Fully Bayesian analysis of multi-exponential decays in low resolution NMR

Proposal for ITT / SAMBa project on applications of "Massive Inference"

Edmund Fordham

Schlumberger Gould Research Cambridge



SGR

N 52°12'40.32"

E 0°4'53.76"

330

270 - W

30

60

Part 1: Nuclear Magnetic Resonance in Porous Media, and what we do with it



NMR can be done at many field strengths

9 T





bore imager

2 T



23 - 12 MHz – rock

0.3 T



permanent magnets

2 MHz – "industry standard" for NMR in petrophysics

> \leq 2 MHz – Well logging

petrophysics & well logging

50 mT



SGR N 52"17 47



50 μT

~ 2 <u>k</u>Hz - Earth's Field NMR

400 MHz – a chemical spectroscopy system

MRI



core imager

Resonance Frequency ------





2D relaxation time distributions - in the same EOR experiment





SGR N S2"17 47

T_1 - T_2 Interpretation in shales



... after Marc Fleury (IFP)

- Measure T_1 AND T_2 relaxation times
- Sensitive to molecular motion
- Frequency (magnet strength) dependent



Schlumberger Resparch

SGR N5217240 E01456

actual examples from organic shales





SGR N 52"17 47

7

2D Diffusion – T_2 distributions



SGR N 52"17 47

Schlumberger ReSparch

US Patent Application 12/731,005

Mitchell and Fordham, J. Magn. Reson (2011) doi: 10.1016/j.jmr.2011.07.020

That EOR experiment again: $D - T_2$ this time





SGR N 52117 40 E 0745 54

More dimensions ...



Sections through an ingenious 4D lab experiment ...



SGR N 52"17 47 FORM

What's wrong with it all?

Poor resolution

- typically only integrals or moments are used
- Rock properties and fluid types the goal
- No error bars !!!! -
 - or any other uncertainty estimates
 - Point values on T₂ almost meaningless
- No easy way to include prior information e.g.
 - 1. Bulk relaxation rates impose ceiling on $T_{1,2} \le T_{1B,2B}$
 - 2. $T_1 \ge T_2$ from physics
- Time/memory prohibitive in dimensions > 2 ... ?



SGR 152117-40

Part 2: The mathematical problem



SGR N 52"17 47

Problem archetype and approaches

The archetype problem is the solution, for an unknown distribution $P(T_2, T_1) \ge 0$, of a

First Kind Fredholm integral equation with exponential kernel :

$$M\left(\tau_{1}^{(i)},\tau_{2}^{(j)}\right) = \int_{0}^{\infty} P(T_{2},T_{1})\exp\left(-\tau_{2}^{(j)}/T_{2}\right) \left[1 - \exp\left(-\tau_{1}^{(i)}/T_{1}\right)\right] dT_{2}dT_{1} + e\left(\tau_{1}^{(i)},\tau_{2}^{(j)}\right)$$

 $M(\tau_1^{(i)}, \tau_2^{(j)})$ experimental magnetization values at discrete sampling times $\tau_1^{(i)}, \tau_2^{(j)}$ for i = 1, ..., N; j = 1, ..., M $e(\tau_1^{(i)}, \tau_2^{(j)})$ experimental noise (Gaussian, white). Signal to Noise Ratio may be poor (0 dB in first experimental prototypes)

Different measurements involve different kernels – may or may not be separable $K(\tau_1, \tau_2; T_1, T_2) = k_1(\tau_1; T_1)k_2(\tau_2; T_2)$ Positivity condition $P(T_2, T_1) \ge 0$ may need to be relaxed (in some 2D contexts).

- Current routine practice: one of many (and evolving) *ad hoc* variants of Tikhonov regularisation [3,4,5,6,7]
- Attempts at uncertainty estimates:
- (i) confidence intervals on the overall data misfit (with explicit *rejection* of Bayesian priors) [8]
- (ii) direct estimation of petrophysically important moments, via the use of Mellin transforms [9, 10]
- (iii) Markov Chain Monte-Carlo (MC²) exploration of posterior probability spaces, but without close attention to the priors [11]
- Not attempted (so far !): Systematic treatment as a problem in Bayesian inference.



Solution by Tikhonov Regularization

Measured magnetization data M_i & unknown distribution f(x) $M_i = \int_{x_a}^{x_b} f(x)k_i(x) dx + n_i$ i = 1, 2, ..., N

Kernel functions exponential decays (or similar variants)

$$k_i = \exp\left(-\tau_i/T_2\right)$$

 $k_i(x) \dots k_N(x)$ are "almost linearly dependent" such that matrix $G_{ij} = \int_{x_a}^{x_b} k_i(x)k_j(x) dx$ i, j = 1, 2, ..., N... is ill-conditioned.

