

Hybrid Samplers for Ill-Posed Inverse Problems

RADU HERBEI

Department of Statistics, The Ohio State University

IAN W. McKEAGUE

Department of Biostatistics, Columbia University

ABSTRACT. In the Bayesian approach to ill-posed inverse problems, regularization is imposed by specifying a prior distribution on the parameters of interest and Markov chain Monte Carlo samplers are used to extract information about its posterior distribution. The aim of this paper is to investigate the convergence properties of the random-scan random-walk Metropolis (RSM) algorithm for posterior distributions in ill-posed inverse problems. We provide an accessible set of sufficient conditions, in terms of the observational model and the prior, to ensure geometric ergodicity of RSM samplers of the posterior distribution. We illustrate how these conditions can be checked in an application to the inversion of oceanographic tracer data.

Key words: advection-diffusion, Bayesian regularization, geometric ergodicity, Markov chain Monte Carlo, ocean circulation, random-scan Metropolis

1. Introduction

Ill-posed inverse problems arise in many scientific fields, and statistical methods to address such problems have been studied extensively, for example, in O’Sullivan (1986), Mair & Ruymgaart (1996), Abramovich & Silverman (1998) and Haario *et al.* (2004). This has often been done in the context of a generalized nonlinear regression model in which a multivariate observation \mathbf{y} has a density $p(\cdot | \mu)$ determined by a transform $\mu = f(\phi)$ of some unknown high-dimensional parameter ϕ . For example, the data could be represented as a vector $\mathbf{y} = f(\phi) + \epsilon$, where ϵ is the zero-mean measurement error and the components of $f(\phi)$ are the values at various discrete locations of the solution to a partial differential equation with coefficients determined by ϕ . When f is non-invertible or invertible with a discontinuous inverse, this gives rise to an ill-posed inverse problem in the sense of Hadamard (1923): estimates of ϕ based on $f^{-1}(\mathbf{y})$ are unstable and some type of regularization is needed (see Kirsch, 1996). In the Bayesian approach to inverse problems, regularization is imposed by specifying a prior distribution on ϕ and Markov chain Monte Carlo (MCMC) samplers are used to extract information about its posterior distribution; see, for example, Andersen *et al.* (2001) and McKeague *et al.* (2005). The aim of the present paper is to investigate the convergence properties of such samplers.

Convergence is a crucial issue for any MCMC procedure. A widely used hybrid sampler for high-dimensional multivariate distributions is the random-scan random-walk Metropolis (RSM) algorithm, which is known to converge geometrically under certain conditions. Jarner & Hansen (2000) show that exponentially light tails are necessary for geometric ergodicity. They also provide sufficient conditions in terms of the geometry of the contour manifold $\{y \in \mathbb{R}^d, p(y) = p(x)\}$ as $|x| \rightarrow \infty$, where p is the stationary distribution of the Markov chain. Subsequently, Fort *et al.* (2003) have shown that geometric ergodicity can hold essentially without any conditions on the geometry of the contour manifold;

instead, they provide a sufficient condition in terms of the rate of decay in the tails of p . In Bayesian inverse problem applications, however, this condition may be difficult to verify.

Our main contribution is to provide a more accessible set of sufficient conditions, directly in terms of $p(\cdot|\mu), f$ and the prior on ϕ , to ensure geometric ergodicity of RSM samplers of the posterior distribution of ϕ . In addition, we demonstrate that these conditions can be checked when $f(\phi)$ is a discrete approximation to the solution of an elliptic partial differential equation (PDE) having coefficients and Dirichlet-type boundary conditions specified by ϕ . The finite difference method of solving PDE is known to preserve certain global properties of the continuous solution, and we are able to exploit such a property (a generalized discrete maximum principle; see Ciarlet, 1970; Al-Mahameed, 2005) to verify the main conditions. We illustrate how this can be done in the context of an application to the inversion of oceanographic tracer data, as described in McKeague *et al.* (2005) and Herbei *et al.* (2008).

The paper is organized as follows. Section 2 contains preliminary material, various assumptions and our main ergodic result. In section 3, we discuss the applicability of this result to the problem of estimating the coefficients of the advection-diffusion equation that arises in the oceanographic application. The relevance of our results to the broader context of ill-posed inverse problems is discussed in section 4. Various properties of the finite difference schemes needed for the inverse problem application are collected in section 5.

