Reflecting Uncertainty in Inverse Problems: A Bayesian Solution using Lévy Processes

Robert L. Wolpert
Duke University ISDS
Durham, NC 27708-0251, USA
wolpert@stat.duke.edu

Katja Ickstadt
Fachbereich Statistik
Universität Dortmund
D-44221 Dortmund, Germany
ickstadt@statistik.uni-dortmund.de

August 4, 2004

Abstract

We formulate the inverse problem of solving Fredholm integral equations of the first kind as a nonparametric Bayesian inference problem, using Lévy random fields (and their mixtures) as prior distributions. Posterior distributions for all features of interest are computed employing novel Markov chain Monte Carlo numerical methods in infinite-dimensional spaces, based on generalizations and extensions of the authors' Inverse Lévy Measure (ILM) algorithm. The method is also well suited for deconvolution problems, for inverting Laplace and Fourier transforms, and for other linear and nonlinear problems in which the unknown feature is high- (or even infinite-) dimensional and where the corresponding forward problem may be solved rapidly. The methods are illustrated for an application to an important problem in rheology: that of inferring the molecular weight distribution of polymers from conventional rheological measurements, in which we achieve not just a point estimate but a posterior probability density plot representing all uncertainty about the weight.

PACS-2003 Classification codes: 02.30.Rz,Zz; 02.50.Ey,Tt; 02.70.Uu; 05.40.Fb; 83.10.-y.
Key words: Bayesian analysis, Lévy processes, Markov chain Monte Carlo methods, nonparametric statistical inference, rheology.

1 Introduction

In inverse problems, we try to make inference about the unknown inputs to some linear or nonlinear system on the basis of knowledge of the system itself and observations of the system's outputs. They arise in a wide range of applications including geophysics, tomography, virology, and rheology.

In this paper, we present the analysis of an example from rheology, the science of the flow and deformation of materials: the estimation of the molecular weight distribution for polymers on the basis of measurements of gross rheological features such as the stress relaxation modulus (ratio of stress to strain) or the storage and loss moduli characterizing the viscoelastic properties of the polymer (see Anderssen and Loy 2001, Maier *et al.* 1998). Our example is the solution of a two-stage inverse problem in which we first infer the stress relaxation modulus from measurements of storage and loss moduli, and then infer the molecular weight distribution from the relaxation modulus.

1.1 Ill-posedness and Uncertainty

Inverse problems are notoriously ill-posed, in the sense that a wide range of possible inputs lead to nearly identical outputs. Most applications feature systems that are nonlinear and often high-dimensional; some, including our rheology application, are even infinite-dimensional. Even low-dimensional inverse problems, including many medical applications, may present daunting computational obstacles.

A range of methods have been proposed to accommodate the ill-posed nature of inverse problems. It is conventional to select a "regular" element from among the wide range of inputs consistent with a given data set of outputs by penalizing those that are in some sense extreme (see Tikhonov (1963) for the introduction of this regularization method or, for an overview, Scales and Tenorio (2001)). A disadvantage of this regularization approach is the difficulty in selecting, interpreting and justifying a specific penalty function or level of regularization, trading off accuracy against smoothness.

In a Bayesian statistical approach, a prior distribution is introduced to express initial uncertainty about the inputs (and any other unknown model parameters) and a measurement-error model is introduced to quantify the probability distribution for measured quantities, given correct values for the inputs and model parameters. The posterior density function is proportional to the product of the prior and the measurement-error model, evaluated at the observed data (and then, when regarded as a function of the inputs and model parameters, called the likelihood function). Its negative logarithm is the sum of the negative log prior, which takes the place of a regularization penalty function, and the negative log likelihood, which plays the role of an optimality criterion. Minimizing this negative log posterior density (or, equivalently, maximizing the posterior density itself) gives the Maximum A Posteriori (MAP) estimate, closely related to the result of Tikhonov's regularization.

One can do more by integrating this posterior density, however, rather than maximizing it. The normalized integral over any set (a ball or product of intervals, for example) gives the probability that the inputs lie in that set, reflecting both the experimental evidence (through the likelihood function) and the desired regularity (through the prior distribution). This offers two advantages over non-statistical regularization methods. Firstly, it is now easy to select, describe, interpret, and defend the degree of regularization (simply by showing the pre-experimental probability of various possible inputs under the prior distribution). Secondly, one gets not only a point estimate of the input parameter vector, but a probability distribution expressing clearly how uncertain the inputs remain after observing the data and reflecting the regularity assumptions.

Bayesian statistical methods have been applied to both linear inverse problems (*e.g.*, Mohammad-Djafari and Idier 1996) and nonlinear ones (*e.g.*, Mosegaard and Tarantola 1995), following discretization to reduce the original problems to finite-dimensional ones. Typically prior distributions are taken to be Gaussian or more generally from the exponential family of distributions, often chosen by maximum entropy methods. Monte Carlo methods usually play an important role in both statistical and non-statistical analyses, either in form of Markov chain Monte Carlo (MCMC) procedures used for sampling from Bayesian posterior distributions or in form of stochastic optimization procedures including simulated annealing and genetic algorithms (see Mosegaard and Sambridge 2002, for an overview).

1.2 Our Method

Many interesting inverse problems are infinite-dimensional, with an unknown function or measure for the uncertain input. All these methods begin by reducing such problems to finite-dimensional ones by some form of discretization. This reduction limits the interpretation of both prior uncertainty (*i.e.*, degree of regularization) and posterior uncertainty (*i.e.*, remaining uncertainty following analysis) to the discrete approximate problem.

