A Metropolis-Hastings-within-Gibbs Sampler for Nonlinear Hierarchical-Bayesian Inverse Problems

Johnathan M. Bardsley† and Tiangang Cui♯

Abstract We investigate the use of randomize-then-optimize (RTO) [3] as a proposal for Metropolis-Hastings (MH) – yielding the so-called RTO-MH method – for sampling from posterior distributions arising in nonlinear, hierarchical-Bayesian inverse problems. Specifically, we extend the hierarchical Gibbs sampler of [2] for linear inverse problems to nonlinear inverse problems by embedding RTO-MH within the hierarchical Gibbs sampler. We test the method on a nonlinear inverse problem arising in differential equations.

1 Introduction

In this paper, we focus on inverse problems of the form

$$\mathbf{y} = \mathbf{F}(\mathbf{u}) + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I}), \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the observed data, $\mathbf{u} \in \mathbb{R}^n$ is the unknown parameter, $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is the forward operator, and λ is known as the measurement precision parameter. The likelihood function then has the form

$$p(\mathbf{y}|\mathbf{u}, \lambda) = (2\pi)^{-\frac{m}{2}} \lambda^{m/2} \exp\left(-\frac{\lambda}{2} \|\mathbf{F}(\mathbf{u}) - \mathbf{y}\|^2\right). \tag{2}$$

Next, we assume that the prior is a zero-mean Gaussian random vector, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, (\delta\mathbf{L})^{-1})$, which has distribution

† Department of Mathematical Sciences, University of Montana, Missoula, MT 59812. e-mail: bardsleyj@mso.umt.edu
♯ School of Mathematical Science, Monash University, VIC 3800, Australia. e-mail: tiangang.cui@monash.edu

$$p(\mathbf{u}|\delta) = (2\pi)^{-\frac{n}{2}}\delta^{\bar{n}/2}\exp\left(-\frac{\delta}{2}\mathbf{u}^T\mathbf{L}\mathbf{u}\right), \tag{3}$$

where $\mathbf{L}$ is defined via a Gaussian Markov random field (GMRF) [1], and $\bar{n}$ is the rank of $\mathbf{L}$. In the one-dimensional numerical example considered at the end of the paper, we choose $\mathbf{L}$ to be a discretization of the negative-Laplacian operator. The hyper-parameter δ, which is known as the prior precision parameter, provides the relative weight given to the prior as compared to the likelihood function.

In keeping with the Bayesian paradigm, we assume hyper-priors $p(\lambda)$ and $p(\delta)$ on λ and δ, respectively. A standard choice in the linear Gaussian case is to choose Gamma hyper-priors:

$$p(\lambda) \propto \lambda^{\alpha_\lambda - 1}\exp(-\beta_\lambda\lambda), \tag{4}$$
$$p(\delta) \propto \delta^{\alpha_\delta - 1}\exp(-\beta_\delta\delta). \tag{5}$$

This is due to the fact that the conditional densities for λ and δ are then also Gamma-distributed (a property known as conjugacy), and hence are easy to sample from. We choose hyper-parameters $\alpha_\lambda = \alpha_\delta = 1$ and $\beta_\lambda = \beta_\delta = 10^{-4}$, making the hyper-priors exponentially distributed with small decay parameters β_λ and β_δ. In the test cases we have considered, these hyper-priors work well, though they should be chosen carefully in a particular situation. Specifically, it is important that they are chosen to be relatively flat over the regions of high probability for λ and δ defined by the posterior density function, so that they are not overly informative.

