# Bayesian NMR Relaxometry

Paul D. Teal

Tuesday, 29 August 2017 —
Friday, 1 September 2017

**Abstract**

This document describes a possible technique for estimating quantities such the fraction of bound fluids from NMR relaxometry measurements.

## 1 The Problem

The data in a NMR relaxometry (and/or diffusion) experiment are generally modelled using an integral equation such as the following (using the case of a 2D $T_1$–$T_2$ measurement as an example):

$$m(\tau_1, \tau_2) = \iint K(\tau_1, \tau_2, T_1, T_2) f(T_1, T_2)\, dT_1\, dT_2 + \eta_{\tau_1,\tau_2} \tag{1}$$

where $M()$ is the measured data, $K()$ is a kernel representing the relaxation processes, and $f()$ is the density of spins in each relaxation environment.

The data are available for discrete values of $\tau_1$, $\tau_2$, $T_1$ and $T_2$, and so this can be represented by the matrix equation

$$m = Kf + \eta \tag{2}$$

There are many situations in which a quantity of interest is given by an integral over the density $f()$:

$$I_A = \iint f(T_1, T_2) g_A(T_1, T_2)\, dT_1\, dT_2 \tag{3}$$

or, as a matrix equation

$$I_A = g^T f \tag{4}$$

$I_A$ is typically estimated by first estimating $f$, and then by performing the integral. The condition of the kernel $K$ is such that the estimation of $f$ is ill-posed, and regularisation must be applied, and one of the difficulties of this approach is choosing the type and level of regularisation.

An approach in which $I_A$ is estimated directly from the data, without having to first estimate $f$ could be advantageous. It is also desirable to provide some indication of the uncertainty in $I_A$. Knowing this uncertainty analytically could lead to some insight into the design of superior pulse sequences.

## 2 Bayesian Estimation

The essence of Bayesian estimation is to treat the parameters of interest as random variables. The problem is to find their distributions, conditioned on the available data.

In this case, we wish to find the density of $I_A$ conditioned on the measurements $m$. The first step is to marginalise this desired distribution by integration over the spin densities $f$:

$$p(I_A|m) = \int p(I_A|m, f) p(f|m)\, df \tag{5}$$

Once $p(I_A|m,f)$ has been conditioned on the density, it is no longer dependent on the measurements, so we can write

$$p(I_A|m,f) = p(I_A|f) = \delta(I_A - g^T f) \tag{6}$$

where $\delta()$ is the Dirac delta function.

We can also use Bayes theorem to find $p(f|m)$:

$$p(f|m) = \frac{p(m|f)p(f)}{p(m)} = \frac{p(m|f)p(f)}{\int p(m|f)p(f)\,df} \tag{7}$$

where $p(f)$ is a prior distribution on the spin density.

Combining the above results, we obtain

$$p(I_A|m) \propto \int \delta(I_A - g^T f)p(m|f)p(f)\,df \tag{8}$$

# 3 Normal Distributions

Many manipulations involving marginalization and Bayes theorem on normal distributions preserve this normality. We will make use of the following results:

**Marginal:**

If

$$x \sim \mathcal{N}(\mu_x, C_x) \tag{9}$$
$$y|x \sim \mathcal{N}(Mx + \mu_y, C_y) \tag{10}$$

then

$$p(y) = \int p(y|x)p(x)\,dx = \mathcal{N}(M\mu_x + \mu_y, MC_xM^T + C_y) \tag{11}$$

**Bayes Theorem:**

If

$$x \sim \mathcal{N}(\mu_x, C_x) \tag{12}$$
$$y|x \sim \mathcal{N}(Mx + \mu_y, C_y) \tag{13}$$

then

$$x|y \sim (R(y - M\mu_x - \mu_y) + \mu_x, C_x - RMC_x^T) \tag{14}$$

where

$$R = C_xM^T(MC_xM^T + C_y)^{-1} \tag{15}$$

To apply these results, we will use as prior distribution for $f$, a normal distribution, with large covariance:

$$f \sim \mathcal{N}(\mu_f, C_f) \tag{16}$$

**Note:** This approach has two important disadvantages:

- First, the results still depend on this covariance.