Recently Wolpert *et al.* (2003) proposed a new nonparametric Bayesian statistical approach for solving infinite-dimensional inverse problems without this approximation step, using gamma process prior distributions to represent the uncertainty about the solution of some Fredholm integral equations of the first kind. In addition to Fredholm integral equations, the method appears to be useful for deconvolution problems, for inverting Laplace and Fourier transforms, and for other inverse problems in which the unknown feature is high- (or even infinite-) dimensional and where the corresponding forward problem may be solved rapidly. Here we implement and extend this approach (to prior distributions expressible as mixtures of arbitrary Lévy processes), and illustrate it in solving a challenging problem in rheology— estimating a polymer's molecular weight distribution (an element of the infinite-dimensional space of positive measures on the real line), on the basis of measurements of gross rheological features.

The specific numerical problem of solving Fredholm integral equations of the first kind arises in a number of application areas. Here we try to impute an unknown measure $Z(ds)$ on the basis of finitely-many observed integrals

$$y_i = \int_{\mathcal{S}} k(x_i, s) Z(ds)\,,\ i \in I \tag{1}$$

of a known kernel $k(x_i, s)$. The problem of inferring Z from $\vec{y} = \{y_i\}_{i\in I}$ is ill-posed.

The conventional approach to solving such systems begins by making a finite-dimensional approximation such as

$$y_i \approx \sum_{j\in J} k_{ij} Z_j,$$

which can be solved (exactly or approximately) with linear algebraic methods. Usually the system remains ill-posed, motivating the use of regularization methods to select from among the (near-) solutions.

Instead of making such a finite-dimensional approximation, we treat the original problem of solving Eqn. (1) as a statistical estimation problem: that of inferring the uncertain positive measure Z upon observing random variables $Y_i \approx y_i$ (measurements subject to measurement-error) at the finite set of points $\{x_i\}_{i\in I}$. We need to specify a measurement-error model for Y_i (leading us to a likelihood function $L(Z \mid \vec{y})$) and a prior distribution $\pi(dZ)$ on the infinite-dimensional space $\mathcal{M}_+(\mathbb{R})$ of all positive measures on the real line.

We use the flexible class of Lévy random fields (and their mixtures) for prior distributions, employing Markov chain Monte Carlo (MCMC) methods based on the Inverse Lévy Measure (ILM) algorithm of Wolpert and Ickstadt (1998a) to calculate the posterior distribution of both the uncertain measure $Z(ds)$ and of all features of interest, fully reflecting all sources of uncertainty that affect inference. Several variations of the algorithm will be discussed for modeling random measures with infinitely many points of support. Different measurement-error models will be considered and compared, such as lognormal and normal models.

2 Rheological Example

Rheology is the study of the flow and deformation of materials. Many physical properties of viscoelastic compounds such as non-Newtonian fluids are determined by the stress relaxation modulus G that relates stress τ to strain γ through Boltzmann's (1876) time-dependent generalization

$$\tau(t) = \int_0^\infty G(s)\dot{\gamma}(t-s)\,ds$$

of Hooke's classic law $\tau = G\gamma$. Recently interest has centered on how the molecular weight distribution $w(dm)$ of polymers might be inferred from rheological features such as $G(t)$, measurable with a stress relaxation rheometer. The stress relaxation modulus G and molecular weight distribution $w(dm)$ are related through the integral equation:

$$G(t) = G_N^0 \left\{ \int_0^\infty e^{-t/r\tilde{\tau}(m)}\, m^{-1}\, w(dm) \right\}^r . \tag{2}$$

The parameter r, called the *reptation number*, reflects the molecular dynamics of how the long polymer molecules interact. It is conventionally taken to be one or two, but for polystyrene data was estimated by Maier *et al.* (1998) to be $r = 3.84 \pm 0.1$ and by Thimm *et al.* (2000) to be $r = 1.9 \pm 0.3$. The function $\tilde{\tau}(m)$ relating normalized molecular weight m to molecular relaxation time $\tilde{\tau}$ is usually taken to be a monomial $\tilde{\tau}(m) = \kappa\, m^\alpha$, with experimentally determined material constants κ and α. The molecular weight distribution is normalized to be a probability distribution on the log scale, *i.e.*, to satisfy $\int_0^\infty m^{-1}\, w(dm) = 1$, so evidently G_N^0 in Eqn. (2) is simply $G(0)$.

Our data set does not include direct oscillatory rheometer measurements of $G(t)$. Therefore, following the suggestion of Maier *et al.* (1998), we first solved (and reported in Wolpert *et al.* 2003) another inverse problem: that of inferring the relaxation spectral measure $H(d\tau)$ from measured values of the storage modulus $G'(\omega_i)$ and loss modulus $G''(\omega_i)$ at selected frequencies ω_i, from the two-dimensional Fredholm equation

$$\big[G'(\omega),\, G''(\omega)\big] = \int_0^\infty \frac{[\omega^2\tau^2,\, \omega\tau]}{1+\omega^2\tau^2}\, \tau^{-1} H(d\tau), \tag{3}$$

from which the stress relaxation modulus G may be computed easily as

$$G(t) = G_N^0 \int_0^\infty e^{-t/\tau}\, \tau^{-1}\, H(d\tau). \tag{4}$$

Following Anderssen and Loy (2001), Anderssen and Mead (1998), and Anderssen and Westcott (2000) we base our analysis on the oscillatory rheometer "frequency sweep" measurements of $G'(\omega)$ and $G''(\omega)$ reported by Berger (1988). We concentrate in this paper on the solution of Eqn. (2) with $G(t)$ assumed known.

3 Bayesian Formulation

To pose the problem of solving the general Fredholm equation

$$Y_i \sim y_i = \int_{\mathcal{S}} k(x_i, s) Z(ds)\,,\ \ i \in I \tag{1}$$

as a Bayesian inference problem, we must identify a likelihood function for the vector $\vec{Y} = \{Y_i\}_{i\in I}$ of observable quantities and construct a prior distribution for the uncertain measure Z (which here plays the role of an infinite-dimensional parameter vector). To solve the problem we must also offer a numerical scheme for evaluating the posterior distribution of Z and of any functionals of interest, such as integrals $Z[\phi] \equiv \int_{\mathcal{S}} \phi(s) Z(ds)$.

