Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 174 (2005) 314–324
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Assigning uncertainties in the inversion of NMR relaxation data

Robert L. Parker [a,*], Yi-Qaio Song [b]

[a] Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, USA
[b] Schlumberger-Doll Research, Ridgefield, USA

Received 29 November 2004; revised 2 March 2005
Available online 5 April 2005

Abstract

Recovering the relaxation-time density function (or distribution) from NMR decay records requires inverting a Laplace transform based on noisy data, an ill-posed inverse problem. An important objective in the face of the consequent ambiguity in the solutions is to establish what reliable information is contained in the measurements. To this end we describe how upper and lower bounds on linear functionals of the density function, and ratios of linear functionals, can be calculated using optimization theory. Those bounded quantities cover most of those commonly used in the geophysical NMR, such as porosity, T_2 log-mean, and bound fluid volume fraction, and include averages over any finite interval of the density function itself. In the theory presented statistical considerations enter to account for the presence of significant noise in the signal, but not in a prior characterization of density models. Our characterization of the uncertainties is conservative and informative; it will have wide application in geophysical NMR and elsewhere.

1. Introduction

NMR signals of T2 decay are routinely recorded in well logs to characterize the porosity, permeability, and saturation of the rock formations [1], and are central in many laboratory measurements for the evaluation of porous rocks [2]. A key quantity of interest is the rate of T2 decay of the signal amplitude as the precessing protons interact with their environment during a CPMG experiment [3,4]. The signal amplitude M as a function of time t after the first pulse is well approximated by a positive sum of exponential decays, which is usually written as an integral

$$M(t) = \int_0^\infty F(T)\exp(-t/T)\,\mathrm{d}T \tag{1}$$

integrating over T, the characteristic relaxation times of the decay mode selected by the initializing pulse train; $F(T)$ is the number density of protons with relaxation time T. There can be a direct relationship (see [1]) between the spectrum of pore sizes to the density spectrum F. But F is not measured; only M is known, and so F must be recovered from this signal.

Eq. (1) is effectively a Laplace transform. If $M(t)$ were known exactly on any dense set of t, Lerch's theorem [5] states that F is uniquely determined by M. Fig. 1 illustrates why Lerch's theorem is inapplicable: it shows a typical NMR record from the Schlumberger NMR logging tool [6], a record used for illustration throughout the paper. Data sets like this one are collected every 0.75 ft in wells that can be many thousands of feet deep (see Section 4.3). Each such record comprises 5000 samples of M taken at intervals of 200 μs, a significant number of data, but not a continuous function. Since F is nonnegative, Eq. (1) shows that M must be positive and decrease monotonically toward zero, yet the observed signal is neither positive nor monotonic because of measurement noise. Since the observed data cannot match Eq. (1) exactly, what can be learned about F from the measured version of M?

* Corresponding author.
E-mail address: rlparker@ucsd.edu (R.L. Parker).

1090-7807/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2005.03.002

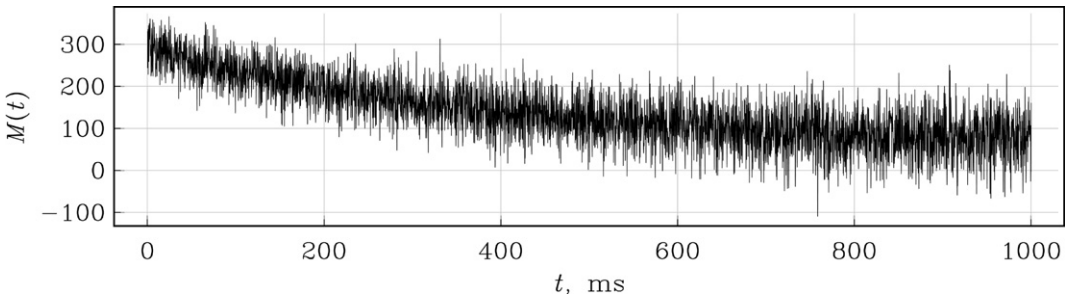

Fig. 1. Typical amplitude record from the Schlumberger NMR logging tool at a single depth in a well. The signal comprises 5000 samples spaced at an interval of 200 μs. This record will be used as a basis for illustration in the calculations.

The standard solution [7–11] to the modeling of F is by means of a regularization which generates a unique density function by minimizing

$$S(\alpha) = \sum_{j=1}^{m} \left[\int_0^\infty F(T) \exp(-t_j/T)\, \mathrm{d}T - M(t_j) \right]^2 + \alpha \int_0^\infty F(T)^2\, \mathrm{d}T \qquad (2)$$

over nonnegative functions F, and selecting the parameter α in a somewhat subjective way. The rationale for the method is that in Eq. (2) the first term represents a measure of model misfit and the second model complexity, both undesirable qualities. The two properties, misfit and complexity, are in conflict and cannot both be made arbitrarily small simultaneously, and so by suitable choice of α, a satisfactory compromise is sought. This kind of regularization provides a plausible density function from which other quantities of interest can be calculated, but it is silent on the question of the range of satisfactory alternative solutions.

An alternative to regularization that does provide uncertainty information is the Bayesian formulation [12,13] but this approach uses a statistical characterization of the density function and an assumption of a prior distribution that we wish to avoid. The associated computational effort can also be quite extensive.

We investigate properties of the set of all models matching the measurements to a specified degree, a much wider class than that of traditional regularized models. If the modeling assumptions and approximations are valid, these must be properties of the true density spectrum and therefore constitute firm information about it. This cautious approach has been developed by Parker [14] for a variety of geophysical inverse problems. Notice that the value $F(T_0)$ at any particular T_0 is not constrained by the available measurements, unless further smoothness assumptions are made about $F(T)$. On the other hand, we will see that linear functionals of F like

$$I_w(a,b) = \int_a^b w(T)F(T)\, \mathrm{d}T \qquad (3)$$

for $0 < a < b < \infty$ and bounded functions $w \geqslant 0$, must lie within strict limits imposed by the values of the observations. Functionals related to I_w, such as the T_2 log-mean, are central to NMR well-log interpretation and therefore we will concentrate on these functionals.

Because of measurement error, the acceptable degree of mismatch between data and the predictions of a density model must be formulated statistically, based on a model of the measurement noise. Fortunately, the system noise in NMR closely approximates an ideal of zero mean, uncorrelated Gaussian random additive noise, which makes its characterization relatively simple. Since a random element is necessarily present, any conclusions we draw must be framed in terms of probabilities, not absolute certainties.

2. Optimization theory

We introduce the set of models Ω that are said to be compatible with a given decay record. Then we examine the question of deciding whether there are such models. This turns out to involve a well-known form of quadratic programming, nonnegative least squares (NNLS), which is the solution to the conventional least-squares problem with the additional condition that every component of the solution vector must be nonnegative. Once we know models exist, we can discover common properties of the members of the set. We show how to find the uncertainties in several important linear functionals of the density and of certain interesting quantities defined by ratios of linear functionals. Each of these calculations can be mapped into a form that can be solved with the algorithm for NNLS.