2. Geometric ergodicity

2.1. Preliminaries

We begin by recalling some background needed to apply the main result of Fort *et al.* (2003). This result applies to Markov chain samplers on finite-dimensional state spaces $\mathcal{X} = \mathbb{R}^d$. In our case, d is the dimension of ϕ and the target of the sampler is the posterior density $p(\phi|\mathbf{y})$, which in this subsection is denoted by $p(\phi)$ and is assumed to be positive and continuous over \mathcal{X} .

A transition kernel K on \mathcal{X} is said to be φ -irreducible if there is a non-zero measure φ on the Borel σ -field $\mathcal{B}(\mathcal{X})$ such that for all $\phi \in \mathcal{X}$, and all measurable sets $B \in \mathcal{B}(\mathcal{X})$ with $\varphi(B) > 0$, the chain has a positive probability of hitting B when started from ϕ . The kernel K is said to be V -uniformly ergodic for some function $V: \mathcal{X} \rightarrow [1, \infty]$ if K is φ -irreducible with invariant probability measure p such that $p(V) < \infty$, and there exist constants $r > 1$ and $R < \infty$ such that for p -almost all $\phi \in \mathcal{X}$,

$$\|K^n(\phi, \cdot) - p(\cdot)\|_V \leq Rr^{-n}V(\phi), \quad n = 1, 2, \dots,$$

where, for any signed measure ν , $\|\nu\|_V = \sup_{|g| \leq V} \nu(g)$.

The transition kernel of the RSM algorithm takes the form $K = (K_1 + \dots + K_d)/d$, where K_i is the transition kernel of a symmetric random-walk Metropolis algorithm on the i th component: starting from ϕ , a component i is selected at random and the proposal is $\phi + ye_i$, where y is sampled from a symmetric density q_i and e_i is the unit vector in the i th direction in \mathbb{R}^d . The proposal is accepted with probability $1 \wedge \{p(\phi + ye_i)/p(\phi)\}$. Each increment density q_i is assumed to be bounded away from zero in a neighbourhood of zero, ensuring that the resulting Markov chain is irreducible and aperiodic (see Tierney, 1994, section 2.3.2). Denote the Euclidean norm by $|\cdot|$. We slightly strengthen Fort *et al.*'s (2003) main assumption on the target density $p(\phi)$ and write it as assumption A.

Assumption A

There exist $0 \leq \delta < \Delta \leq \infty$ such that

$$\xi = \min_{1 \leq i \leq d} \int_{\delta}^{\Delta} q_i(y) dy > 0$$

and, for any sequence $\{\phi^n\}$ in \mathbb{R}^d with $\lim_n |\phi^n| = \infty$, we may extract a subsequence $\{\tilde{\phi}^n\}$ with the property that for some $l \in \{1, \dots, d\}$ we have $\lim_n |\tilde{\phi}_l^n| = \infty$ and for all $y \in [\delta, \Delta]$,

$$\lim_n \frac{p(\tilde{\phi}^n)}{p(\tilde{\phi}^n - \text{sign}(\phi_l^n)ye_l)} = 0 \quad \text{and} \quad \lim_n \frac{p(\tilde{\phi}^n + \text{sign}(\phi_l^n)ye_l)}{p(\tilde{\phi}^n)} = 0. \tag{1}$$

The slight strengthening here is to require that $\lim_n |\tilde{\phi}_l^n| = \infty$. The assumption can be interpreted as follows. For $d = 1$ and $y > 0$, if $p(\phi) \approx e^{-|\phi|}$ for large $|\phi|$, then $p(\phi + \text{sign}(\phi)y)/p(\phi) \approx e^{-y}$ and $p(\phi)/p(\phi - \text{sign}(\phi)y) \approx e^{-y}$. Condition (1) is therefore tantamount to assuming that the tails of p decrease faster than $e^{-|\phi|}$. The scalar $y > 0$ plays the role of a generic step size, while the sequence ϕ^n acts as a realization of the resulting Markov chain. Assumption A states that if the sampler happens to wander into the tails of the target distribution ($|\phi^n| \rightarrow \infty$), the Metropolis acceptance probability of going farther out in the tails, which appears in the two limits in (1), decreases to zero.