3.1 Measurement Error Models

Both normal and lognormal measurement-error models are commonly used in statistical applications. A normal measurement-error model $Y_i \sim \mathsf{No}(y_i, \sigma_i^2)$ (σ_i^2 denotes variance) leads to a log likelihood function from Eqn. (1) of the form

$$\log f(\vec{Y} \mid Z) = -\frac{1}{2} \sum_{i\in I} \big\{\log(2\pi\sigma_i^2) + (Y_i - y_i)^2/\sigma_i^2\big\}$$

or, since likelihood functions are only determined up to an arbitrary scale factor, to

$$\log f(\vec{Y} \mid Z) = -\frac{1}{2} \sum_{i\in I} (Y_i - y_i)^2/\sigma_i^2. \tag{5}$$

If the variances $\{\sigma_i^2\}_{i\in I}$ (which we take to be known) are equal this model is homoskedastic, meaning that all the squared errors $(Y_i - y_i)^2$ have the same expected value σ^2.

A lognormal measurement-error model $Y_i \sim \mathsf{LN}(y_i, \sigma_i^2)$, on the other hand, leads to a log likelihood function of the form

$$\log f(\vec{Y} \mid Z) = -\frac{1}{2} \sum_{i\in I} \left(\log \frac{y_i}{Y_i} \right)^2 / \sigma_i^2. \tag{6}$$

In such models, larger squared errors $(Y_i - y_i)^2$ are associated with larger predicted values y_i.

In our application we employ a lognormal measurement-error model for solving Eqn. (3), because observations at low angular frequencies ω where $G'(\omega)$ and $G''(\omega)$ are higher require a longer series of measurements, leading to proportionately smaller measurement-errors. We considered both normal and lognormal measurement-error models for solving Eqn. (2) and, because $G(t)$ varies over four orders of magnitude, base our inference on a lognormal model.

3.2 Lévy Prior Distributions

Our nonparametric Bayesian approach begins by modeling the measure $Z(ds)$ *a priori* as an increasing Lévy process $Z \sim \mathsf{Lv}\big(\nu(du, ds)\big)$ (see Rogers and Williams 2000, *p.* 310) with Lévy measure $\nu(du, ds)$. Such a random measure is expressible as a finite or countable sum

$$Z(ds) = \sum_j \upsilon_j \, \delta_{\sigma_j}(ds) \tag{7}$$

of point masses of random sizes $\upsilon_j \ge 0$ at random locations $\sigma_j \in \mathcal{S}$ drawn jointly from a Poisson point process on $\mathbb{R}_+ \times \mathcal{S}$ with intensity measure $\nu(du, ds)$ satisfying the integrability condition

$$\iint_{\mathbb{R}_+ \times \mathcal{S}} \min(1, u) \, \nu(du, ds) < \infty. \tag{8}$$

The sum in Eqn. (7) will be finite if and only if $\nu(\mathbb{R}_+ \times \mathcal{S}) < \infty$, a condition not guaranteed by Eqn. (8).

The random measure $Z(ds)$ will assign independent random variables $Z(A_i)$ to disjoint sets $A_i \subset \mathcal{S}$. Each random variable $Z(A)$ has the characteristic function

$$\mathsf{E}[e^{i\omega Z(A)}] = \exp\left[\iint_{\mathbb{R}_+ \times A} (e^{i\omega u} - 1) \, \nu(du, ds) \right]$$

and hence (finite or infinite) prior mean $\mathsf{E}[Z(A)] = \iint_{\mathbb{R}_+\times A} u\,\nu(du,ds)$ and variance $\mathsf{V}[Z(A)] = \iint_{\mathbb{R}_+\times A} u^2\,\nu(du,ds)$. The mean and variance of stochastic integrals $Z[\phi] \equiv \int \phi(s)\,Z(ds)$ (needed, *e.g.*, for the truncation bounds in Section 4) are similarly available as

$$\mathsf{E}(Z[\phi]) = \iint_{\mathbb{R}_+\times\mathcal{S}} u\,\phi(s)\,\nu(du,ds), \quad \mathsf{V}(Z[\phi]) = \iint_{\mathbb{R}_+\times\mathcal{S}} u^2\,\phi(s)^2\,\nu(du,ds). \tag{9}$$

Familiar examples of Lévy processes on $\mathcal{S}$ include the Poisson process, with Lévy measure $\nu(du,ds) = \delta_1(du)\,\gamma(ds)$, assigning Poisson random variables $Z(A) \sim \mathsf{Po}\big(\gamma(A)\big)$ with mean $\gamma(A)$; the gamma process, with Lévy measure $\nu(du,ds) = e^{-u\beta}u^{-1}\,du\,\gamma(ds)$, assigning gamma random variables $Z(A) \sim \mathsf{Ga}(\gamma(A),\beta)$ with shape $\gamma(A)$ and rate β; and the (one-sided) α-stable process, with $\nu(du,ds) = c_\alpha u^{-\alpha-1}\,du\,\gamma(ds)$, $c_\alpha \equiv \frac{2}{\pi}\Gamma(1+\alpha)\sin\frac{\pi\alpha}{2}$, assigning fully-skewed α-stable random variables $Z(A) \sim \mathsf{St}\big(\alpha, 1, \gamma(A), \tan\frac{\pi\alpha}{2}\gamma(A)\big)$ with $0<\alpha<1$ and $\beta = 1$. Our continuous parametrization of the α-stable $\mathsf{St}(\alpha,\beta,\gamma,\delta)$ distribution has α-stable index $\alpha \in (0,2]$, skewness $\beta \in [-1,1]$, rate $\gamma \geq 0$, shift $\delta \in \mathbb{R}$, and log characteristic function

$$\log \mathsf{E}[e^{i\omega X}] = \begin{cases} i\delta\omega - \gamma|\omega|^\alpha - i\beta\gamma\tan\frac{\pi\alpha}{2}\{\omega - |\omega|^\alpha\operatorname{sgn}\omega\} & \alpha \neq 1 \\ i\delta\omega - \gamma|\omega|\{1 + \frac{2i\beta}{\pi}\operatorname{sgn}\omega\log|\omega|\} & \alpha = 1, \end{cases} \tag{10}$$

taken from Cheng and Liu (1997) but modified slightly to make it extensible in the (third) rate parameter γ. For more about α-stable random variables and processes see Samorodnitsky and Taqqu (1994) or Sato (1999).