- Secondly, the prior is quite unrealistic, since it is known (in many situations) that the density cannot be negative.

However, to use a more realistic prior could quickly become analytically intractable.

Using $m = Kf + \eta$, we deduce that

$$m|f \sim \mathcal{N}(Kf, C_\eta) \tag{17}$$

Note that both $C_\eta$ and $C_f$ in most situations are appropriately chosen to be diagonal matrices, although the diagonal entries may not always be identical.

We first apply Bayes theoreom to obtain the density of $f|m$:

$$f|m \sim \mathcal{N}(\ R(m - K\mu_f) - \mu_f\ ,\ C_f - RKC_f^T\ ) \tag{18}$$

where $R = C_f K^T (KC_f K^T + C_\eta)^{-1}$

We then marginalise over $f$. We can apply the marginal result for normal densities by considering the Dirac delta as equivalent to a normal density with zero variance.

We thus obtain

$$I_A|m \sim \mathcal{N}(\quad g^T(R(m - K\mu_f) - \mu_f) \quad ,\quad g^T(C_f - RKC_f^T)g \quad ) \tag{19}$$

Note that

$$\begin{aligned} C_f - RKC_f^T =& C_f - C_f K^T (KC_f K^T + C_\eta)^{-1} KC_f^T \\ =& (C_f^{-1} + K^T C_\eta^{-1} K)^{-1} \end{aligned} \tag{20}$$

This form is not only simpler in form, but also numerically advantageous if the number of points in the distribution $f$ is fewer than the number of measurements, which (at least for measurements involving $T_2$) is often the case.

Note that the rapid decay of terms in the kernel $K$ means that in general $K^T C_\eta^{-1} K$ will not be invertible, so it is not possible to set the covariance $C_f$ to be infinite. Thus the results will depend on the choice of the covariance of the prior. This is exactly analogous to the choice of regulariser in finding the spin density $f$.

# 4 An analogy

If we write the density of the measurements $m$, given the $T_2$ distribution $f$, as

$$p(m|f) \propto \exp\left(-\frac{1}{2}(m - Kf)^T C_\eta^{-1}(m - Kf)\right) \tag{21}$$

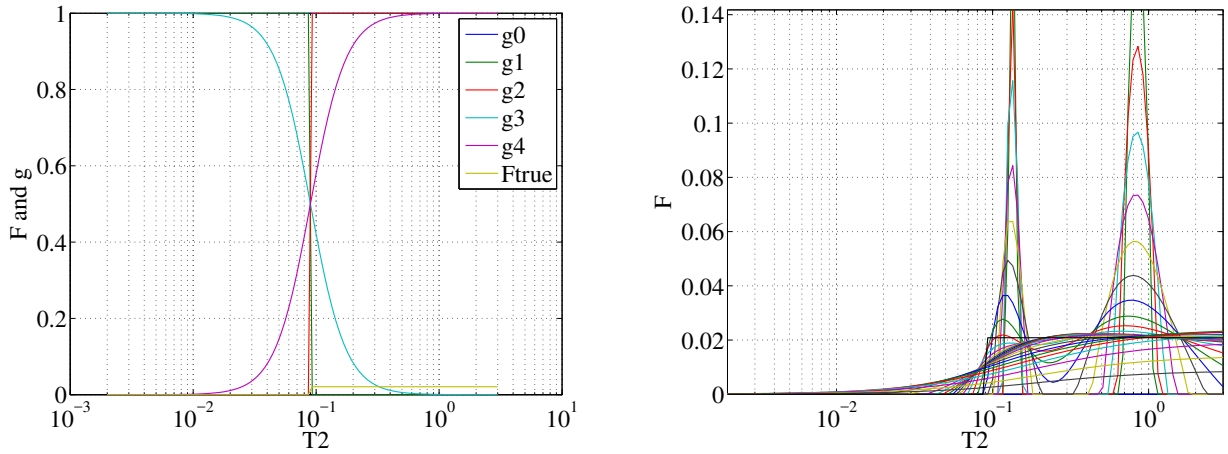
and the prior $p(f)$ as

$$p(f) \propto \exp\left(-\frac{1}{2}(f - \mu_f)^T C_f^{-1}(f - \mu_f)\right) \tag{22}$$

