January 1998.