In (Wolpert and Ickstadt 1998a,b) we present an algorithm, called the Inverse Lévy Measure (ILM) algorithm, for constructing these processes. Note that these processes have increments which are independent and infinitely divisible but perhaps not stationary, so they are somewhat more general than the usual stationary independent-increment (SII) Lévy processes; see Jacod and Shiryaev (1987, Chapter II, §4c) for mathematical details.

For implementation purposes we truncate if necessary (see Section 4) to achieve a Lévy measure with finite total mass $\nu(\mathbb{R}_+\times\mathcal{S}) < \infty$, so that $Z(ds)$ will be supported (almost surely) on only finitely many points. Thus a realization can be represented as $Z = Z^\theta \equiv \sum_{j=1}^{M} u_j\,\delta_{s_j}$, a sum of point-masses of size $u_j \geq 0$ at points $s_j \in \mathcal{S}$, for $\theta = \big((u_1,s_1),\ldots,(u_M,s_M)\big)$ in the disjoint union $\Theta = \cup_{M<\infty}\big(\mathbb{R}_+\times\mathcal{S}\big)^M$. In each of the examples we consider here, the set $\mathcal{S}$ will be a subset of Euclidean space $\mathbb{R}^d$ and the Lévy measure $\nu(du,ds)$ will have a density function $\nu(u,s)$ with respect to a Lebesgue product reference measure $m(du,ds) = du\,ds$. This makes the prior distribution $\pi(dZ)$ absolutely continuous with respect to the reference Poisson random field distribution on Θ, $P(dZ) = \mathsf{Po}\big(m(du, ds)\big)$, with prior log density function $\pi(Z) \equiv \pi(dZ)/P(dZ)$ given by

$$\log \pi(Z) = c + \sum_{j=1}^{M} \log \nu(u_j, s_j), \tag{11}$$

where $c = m(\mathbb{R}_+ \times \mathcal{S}) - \nu(\mathbb{R}_+ \times \mathcal{S})$ is constant.

To illustrate our methodology, we solve Eqn. (3) using a gamma distribution truncated to have no jumps smaller than $\epsilon_H > 0$, with $\nu(du, d\tau) = e^{-u\beta} u^{-1}\, du\, \gamma(d\tau)$ on $u > \epsilon_H$. We choose $\gamma(d\tau)$ to be log-scale uniform on the range of times τ supported by the data (see Davies and Anderssen 1997 for how that range is determined, and Wolpert *et al.* 2003 for implementation details). We then solve Eqn. (2) using a fully-skewed α-stable distribution truncated to have no jumps smaller than $\epsilon_w > 0$, with $\nu(du, dm) = c_\alpha u^{-\alpha-1}\, du\, \gamma(dm)$, with $\gamma(dm)$ taken to be log-scale uniform on the range (m_0, m_1) of scaled molecular weights m. Both ϵ_H and ϵ_w are chosen to ensure that the truncation errors are less than $\frac{1}{2}\%$ under the prior distribution (see Section 4 for details).

3.3 Mixtures of Lévy Distributions

In some applications, we may have $\mathcal{S} \subseteq \mathcal{T} \equiv \mathbb{R}^d$ and there may be reason to expect the uncertain measure Z in Eqn. (1) to be absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, with a (perhaps continuous) density function $Z(t) = Z(dt)/dt$. If $Z(ds)$ is any approximate solution then so is $Z^\circ \equiv \eta * Z$ for any smooth approximate identity η (for example, a Gaussian kernel with suitably high precision), and Z° will have the smooth density $Z^\circ(t) = \eta' * Z$; moreover,

$$\begin{aligned}
& y_i && \equiv \int_{\mathcal{S}} k(x_i, s) Z(ds)\,,\ \ i \in I \\
\approx\ & y_i^\circ && \equiv \int_{\mathcal{T}} k(x_i, t) Z^\circ(dt) \\
& && = \iint_{\mathcal{S}\times\mathcal{T}} k(x_i, t)\eta(t-s) Z(ds) dt \\
& && = \int_{\mathcal{S}} k^\circ(x_i, s) Z(ds)
\end{aligned}$$

where $k^\circ \equiv k * \eta$. Thus we can find a smooth approximate solution, as $Z^\circ = \eta * Z$ for an approximate solution Z to

$$Y_i \sim y_i^\circ = \int_{\mathcal{S}} k^\circ(x_i, s) Z(ds)\,,\ i \in I, \tag{1°}$$

and our approach can accommodate smooth processes of the form $Z(ds) = Z(s)ds$ if required by the application.

4 Computational Framework

We solve Eqn. (1) by using the Metropolis-Hastings variation of the Markov chain Monte Carlo (MCMC) method (see Metropolis *et al.* 1953, Hastings 1970, Tierney 1994, Gilks *et al.* 1996) to construct an ergodic Markov chain of draws Z^t $(t = 0, 1, ...)$ from the posterior distribution of Z, given $\vec{Y}$, and evaluate the means of Z and its functionals from ergodic averages of $\{Z^t\}$.

To implement this approach, we must construct the proposal transition kernel $Q(dZ^*|Z)$ for a transitive Markov chain on the space Θ representing finite sums of point masses on $(\mathbb{R}_+ \times \mathcal{S})$ and evaluate the Hastings acceptance probabilities from the transition probabilities Q, the likelihood function from Eqns. (5, 6), and the prior distribution from Eqn. (11).