Taking into account the likelihood, the prior, and the hyper-priors, the posterior probability density function over all of the unknown parameters is given, by Bayes' law, as

$$\begin{aligned} &p(\mathbf{u},\lambda,\delta|\mathbf{y}) \\ &= p(\mathbf{y}|\mathbf{u},\lambda)p(\mathbf{u}|\delta)p(\lambda)p(\delta)\big/\, p(\mathbf{y}) \\ &\propto \lambda^{m/2+\alpha_\lambda-1}\delta^{\bar{n}/2+\alpha_\delta-1}\exp\left(-\frac{\lambda}{2}\|\mathbf{F}(\mathbf{u})-\mathbf{y}\|^2 - \frac{\delta}{2}\mathbf{u}^T\mathbf{L}\mathbf{u} - \beta_\lambda\lambda - \beta_\delta\delta\right), \end{aligned} \tag{6}$$

where $p(\mathbf{y})$ is the normalizing constant for the posterior. Our focus in this paper is to develop a Gibbs sampler for sampling from the full posterior (6). For this, we need the full conditionals, which are given by

$$p(\lambda|\mathbf{b},\mathbf{u},\delta) \propto \lambda^{m/2+\alpha_\lambda-1}\exp\left(\left[-\frac{1}{2}\|\mathbf{F}(\mathbf{u})-\mathbf{y}\|^2 - \beta_\lambda\right]\lambda\right), \tag{7}$$
$$p(\delta|\mathbf{y},\mathbf{u},\lambda) \propto \delta^{\bar{n}/2+\alpha_\delta-1}\exp\left(\left[-\frac{1}{2}\mathbf{u}^T\mathbf{L}\mathbf{u} - \beta_\delta\right]\delta\right), \tag{8}$$
$$p(\mathbf{u}|\mathbf{y},\lambda,\delta) \propto \exp\left(-\frac{\lambda}{2}\|\mathbf{F}(\mathbf{u})-\mathbf{y}\|^2 - \frac{\delta}{2}\mathbf{u}^T\mathbf{L}\mathbf{u}\right). \tag{9}$$

The Gamma-hyper priors are conjugate, and hence the conditional densities for λ and δ are also Gamma-distributed:

$$\lambda|\mathbf{u},\delta,\mathbf{b} \sim \Gamma\left(m/2+\alpha_\lambda, \frac{1}{2}\|\mathbf{F}(\mathbf{u})-\mathbf{b}\|^2+\beta_\lambda\right), \tag{10}$$

$$\delta|\mathbf{u},\lambda,\mathbf{b} \sim \Gamma\left(\overline{n}/2+\alpha_\delta, \frac{1}{2}\mathbf{u}^T\mathbf{L}\mathbf{u}+\beta_\delta\right). \tag{11}$$

The distributions (10) and (11) are independent so that $p(\lambda,\delta|\mathbf{b},\mathbf{x}) = p(\lambda|\mathbf{b},\mathbf{x})p(\delta|\mathbf{b},\mathbf{x})$. Hence, computing independent samples from (10) and (11) yields a sample from $p(\lambda,\delta|\mathbf{b},\mathbf{x})$. Moreover, in the linear case, $\mathbf{F}$ is a matrix and the Gaussian prior is also conjugate, leading to a Gaussian conditional (9), which can be equivalently expressed

$$\mathbf{u} \sim \mathscr{N}\left((\lambda\mathbf{F}^T\mathbf{F}+\delta\mathbf{L})^{-1}\lambda\mathbf{F}^T\mathbf{y}, (\lambda\mathbf{F}^T\mathbf{F}+\delta\mathbf{L})^{-1}\right).$$

Taking these observations all together leads to the two-stage Gibbs sampler given next, which is also presented in [1, 2].