If we assume that $p(f)$ has zero mean, and that the components of $m$ and $f$ are independent and identically distributed, and that the prior distribution $p(f)$ has zero mean then

$$p(m|f)p(f) \propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_\eta^2}\|m - Kf\|_2^2 + \frac{1}{\sigma_f^2}\|f\|_2^2\right)\right) \tag{23}$$

$$\propto \exp\left(-\frac{1}{2\sigma_\eta^2}\left(\|m - Kf\|_2^2 + \frac{\sigma_\eta^2}{\sigma_f^2}\|f\|_2^2\right)\right) \tag{24}$$




(a) (b)

Figure 1: (a) Five different integration windows, and the true $f$ distribution. (b) An example recovered distribution, for various values of the regularisation parameter.

The expression in the innermost parentheses is the expression minimised by the regularised least squares criterion, if

$$\sigma_f^2 = \frac{\sigma_\eta^2}{\alpha} \tag{25}$$

Accordingly, in the numerical study of Section 5, the prior variance of $p(f)$ is presented as $\alpha$, the regularisation to which $\sigma_f^2$ corresponds.

# 5 A numerical study

A numerical example was studied to investigate the properties of this Bayesian approach. A simple $T_2$ distribution was generated, with the measurement objective being to determine the proportions of spins above and below a cut off of $T_{2c} = 90\,\text{ms}$. The true distribution was a step function, with zero values for small $T_2$, and the same positive value for large $T_2$, with the step at $T_{2c} = 90\,\text{ms}$. The large $T_2$ values were normalised so that the total porosity was equal to 1.

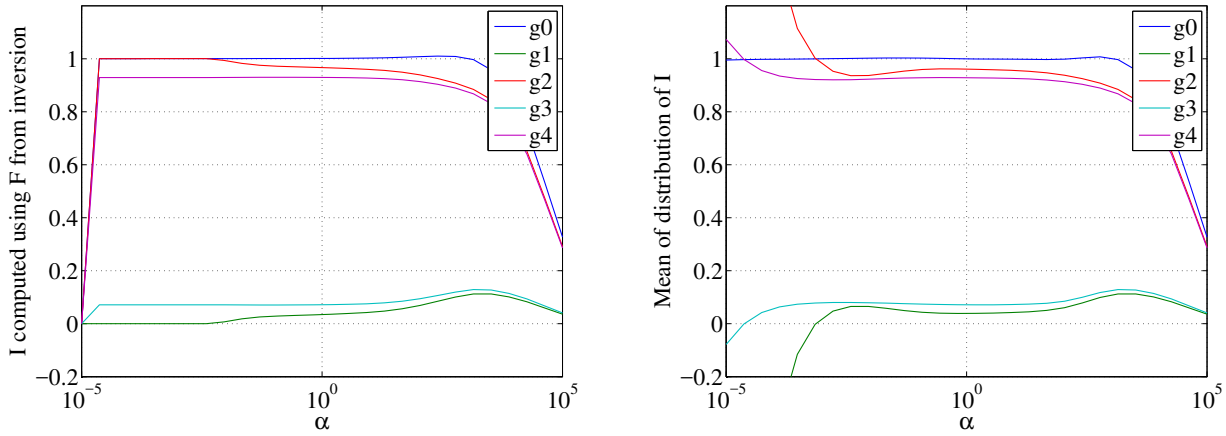
The signal to noise ratio corresponded to a porosity to noise standard deviation of 50. Qualitatively similar results (not shown) were obtained for a porosity to noise level of 4.6, and with the true distribution being a narrow peak of magnitude 1 at $0.25T_{2c}$ and a narrow peak of magnitude 2 at $4T_{2c}$.

Five different "integration" windows were used: $g_0$ which sums the entire $T_2$ distribution, $g_1$ and $g_2$ which sum the contributions above and below $T_{2c}$ respectively, while $g_3$ and $g_4$ are similar, but have a gradual transition, generated using a tanh() function. The 5 windows and the true $f$ distribution are shown in Figure 1, along with some example distributions recovered using a "classical" method (the FLINT algorithm was used for this estimation).