2.1. The set of adequate models

The range of relaxation times T encountered is so large that common practice displays results on a log time axis; we follow the same convention in the theory and introduce the log time $\lambda = \ln T$. Then

$$M(t_j) = \int_{-\infty}^{\infty} \widetilde{F}(\lambda) \exp(-t_j e^{-\lambda})\, d\lambda, \tag{4}$$

where $\widetilde{F}(\lambda) = F(e^{\lambda})e^{\lambda}$. We use the abbreviation

$$G_j(\lambda) = \exp(-t_j e^{-\lambda}). \tag{5}$$

In the log domain the kernels G_j are shifted versions of the same function of λ: the continuous version of Eq. (4) is a convolution in log time.

We define a set Ω of density functions as those $\widetilde{F}$ obeying

$$\chi^2 = \sum_{j=1}^{m} \frac{1}{\sigma_j^2} \left[M_j - \int_{-\infty}^{\infty} \widetilde{F}(\lambda) G_j(\lambda)\, d\lambda \right]^2 \leqslant X_0^2 \quad \text{and} \quad \widetilde{F} \geqslant 0, \tag{6}$$

where σ_j is the standard error of the measurement made at time t_j. The expression on the left of Eq. (6) will be recognized as the model's misfit statistic χ^2. For an appropriate choice of X_0 we will say that the set Ω contains models that match the measurements satisfactorily; in other words when $\widetilde{F} \in \Omega$, the function $\widetilde{F}$ fits the data. We will return to the question of how to determine X_0^2 in Section 3. Notice the case of noise-free measurements is covered in Eq. (6) by setting $X_0^2 = 0$ and $\sigma_j = 1$.

First, we address the existence of solutions: Are there any models at all fitting the data? Systematic errors or unusually low signal environments can cause a failure of the model equation (see Section 4.3); we must detect this situation, because then we will not be able to make valid inferences about $\widetilde{F}$. As a first step we construct the best-fitting solution, $\widetilde{F}_0$: we solve

$$\widetilde{F}_0 = \arg\min_{\widetilde{F} \geqslant 0} \sum_{j=1}^{m} \frac{1}{\sigma_j^2} \left[M_j - \int_{-\infty}^{\infty} \widetilde{F}(\lambda) G_j(\lambda)\, d\lambda \right]^2. \tag{7}$$

If the minimum value is less than the adopted X_0^2, there are models, otherwise the set Ω is empty.

Mathematically, Eq. (7) presents a problem in semi-infinite quadratic programming; the proper setting for which is NBV, the normed space of functions of bounded variation; the integrals must to be written as Stieltjes integrals, since the minimizing elements are not ordinary functions, but delta functions. Readers interested in setting up a rigorous treatment of the continuous time problem should consult Luenberger [15]. Here however, we will replace the integrals by sums in a finite-dimensional version. In obvious matrix notation Eq. (7) becomes

$$\mathbf{f}_0 = \arg\min_{\mathbf{f} \geqslant 0} \|\Sigma^{-1}(\mathbf{m} - G\mathbf{f})\|^2, \tag{8}$$

where the norm is the Euclidean norm; the matrix $G \in \mathcal{R}^{m \times n}$ is composed of rows sampling the function $G_j(\lambda)$ at n evenly spaced points with $\lambda_{\min} \leqslant \lambda \leqslant \lambda_{\max}$; $\Sigma \in \mathcal{R}^{m \times m}$ is the diagonal matrix of standard errors; $\mathbf{f} \in \mathcal{R}^n$ is a vector representing the unknown density function, sampled in λ in the same way as the rows of G; and $\mathbf{m} \in \mathcal{R}^m$ is the vector of amplitude measurements. The integral over the real line has been replaced by a finite sum, and so values must be provided for $\lambda_{\min}$ and $\lambda_{\max}$, lower and upper limits for the permitted relaxation times. Obviously in practical computations the number n must be chosen to be much smaller than the actual number of protons involved.

Eq. (8) is a convex optimization problem, and therefore there are no spurious local minima [15]. It is an example of a quadratic program called nonnegative least squares (NNLS); see Lawson and Hanson [16]. We will transform all the numerical optimizations encountered later on into NNLS problems, because the numerical algorithm for NNLS is fast and stable. At the minimizing vector, $\mathbf{f}_0$, the Kuhn–Tucker conditions [17] require a solution with no more than m positive components; in principle $n \gg m$ (the number of protons is much larger than the number of observations) and so $\mathbf{f}_0$ consists of at most m positive entries in a vector whose elements are otherwise zeros. The positive components would become delta functions in the continuous representation Eq. (7). In calculations on well-log records like the one shown in Fig. 1 the number of positive elements in $\mathbf{f}_0$ rarely exceeds ten. Further details of a numerical implementation will be described in Section 4, after we have examined the statistical issues. Notice that within Ω the best-fitting density function $\widetilde{F}_0$ is not given privileged status, such as that of most probable candidate model.

2.2. Bounding linear functionals

Having confirmed the existence of solutions by showing that the smallest attainable misfit in Eq. (8) is less than X_0^2, we can proceed to the investigation of properties common to all the densities in Ω. For example, an important quantity determined from the NMR record in wells is the local porosity, given by $\gamma M(0)$, the amplitude at $t = 0$ times a calibration constant; see [1]. $M(0)$ cannot be measured directly, but from Eqs. (1) and (4) it is

$$M(0) = \int_0^{\infty} F(T)\, dT = \int_{-\infty}^{\infty} \widetilde{F}(\lambda)\, d\lambda. \tag{9}$$

Our goal is to discover the range of permitted values of $M(0)$. To find the largest value consider

$$M(0)^+ = \max_{\widetilde{F} \in \Omega} \int_{-\infty}^{\infty} \widetilde{F}(\lambda)\, d\lambda. \tag{10}$$

The finite-dimensional version of Eq. (10) is the constrained optimization problem

$$M(0)^+ = \max \mathbf{u}^{\mathrm{T}}\mathbf{f} \tag{11}$$

over vectors $\mathbf{f} \in \mathcal{R}^n$ subject to

$$\|\Sigma^{-1}(\mathbf{m}-G\mathbf{f})\|^2 \leqslant X_0^2 \quad \text{and} \quad \mathbf{f} \geqslant \mathbf{0}, \tag{12}$$

where $\mathbf{u} \in \mathcal{R}^n$ is the vector $\Delta\lambda[1,1,1,\ldots 1]^{\mathrm{T}}$ and $\Delta\lambda$ is the spacing of the samples in λ. The matrix G can be shown to be of full rank, and hence Eq. (12) describes a bounded convex region, the intersection of an ellipsoid and the positive orthant in $\mathcal{R}^n$, which is not empty by hypothesis. For computation we convert the convex optimization given by Eqs. (11) and (12) into an NNLS problem.

Rather than maximizing the dot product in Eq. (11), we use it as a constraint, along with positivity of $\mathbf{f}$, when we minimize the data misfit: with $\mu \geqslant 0$

$$X_1(\mu)^2 = \min_{\mathbf{f}\geqslant 0} \|\Sigma^{-1}(\mathbf{m}-G\mathbf{f})\|^2, \quad \text{subject to } \mathbf{u}^{\mathrm{T}}\mathbf{f} = \mu. \tag{13}$$

Consider solving Eq. (13) for a series of values of μ. Every nonnegative density $\mathbf{f}$ must give rise to a point lying in the shaded region in Fig. 2; those in the shaded region below the line $\chi^2 = X_0^2$ fit the data adequately. Because all the entries of $G > 0$, the value of X_0^2 increases without bound as μ tends to infinity, and since $X_1(\mu)^2$ is continuous, the curve must intersect X_0^2; the largest value of μ for which $X_1(\mu)^2 = X_0^2$ is the upper bound on $\mathbf{u}^{\mathrm{T}}\mathbf{f} = M(0)$ for all $\mathbf{f}$ in Ω. In practice the curve is monotonically increasing away from its minimum (which is at $\mu = \mathbf{u}^{\mathrm{T}}\mathbf{f}_0$) as illustrated in Fig. 2. Hence to bound $M(0)$, we solve the equation $X_1(\mu)^2 = X_0^2$ with $\mu > \mathbf{u}^{\mathrm{T}}\mathbf{f}_0$, and the root is $M(0)^+$, the desired upper bound on $M(0)$.