The Hierarchical Gibbs Sampler, Linear Case

0. Initialize (λ_0,δ_0), $\mathbf{u}^0 = (\lambda_0\mathbf{F}^T\mathbf{F}+\delta_0\mathbf{L})^{-1}\lambda_0\mathbf{F}^T\mathbf{y}$, set $k=1$, define k_{total}.
1. Compute $(\lambda_k,\delta_k) \sim p(\lambda,\delta|\mathbf{y},\mathbf{u}^{k-1})$ as follows.
   a. Compute $\lambda_k \sim \Gamma\left(m/2+\alpha_\lambda, \frac{1}{2}\|\mathbf{F}\mathbf{u}^{k-1}-\mathbf{y}\|^2+\beta_\lambda\right)$.
   b. Compute $\delta_k \sim \Gamma\left(\overline{n}/2+\alpha_\delta, \frac{1}{2}(\mathbf{u}^{k-1})^T\mathbf{L}\mathbf{u}^{k-1}+\beta_\delta\right)$.
2. Compute $\mathbf{u}^k \sim \mathscr{N}\left((\lambda_k\mathbf{F}^T\mathbf{F}+\delta_k\mathbf{L})^{-1}\lambda_k\mathbf{F}^T\mathbf{y}, (\lambda_k\mathbf{F}^T\mathbf{F}+\delta_k\mathbf{L})^{-1}\right)$.
3. If $k=k_{\text{total}}$ stop, otherwise, set $k=k+1$ and return to Step 1.

When $\mathbf{F}$ is nonlinear, the conditional density $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$, defined in (9), is no longer Gaussian and cannot be sample from it directly. To overcome this, we embed a Metropolis-Hastings (MH) step within step 2 of hierarchical Gibbs, as advocated in [4, Algorithm A.43]. For the MH proposal, we use the randomize-then-optimize (RTO) [3], and thus we begin in Section 2 by describing the RTO proposal. In Section 3, we describe RTO-MH and its embedding within hierarchical Gibbs for sampling from the full posterior (6). Finally, we use RTO-MH-within-hierarchical Gibbs to sample from (6) in a specific nonlinear inverse problem arising in differential equations. Concluding remarks are provided in Section 5.

2 The Randomize-Then-Optimize Proposal Density

We first define the augmented forward model and observation taking the form

$$\mathbf{F}_{\lambda,\delta}(\mathbf{u}) \stackrel{\text{def}}{=} \begin{bmatrix} \lambda^{1/2}\mathbf{F}(\mathbf{u}) \\ \delta^{1/2}\mathbf{L}^{1/2}\mathbf{x} \end{bmatrix} \quad \text{and} \quad \mathbf{y}_{\lambda,\delta} \stackrel{\text{def}}{=} \begin{bmatrix} \lambda^{1/2}\mathbf{y} \\ \mathbf{0} \end{bmatrix}.$$

For motivation, note that in the linear case, $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$ is Gaussian and can be sampled by solving the stochastic least squares problem

$$\mathbf{u}|\mathbf{y},\lambda,\delta = \arg\min_{\psi} \|\mathbf{F}_{\lambda,\delta}\psi - (\mathbf{y}_{\lambda,\delta}+\varepsilon)\|^2, \quad \varepsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I}). \tag{12}$$

This follows from the fact that if $\mathbf{F}_{\lambda,\delta} = \mathbf{Q}_{\lambda,\delta}\mathbf{R}_{\lambda,\delta}$ is the thin (or condensed) **QR**-factorization of $\mathbf{F}_{\lambda,\delta}$, and $\mathbf{F}_{\lambda,\delta}$ has full column rank, then $\mathbf{Q}_{\lambda,\delta} \in \mathbb{R}^{(M+N)\times N}$ has orthonormal columns spanning the column space of $\mathbf{F}_{\lambda,\delta}$; $\mathbf{R}_{\lambda,\delta} \in \mathbb{R}^{N\times N}$ is upper-triangular and invertible; and the solution of (12) is unique and can be expressed

$$\mathbf{Q}_{\lambda,\delta}^T\mathbf{F}_{\lambda,\delta}(\mathbf{u}|\mathbf{b},\lambda,\delta) = \mathbf{Q}_{\lambda,\delta}^T(\mathbf{y}_{\lambda,\delta}+\varepsilon), \quad \varepsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I}). \tag{13}$$

Note that in the linear case $\mathbf{Q}_{\lambda,\delta}^T\mathbf{F}_{\lambda,\delta} = \mathbf{R}_{\lambda,\delta}$, and it follows that (13) yields samples from $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$.