Verifying this assumption, however, poses additional challenges. When p is continuously differentiable, Fort *et al.* (2003) provide an alternative criterion to be checked, which involves taking partial derivatives of p . In our setting, the target density takes the form of Bayes formula $p(\phi) = p(\phi | \mathbf{y}) \propto p(\mathbf{y} | f(\phi))\pi(\phi)$, where $\pi(\phi)$ is the prior density for ϕ . Both criteria (assumption A and the alternative proposed by Fort *et al.*, 2003) are not explicit enough for a posterior density, sometimes even unusable. For instance, apart from trivial cases, it is even hard to check that $p(\phi)$ is differentiable, let alone compute its partial derivatives. In what follows, we provide a set of assumptions in terms of the two components of the posterior density that will imply assumption A.

2.2. Observation model assumptions

Suppose the forward map is $f : \mathbb{R}^d \rightarrow \mathbb{M}$, where \mathbb{M} is a set equipped with a partial preorder \lesssim (a reflexive and transitive binary relation). For any two vectors ϕ, ϕ' , we say $\phi \leq \phi'$ when $|\phi_i| \leq |\phi'_i|$ for each i .

Assumption B

For each fixed \mathbf{y} , (i) the function $\phi \mapsto p(\mathbf{y} | f(\phi))$ is positive, continuous and bounded; (ii) there exists $\mu_0 \in \mathbb{M}$ such that $p(\mathbf{y} | \cdot)$ is decreasing beyond μ_0 in the sense that $p(\mathbf{y} | \mu) \geq p(\mathbf{y} | \mu')$ if $\mu_0 \lesssim \mu \lesssim \mu'$; and (iii) $\inf_{\mu \lesssim \mu_1} p(\mathbf{y} | \mu) > 0$ for any $\mu_1 \in \mathbb{M}$.

Part (ii) of this assumption, that $p(\mathbf{y} | \cdot)$ is ‘eventually’ monotone decreasing, will be used in conjunction with the following ‘partial monotonicity’ of $f(\phi)$ along some subvector ϕ_D of ϕ .

Assumption C

$\mu_0 \lesssim f(\phi) \lesssim f(\phi')$ when $\phi_D \leq \phi'_D$ and $|\phi_D|$ is sufficiently large.

Our final assumption states that $f(\phi)$ is dominated by a monotone function of ϕ_D that does not depend on the remaining components ϕ_{-D} of ϕ .

Assumption D

There exists a function $g(\phi_D)$ taking values in \mathbb{M} such that $g(\phi_D) \lesssim g(\phi'_D)$ if $\phi_D \leq \phi'_D$, and $f(\phi) \lesssim g(\phi_D)$ for all $\phi \in \mathbb{R}^d$.

The parameters ϕ_D will be called *dominant parameters*. We note that both μ_0 and the choice of dominant parameters ϕ_D may depend on the data \mathbf{y} . In our setting, \mathbf{y} is fixed throughout and thus we will not alter the notation to reflect this dependence. In the application we analyse later, μ_0 will depend on the data; however, the selection of dominant parameters will not. This completes the structure of the observation model.

2.3. *Prior model assumptions*

In the prior model, we specify ϕ_D to be independent of the remaining parameters ϕ_{-D} . In addition, we also require that assumption A holds for the priors on ϕ_D and ϕ_{-D} . That is, now we have assumption E.

Assumption E

The prior density factors as $\pi(\phi) = \pi_D(\phi_D)\pi_{-D}(\phi_{-D})$, where $\pi_D(\cdot)$ and $\pi_{-D}(\cdot)$ are positive densities that are continuous functions of their respective arguments and assumption A holds with $p = \pi_D$ and $p = \pi_{-D}$ and a common δ and Δ .

In the trivial case of no data, the posterior density is the prior density, hence, under assumption E, by the result of Fort *et al.* (2003), the RSM algorithm for the posterior density is immediately seen to be V-uniformly ergodic. Our aim is to extend this result to the posterior density

$$p(\phi | \mathbf{y}) \propto p(\mathbf{y} | f(\phi))\pi_D(\phi_D)\pi_{-D}(\phi_{-D}). \tag{2}$$

2.4. *Ergodic result*

We now state the main result of the paper.