Formal statement of (ill-conditioned) problem: $f = \underset{f \ge 0}{\operatorname{argmin}} ||Kf - \mathbf{d}||$

Regularized problem: $f^{(\alpha)} = \operatorname*{argmin}_{f \ge 0} \left(||Kf - \mathbf{d}||^2 + \alpha ||f||^2 \right)$

Seek f(x) minimising mean square misfit with regularising term, for some α selected by one of many *ad hoc* criteria

Key papers and reviews

IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 50, NO. 5, MAY 2002

101

Contents lists available at SciVerse ScienceDirect

Progress in Nuclear Magnetic Resonance Spectroscopy

journal homepage: www.elsevier.com/locate/pnmrs

Solving Fredholm Integrals of the First Kind With Tensor Product Structure in 2 and 2.5 Dimensions

Lalitha Venkataramanan, Member, IEEE, Yi-Qiao Song, and Martin D. Hürlimann

The "Schlumberger algorithm" for separable kernels in 2D:

$$M(\tau,\tau_2) = \int \int k_1(x,\tau_1)k_2(y,\tau_2)\mathcal{F}(x,y)\,\mathrm{d}x\,\mathrm{d}y + e(\tau_1,\tau_2)$$

or in discretised matrix form:

 $M = K_1 F K_2' + E$

Separability applies to most 2D relaxation experiments but *not* all diffusion ones

Numerical estimation of relaxation and diffusion distributions in two dimensions J. Mitchell*, T.C. Chandrasekera, L.F. Gladden

Department of Chemical Engineering and Biotechnology, University of Cambridge, New Museums Site, Pembroke Street, Cambridge CB2 3RA, United Kingdom





Innocent questions that don't go away

The story:

- The problem is very ill-conditioned.
- We distort it to make it less so.
- The form of the regularising functional is a matter of taste.
- We fiddle the value of the regularising parameter α according to criteria which are also a matter of taste.
- The results "look" reasonable.
- They agree well enough with known distributions for synthetic data.

Innocent questions:

- We all know the inversion is highly ill-conditioned. What do these distributions really mean ?
- How reliable / accurate are they ?
- How do you choose the value of the "tuning knob" α ?
- How do you choose the regularising functional ? (square norm, curvature, entropy, UPEN ...)
- How much information do they contain ?
- How many measurements can you extract ?



Bayesian interpretation of Regularisation

Regularization problem:

$$f^{(\alpha)} = \operatorname*{argmin}_{f \ge 0} \left(||Kf - \mathbf{d}||^2 + \alpha ||f||^2 \right)$$

in which:
$$||Kf - \mathbf{d}||^2 = \sum_k (K_{kj}f_j - d_k)^2 = \sigma^2 \chi^2$$

Noise known to be white Gaussian, so *Likelihood* is:

$$\mathcal{P}(\mathbf{d}|\mathbf{f},I) = \exp\left(-\chi^2/2\right)/Z_L$$

Hence identify: $||Kf - \mathbf{d}||^2 = -2\sigma^2 L(\mathbf{f}; \mathbf{d}, \sigma)$ log-likelihood

log-prior

Similarly identify:

i.e. prior on distribution ${f f}$:

 $\mathcal{P}(\mathbf{f}|I) \propto \exp\left(\alpha S(\mathbf{f})\right)$

 $\alpha ||f||^2 = \alpha S(\mathbf{f})$

So minimising the argument $-Q = L - \alpha S$ or maximising $Q = \alpha S - L$

maximises posterior "inference" $\mathcal{P}(\mathbf{f}|\mathbf{d},I) \propto \exp{(Q)}$

 $\mathcal{P}(\mathbf{f}|\mathbf{d}, I) \propto \exp\left(-(L(\mathbf{f}; \mathbf{d}) - \alpha S(\mathbf{f}))\right)$

"Best" distribution maximises ${m Q}$ for ${f f}=\hat{f f}=f^{(lpha)}$

Then by Bayes' Theorem:

Regularised distribution is "most probable" (maximum a posteriori probability or MAPP) – assigning αS as log-prior



17

The importance of the Evidence



- "Evidence" statistics in principle should be computed (though rarely done)
 - from data analysis using competing models
 - **BayeSys** and **MassInf** generate Evidence values
- Competing models then rationally ranked by a probability ratio
- No experience anywhere ! but model adequacy questions include:
 - 1. Discrete vs continuous T_2 distributions in "conventional" reservoir rocks
 - 2. Exponential vs Gaussian decay models in "unconventional" rocks (organic shales)



SGR N 52"17 47

Apparatus for exploration of the problem:

• Potential apparatus:

Constraints on priors from independence of discretization [12]:

Random samples from a wide class of uncorrelated measures, appropriate as priors, are known to be "atomic" (i.e. encodable as a finite set of discrete masses) [12,13,14].