The bound and free volumes obtained via the $f$ distribution and via the Bayesian approach are shown in Figure 2.

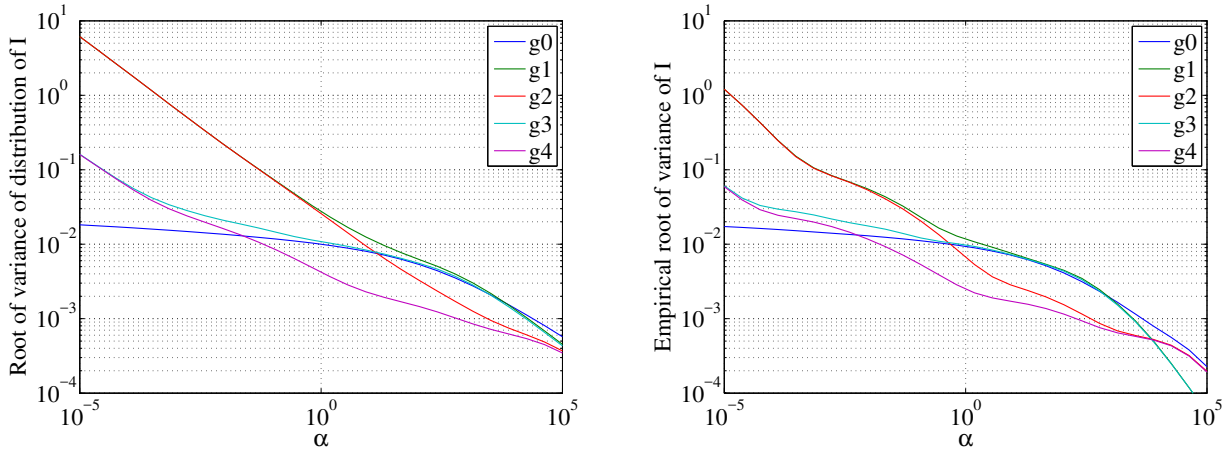
The variance was calculated using the Bayesian approach. Also 1000 trials were run, and the empirical variance of the results was obtained. These two results are compared in Figure 3.

The mean squared error for the 1000 runs using the classical and Bayesian approaches are compared in Figure 4. A striking feature of the comparison is that both methods achieve very similar mean squared error levels, although the regularisation parameter or prior variance required to achieve that level is different for the various windows. The classical approach appears to be somewhat more robust in that the range of regularisation parameters for which acceptable values are found is somewhat more broad.




(a) (b)

Figure 2: Example estimated volumes using the classical approach for various regularisation factors (a), and using the Bayesian approach, for various prior variances. (b)




(a) (b)

Figure 3: Comparison of the computed variance (a) and empirical variance (b) of the estimates as a function of the prior variance. The two sets of curves are similar, but they do differ because the computed result assumes that all values of the prior are Gaussian distributed, having zero mean, whereas for the empirical result, the actual distribution is fixed as one having just two two fixed non-zero values.

# 6 Adaptivity

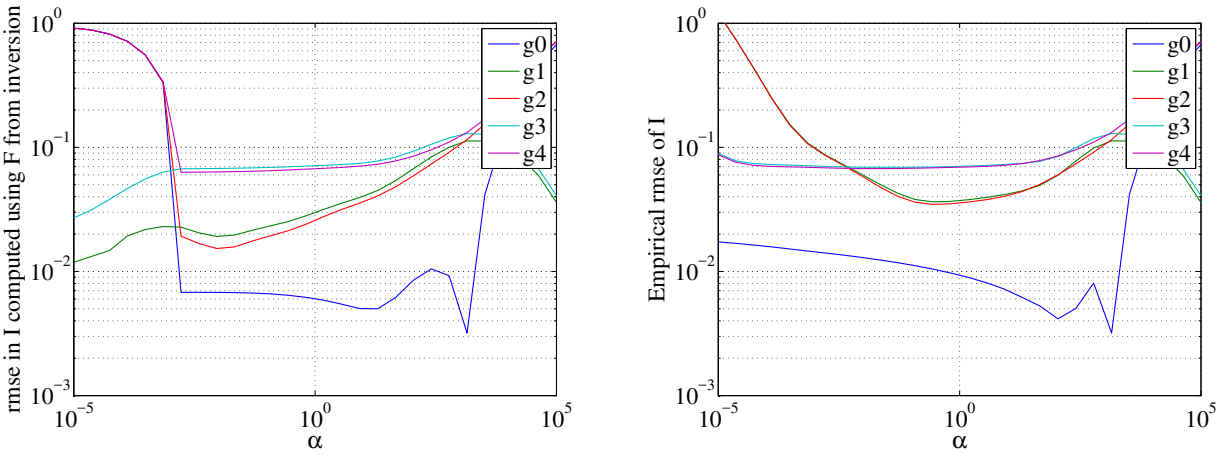
The Bayesian framework is appropriate for the design of an adaptive measurement system. Suppose that we already have measurements $m$ at some set of delays. The measurements can be modelled as $m = Kf + \eta$ where $f$ is the underlying spin distribution. These measurements are then to be augmented to $m' = K'f + \eta$ by the addition of another row $k^T$ to kernel matrix $K'$.