Theorem 1

Under assumptions B–E, with each increment density q_i assumed to be symmetric and bounded away from zero in a neighbourhood of zero, the RSM sampler for the posterior density $p(\phi | \mathbf{y})$ is V-uniformly ergodic.

Proof. It suffices to show that $p(\phi | \mathbf{y})$ satisfies assumptions 1–3 of theorem 3 in Fort *et al.* (2003). By assumptions B and E and (2), the function $\phi \mapsto p(\phi | \mathbf{y})$ is continuous and supported on \mathbb{R}^d , so assumption 1 holds. The increment densities q_i satisfy assumption 2 by hypothesis. It remains to check that $p(\phi | \mathbf{y})$ has exponentially decreasing tails in the sense of (1) (assumption 3).

Let $\{\phi^n\}$ be a sequence in \mathbb{R}^d with $|\phi^n| \rightarrow \infty$. Clearly, there exists an index $j \in \{1, \dots, d\}$ such that $|\phi^n_j| \rightarrow \infty$. The proof is now split into two cases, depending on whether the parameter indexed by j is dominant (case 1) or not (case 2).

Case 1. Suppose j corresponds to a dominant parameter. Then $|\phi^n_D| \rightarrow \infty$ and by assumption E for $\pi_D(\cdot)$, there exist $0 \leq \delta < \Delta \leq \infty$ and a subsequence $\{\tilde{\phi}^n_D\}$ of $\{\phi^n_D\}$ with the property that for some index l , corresponding to a dominant parameter, we have $|\tilde{\phi}^n_l| \rightarrow \infty$ and (1) holds with $p = \pi_D$. For this particular l , in what follows, we analyse both limits in (1), now with $p = p(\phi | \mathbf{y})$.

First consider the second limit in (1), where p is now the posterior density. Using the Bayes formula (2), for $y \in [\delta, \Delta]$

$$R_n \equiv \frac{p(\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l | \mathbf{y})}{p(\tilde{\phi}^n | \mathbf{y})} = \frac{p(\mathbf{y} | f(\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l)) \pi_D((\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l)_D)}{p(\mathbf{y} | f(\tilde{\phi}^n)) \pi_D(\tilde{\phi}_D^n)},$$

where the term in (2) involving π_{-D} cancels by the assumption of prior independence between the dominant and non-dominant parameters. Note that the prior ratio in the last line of the previous display converges to zero (by assumption E), so it suffices to show that the likelihood ratio

$$L_n \equiv \frac{p(\mathbf{y} | f(\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l))}{p(\mathbf{y} | f(\tilde{\phi}^n))}$$

is bounded above in order to have $R_n \rightarrow 0$. To see that L_n is bounded above, note that $|\tilde{\phi}_D^n| \rightarrow \infty$, so when n is sufficiently large ($n \geq n_0$), the following hold:

$$\tilde{\phi}_D^n \leq (\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l)_D \quad \text{as } y > 0.$$

Thus

$$\mu_0 \lesssim f(\tilde{\phi}^n) \lesssim f(\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l) \quad \text{by assumption C.}$$

Use assumption B(ii) to obtain that

$$p(\mathbf{y} | f(\tilde{\phi}^n)) \geq p(\mathbf{y} | f(\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l)) \quad \text{for } n \geq n_0,$$

which shows that L_n is bounded above.

The first part of (1) with p as the posterior density is analysed in a similar way. The posterior ratio

$$R'_n \equiv \frac{p(\tilde{\phi}^n | \mathbf{y})}{p(\tilde{\phi}^n - \text{sign}(\tilde{\phi}_l^n)ye_l | \mathbf{y})} \tag{3}$$

is again split into a product of likelihood and prior ratios. The prior ratio will converge to zero (by assumption E), so it suffices to show that the likelihood ratio

$$L'_n \equiv \frac{p(\mathbf{y} | f(\tilde{\phi}^n))}{p(\mathbf{y} | f(\tilde{\phi}^n - \text{sign}(\tilde{\phi}_l^n)ye_l))} \tag{4}$$

is bounded above. When n is sufficiently large, $|\tilde{\phi}_l^n| > y$ implying that $(\tilde{\phi}_l^n - \text{sign}(\tilde{\phi}_l^n)ye_l)_D \leq \tilde{\phi}_D^n$, hence

$$\mu_0 \lesssim f(\tilde{\phi}^n - \text{sign}(\tilde{\phi}_l^n)ye_l) \lesssim f(\tilde{\phi}^n) \quad \text{by assumption C.}$$

Use assumption B(ii) to conclude again that the likelihood ratio is bounded above, hence the first limit in (1) with p as the posterior density is equal to zero.