Numerically, the constrained optimization problem in Eq. (13) can be solved by taking the equality constraint into the penalty, as a heavily weighted row: we solve

$$Z_1(\mu)^2 = \min_{\mathbf{f}\geqslant 0}[\|\Sigma^{-1}(\mathbf{m}-G\mathbf{f})\|^2 + W^2(\mathbf{u}^{\mathrm{T}}\mathbf{f}-\mu)^2] \tag{14}$$

$$= \min_{\mathbf{f}\geqslant 0} \left\| \begin{bmatrix} \Sigma^{-1}G \\ W\mathbf{u}^{\mathrm{T}} \end{bmatrix} \mathbf{f} - \begin{bmatrix} \Sigma^{-1}\mathbf{m} \\ W\mu \end{bmatrix} \right\|^2. \tag{15}$$

Here W is a fixed positive weight, large enough to insure almost exact satisfaction of the constraint condition; the optimal $\mathbf{f}$ for Eqs. (14) and (15) solves (13) too. As Lawson and Hanson [16] observe, proper weighting can achieve solutions as accurate as those found by techniques that enforce linear equality constraints. Observe that Eq. (15) is in the form of NNLS.

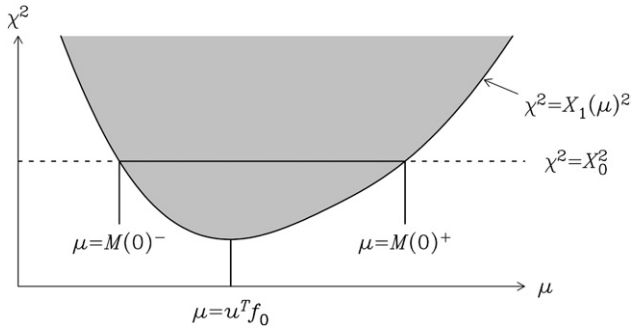

Fig. 2. Data misfit as a function of the constraint parameter μ in Eq. (13). The shaded region corresponds to solution vectors obeying $\mathbf{f} \geqslant 0$ and $\mathbf{u}^{\mathrm{T}}\mathbf{f} = \mu$.

The greatest lower bound, $M(0)^-$, is found in the same way, but by searching in $0 < \mu < \mathbf{u}^T\mathbf{f}_0$. The only difference is that $X_1(0)^2$ might be less than X_0^2, in which case $M(0)^-$ is zero. In this way we have obtained best possible values for the limits in $M(0)^- \leqslant M(0) \leqslant M(0)^+$ for all density models in satisfactory agreement with the measurements: this is a statement of the uncertainty in our knowledge of $M(0)$.

Exactly the same ideas work for other linear functionals of $\widetilde{F}$. For example, we can determine the interval containing smoothed versions of $\widetilde{F}$, a variant of the analysis of resolution introduced by Backus and Gilbert [18]. We average the density over the interval $(\lambda-\delta, \lambda+\delta)$

$$\langle \widetilde{F}(\lambda)\rangle_{2\delta} = \frac{1}{2\delta}\int_{\lambda-\delta}^{\lambda+\delta} \widetilde{F}(\mu)\,\mathrm{d}\mu. \tag{16}$$

This is a bounded linear functional of $\widetilde{F}$, and thus upper and lower bounds can be computed for it at each λ just as we found those for $M(0)$. In this way we can discover the corridor of permitted values of the smoothed functions $\langle \widetilde{F}(\lambda)\rangle_{2\delta}$; the quantity 2δ is the resolution of the smoothed models.

2.3. Bounding ratios of linear functionals

Another number commonly used in well-log interpretation derived from the density $\widetilde{F}$ is τ, the T_2 log-mean, which measures the logarithmic center of mass of the density

$$\ln \tau = \frac{\int_{-\infty}^{\infty} \lambda\widetilde{F}(\lambda)\,\mathrm{d}\lambda}{\int_{-\infty}^{\infty} \widetilde{F}(\lambda)\,\mathrm{d}\lambda} = \beta(\widetilde{F}). \tag{17}$$

We would like to place bounds on β, and hence on τ, but because β is a ratio, it is not a linear functional of $\widetilde{F}$ nor even a convex one. Nonetheless, the same approach used for calculating bounds on $M(0)$ is effective here also. Rather than treating β in Eq. (17) as the objective function in an optimization problem, we declare it to be a constraint when we minimize the misfit function. As a constraint Eq. (17) becomes linear after we rationalize

$$\beta\int_{-\infty}^{\infty} \widetilde{F}(\lambda)\,\mathrm{d}\lambda = \int_{-\infty}^{\infty} \lambda\widetilde{F}(\lambda)\,\mathrm{d}\lambda. \tag{18}$$

In matrix form we must solve the optimization problem

$$X_2(\beta)^2 = \min_{\mathbf{f}\geqslant 0} \|\Sigma^{-2}(\mathbf{m}-G\mathbf{f})\|^2, \quad \text{subject to } \beta\mathbf{u}^{\mathrm{T}}\mathbf{f} = \mathbf{l}^{\mathrm{T}}\mathbf{f} \tag{19}$$

or, with heavy weighting to apply the linear constraint, the NNLS problem with the same solution vector:

$$Z_2(\beta)^2 = \min_{\mathbf{f}\geqslant 0} \left\| \begin{bmatrix} \Sigma^{-1}G \\ W(\beta\mathbf{u}^{\mathrm{T}} - \mathbf{l}^{\mathrm{T}}) \end{bmatrix} \mathbf{f} - \begin{bmatrix} \Sigma^{-1}\mathbf{m} \\ 0 \end{bmatrix} \right\|^2. \tag{20}$$

Here $\mathbf{l} \in \mathcal{R}^n$ is the vector of sampled values of the function λ. As with $M(0)$, we sweep through values of β, seeking the intersection of $X_2(\beta)^2$ with the target misfit X_0^2: the smaller solution corresponds to β^-, the greatest lower bound on β, and the larger one to the least upper bound, β^+; and from Eq. (17) we may conclude: $\tau^- = \exp\beta^- \leqslant \tau \leqslant \exp\beta^+ = \tau^+$.

A complication arises in the case of the lower bound, however. In the continuous version of the problem there is no finite lower bound on the value of β for the members of Ω because it is always possible to add to any density adequately fitting the observations a component with a very large amplitude, but at very small λ. In this way we can arrange for β to be arbitrarily small while the model misfit is essentially unchanged owing to the vanishingly small amplitudes of all the kernels G_j when λ is small enough (see Eq. (5)). This behavior corresponds to a high density of protons with very rapid relaxation rates whose presence will not show up in the data because, by the time the first echo is measured, their signals have all vanished. Thus, based on the measurements alone, there is no positive lower bound on τ. In the finite-dimensional version the difficulty is evaded by assigning $\lambda_{\min}$, the lower cut-off in the sampling of the λ axis. A plausible default value for $\lambda_{\min}$ might be the log of the shortest measured T_2, but as we would expect, the choice of the cut-off has a decisive influence on the lower bound in τ. It is sometimes possible to estimate $\lambda_{\min}$ on the basis of geological knowledge (prior information), and when this can be done it may improve the τ^- considerably; see Section 5.