In the nonlinear case, equation (13) can still be used, but the resulting samples do not have distribution $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$. To derive the form of the distribution, we first define

$$\mathbf{r}_{\lambda,\delta}(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{F}_{\lambda,\delta}(\mathbf{u}) - \mathbf{y}_{\lambda,\delta}$$

and denote the Jacobian of $\mathbf{F}_{\lambda,\delta}$, evaluated at $\mathbf{u}$, by $\mathbf{J}_{\lambda,\delta}(\mathbf{u})$. Then, provided $\mathbf{Q}_{\lambda,\delta}^T\mathbf{F}_{\lambda,\delta}$ is a one-to-one function with continuous first partial derivatives, and its Jacobian, $\mathbf{Q}_{\lambda,\delta}^T\mathbf{J}_{\lambda,\delta}$, is invertible, the probability density function for $\mathbf{u}|\mathbf{b},\lambda,\delta$ defined by (13) is

$$\begin{aligned} p_{\text{RTO}}(\mathbf{u}|\mathbf{b},\lambda,\delta) &\propto \left|\det(\mathbf{Q}_{\lambda,\delta}^T\mathbf{J}_{\lambda,\delta}(\mathbf{u}))\right| \exp\left(-\frac{1}{2}\|\mathbf{Q}_{\lambda,\delta}^T\mathbf{r}_{\lambda,\delta}(\mathbf{u})\|^2\right) \\ &= c_{\lambda,\delta}(\mathbf{x})p(\mathbf{u}|\mathbf{b},\lambda,\delta), \end{aligned} \tag{14}$$

where

$$c_{\lambda,\delta}(\mathbf{u}) = \left|\det(\mathbf{Q}_{\lambda,\delta}^T\mathbf{J}_{\lambda,\delta}(\mathbf{u}))\right| \exp\left(\frac{1}{2}\|\mathbf{r}_{\lambda,\delta}(\mathbf{u})\|^2 - \frac{1}{2}\|\mathbf{Q}_{\lambda,\delta}^T\mathbf{r}_{\lambda,\delta}(\mathbf{u})\|^2\right). \tag{15}$$

There is flexibility in how to choose $\mathbf{Q}_{\lambda,\delta} \in \mathbb{R}^{(M+N)\times N}$, though $\mathbf{Q}_{\lambda,\delta}^T\mathbf{F}_{\lambda,\delta}$ must satisfy the conditions mentioned in the previous sentence. In our implementations of RTO, we have used $\mathbf{Q}_{\lambda,\delta}$ from the thin **QR**-factorization $\mathbf{J}_{\lambda,\delta}(\mathbf{u}_{\lambda,\delta}) = \mathbf{Q}_{\lambda,\delta}\mathbf{R}_{\lambda,\delta}$, where $\mathbf{u}_{\lambda,\delta}$ is the MAP estimator, i.e., $\mathbf{u}_{\lambda,\delta} = \arg\min_{\mathbf{u}} \|\mathbf{F}_{\lambda,\delta}(\mathbf{u}) - \mathbf{y}_{\lambda,\delta}\|^2$.

In practice, we compute samples from (13) by solving the stochastic optimization problem

$$\mathbf{u}^* = \arg\min_{\psi} \frac{1}{2}\|\mathbf{Q}_{\lambda,\delta}^T(\mathbf{F}_{\lambda,\delta}(\psi) - (\mathbf{y}_{\lambda,\delta}+\varepsilon^*))\|^2, \quad \varepsilon^* \sim \mathcal{N}(\mathbf{0},\mathbf{I}). \tag{16}$$

The name randomize-then-optimize stems from (16), where $\mathbf{y}_{\lambda,\delta}$ is first 'randomized', by adding ε^*, and then 'optimized', by solving (16). Finally, we note that if the cost function minimum in (16) is greater than zero, (13) has no solution, and we must discard the corresponding sample. In practice, we discard solutions $\mathbf{x}^*$ of (16) with a cost function minimum greater than $\eta = 10^{-8}$, though we have found this to occur very rarely in practice.