Case 2. Now suppose j corresponds to a non-dominant parameter. We can assume that $|\phi_D^n|$ is bounded, for if $|\phi_D^n|$ is unbounded we are back in case 1. Hence, $|\phi_{-D}^n| \rightarrow \infty$ and, again using assumption E for π_{-D} , there exists a subsequence $\tilde{\phi}_{-D}^n$ such that (1) holds for π_{-D} , for some index l (possibly different than j) corresponding to a non-dominant parameter and $y \in [\delta, \Delta]$. Let $\tilde{\phi}^n$ be the subsequence corresponding to $\tilde{\phi}_{-D}^n$.

Again we start by analysing the second limit in (1) with p as the posterior density. For $y \in [\delta, \Delta]$,

$$\begin{aligned} R_n &= \frac{p(\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l | \mathbf{y})}{p(\tilde{\phi}^n | \mathbf{y})} \\ &= \frac{p(\mathbf{y} | f(\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l)) \pi_{-D}((\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l)_{-D})}{p(\mathbf{y} | f(\tilde{\phi}^n)) \pi_{-D}(\tilde{\phi}_{-D}^n)} \\ &= L_n \frac{\pi_{-D}((\tilde{\phi}^n + \text{sign}(\tilde{\phi}_l^n)ye_l)_{-D})}{\pi_{-D}(\tilde{\phi}_{-D}^n)}. \end{aligned}$$

Similarly, it suffices to show that L_n is bounded above, as the prior ratio converges to zero by assumption E. As $|\tilde{\phi}_D^n|$ is bounded in this case, there exists $\phi^0 \in \mathbb{R}^d$ such that $\tilde{\phi}_D^n \leq \phi_D^0$ for all n . Use assumption D to conclude that $f(\tilde{\phi}^n) \lesssim g(\tilde{\phi}_D^n) \lesssim g(\phi_D^0)$ for each n . Furthermore, by assumption B(iii) and transitivity of the preorder,

$$\inf_{n \geq 1} p(\mathbf{y} | f(\tilde{\phi}^n)) \geq \inf_{\mu \lesssim g(\phi_D^0)} p(\mathbf{y} | \mu) > 0,$$

so the denominator of L_n is bounded away from zero. As the numerator of L_n is bounded above (by assumption B(i)), we conclude that L_n is bounded above and thus $R_n \rightarrow 0$.

For the first part of (1) with p as the posterior density, again split (3) into a product of likelihood and prior ratios, L'_n and R'_n , respectively. It suffices to show that L'_n is bounded above. When l corresponds to a non-dominant parameter, assumption D and transitivity give

$$f(\tilde{\phi}^n - \text{sign}(\tilde{\phi}_l^n)ye_l) \lesssim g(\tilde{\phi}_D^n) \lesssim g(\phi_D^0),$$

for each n . Using assumption B(iii) again, we get

$$\inf_{n \geq 1} p(\mathbf{y} | f(\tilde{\phi}^n - \text{sign}(\tilde{\phi}_l^n)ye_l)) \geq \inf_{\mu \lesssim g(\phi_D^0)} p(\mathbf{y} | \mu) > 0.$$

We conclude that the denominator in the likelihood ratio (4) is bounded away from zero, hence the ratio itself is bounded above (by assumption B(i)).

Thus, $p(\phi | \mathbf{y})$ satisfies assumption 3 of Fort *et al.* (2003) and the proof is complete.

3. Application

In this section, we illustrate how our main result applies to the inverse problem of estimating the coefficients and boundary conditions of an elliptic PDE from noisy and sparse measurements of its solution. Specifically, we show that the RSM algorithm used by McKeague *et al.* (2005) and Herbei *et al.* (2008) in an ocean circulation inverse problem is geometrically ergodic.