Finally, the bound fluid volume fraction below a certain critical relaxation time can be estimated from $\widetilde{F}$ via integrals in the form

$$C(T_c) = \frac{\int_{-\infty}^{\infty} K(T_c;\lambda)\widetilde{F}(\lambda)\,\mathrm{d}\lambda}{\int_{-\infty}^{\infty} \widetilde{F}(\lambda)\,\mathrm{d}\lambda}, \tag{21}$$

where the function $K(T_c;\lambda)$ is constant up to $\lambda = \ln T_c$ then drops abruptly to zero, or decays smoothly after that; see Kleinberg and Boyd [19]. These quantities can be bounded in exactly the same way as τ. The upper bound of C is subject to the same problems as those on τ^-, but since the kernel function K is constant for small enough λ the effect is not as severe.

3. Noise and statistics

The signal is contaminated by noise of considerable magnitude. To accept or reject a density model as an adequate predictor of the observations requires a statistical basis. First, we need a mathematical model for the noise, and then we must have a theory for testing the models we calculate.

3.1. Characterization of the measurement noise

A standard mathematical model for the noise consists of additive, zero-mean, uncorrelated Gaussian random variables. This choice is supported by laboratory experiments on the instrumentation, and by a "noise channel" collected during the logging; it will be confirmed by tests on the record itself. To use Eq. (6), we need a value for σ_j, the standard error of the noise at the jth signal point, or echo. We must allow for σ to vary slowly in time (see Fig. 1), an effect due to processing to correct for the motion of the tool. Assume that σ increases linearly with time so that for CMR signals we can write

$$\sigma_j = \sigma_0 + (j-1)\kappa, \tag{22}$$

where the parameters σ_0 and κ are obtained from the record. One approach would be to find the best-fitting theoretical model and subtract it but, as discussed in Section 3.2, the resultant noise estimate would be biased downwards. Instead we form first differences

$$Y_j = M(t_{j+1}) - M(t_j), \quad j = 1,2,\ldots,m-1. \tag{23}$$

Anticipating the numerical results, we find that σ lies between 35 and 50 U, and $M(0) = 350$. Then the variance of Y_j is about $2\sigma_j^2 \geqslant 2450$, while the signal component is the drop between consecutive echos, on average $(350/5000)^2 = 0.005$, utterly negligible in comparison. Thus, the first differences are essentially pure noise. However, the variables Y_j are correlated, even when the noise in $M(t)$ is not. We can get an uncorrelated set by deleting every other one, wasteful, but in view of the large number of observations, not a serious loss. Now σ_0 and κ can be estimated from the Y_j series using the method of maximum likelihood [20]; we omit the details.

For the signal shown in Fig. 1 we find $\sigma_0 = 36.1$ and $\kappa = 3.76\times10^{-3}$, so that in the 1-s record the noise increases by about a factor of 1.5. The noise model predicts the normalized differences $V_j = Y_{2j}/\sigma_{2j} \approx N(0,2)$ and that V_j should be uncorrelated. We test these predictions. Figs. 3 and 4 show the autocorrelation function of

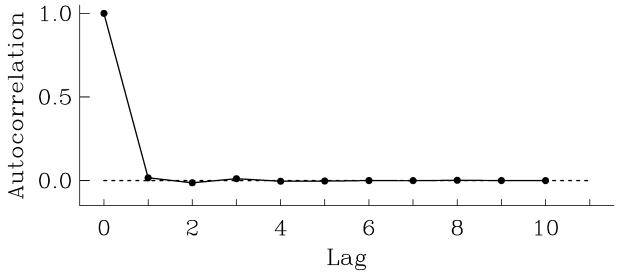

Fig. 3. Autocorrelation function for the series of normalized first differences $V_j = Y_{2j}/\sigma_{2j}$ based on data shown in Fig. 1. The series is essentially uncorrelated, as predicted by the noise model.

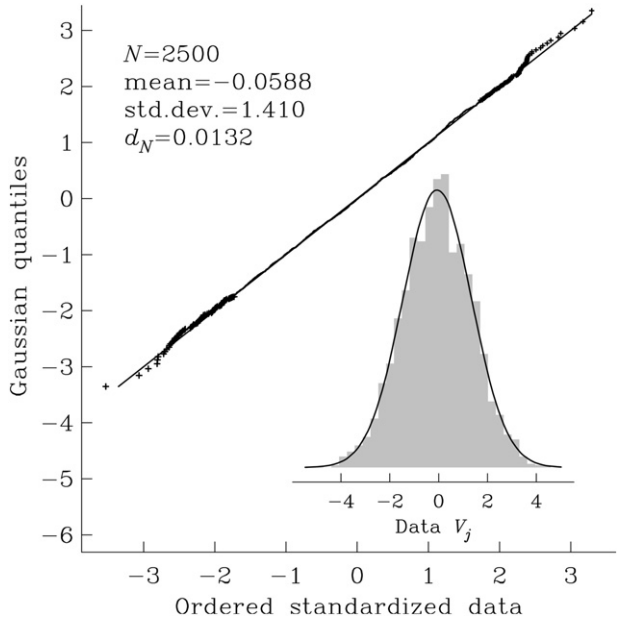

Fig. 4. Q–Q plot and histogram of standardized first differences V_j calculated from record in Fig. 1. The comparison distribution is $N(0,2)$.

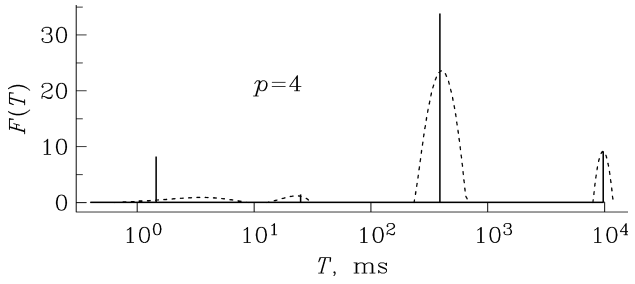

Fig. 5. Best-fitting density model $\mathbf{f}_0$ found by solving Eq. (8). The dashed line is the solution regularized by Eq. (2) with the expected misfit. See Section 4 for details.

normalized first differences, obtained by Fourier transforming the power spectrum [21]; the calculated correlation is negligible. To check the distribution of the normalized differences we calculated the quantile–quantile or Q–Q plot [20] and histogram shown in Fig. 5, which perfectly verifies the Gaussian distribution. The Kolmogorov–Smirnov statistic [22] confirms there is no significant departure from the Gaussian probability density function: $P(d_N > 0.0132) = 0.78$.