3 RTO-Metropolis-Hastings and its Embedding within Hiererichical Gibbs

Although RTO does not yield samples from $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$ for nonlinear problems, it can be used as a proposal for MH. At step k of the MH algorithm, given the current sample $\mathbf{u}^{k-1}$, one can use (16) to compute $\mathbf{u}^* \sim p_{\text{RTO}}(\mathbf{u}|\mathbf{b},\lambda,\delta)$ and then set $\mathbf{u}^k = \mathbf{u}^{k-1}$ with probability

$$\begin{aligned} r_{\lambda,\delta} &= \min\left(1, \frac{p(\mathbf{u}^*|\mathbf{y},\lambda,\delta)p_{\text{RTO}}(\mathbf{u}^{k-1}|\mathbf{y},\lambda,\delta)}{p(\mathbf{u}^{k-1}|\mathbf{y},\lambda,\delta)p_{\text{RTO}}(\mathbf{u}^*|\mathbf{y},\lambda,\delta)}\right) \\ &= \min\left(1, \frac{p(\mathbf{u}^*|\mathbf{y},\lambda,\delta)c_{\lambda,\delta}(\mathbf{u}^{k-1})p(\mathbf{u}^{k-1}|\mathbf{y},\lambda,\delta)}{p(\mathbf{u}^{k-1}|\mathbf{y},\lambda,\delta)c_{\lambda,\delta}(\mathbf{u}^*)p(\mathbf{u}^*|\mathbf{y},\lambda,\delta)}\right) \\ &= \min\left(1, \frac{c_{\lambda,\delta}(\mathbf{u}^{k-1})}{c_{\lambda,\delta}(\mathbf{u}^*)}\right). \end{aligned} \tag{17}$$

Note that it is often advantageous, for numerical reasons, to replace the ratio in (17) by the equivalent expression

$$c_{\lambda,\delta}(\mathbf{u}^{k-1})/c_{\lambda,\delta}(\mathbf{u}^*) = \exp\Big(\ln c_{\lambda,\delta}(\mathbf{u}^{k-1}) - \ln c_{\lambda,\delta}(\mathbf{u}^*)\Big),$$

where

$$\ln c_{\lambda,\delta}(\mathbf{u}) \simeq \ln\left|\mathbf{Q}_{\lambda,\delta}^T \mathbf{J}_{\lambda,\delta}(\mathbf{u})\right| + \frac{1}{2}\|\mathbf{r}_{\lambda,\delta}(\mathbf{x})\|^2 - \frac{1}{2}\|\mathbf{Q}_{\lambda,\delta}^T \mathbf{r}_{\lambda,\delta}(\mathbf{x})\|^2,$$

and '$\simeq$' denotes 'equal up to an additive, unimportant constant.'

The RTO-MH Algorithm

1. Choose initial vector $\mathbf{u}^0$, parameter $0 < \eta \ll 0$, and samples N. Set $k = 1$.
2. Compute $\mathbf{u}^* \sim p_{\text{RTO}}(\mathbf{u}|\mathbf{y},\lambda,\delta)$ by solving (16) for a fixed realization $\varepsilon^* \sim \mathcal{N}(\mathbf{0},\mathbf{I})$. If
$$\|\mathbf{Q}_{\lambda,\delta}^T(\mathbf{F}_{\lambda,\delta}(\mathbf{u}^*) - (\mathbf{y}_{\lambda,\delta} + \varepsilon^*))\|^2 > \eta,$$
then repeat step 2.
3. Set $\mathbf{u}^k = \mathbf{u}^*$ with probability $r_{\lambda,\delta}$ defined by (17). Else, set $\mathbf{u}^k = \mathbf{u}^{k-1}$.
4. If $k < N$, set $k = k+1$ and return to Step 2, otherwise stop.