4.1 Reversible-Jump MCMC Algorithm

In order to allow for a varying number M^t of mass points, we employ a "reversible jump" MCMC (RJ-MCMC) algorithm (Green 1995). Set $\mathcal{U} = (\epsilon, \infty)$ with $\epsilon \geq 0$ chosen (by Eqn. (8)) so that $\nu\big(\mathcal{U} \times \mathcal{S}\big) < \infty$. Starting from any $Z = Z^\theta$ with $\theta = \big((u_1, s_1), \ldots, (u_M, s_M)\big) \in \Theta$, with each $(u_i, s_i) \in \mathcal{U} \times \mathcal{S}$, we propose three distinct kinds of moves to generate a new $\theta^* \in \Theta$ to determine $Z^* = Z^{\theta^*}$:

ADD Draw a new point (u_{M+1}, s_{M+1}) in $\mathcal{U} \times \mathcal{S}$ from any fixed absolutely continuous distribution with full support. Let θ^* be θ with the new point added, and increment $M \leftarrow M+1$ by one.

MOV Choose $1 \leq j \leq M$ from the discrete uniform distribution and move the j^{th} point $(u_j, s_j) \in (\mathbb{R}_+ \times \mathcal{S})$ to a proposed new point (u_j^*, s_j^*). We use a Gaussian random walk on the log scale, with reflecting boundary conditions at $\partial\mathcal{S}$. If $u_j^* \in \mathcal{U}$, we let θ^* be θ, with (u_j^*, s_j^*) replacing (u_j, s_j). M remains unchanged.
If instead $u_j^* \notin \mathcal{U}$, we construct θ^* from θ by deleting (u_j, s_j), decrementing $M \leftarrow M-1$ by one, as in the REM step below.

REM Choose $1 \le j \le M$ from the discrete uniform distribution and remove the j^{th} point (u_j, s_j), decrementing $M \leftarrow M-1$ by one.

Writing $Z = Z^\theta$ and $Z^* = Z^{\theta^*}$, it is tedious but straightforward to evaluate the likelihood $f(\vec{Y} \mid Z)$ for Z (from Eqns. (5, 6)) and the Hastings acceptance probability ratio

$$h(Z^*|Z) = \frac{\pi(dZ^*)\ f(\vec{Y} \mid Z^*)\ Q(dZ|Z^*)}{\pi(dZ)\ f(\vec{Y} \mid Z)\ Q(dZ^*|Z)}.$$

In summary, our algorithm proceeds in the following fashion:

0. Select burn-in period $T_B \ge 0$, run length $T_R \ge 1$, and thinning rate $\Delta \ge 1$. Set $t \leftarrow -T_B$ and initialize Z^t;
1. Choose (randomly) a move of type ADD, MOV, or REM, and draw a proposed new Z^* from the corresponding conditional distribution;
2. With probability $1 \wedge h(Z^*|Z^t)$ accept the proposal and set $Z^{t+1} \leftarrow Z^*$; otherwise set $Z^{t+1} \leftarrow Z^t$;
3. If $0 \le t$ and $t \equiv 0 \pmod{\Delta}$, save state for later analysis.
4. Increment $t \leftarrow t+1$. If $t \le T_R$, return to Step 1 above.

Following this we have $1+\lfloor T_R/\Delta \rfloor$ saved states from which we may calculate ergodic averages to approximate the posterior means of any functions of Z, such as indicator functions $1_A(Z)$ (whose posterior expectation is the probability that Z lies in a set A) or Z itself (whose posterior mean is the point estimate of Z).

4.2 Relation to the Inverse Lévy Measure Algorithm

The possibility that the number of mass points $M^t = M(\theta^t)$ for the random measure Z will change over the simulation run makes this an RJ-MCMC scheme. In (Wolpert and Ickstadt 1998a) we introduced a related computational approach, based on the Inverse Lévy Measure (ILM) algorithm, in which data augmentation is employed to achieve a conjugate model that allows the entire process $Z^t(\cdot)$ to be drawn from its complete conditional distribution at each time step in the Gibbs Sampling variation of MCMC (Gelfand and Smith 1990). The present (RJ-MCMC) approach is more widely applicable, suitable for any Lévy random field (*i.e.*, without a need for conjugacy), but the ILM approach is faster when it is available.

4.3 Infinite Lévy Measure Considerations

When the Lévy measure is infinite, $\nu(\mathbb{R}_+ \times \mathcal{S}) = \infty$, the representation of $Z(ds)$ in Eqn. (7) would (almost surely) include infinitely many terms. While we cannot represent them all in the finite memory of a computer, this algorithm does represent the exact posterior distribution of all the mass in $\mathcal{U} \times \mathcal{S}$. The image underlying our algorithm is that of infinitely many particles (u_j^t, s_j^t) undergoing independent diffusions in $\mathbb{R}_+ \times \mathcal{S}$. We model explicitly only the finitely-many particles in $\mathcal{U} \times \mathcal{S}$, and only at intermittent times t; at equilibrium, the number of particles that drift across the boundary $\partial\mathcal{U}$ (*i.e.*, fall below $u_j < \epsilon$) and disappear from $\mathcal{U} \times \mathcal{S}$ in any time interval is just offset by the number that re-enter from outside that window, modeled in our algorithm by the re-entry distribution. We compute and bound the prior mass outside $\mathcal{U} \times \mathcal{S}$, choosing ϵ to ensure that it is miniscule under the *prior* distribution, and we employ a sensitivity analysis in ϵ to confirm that the omitted *posterior* mass remains negligible.