3.2. Expected misfit of best-fitting model

Now we are in a position to compute the best-fitting model, by solving Eq. (8). As we have already noted, the quantity being minimized is the familiar misfit measure, χ^2. How large should χ^2 be? Naively we might say that its expected value (statistically) must be m, the expected value of the sum of the squares of m standardized Gaussian random variables. But we have deliberately sought the model with the smallest possible misfit which must result in a downward bias. If the fit were a linear least-squares problem with P parameters to be estimated, it is well known [21] that the expected χ^2 is reduced by the number of degrees of freedom, P, giving $E[\chi^2] = m - P$; further, the misfit itself is distributed according to χ^2_{m-P}. Ours is not a simple linear problem because of the positivity constraint, and moreover, $P > m$ since there are far more free parameters (the number of protons, or in the continuous approximation, infinitely many) than measurements. As we noted Section 2, the solution vector of the NNLS problem contains no more than m positive components, and empirically, in the limit of large n, the model tends to a definite function with a finite number of peaks, say $p \leqslant m$. Fig. 5 shows the result for our standard example, where $p = 4$. It requires $2p$ parameters to specify such a model, one each for the amplitude and the location of the peak; we propose that the effective number of degrees of freedom is $P = 2p$. Then the expected value of misfit is $m - 2p$, and the variance $2m - 4p$. This guess is based on the idea that near the true solution, the system behaves linearly under small perturbations. We tested the conjecture with a Monte Carlo simulation, adding pseudo-random Gaussian noise to the signal from a known spectrum, then solving Eq. (8): the results correspond well to the predictions both in mean and variance.

It should be noted that in actual calculations the number of fitted observations is typically 120, considerably fewer than 5000. The measurements are aggregated, or binned, to reduce computational work; see Section 4.

3.3. Testing models against observation

If we accept the conjecture, we can settle the existence question by comparing the minimum misfit found by solving Eq. (8) with $m - 2p$; given that the variance of the smallest misfit is $2m - 4p$, we may reject the model if the probability of matching or exceeding the observed value by chance is too large. In that case something in the theoretical edifice is inconsistent with observation. One option is simply to go on to the next record. Alternatively, one might wish to repair the model in some way, and the most tractable candidate for modification is the noise model. In many situations, noise closely approximated by a Gaussian process is mixed with a process that adds very occasional events with large amplitude, outliers. A quadratic misfit measure like χ^2 gives outliers considerable weight due to squaring, even though they are the least reliable sources of information. Therefore, a common strategy [23], which we adopt, is to delete the largest deviating data if their misfit forces χ^2 above an acceptable level. Removing one or two measurements out of 120 is a minor modification of the data set. When, as sometimes happens, more than a few dele-

tions are required, it is best to reject the record as incompatible with the theory.

The next question is the choice of target misfit, X_0^2 in Eq. (6) or Eq. (12), the definition of an acceptable density function. If we choose $X_0^2 = m$, which is the expected value given m observations, then roughly 50% of the time the true signal will have a larger misfit than the one assigned. It would be unwise to draw conclusions about $\widetilde{F}$ from Ω when the actual density function belongs to that set only half the time. To be reasonably certain that the proper density belongs to the set we must choose X_0^2 larger than m. The chi-squared distribution of misfit can be employed to calculate the of value X_0^2 that includes the actual misfit with a probability P_Ω say 0.9 or 0.99. Notice, however, that bounds on functionals like $M(0)$ based on the value 0.99 do not correspond to the standard 99% confidence interval for $M(0)$. We have not made a statistical estimate of $M(0)$ itself.

To clarify the issue consider the situation illustrated in Fig. 6, where bounds on $M(0)$ have been computed at a series of depths. Imagine that exactly measured amplitude values were also available. A finite set of exact amplitudes does not meet the conditions of Lerch's theorem, and so even with such measurements we cannot calculate the true $M(0)$ at each depth, only strict bounds such that $M(0)^- \leqslant M(0) \leqslant M(0)^+$. Here the bounds $M(0)^-$ and $M(0)^+$ are exact numbers. When noise contaminates the measurements, those numbers are replaced by statistical estimates. Thus when $P_\Omega = 0.99$ we expect that 99% of the time the estimated bounds will encompass the bounds generated by precise measurements, but run inside them (that is, be in error) 1% of the time. Contrast this with the conventional 99% confidence interval for $M(0)$, where the true value of $M(0)$ itself will lie within the limits 99% of the time.

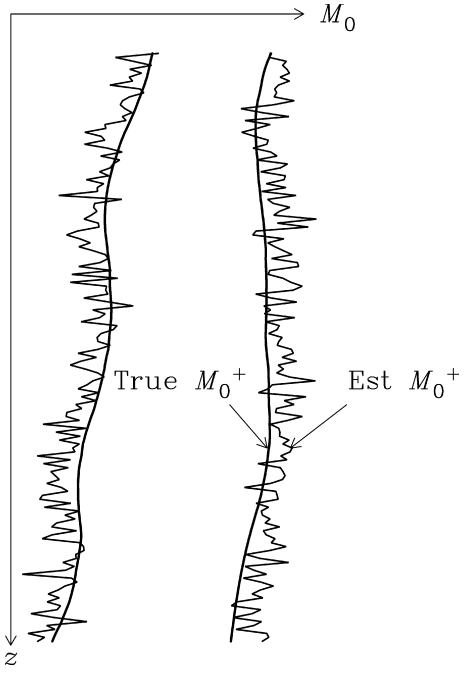

Fig. 6. Because of observational noise, the bounds found by the optimization are themselves noisy and, depending on the choice of P_Ω are incorrect a certain fraction of the time. A value of $P_\Omega = 0.99$ does not generate a 99% confidence interval for $M(0)$.

4. Applications

To build a practical program from the theory we need to set a number of parameters. The use of Eqs. (8) and (12) requires a sampling scheme for λ. The interval of λ should at least cover that of $\ln T$ in the observations: initially, we take $\lambda \in (\ln T_{\min}, \ln 10 T_{\max})$, although the lower limit may be raised if there is reason to believe short relaxation times are largely absent. To create the solution vector $\mathbf{f} \in \mathcal{R}^n$, the interval is sampled uniformly in λ. Usually we set $n = 150$. In theory, the larger n, the more accurate the computed results, but there is no significant improvement for values above 150; in fact, $n = 60$ yields results typically only a few percent different from those found with 150.

It might appear we have no choice about m, the number of measurements. However, a common practice [24] is to bin the observations, that is, group them into sets of k consecutive values, taking the mean of the group, and presenting the mean as the datum to be matched; the uncertainty of the mean is decreased by the factor $k^{1/2}$. Furthermore, the size of the group varies along the record, getting bigger exponentially for later times (log binning), in this way preserving higher resolution of early times dominated by rapid decays. The purpose of this procedure is to reduce the size of the matrix $G \in \mathcal{R}^{m \times n}$ and the concomitant computational costs. We have followed this practice, normally reducing m to 120, excluding the first 20 echos from the binning. We remark that matrix size in the NNLS calculations can also be reduced without the approximation of binning by preliminary use of the QR factorization [16] of $\Sigma^{-1}G$.