The proposed sample $\mathbf{u}^*$ is independent of $\mathbf{u}^{k-1}$, making RTO-MH an independence MH method. Thus, we can apply [4, Theorem 7.8] to obtain the result that RTO-MH will produce a uniformly ergodic chain that converges in distribution to $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$ provided there exists $M > 0$ such that $p(\mathbf{u}|\mathbf{y},\lambda,\delta) \leq M \cdot p_{\text{RTO}}(\mathbf{u}|\mathbf{y},\lambda,\delta)$, for all $\mathbf{u} \in \mathbb{R}^N$. Given (14), this inequality holds if and only if $c_{\lambda,\delta}(\mathbf{u})$, defined by (15), is bounded away from zero for all $\mathbf{u}$.

3.1 RTO-MH-within-Hierarchical Gibbs

In the hierarchical setting, we embed a single RTO-MH step within the hierarchical Gibbs sampler, to obtain the following MCMC method.

RTO-MH-within-Hierarchical Gibbs

0. Initialize (λ_0, δ_0), set $k = 1$, define k_{total}, and set

$$\mathbf{u}^0 = \arg\min_{\mathbf{u}} \|\mathbf{F}_{\lambda_0,\delta_0}(\mathbf{u}) - \mathbf{y}_{\lambda_0,\delta_0}\|^2.$$

1. Simulate $(\lambda_k, \delta_k) \sim p(\lambda, \delta | \mathbf{y}, \mathbf{u}^{k-1})$ as follows.
   a. Compute $\lambda_k \sim \Gamma\left(m/2 + \alpha_\lambda,\ \frac{1}{2}\|\mathbf{F}(\mathbf{x}^{k-1}) - \mathbf{y}\|^2 + \beta_\lambda\right)$.
   b. Compute $\delta_k \sim \Gamma\left(\bar{n}/2 + \alpha_\delta,\ \frac{1}{2}(\mathbf{x}^{k-1})^T \mathbf{L}\mathbf{x}^{k-1} + \beta_\delta\right)$.
2. Simulate ub^k using RTO as follows.
   a. Compute $\mathbf{u}^* \sim p_{\text{RTO}}(\mathbf{u}|\mathbf{y}, \lambda_k, \delta_k)$ by solving (16) with $(\lambda, \delta) = (\lambda_k, \delta_k)$.
   b. Set $\mathbf{u}^k = \mathbf{u}^*$ with probability r_{λ_k,δ_k} defined by (17), else set $\mathbf{u}^k = \mathbf{u}^{k-1}$.
3. If $k = k_{\text{total}}$ stop, otherwise, set $k = k+1$ and return to Step 1.

In step 2a, note that two optimization problems must be solved. First, the MAP estimator $\mathbf{u}_{\lambda_k,\delta_k}$ is computed; then the **QR**-factorization $\mathbf{J}(\mathbf{u}_{\lambda_k,\delta_k}) = \mathbf{Q}_{\lambda_k,\delta_k}\mathbf{R}_{\lambda_k,\delta_k}$ is computed; and finally, the stochastic optimization problem (16) is solved, with $(\lambda, \delta) = (\lambda_k, \delta_k)$, to obtain the RTO sample $\mathbf{u}^*$. One could take multiple RTO-MH steps in Step 2, within each outer loop, to improve the chances of updating $\mathbf{u}^{k-1}$, but we do not implement that here.

4 Numerical Experiment

To test RTO-MH-within-hierarchical Gibbs, we consider a nonlinear inverse problem from [1, Chapter 6]. The inverse problem is to estimate the diffusion coefficient $u(s)$ from measurements of the solution $x(s)$ of the Poisson equation

$$-\frac{d}{ds}\left(u(s)\frac{dx}{ds}\right) = f(s), \quad 0 < s < 1, \tag{18}$$

with zeros boundary conditions $x(0) = x(1) = 0$. Assuming a uniform mesh on $[0,1]$, after numerical discretization, (18) takes the form