5 Results

Using Bayesian nonparametric statistical methods we solve two related Fredholm equations,

$$G(t) = G_N^0 \left\{ \int_0^\infty e^{-t/r\tilde{\tau}(m)}\, m^{-1}\, w(dm) \right\}^r \tag{2}$$

for the molecular weight distribution $w(dm)$ and

$$\left[G'(\omega),\, G''(\omega)\right] = \int_0^\infty \frac{[\omega^2\tau^2,\, \omega\tau]}{1+\omega^2\tau^2}\, \tau^{-1} H(d\tau) \tag{3}$$

for the spectral distribution measure $H(d\tau)$, coupled by the relation $G(t) = G_N^0 \int_0^\infty e^{-t/\tau}\, \tau^{-1}\, H(d\tau)$, based on values of $\left[G'(\omega_i),\, G''(\omega_i)\right]$ reported in Berger (1988). We use lognormal measurement-error models, with an α-stable Lévy prior distribution for $w(dm)$ and a gamma process Lévy prior distribution for $H(d\tau)$.

5.1 Molecular Weight Distribution

We model the molecular weight distribution $w(dm)$ as the normalized ratio $w(dm) = Z(dm)/Z(\mathbb{R}_+)$ for a fully-skewed α-stable random field $Z(dm) \sim \mathsf{St}\big(\alpha,\, 1,\, \gamma(dm),\, \tan\frac{\pi\alpha}{2}\gamma(dm)\big)$, with an intermediate value of $\alpha = 0.40$ and with $\gamma(dm)$ chosen to be uniform on an interval $(m_0, m_1) \subset \mathbb{R}_+$ of standardized molecular weights. We employ a Pareto $\mathsf{Pa}(\alpha, \epsilon)$ re-entry distribution (this is the exact distribution of our truncated jump sizes $u_j > \epsilon$ under the α-stable random field prior distribution).

Fig. (1) illustrates the posterior of the molecular weight cumulative distribution $w(0, m]$, showing its posterior mean as a thick solid curve, along with the quartiles (25%, 50% and 75% percentiles) as thinner curves and faint curves for replicates from the posterior distribution, with the prior mean distribution shown as a dotted line for contrast.

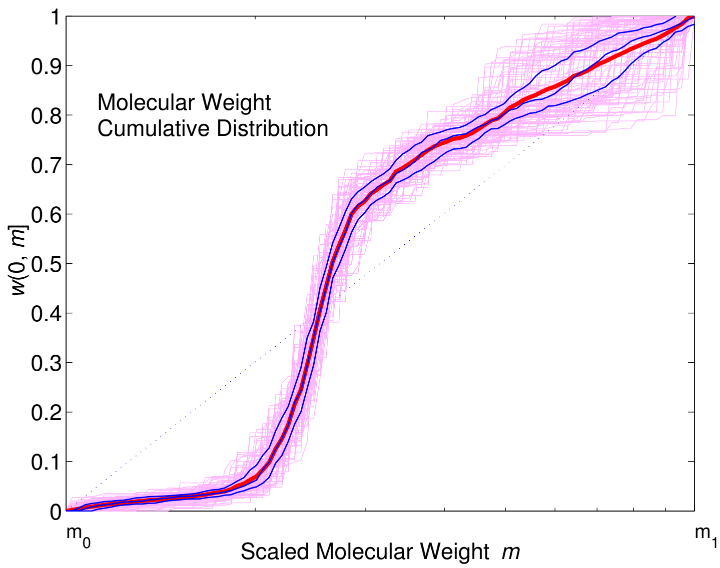

Figure 1: Cumulative Molecular Weight Distribution $w(0, m]$: Posterior mean distribution (thick solid), quartiles (thin solid), 100 replicates from posterior distribution (faint solid), and prior mean distribution (dotted).

Fig. (2) shows prior and posterior mean estimates of the molecular weight density function $w(dm)/dm$. The distribution appears to be unimodal, with some skewness to the right; these qualitative features of the inverse problem solution are of interest to rheologists. The MCMC chain mixes well; convergence was attained after only a few tens of thousands of replicates, and replicates separated by more than a few thousand steps appear to be nearly independent. We treat the first 100,000 steps as a burn-in period.

Fig. (3) illustrates the "input" for our molecular weight reconstruction, $G(t) = G_N^0 \int_0^\infty e^{-t/\tau}\, \tau^{-1}\, H(d\tau)$, based on the reconstruction of $H(d\tau)$ from

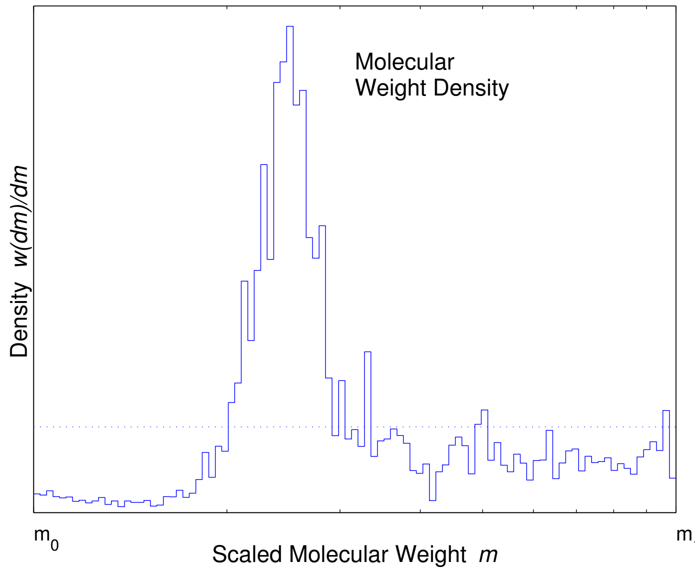

Figure 2: Posterior (solid) and prior (dashed) Molecular Weight Density $w(dm)/dm$.

the solution of Eqn. (3) described below in Section 5.2. The thick solid curve is the posterior mean estimate of $G(t)$, while the dashed curves are samples from the posterior distribution of $G(t)$ based on 100 evenly-spaced samples from a stream of one million simulations of $H(d\tau)$. Model fit is exceptionally good, leaving very little uncertainty about $G(t)$.