4.1. Minimum misfit

For the first example, we analyze the record of Fig. 1. The reduced data set after log binning is shown in Fig. 7; the first 20 echos are taken individually to retain the highest possible resolution in time, the rest (100 groups) grouped together into sets of increasing size, ranging from one member up to 272. The total number of data in the reduced set is 120. As can be seen in the figure, log binning generates data with approximately constant relative error in the tail. The best-fitting density is computed from the reduced set by solving Eq. (8) and $\chi^2_{\min} = 128.2$. The number of positive components, $p = 5$, and so the expected value for the misfit is $120 - 10 = 110$ according to the conjecture of Section 3. Is that misfit acceptable? The probability of so large

a misfit is 0.11, which we regard as too small (our cut-off is 0.2). The strategy described in Section 3.2 prescribes removal of the worst fitting points until the misfit is acceptable: we delete the two biggest contributors to the misfit, the points indicated in Fig. 7 which, if the model were unchanged, would reduce χ^2 to 116.7; but we must recompute the value with new data set. Now the best-fitting model achieves $\chi^2_{\min} = 115.7$, but with $p = 4$ and $m = 118$, so that the expected value is 110. The probability of $\chi^2 \geqslant 115.7$ is 0.34, which satisfies the acceptance criterion. The best-fitting density with four positive elements has already been plotted in Fig. 5; the corresponding signal is shown as the smooth curve in Fig. 7.

In support of the conventional form of regularization, which penalizes the solution 2-norm, it is sometimes asserted [1,3] that optimal solutions (comprised of delta functions) are unphysical, from which it would follow that the set Ω is too large, and ought to be further restricted. We argue that smoother models can achieve almost the same values of bounded parameters as the extremal solutions. For example, we have already computed a bound: the smallest possible chi-squared misfit. How much larger would the demand of smoothness make that bound? To avoid an elaborate calculation we simply convolved the best-fitting solution with a smooth template, and computed chi-squared for resulting densities. Two examples are shown in Fig. 8; the increase in misfit is 1.3% for the solid curve and 6.1% for the dashed one. In our judgement, these two functions are perfectly physical yet the increase in chi-squared is trivial in one case, and modest in the other. We conclude that attempting to restrict attention to a class of smooth models would improve the bounds only slightly, at the expense of introducing a standard of solution acceptability that is difficult to justify or quantify.

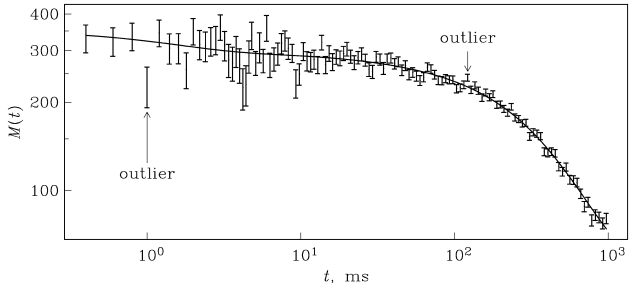

Fig. 7. Binned data set with ± 1-σ uncertainties. The two outliers shown must be removed to attain an acceptable fit to the model equation. Best fitting solution generates the smooth amplitude curve with $\chi^2 = 115.7$, and $n = 118$. Note both axes are logarithmic.

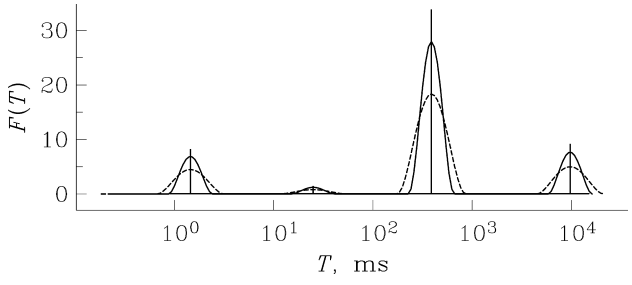

Fig. 8. Minimum misfit solution shown in Fig. 5 convolved with two smoothing functions to map them into ordinary continuous densities.

4.2. Bounds on the T_2 log-mean

Next we illustrate the bounding of useful functionals of $\widetilde{F}$. We focus on the estimation of upper and lower bounds of τ, the T_2 log-mean of the distribution, given by Eq. (17). Again the measurements of Fig. 1 will be used, having first been binned and edited as shown in Fig. 7. A series of solutions to the NNLS problem in Eq. (20) must be found for trial values of τ, to secure a match between $X_2(\ln\tau)^2$ and X_0^2. First, we need a value for W, the factor applying a heavy weight to equality constraint; we insure that the norm of the heavily weighted row is 1000 times larger than that of any row in $\Sigma^{-1}G$; Lawson and Hanson [16] suggest a value of roughly $\eta^{-1/2}$, where η is the relative computer precision, in our calculations about 10^{-14}. We have chosen a smaller number because making W too big risks numerical instability; another stabilizing device is to put the constraint row first, rather than last as shown in Eq. (20), because of the way the computer code uses QR transformations. Next the target misfit X_0^2 must be assigned, which governs the size of the set Ω. As discussed in Section 3, we choose a value to give a reasonably high chance of including the misfit of the true signal. A probability $P_\Omega = 0.9$, with $m = 118$ (recall we deleted two of the original 120 reduced data), gives $X_0^2 = 138.06$ from the chi-squared distribution.

The situation for our data set is shown in Fig. 9. The upper bound is 810.9 ms and the lower bound 6.50 ms. The very steep gradient of misfit at the right means that the bound is almost indifferent to the choice of probability; when $P_\Omega = 0.99$, $X_0^2 = 156.66$ which raises τ^+ by only 3.25%, to 836.2. The lower bound is more sensitive:

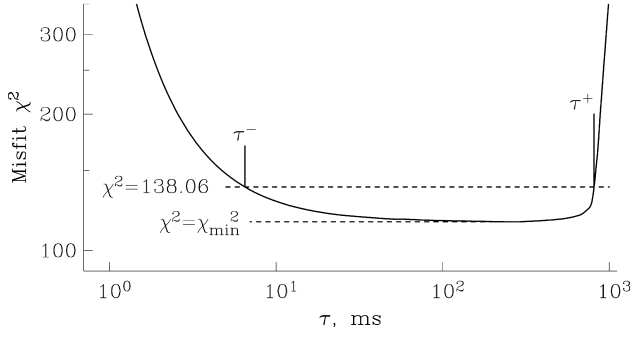

Fig. 9. Data misfit as a function of $\tau = \exp\beta$ in Eq. (20). Misfits less than 138.06 are acceptable with a probability of 0.9. Note both axes are logarithmic.

τ^- becomes 4.178, a decrease of 36%. Sensitivity in τ^- is not surprising, because τ^- is not determined by the measurements alone, but depends on the choice of the lower log-time cut-off, $\lambda_{\min}$. For either choice of P_Ω this is a wide interval, indicating that the observations are consistent with a great range of T_2 log-mean values. Another choice is the level at which to reject outliers in the preliminary data screening; for routine processing we have adopted a relatively harsh standard, deleting misfits that might arise 20% of the time by chance. If we accept instead both the rejected points indicated in Fig. 7, the bounding interval shrinks: τ^- becomes 10.9 and τ^+ 795.8, with the lower bound being much more sensitive as we would expect. Rejection of the outliers leads to a more conservative bound in each case, consistent with our generally cautious approach.

A mathematical purist will say that the lower bound on T_2 log-mean is simply an unsuitable quantity for estimation, and that one ought not to attempt the impossible. Nonetheless, a value is desirable for petrophysical analysis. A potential solution to the dilemma is to introduce information from outside the set of T_2 relaxation measurements. One option is to restrict even further the λ interval for modeling. We know from the geology of the site that the well is drilled in a carbonate sequence, and that carbonates cannot exhibit such short decay times as implied by the τ^- density model, where most of the protons have decay constants of 400 μs. We might move the lower cut-off in T up a decade to 5 ms on the basis of this expectation. With $P_\Omega = 0.9$, the lower bound increases sharply to 237.2, a factor of 36 increase, while the upper bound is essentially unaltered. The risk is that we are substituting an assumption for a measurement that we do not have, yet we tend to invest subsequent conclusions with the same status as those founded on genuine observations.