"Atomic" property is arguably reasonable in the context:

 T_2 is a routine proxy for pore "size" – but derives from "isolated" pore assumptions – clearly non-physical ! Dominant diffusion eigenmodes of fully inter-connected pore network a better model – and a finite discrete set may describe them

Practical implementations [15] use various "engines" for MC² exploration of posterior probability spaces. Concomitant estimates of the Bayesian "evidence" ("prior predictive") are also available.

LGPL codes "**BayeSys**" and "**MassInf**" available with documentation (John Skilling, 2004, Maximum Entropy Data Consultants Ltd)



So how *should* we assign the prior?

Entropic Prior

$$\mathcal{P}(\mathbf{f}|\alpha, \mathbf{m}) = \left(\frac{\alpha}{2\pi}\right)^{M/2} \exp\left(\alpha S(\mathbf{f}, \mathbf{m})\right)$$

where $S(\mathbf{f}, \mathbf{m}) = -\sum_{i=1}^{M} f_i \log\left(\frac{f_i}{m_i}\right)$

- Highly structured **f** highly informative (low *P* without data)
- Featureless f uninformative (high p without data)
- Shannon entropy should be a candidate
- Entropy the only form without spurious correlations
- Over-riding consideration should be *consistency*

Hence: "Quantified Maximum Entropy" or QME

Problems emerge:

- *Not* independent of pixelation / discretisation of domain
- Consistency requires *independence* of pixelation ... For *any* disjoint sub-domains $A \cap B = \emptyset$ Additivity: Positivity: $\Phi_{A\cup B} = \Phi_A + \Phi_B$ and $0 \le \Phi_C < \infty \quad \forall C$

Surprising consequences !



"Atomic" priors

J. R. Statist. Soc. B (1997) 59, No. 1, pp. 217–235

Prior Distributions on Measure Space

By SIBUSISO SIBISI† and JOHN SKILLING

University of Cambridge, UK

[Received March 1994. Final revision March 1996]



Fig. 1. Evolution of a sample from the Dirichlet prior

SUMMARY

A measure is the formal representation of the non-negative additive functions that abound in science. We review and develop the art of assigning Bayesian priors to measures. Where necessary, spatial correlation is delegated to correlating kernels imposed on otherwise uncorrelated priors. The latter must be infinitely divisible (ID) and hence described by the Lévy–Khinchin representation. Thus the fundamental object is the Lévy measure, the choice of which corresponds to different ID process priors. The general case of a Lévy measure comprising a mixture of assigned base measures leads to a prior process comprising a convolution of corresponding processes. Examples involving a single base measure are the gamma process, the Dirichlet process (for the normalized case) and the Poisson process. We also discuss processes that we call the supergamma and super-Dirichlet processes. Examples of multiple and continuum base measures are also discussed. We conclude with numerical examples of density estimation.



More on atomic priors

Proc. 15th Workshop on Maximum Entropy and Bayesian Methods (1996) pp 261-270 Kluwer

JOHN SKILLING AND SIBUSISO SIBISI University of Cambridge, Cavendish Laboratory Madingley Road, England CB3 0HE

Abstract. A "measure" is the mathematical concept underlying distributed quantities such as images, spectra, and probability distributions. Inference about a measure requires a suitable Bayesian prior. If the prior is to remain valid on indefinitely small cells, it must be "infinitely divisible", and in consequence samples from it will be "atomic" (being essentially a limited sum of delta functions). Infinitely divisible priors are described in terms of the distribution of delta function strengths via the Lévy-Khinchin representation. Such priors include the Poisson process, but not the Quantified Maximum Entropy prior that has been used as its continuous reformulation.

Figure 4. Subdivision of an *id*-process into atoms. A sample from a Gamma process with measure $\alpha = 5$ per cell is shown (top) with successive five-fold expansions of the central fraction.