Deep ocean circulation is difficult to measure directly, but it is possible to infer steady-state flow indirectly from tracer measurements. Data on various tracers (e.g. oxygen, salinity, silica) can be used in the inversion, but here, for simplicity, we only consider oxygen, which is a transient tracer (unlike salinity and silica). Let $C = C(x, y)$ denote the concentration of dissolved oxygen at a location (x, y) in a rectangular domain $\Omega \subset \mathbb{R}^2$ representing the layer of the ocean being studied. The statistical problem is to estimate the (horizontal) water velocities and diffusion coefficients based on noisy measurements of C at a sparse set of locations in Ω .

The left panel in Fig. 1 shows the posterior mean of the oxygen concentration and steady-state water velocity at a depth of 2 km in a region of the South Atlantic. These results

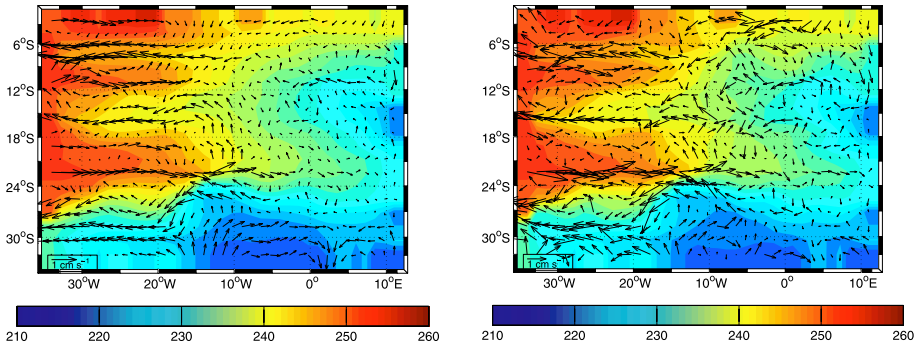


Fig. 1. Posterior mean of the oxygen concentration (in $\mu\text{mol/kg}$) and steady-state water velocity (left panel); a draw from the sampler after burn-in (right panel). The inversion domain Ω is a neutral density layer in the South Atlantic ocean, roughly corresponding to depths within 100 or 200 m of 2000 m, with lateral boundaries determined by 34°W , 11°E , 32°S and 3°S .

are based on a run of the RSM sampler used by McKeague *et al.* (2005). The alternating ‘jets’ of water flow are thought to cause the ‘tongues’ of high oxygen concentration in the northwest part of the inversion region, and are of particular interest to oceanographers. The right panel of Fig. 1 shows a single draw from the sampler after burn-in, indicating similar features. Trace plots given in McKeague *et al.* (2005) suggest that the sampler converges rapidly, but by establishing geometric ergodicity, we can now place their results on firmer ground.

The connection between the oxygen concentration C and the velocities and diffusion coefficients is given by the steady-state advection-diffusion equation as:

$$\begin{cases} \left[k_1 \frac{\partial^2}{\partial x^2} + k_2 \frac{\partial^2}{\partial y^2} - \lambda \right] C(x, y) = \left[v_1(x, y) \frac{\partial}{\partial x} - v_2(x, y) \frac{\partial}{\partial y} \right] C(x, y) & \text{for } (x, y) \in \Omega, \\ C(x, y) = b(x, y) & \text{for } (x, y) \in \partial\Omega, \end{cases} \quad (5)$$

where $k_1, k_2 \geq 0$ are diffusion coefficients not depending on location, and v_1, v_2 are the zonal and meridional water velocities (which depend on location), respectively. The oxygen consumption rate $\lambda > 0$ is assumed to be fixed and known; the values of C on the boundary $\partial\Omega$ of Ω are constrained by a Dirichlet boundary condition; and $b = b(x, y) \geq 0$. The forward problem (5) does not have a closed-form solution in terms of the coefficients and boundary values, but an approximation to the solution can be obtained using a finite difference scheme over a regular grid $\Omega_G \subset \Omega$; this scheme is described in section 5. A sufficient condition for the finite difference approximation of (5) to have a unique solution (see proposition 1) is:

$$2k_1 > \Delta_1 |v_1| \quad \text{and} \quad 2k_2 > \Delta_2 |v_2| \quad \text{on } \Omega_G, \quad (6)$$

where Δ_1 and Δ_2 are the grid mesh sizes in the x - and y -directions, respectively. This condition means that the tracer transport is dominated by diffusion, or that the Péclet numbers are strictly less than 2.