$$\mathbf{B}(\mathbf{u})\mathbf{x} = \mathbf{f}, \qquad \mathbf{B}(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{D}^T \operatorname{diag}(\mathbf{u})\mathbf{D}, \tag{19}$$

where $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{D} \in \mathbb{R}^{n\times n-1}$ is a discrete derivative matrix. To generate data, we compute numerical solutions corrresponding to two discrete Dirac delta forcing functions, $f_1(s)$ and $f_2(s)$, centered at $s = 1/3$ and $s = 2/3$, respectively. After discretization, f_1 and f_2 become $(n-1)\times 1$ Kronecker delta vectors $\mathbf{f}_1$ and $\mathbf{f}_2$, and the measurement model takes the form of (1) with

$$\mathbf{y} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}_{2n-2} \quad \text{and} \quad \mathbf{F}(\mathbf{u}) \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{B}(\mathbf{u})^{-1}\mathbf{f}_1 \\ \mathbf{B}(\mathbf{u})^{-1}\mathbf{f}_2 \end{bmatrix}_{2n-2,}$$

so that $m = 2n-2$. We generate data using (1) with $n = 50$ and $\mathbf{u}_{\text{true}}$ obtained by discretizing

$$u(s) = \min\{1, 1-0.5\,\sin(2\pi(s-0.25))\},$$

and λ^{-1} chosen so that signal-to-noise ratio, $\|\mathbf{F}(\mathbf{u}_{\text{true}})\|/\sqrt{m\lambda^{-1}}$, is 100. The data vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ are plotted in Figure 1(a) together with the noise-free data $\mathbf{B}(\mathbf{x})^{-1}\mathbf{f}_1$ and $\mathbf{B}(\mathbf{x})^{-1}\mathbf{f}_2$.

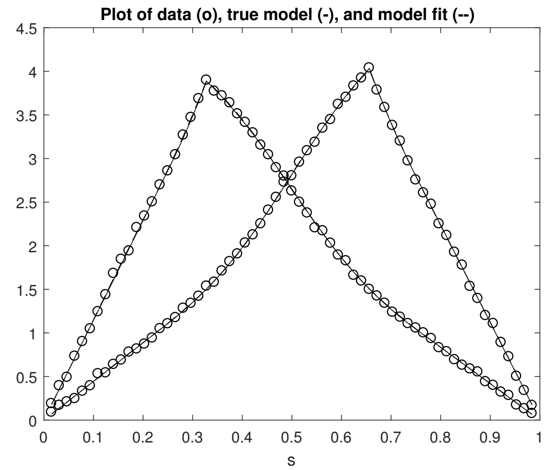
With the measurements in hand, we implement RTO-MH-within-hierarchical Gibbs for sampling from (6). The results are plotted in Figures 1 and 2. In Figure 1(b), we see the sample median together with 95% credibility intervals computed from the $\mathbf{u}$-chain generated by the MCMC method. In Figure 2(a), we plot the individual chains for λ, δ, and a randomly chosen element of the $\mathbf{u}$-chain. And finally, in Figure 2(b), we plot the auto correlation functions and associated integrated autocorrelation times (τ_{int}) for these three parameters [1].

5 Conclusions

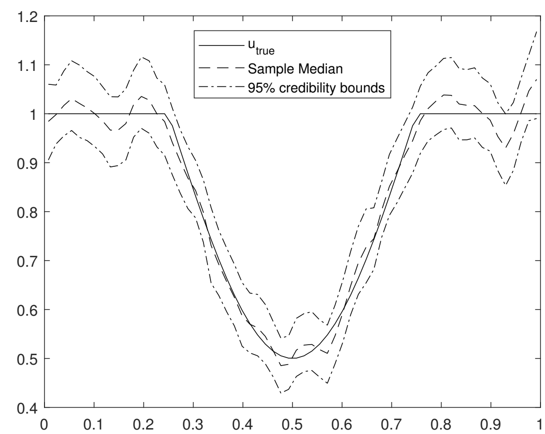
In this paper, we have tackled the problem of sampling from the full posterior (6) when $\mathbf{F}$ is a nonlinear function. To do this, we followed the same approach as the hierarchical Gibbs algorithm of [2], however in that algorithm, $\mathbf{F}$ is linear, the conditional density $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$ is Gaussian, and hence samples from $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$ can be computed by solving a linear system of equations. In the nonlinear case, $p(\mathbf{u}|\mathbf{y},\lambda,\delta)$ is non-Gaussian, but we can use RTO-MH to obtain samples, as described in [3]. We obtain a MH-within-Gibbs method for sampling from (6) by embedding a single RTO-MH step with hierarchical Gibbs. We then tested the method on a nonlinear inverse problem arising in differential equations and found that it worked well.