If we assume that the noise of the new measurement is uncorrelated with that of the previous measurements, then it is easy to see that the estimated variance of $I_A$ is given by

$$\sigma_y = g^T \left( C_f^{-1} + \begin{bmatrix} K^T & k \end{bmatrix} \begin{bmatrix} C_\eta & \mathbf{0} \\ \mathbf{0}^T & \sigma_\eta^2 \end{bmatrix}^{-1} \begin{bmatrix} K \\ k^T \end{bmatrix} \right) g \tag{26}$$

$$= g^T \left( C_f^{-1} + K^T C_\eta^{-1} K + \frac{1}{\sigma_\eta^2} k k^T \right) g \tag{27}$$

It is a relatively simple calculation to determine from a set of candidate vectors $k$, which one of them most reduces the variance of the resulting estimate.




(a) (b)

Figure 4: Comparison of the root mean squared error using the classical (a) and Bayesian (b) approaches.

# A Multidimensional distributions

This appendix shows how, in the case of a separable kernel (such as $K = K_2 \otimes K_1$) it is possible to compute the variance

$$\sigma_I^2 = g^T (C_f^{-1} + K^T C_\eta^{-1} K)^{-1} g \tag{28}$$

without having to compute the large Kronecker product $K$, which may be prohibitively large, so that the effect on this variance of the prior (and of other parameters of interest) can be investigated.

If the the singular value decomposition (SVD) of the kernel matrices are $K_{1N_1 \times N_x} = U_1 \Sigma_1 V_1^T$ and $K_{2N_2 \times N_y} = U_2 \Sigma_2 V_2^T$, and $C_\eta = \sigma_\eta^2 I_{N_1 N_2}$ and $C_f = \sigma_f^2 I_{N_x N_y}$, and $g$ is arranged into matrix $G$ of the same size $(N_1 \times N_2)$ as the distribution $F$, then we can define

$$D = \frac{1}{\frac{(S_1 \odot S_1)^T (S_2 \odot S_2)}{\sigma_\eta^2} + \frac{1}{\sigma_f^2}} \tag{29}$$

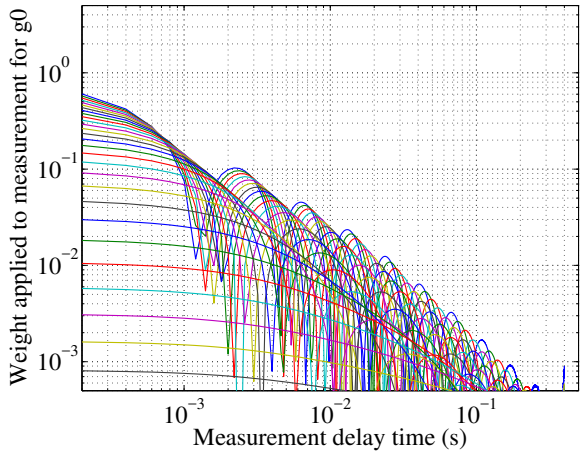
where all the divisions and additions are performed element-wise. Note that this method will only work if the two covariance matrices $C_f$ and $C_\eta$ are both scaled identity matrices.