Small probabilities of O(1) "flux" in a small cell *Not* O(1) probability of *small* flux





The "Massive Inference" method

The "Massive Inference" prior for measure *F*:

$$\mathcal{P}(F|\alpha,q) = e^{-\alpha} \left[\delta(F) + e^{-F/q} \sqrt{\frac{\alpha}{qF}} I_1\left(2\sqrt{\frac{\alpha F}{q}}\right) \right]$$

Properties:

- Satisfies requirement of "infinite divisibility"
- Delta-function at F = 0 forces posterior mode to be null ! Counter-example to MAPP: mode is not useful
- Mean values instead used for display / "best" distribution
- Two hyper-parameters:
 - α is Poisson Expectation number of "atoms" over domain range on which F is estimated
 - q is a dimensional flux unit (expected scale)



The "Massive Inference" method: sketch derivation



How is total flux ζ distributed ? (given r atoms) $\mathcal{P}(\zeta|r) = \begin{cases} \delta(\zeta) & r = 0\\ e^{-\zeta} \zeta^{r-1}/(r-1)! & r > 0 \end{cases}$

Hint: pd of sum is (Laplace) convolution of pd's: Use product of Laplace Transforms

SGR N 52117 40

How many atoms *r* in cell δx ?

Choose Poisson prior of expectation / mean $\mu = \alpha \, \delta x$

 $\mathcal{P}(r) = e^{-\mu} \mu^r / r!$ r = 0, 1, 2, ...

How large ($\zeta_i = F_i/q$) are each of the *r* atoms ?

$$\mathcal{P}(\zeta_i) = \exp(-\zeta_i)$$
 $i = 1, 2, ..., r$
Total flux in cell width δx is $\zeta = \sum_{i=1}^r \zeta_i$

How is total flux ζ distributed ? (any *r* atoms in cell δx)

$$\mathcal{P}(\zeta) = \sum_{r=0}^{\infty} \mathcal{P}(\zeta|r) \mathcal{P}(r)$$
Number of atoms r
in cell summed
away analytically
$$= e^{-\mu} \left[\delta(\zeta) + e^{-\zeta} \sqrt{\mu/\zeta} I_1 \left(2\sqrt{\mu\zeta} \right) \right]$$

Hence previous page ($F = \text{dimensionless flux} \times \text{flux quantum}$)

$$\mathcal{P}(F|\alpha,q) = e^{-\alpha} \left[\delta(F) + e^{-F/q} \sqrt{\frac{\alpha}{qF}} I_1 \left(2\sqrt{\frac{\alpha F}{q}} \right) \right]$$



Computational engines



- Markov Chain Monte Carlo algorithms
- exploring Hilbert (space-filling) curves
- Implemented in Public Domain (LGPL) codes BayeSys and Massinf
- © John Skilling 2004
- Further development since 2004 bibliography incomplete !
- May be more advanced codes available commercially

(Maximum Entropy Data Consultants Ltd)

How Far Can We Go?

كــم لــما பி







Bibliography

- [1] W E Kenyon (1997). Petrophysical principles of applications of NMR Logging. *The Log Analyst*, March-April 1997, 21-43
- [2] J Mitchell, L F Gladden, T C Chandrasekera & E J Fordham (2014). Low-field permanent magnets for industrial process and quality control. *Prog Nucl Mag Res Spectros*, **76**, 1-60
- [3] A N Tikhonov & V Y Arsenin (1977). *Solutions of ill-posed problems.* London, Wiley.
- [4] E J Fordham, A Sezginer & L D Hall (1995). Imaging multiexponential relaxation in the $(y, \log T_1)$ plane, with application to clay filtration in rock cores. J Mag Res Ser A 113 139-150
- [5] L Venkataramanan, Y-Q Song & M D Hürlimann (2002). Solving Fredholm integrals of the first kind with tensor product structure in 2 and 2.5 dimensions. *IEEE Trans Signal Process* **50** 1017–1026
- [6] J Mitchell, T Chandrasekera & L F Gladden (2012).
 Numerical estimation of relaxation and diffusion distributions in two dimensions.
 Prog Nucl Mag Res Spectros 62 34–50
- [7] V Bortolotti, R J S Brown, P Fantazzini, G Landi & F. Zama (2016).
 Improved inversion of two-dimensional NMR relaxation data with the UPEN principle.
 Proc. 13th Int. Conf. on Magnetic Resonance in Porous Media

- [8] R L Parker & Y-Q Song (2005). Assigning uncertainties in the inversion of NMR relaxation data. *J Mag Res* **174** 314-324
- [9] L Venkataramanan T M Habashy D E Freed & F K Gruber (2012). Continuous moment estimation of CPMG data using Mellin transform. *J Mag Res* **216** 43-52
- [10] F K Gruber, L Venkataramanan, T M Habashy & D E Freed (2013). Estimation of petrophysical and fluid properties using integral transforms in NMR. J Mag Res 228 104–115
- [11] M Prange & Y-Q Song (2010). Understanding NMR *T*₂ spectral uncertainty. *J Mag Res* **204** 118–123
- [12] S Sibisi & J Skilling (1997). Prior distributions on measure space. *J Roy Stat Soc B* **59**(1) 217-235
- [13] T S Ferguson (1973). A Bayesian analysis of some nonparametric problems. Ann. Statistics 1(2) 209-230
- [14] J F C Kingman (1993). Poisson Processes, OUP.
- [15] Skilling, J. (2004) **BayeSys** and **MassInf** (2004).