Another possible source of information is porosity, which is often measured independently by nuclear physics techniques [25]. Porosity provides a value for $M(0)$, which can act as a constraint through Eq. (9). It might be supposed that, because the solution regularized by (2) has been observed to track porosity quite accurately, one could apply the $M(0)$ obtained as a constraint, or perhaps even the value of $M(0)$ associated with the best-fitting solution. Both these estimates are derived from the NMR record, not outside information, and agreement between the regularized porosity and the true one, for example, must result from the absence of small T terms; that fact cannot be inferred from the T_2 data themselves.

Yet another strategy for improving the lower bound is to modify the misfit criterion, to apply more weight to the earliest echos. Measurements at early times influence the lower bound much more than the later ones and with chi-squared, their misfits are systematically larger. Various weighting schemes were investigated. Even in the simplest case, which involves multiplying the first term in the sum for χ^2 by $w_1 > 1$, the definition of P_Ω requires a nonstandard probability distribution found numerically or by asymptotic analysis. We omit the tedious details, and report that the increases in the lower bound obtained by this approach were all fairly modest, usually a small fraction of what can be realized by a change in $\lambda_{\min}$, for example. It is unrealistic to expect mere adjustment in the fitting process will be capable of curing a fundamental problem, the fact that β^- is not bounded below if Eq. (17) is used.

4.3. A longer section

Fig. 10 shows the results of calculations under two parameter settings performed on series of 439 records measured in a 332 ft section of a well drilled in carbonates. The record shown in Fig. 1 used throughout for illustration comes from the depth 8550 ft in this series.

With a larger data set we can examine some questions not amenable to test with a single record. For example, in standard analysis of T_2 data, the first echo at 200 μs is

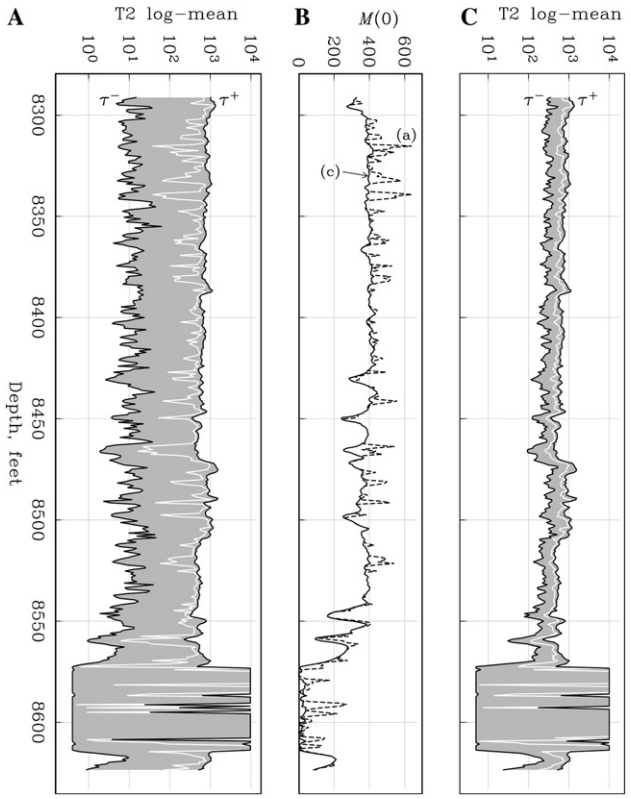

Fig. 10. (A) Upper and lower bounds in milliseconds on T_2 log-mean from T_2 NMR data taken with the Schlumberger logging tool in a well drilled through a carbonate sequence; the white line is τ calculated from the best-fitting model. The lower cut-off of decay times is 0.4 ms and $P_\Omega = 0.9$. (B) Zero-time amplitudes from the best-fitting solutions and estimate porosity; the two curves correspond to models with the cut-offs of 5 ms and, shown dashed, 0.4 ms. (C) Same as in (A) but calculations are based on the 5 ms lower cut-off. Depth scale in feet.

often discarded because of suspected systematic instrumental effects. We confirm the existence of an anomaly by looking at the statistics of the residuals of the first echo after subtracting the prediction of the best-fit solution. If the standardized residual is formed

$$R_k(t_1) = \frac{1}{\sigma_1}\left[M(t_1) - \int \widetilde{F}_k(\lambda) G_1(\lambda)\,\mathrm{d}\lambda\right], \tag{24}$$

where $\widetilde{F}_k$ is the best fit density of the kth record, under the assumptions of the theory, we expect R_k to be drawn from a Gaussian population with zero mean and variance one. From the 439 records we estimate the mean of the population of $R_k(t_1)$ to be -1.309 and the standard error 0.958. The standard error of the mean is $0.958/\sqrt{439} = 0.0457$. Thus the observed mean is more than 30 standard deviations away from the predicted value, leading us to firmly reject the proposition that the first echo conforms to the statistical model: it is biased downward. In comparison the same calculations on the second echo yield a population mean of -0.0408 and standard deviation 1.009, and similarly satisfactory values for later signals. Based on this analysis, we have always deleted the amplitude at 200 μs from the observations.

The calculations for Fig. 10 were performed with the following parameters. For each depth the original 5000 amplitudes were log binned, reducing the total number for analysis to 120. Outliers were deleted from the data if $\chi^2_{\min}$ was so large that it would be exceeded by chance with probability 0.2; this is rather aggressive pruning. The probability defining the target misfit X_0^2, was $P_\Omega = 0.9$. The number of samples in λ was 150. No additional weight in the misfit criterion was placed on earlier echos. In panel (A) the bounds for τ are based on a lower limit $\lambda_{\min} = \ln 0.4$, the log of the earliest time in the record (that is 0.4 ms; the echo at $t = 0.2$ ms has been deleted). Calculations for graph (C) used a cut-off of 5 ms, a figure based on the likely shortest decay time to be found in carbonates. The choice of upper limit $\lambda_{\max}$ is set at $\ln 10 T_{\max} = \ln 10^4$; making the upper limit larger had no effect on the results, except in those parts of the record ($8620 < z < 8570$ ft) where bounds on τ could not be established from the decay times, and then the bound is supplied by the limit chosen.

The bounds on τ in milliseconds are the outer lines in graphs (A) and (C); the grey middle curve is the value of τ derived from the best fit model. In the center plot (B) we show curves for $M(0)$ from the best-fitting model, the higher curve always resulting from the lower cut-off $\lambda_{\min}$. As the reader will recall from Section 2, $M(0)$ is proportional to porosity. At the base of the section there is an interval of low porosity, which causes the signal amplitudes to become very small; the noise is not similarly reduced, and so that the reliability of the models degrades, something clearly reflected in the bounds, which expand to the extremes allowed by the program. For five records in this low porosity segment, the signal to noise ratio was so poor that the best-fitting density was the function zero. The very large uncertainties calculated in the lower segment demonstrate that no reliance can be placed on τ in this interval.