5.2 Spectral Density Function

Fig. (4) shows the posterior mean spectral density function $H'(\tau)$ overlaid with the relaxation modulus $G(t) = G_N^0 \int_0^\infty \exp(-t/\tau)\, H(d\tau)/\tau$ from the same reconstruction. The density $H'(t)$ appears to be unimodal and symmetric on a log-log scale.

Fig. (5) shows the entire curves $[G'(\omega), G''(\omega)]$ (solid lines are posterior means, dashed lines are multiple draws from the posterior distributions intended to illustrate the posterior uncertainty) reconstructed by solving Eqn. (4) for $H(d\tau)$, along with the raw data in the form of measured values of the storage and loss moduli $G'(\omega_i)$ and $G''(\omega_i)$ (as $\times$ and $+$ symbols, respectively) at seventeen frequencies ω_i from a frequency sweep experiment reported by Berger (1988). The method seems to work very well, as demonstrated by the close model fit. The details of this reconstruction of $H(d\tau)$

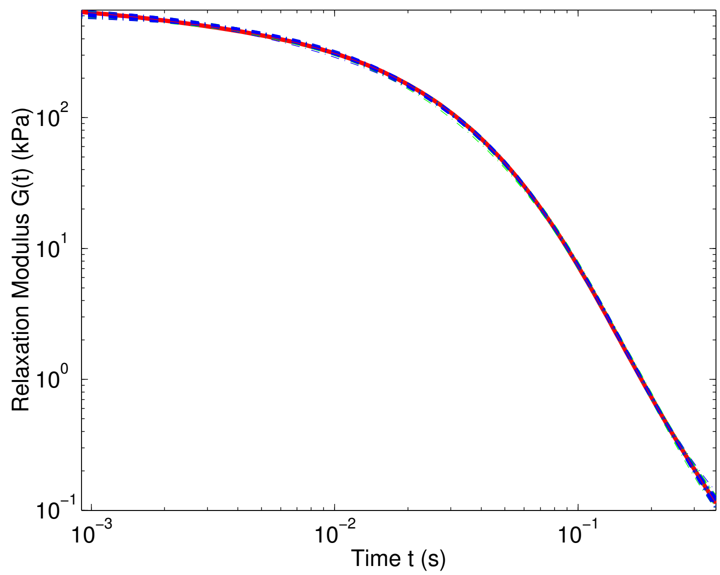

Figure 3: Relaxation Modulus $G(t)$: Posterior mean (thick solid curve) and 100 replicates from posterior distribution (dash-dot curves, faint for first 50 and solid for last 50) from Eqn. (2).

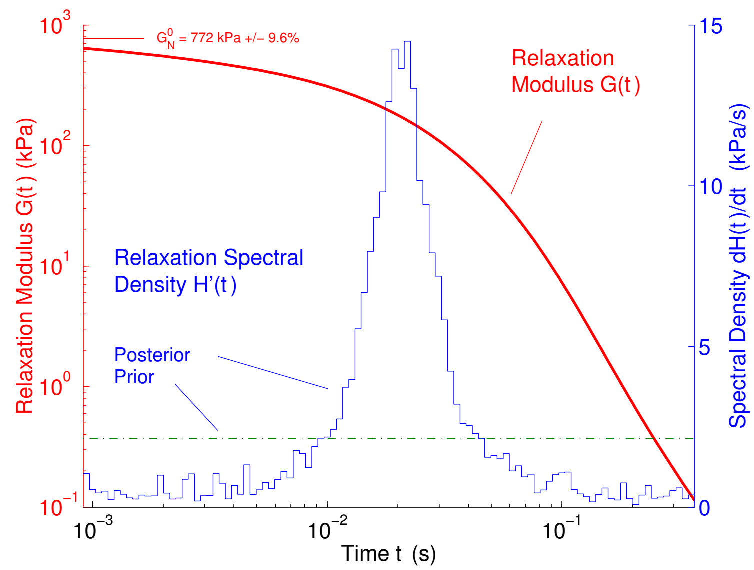

Figure 4: Posterior Mean Relaxation Modulus $G(t)$ and Spectral Density Function $H'(t)$.

are reported in (Wolpert *et al.* 2003).

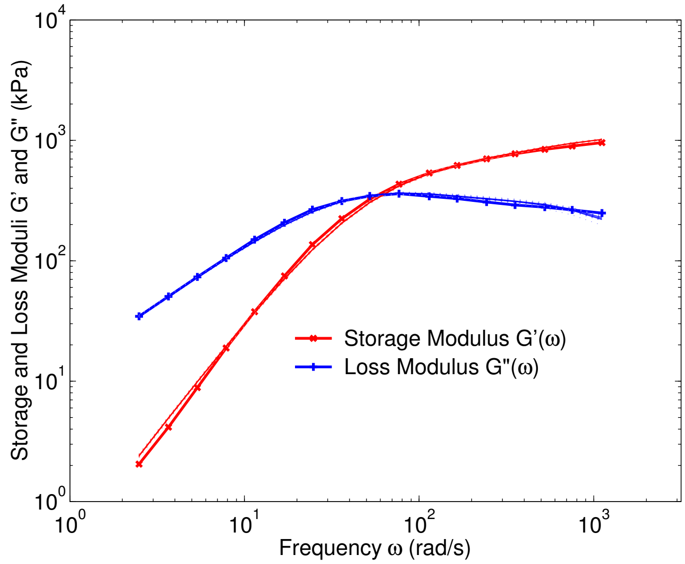

Figure 5: Observed and Model-Fit Storage and Loss Moduli $G'(\omega)$ and $G''(\omega)$.