We further define

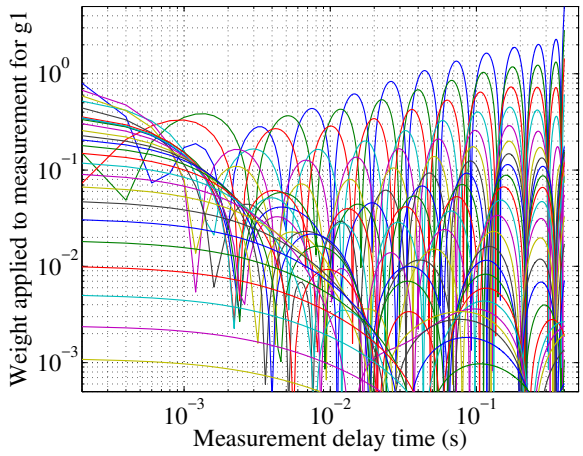
$$W = V_1^T G V_2 \tag{30}$$

and then $\sigma_I^2$ is given by

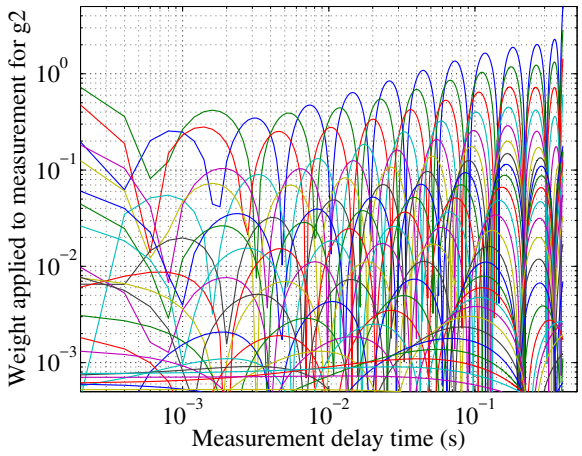
$$\sigma_I^2 = \sum_{\text{rows}} \sum_{\text{cols}} W \odot W \odot D \tag{31}$$



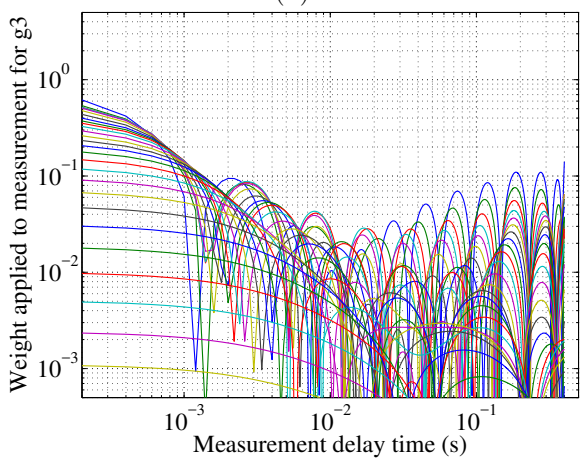

(a)



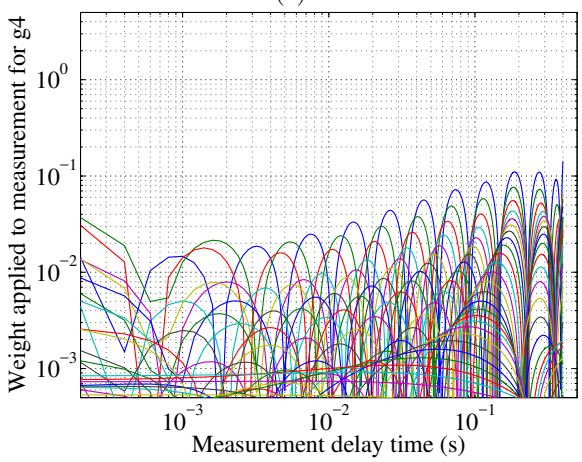

(b)




(c)




(d)




(e)

Figure 5: The weights applied to the measurements for the various windows. (a) $g_0$ (b) $g_1$ (c) $g_2$ (d) $g_3$ (e) $g_4$.