In panel (A) the bounds on τ are separated typically by two decades. In (C), where the lower limit λ has been raised, the lower bound is usually a decade and a half higher, while, the upper bound has not changed perceptibly. In no case did the assumption of the restricted range in λ cause a failure to model the measurements (indicated by $\chi^2_{\min}$ so large that more than two data points would have to be discarded), so the assumption of a higher log time cut-off is completely compatible with the observations. If it is true that in this formation there are essentially no protons with T_2 relaxation times below 5 ms, we can confidently constrain τ to a tight corridor, except where the porosity is so low the measurements become uninformative.

The code to perform these calculations, written in Fortran, is quite fast: on a Sun Ultra 50 workstation the 439 records were processed in 87 s for minimum misfit and bounds on τ, that is, about a fifth of a second each.

5. Discussion and conclusions

It is a truism of quantitative science that a numerical estimate for a quantity has little value unless it is accompanied by some idea of its accuracy. In this paper, we describe a methodology for assessing the uncertainty of relaxation-time density functions calculated from geophysical NMR measurements. The theory gives a range within which certain properties of the density function must lie by exploiting the fact that the density must be positive. The properties treated are bounded linear functionals of the density, and ratios of linear functionals, and include all of the commonly used diagnostics used for geophysical well-logging with NMR, except the density function $F(t)$ itself: we find that the density for any particular decay time has an arbitrarily large uncertainty.

A key consideration in the development has been respect for the statistical aspects of the problem: statistics are necessary because of the very significant random component in the measurements, which cannot be ignored. The usual regularization approach, in our opinion, pays insufficient attention to these matters. On the other hand, statistical assumptions about the density function itself are not required in our treatment. We have chosen a conservative approach to the uncertainty problem: we wish to avoid assumptions about the smoothness of the density or its statistical properties, things that we believe are poorly determined, and whose introduction will generate false confidence in the results.

The mathematical techniques we use are those of optimization theory in finite-dimensional spaces. In particular we appeal to the theory of quadratic programming of positive definite forms, a subject with a fully worked out theoretical basis. Furthermore, there are efficient and stable numerical methods to implement the theory.

For illustration purposes we have concentrated on the T_2 log-mean, a logarithmic measure of the center of gravity of the density distribution. We find that T_2 log-mean has no positive lower bound without the introduction of some kind of additional assumption coming from outside the set of NMR data themselves, an assumption such as the exclusion of protons with decay rates faster than some limit. If a lower cut-off in relaxation times can be reliably assigned, by rock type, for example, this prior information helps generate sharp lower bounds. Other potential sources of information include porosity, although this possibility has not been examined in detail. Constraining the initial amplitude to match that of the regularized solution, or the best-fit solution, is equivalent to assuming the absence of small T terms in the density, and we believe that such an assumption is better made explicitly. In a section of well-log data we find that without an assigned lower cut-off in decay times, the range of permitted τ values is too large to be much use, but once a plausible value is specified, the corridor of permitted values for τ becomes gratifyingly narrow, except in one zone, where the signal-to-noise ratio is apparently very poor. The ability to identify unambiguously intervals where the NMR data are completely untrustworthy is valuable in itself.

Uncertainties can be found for other interesting quantities: in the geophysical NMR problem, we can treat the bound fluid volume fraction, and porosity. As we have seen, the same methods also permit the calculation of uncertainty in a smoothed version of the density function itself, something that will be most valuable in the analysis of laboratory NMR records, where the signal-to-noise ratio is much higher than for field measurements. Other possible applications to laboratory work include the ability to discriminate between density functions possessing several peaks or only one. These are topics for further research.

Acknowledgments

We thank Pabrita Sen for suggesting that we look at this problem, and for his helpful review of a first draft. We are extremely grateful to David Allen who provided invaluable advice as well as all the well-log data discussed here. Finally, we have benefited from Albert Malinverno's insights into inverse problems and statistical matters, and for his thorough reading of the manuscript; our special gratitude to him.

References

[1] R.L. Kleinberg, Well logging, Encyclopedia of Nuclear Magnetic Resonance, vol. 8, Wiley, New York, 1996, pp. 4960–4969.
[2] P.Z. Wong, Methods in the Physics of Porous Media, Academic Press, London, 1999.
[3] H.Y. Carr, E.M. Purcell, Effects of diffusion on free precession in NMR experiments, Phys. Rev. 94 (1954) 630.
[4] S. Meiboom, D. Gill, Modified spin-echo method for measuring nuclear relaxation times, Rev. Sci. Instrum. 29 (1958) 688–691.
[5] T.W. Körner, Fourier Analysis, Cambridge University Press, Cambridge, 1989.
[6] R.L. Kleinberg, A. Sezginer, D.D. Griffin, M. Fukuhara, Novel NMR apparatus for investigating an external sample, J. Magn. Reson. 97 (1992) 466–485.
[7] S.W. Provencher, A constrained regularization method for inverting data represented by linear algebraic or integral equations, Comput. Phys. Commun. 27 (1982) 213.
[8] E.J. Fordham, A. Sezginer, L.D. Hall, Imaging multiexponential relaxation in the (y, $\log T_1$) plane, with application to clay filtration in rock cores, J. Magn. Reson. Ser. A 113 (1995) 139–150.
[9] R.M. Kroeker, R.M. Henkelman, Analysis of biological NMR relaxation data with continuous distributions of relaxation times, J. Magn. Reson. 69 (1986) 218–235.
[10] G.C. Borgia, R.J.S. Brown, P. Fantazzini, Example of uniform-penalty inversion of multiexponential relaxation data, Magn. Reson. Imag. 16 (1998) 549–552.
[11] J.P. Butler, J.A. Reeds, S.V. Dawson, Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing, SIAM J. Numer. Anal. 18 (1981) 381–397.
[12] A. Tarantola, Inverse Problem Theory: Methods for Fitting and Model Parameter Estimation, Elsevier, New York, NY, 1987.
[13] M. Andrec, J.H. Prestegard, A Metropolis Monte Carlo implementation of Bayesian time-domain parameter estimation: application to coupling constant estimation from antiphase multiplets, J. Magn. Reson. 130 (1998) 217–232.
[14] R.L. Parker, Geophysical Inverse Theory, Princeton University Press, Princeton, 1994.
[15] D.G. Luenberger, Optimization by Vector Space Methods, Wiley, New York, 1969.
[16] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
[17] R. Fletcher, Practical Methods of Optimization, Wiley, New York, 1987.
[18] G. Backus, F. Gilbert, The resolving power of gross Earth data, Geophys. J. R. Astron. Soc. 16 (1968) 169–205.
[19] R.L. Kleinberg, A. Boyd, Tapered cutoffs for magnetic resonance bound water volume, SPE 38737 (1997).
[20] J.A. Rice, Mathematical Statistics and Data Analysis, Brooks-Cole Pub., Monterey, CA, 1988.
[21] M.B. Priestley, Spectral Analysis and Time Series, Academic Press, New York, 1981.
[22] M.G. Kendall, A. Stuart, Advanced Theory of Statistics, Griffin, London, 1966.
[23] F.R. Hampel, E.M. Ronchetti, P.J. Rouseenuv, W.A. Stahel, Robust Statistics: The Approach Based on Influence Functions, Wiley, New York, 1986.
[24] R. Freedman, Processing method and apparatus for processing spin echo in-phase and quadrature amplitudes from a pulsed nuclear magnetism tool and producing new output data to be recorded on an output record, United States Patent US 5,291,137, 1994.
[25] J.R. Hearst, P.H. Nelson, Well Logging for Physical Properties, McGraw-Hill Book Company, New York, 1985.