Manual for the eponymous softwares. Trademarked and Copyright by Maximum Entropy Data Consultants Ltd. Distributed in the public domain under the GNU Lesser General Public License.



Part 3: Updates !!



SGR N52'12 40' E 0'4'50'

... recent visit to Steve Gull

- Long-term collaborator with John Skilling on "Maximum Entropy and Bayesian Methods ..."
- Co-founder (with Skilling) of "Maximum Entropy Data Consultants Ltd" (MEDC Ltd) (not seeking new clients)
- MEDC released BayeSys and MassInf under LGPL (but © MEDC)
- Still active in Cambridge (Astrophysics) offered to run (compare ?) proprietary MemSys code (QME) on same data
- Skilling retired but still active has novel MC² approach "Nested Sampling" Bayesian Analysis (2006) 1, Number 4, pp. 833–860
 Mested Sampling for General Bayesian directly addressing the Evidence statistic: Computation
 Nested Sampling for General Bayesian Computation

Abstract. "The evidence Z is often the single most important number in the [Bayesian] problem and I think every effort should be devoted to calculating it" (MacKay 2003)[1]. Nested sampling does this by giving a direct estimate of the density of states. Posterior samples are an optional by-product.



29

and after last time ... Evren Yarman's discrete decay model

 Based on svd of Hankel matrix representation of signal to determine set of discrete decays

$$h_k = \sum_{n=1}^{N} w_n \gamma_n^k + \sigma d_k^{(L)} \quad \text{for all } k \quad 0 \le k \le 2N$$

signal approximation error

 $\gamma_n = \mathrm{e}^{-T_\mathrm{e}/T_2^{(j)}}$

• or in matrix form ...

 ${f H}={f H}_{f a}+\sigma{f H}_{f d}$ (lpha is an L'th root of unity)

• error term
$$d_k^{(L)} = \frac{1}{L} \sum_{\ell=0}^{L-1} \tilde{d}_\ell \alpha^{\ell k}$$
 where $\tilde{d}_k = \lim_{z \to \alpha^k} \frac{\mathbf{P}_{\bar{\mathbf{u}}}(z^{-1})}{\mathbf{P}_{\mathbf{u}}(z)}$

u is con-eigenvector of H and σ the con-eigenvalue:

 $\mathbf{H}\mathbf{u}=\sigma\bar{\mathbf{u}}$

- **P** are the polynomials of con-eigenvectors: $\mathbf{P}_{\mathbf{u}}(z) = \sum u_k z^k$
- γ_n found as the roots of the eigenpolynomial $\mathbf{P}_{\mathbf{u}}(z)$

On approximation of functions by exponential sums *

Gregory Beylkin*, Lucas Monzón

Appl. Comput. Harmon. Anal. 19 (2005) 17-48

A new inversion method for NMR signal processing

C. E. Yarman¹, L. Monzón^{2,4}, M. Reynolds^{2,4}, N. Heaton³

$$M_{n}(k) = \sum_{j=1}^{J} a_{j} \left(1 - e^{-\frac{T_{W,n}}{T_{1,j}}} \right) e^{-\frac{k T_{E}}{T_{2,j}}} + \epsilon_{n}(k)$$



Fig. 2. [Top] Noisy measurement (blue) and denoised approximation (red). [Bottom] The logarithm of the absolute value of the difference between them.

Brine calibration (data thanks to Dr Jon Mitchell)



Schlumberger Research

Data fitting example - Bentheimer





Telling the algorithm – try one more decay





Sandstones











Sandstones





10

0.5

0

0.5

0

0.5

0

0.5

0

0.5

0

M(t)/M(0)

Carbonates





SGR NS21747 E9455

Carbonates

- Raw data
- Tikhonov inversion
 Hankel matrix
 Tri-exponential fit

0.6

....ul

0.01

0.1

 T_2 / s







SGR NS21747 E9455



Porosity and its errors – (complete) integral of the distribution



Computational results compared



Errors against independent (dry gas) reference



N

SGR N 52"17 40 E 0"4" 51"

Log-mean T_2 – (first-order) *moment* of the distribution



Used in permeability estimators ...





SGR N 52"17 40 E 0"4" 51"



SGR N 52"17 40" E 0"4" 54"