6 Discussion and Conclusions

We have illustrated our method for posing inverse problems as nonparametric Bayesian estimation problems with Lévy random fields as prior distributions, and of solving them using RJ-MCMC simulation techniques based on the ILM algorithm, in two related applications: the nonlinear problem of inferring the molecular weight distribution measure $w(dm)$ from the relaxation modulus $G(t)$ through Eqn. (2), and the linear problem of inferring the spectral relaxation measure $H(d\tau)$ (and, from it, $G(t)$) from frequency-sweep data, solving Eqn. (3) for $H(d\tau)$.

Our illustrations entailed the simulation of random fields whose prior distributions are Lévy with infinite Lévy measures, and hence which (under both prior and posterior) have infinitely-many random points of support. By including only jumps larger than a threshold $\epsilon > 0$ we are able to approximate the full Lévy distribution arbitrarily well with finitely many $M < \infty$ mass points, omitting an arbitrarily small fraction ($\frac{1}{2}$% in our illustration) of the prior mass. In our simulation 90% of the numbers M^t of mass points

(u_j, s_j) were in the range $[27, 53]$, with a median value of 35.

We chose random-walk step sizes (0.25 on a log scale) and move probabilities (0.07 each for ADD and REM moves, 0.86 for MOV moves) to ensure approximately a 40% acceptance rate for each form of Metropolis step. Our implementation in the MatLab (The MathWorks 2002) computing environment executes about two hundred iterations per second on a contemporary (approx. 3GHz dual Xeon) computer, completing one million replications in about eighty minutes.

We have shown how the technique can be used to generate random measure estimates of Fredholm solutions with whatever degree of smoothness is required, ranging from measures with smooth density functions (see Section 3.3) to those with finite Lévy measures and hence a small number of mass points (appropriate if it were known that the molecular weight distribution must be concentrated on only a few points). We have illustrated how to select among different measurement-error models.

In future work in this application area it will be important to treat the "reptation index" r as uncertain, and to use the data to help discover what values the data support. We would also like to explore the usefulness of this approach to expressing uncertainty about the solutions to inverse problems in other situations, such as inverting Laplace and Fourier transforms, and in other application areas, such as tomography.

Acknowledgments

The authors would like to thank Martin Hansen for introducing us to rheology and Robert Anderssen for many helpful suggestions. This work was supported by the Deutsche Forschungsgemeinschaft collaborative research center SFB 475, U.S. Environmental Protection Agency grants CR-822047-01-0 and R828686-01-0, and U.S. National Science Foundation grants DMS-9626829 and DMS-0112069. It has not been reviewed by the EPA or NSF and no official endorsement should be inferred.

References

Anderssen, R. S. and Loy, R. J. (2001). On the scaling of molecular weight distribution functionals. *Journal of Rheology* **45**, 891–901.

Anderssen, R. S. and Mead, D. W. (1998). Theoretical derivation of molecular weight scaling for rheological parameters. *Journal of Non-Newtonian Fluid Mechanics* **76**, 299–306.

Anderssen, R. S. and Westcott, M. (2000). The molecular weight distribution problem and reptation mixing rules. *ANZIAM Journal* **42**, 26–40.

Berger, L. (1988). *Untersuchung zum rheologischen Verhalten von Polybutadienen mit bimodaler Molmassenverteilung*. Ph.D. dissertation, Eidgenössische Technische Hochschule (ETH), Zürich, CH.

Boltzmann, L. (1876). Zur Theorie der elastischen Nachwirkung. *Annals of Physical Chemistry* **7**, 624–654.

Cheng, R. C. H. and Liu, W. B. (1997). A continuous representation of the family of stable law distributions. *Journal of the Royal Statistical Society, Ser.* B *(Statistical Methodology)* **59**, 137–145.

Davies, A. R. and Anderssen, R. S. (1997). Sampling localization in determining the relaxation spectrum. *Journal of Non-Newtonian Fluid Mechanics* **73**, 163–179.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds.) (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, New York, NY.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, DE.

Maier, D., Eckstein, A., Friedrich, C. and Honerkamp, J. (1998). Evaluation of models combining rheological data with the molecular weight distribution. *Journal of Rheology* **42**, 1153–1173.

The MathWorks, Inc. (2002). MatLab. Version 6.5, Release 13.

Metropolis, N. C., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091.

Mohammad-Djafari, A. and Idier, J. (1996). A scale-invariant Bayesian method to solve linear inverse problems. *Maximum Entropy and Bayesian Methods*, G. R. Heidbreder, ed. Kluwer Academic Publishers, pp. 121–134.

Mosegaard, K. and Sambridge, M. (2002). Monte Carlo analysis of inverse problems. *Inverse Problems* **18**, R29–R54.

Mosegaard, K. and Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research* **100**, 12431–12447.

Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov Processes, and Martingales*, volume 2. Cambridge University Press, Cambridge, UK, second edition.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, volume 1 of *Stochastic Modeling Series*. Chapman & Hall, New York, NY.

Sato, K.-i. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, UK.

Scales, J. A. and Tenorio, L. (2001). Prior information and uncertainty in inverse problems. *Geophysics* **66**, 389–397.

Thimm, W., Friedrich, C., Marth, M. and Honerkamp, J. (2000). On the Rouse spectrum and the determination of the molecular weight distribution from rheological data. *Journal of Rheology* **44**, 429–438.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**, 1701–1762.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Doklady* **4**, 1035–1038.

Wolpert, R. L. and Ickstadt, K. (1998a). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.

Wolpert, R. L. and Ickstadt, K. (1998b). Simulation of Lévy random fields. *Practical Nonparametric and Semiparametric Bayesian Statistics*, D. Dey, P. Müller and D. Sinha, eds. Springer-Verlag, New York, NY, volume 133 of *Lecture Notes in Statistics*, pp. 227–242.

Wolpert, R. L., Ickstadt, K. and Hansen, M. B. (2003). A nonparametric Bayesian approach to inverse problems (with discussion). *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds. Oxford University Press, Oxford, UK, pp. 403–418.