References

1. J. M. Bardsley, *Computational Uncertainty Quantification for Inverse Problems*, to be published by SIAM.
2. J. M. Bardsley, “MCMC-Based Image Reconstruction with Uncertainty Quantification,” *SIAM Journal on Scientific Computing*, Vol. 34, No. 3, 2012, pp. A1316–A1332.
3. J. M. Bardsley, Antti Solonen, Heikki Haario, and Marko Laine, “Randomize-then-Optimize: a method for sampling from posterior distributions in nonlinear inverse problems,” *SIAM Journal on Scientific Computing*, Vol. 36, No. 4, 2014, pp. A1359-C399.



(a)



(b)

Fig. 1 (a) Plots of the measured data $\mathbf{b}_1$ and $\mathbf{b}_2$, the true state $\mathbf{u}$, and the model fits. (b) Plots of the true diffusion coefficient $\mathbf{x}$ together with the RTO-MH sample median and the element-wise 95% credibility bounds.

4. C. P. Robert and G. Casella, *Monte Carlo Statistical methods*, Springer, 2004.

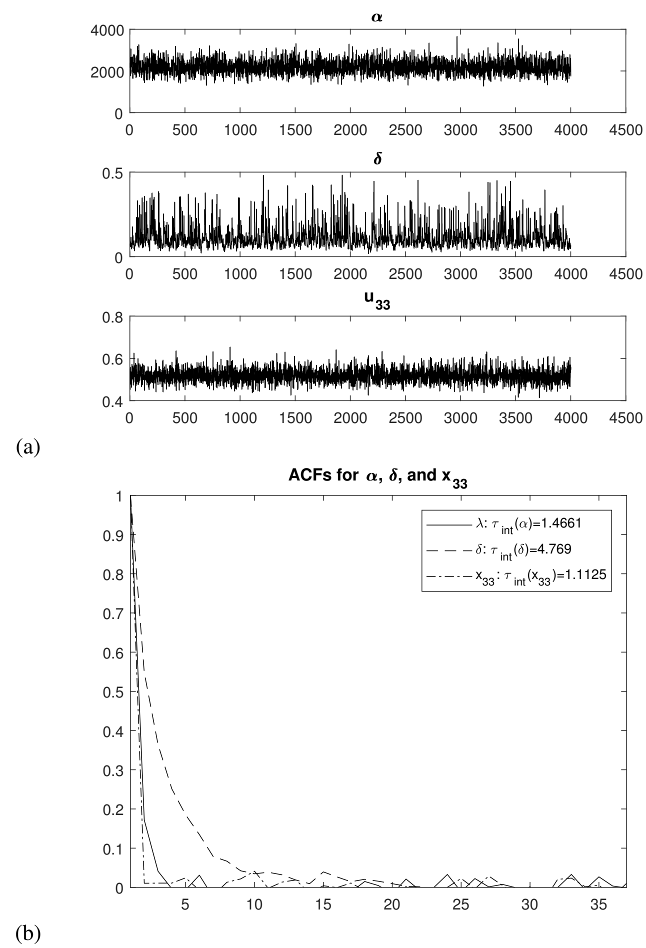


Fig. 2 (a) Plots of the chains for three randomly selected elements of **x**. (b) Autocorrelation times associated with these chains.