From now on, we treat v_1, v_2 and C as restricted to the grid Ω_G , and b as restricted to the grid boundary $\partial\Omega_G$. We exclude the four corners from $\partial\Omega_G$ (without changing notation) because the finite difference scheme does not involve them. To apply our ergodic result we need the support of the posterior density to be an entire Euclidean space, but k_1, k_2 and the components of b are restricted to be non-negative. Thus, we need to re-express the target parameters of interest (b, v_1, v_2, k_1, k_2) in terms of a vector of unrestricted parameters $\phi \in \mathbb{R}^d$

over which the prior density will be defined. For b , this can be done by setting $b(x, y; \phi) = |\phi_l|$, where l indexes the particular location (x, y) on the grid boundary $\partial\Omega_G$.

The subvector of ϕ specifying the boundary conditions b will be used in the role of the dominant parameters ϕ_D . The other target parameters are expressed as continuous functions of the remaining components of ϕ , the non-dominant parameters ϕ_{-D} , in such a way that the following ‘uniform Péclet’ condition holds:

$$\begin{cases} \sup_{\phi \in \mathbb{R}^d} \Delta_j |v_j(x, y; \phi)| / k_j(\phi) < 2, \\ \inf_{\phi \in \mathbb{R}^d} k_j(\phi) > 0, \quad k_1(\phi) \asymp k_2(\phi), \end{cases} \tag{7}$$

for all $(x, y) \in \Omega_G, j = 1, 2$, and the last part indicates that k_1 and k_2 are of the same order (in the ‘big-O’ sense). We refer the interested reader to the remark at the end of this section for discussion of how (7) can be arranged.

The unique solution of the finite difference approximation to the forward problem (5) at fixed ϕ is now viewed as a $|\Omega_G|$ -dimensional vector $(C(x, y; \phi))$, where each component corresponds to a location $(x, y) \in \Omega_G$, and $|A|$ denotes the cardinality of a set A . We sometimes suppress the dependence on ϕ or (x, y) to simplify the notation.

The data consist of sparse measurements of C at various locations on the grid, and are represented in vector form as $\mathbf{y} = f(\phi) + \epsilon$, where $f(\phi)$ is the subvector of $C(\phi)$ corresponding to m grid sites where data are available, and ϵ is a vector of m independent $N(0, \sigma^2)$ random variables with fixed and known variance $\sigma^2 > 0$.

The prior distribution on the dominant parameters ϕ_D (which determine b) is specified by a Gaussian Markov random field (GMRF) with an unnormalized density of the form

$$\pi_D(\phi_D) \propto \exp \left\{ -\delta_1 \sum_{l \sim l'} (\phi_l - \phi_{l'})^2 - \delta_2 \sum_{l=1}^{|\partial\Omega_G|} \phi_l^2 \right\}, \tag{8}$$

where $\delta_1 > 0, \delta_2 > 0$ are tuning parameters that control the prior smoothness and variance of ϕ_D . Here $l \sim l'$ indicates that the grid sites indexed by l and l' are nearest neighbours on $\partial\Omega_G$ and the first sum runs over all such pairs. The non-dominant parameters representing the velocities and diffusion coefficients are given a prior density of the same form, independent of the dominant parameters. For background on the use of GMRF in complex hierarchical models, see Rue & Held (2005).

Bayes formula gives the unnormalized posterior density

$$p(\phi | \mathbf{y}) \propto \left[\prod_{i=1}^m \exp \left\{ -\frac{(\mathbf{y}_i - f_i(\phi))^2}{2\sigma^2} \right\} \right] \pi_D(\phi_D) \pi_{-D}(\phi_{-D}), \tag{9}$$

where i indexes the grid sites where data are available; the expression in square brackets is the likelihood $p(\mathbf{y} | f(\phi))$. Although both the observation and prior models are Gaussian, the posterior distribution is not Gaussian owing to the nonlinearity of $C(\phi)$. The RSM increment densities $q_l, l = 1, \dots, d$ are taken to be uniform over a prespecified symmetric interval centred at zero.

Theorem 2

Under the uniform Péclet condition (7) and the continuity of b, v_1, v_2, k_1, k_2 in ϕ , we have that the RSM algorithm for exploring the posterior density $p(\phi | \mathbf{y})$ in (9) is V -uniformly ergodic.

Proof. The proof consists in checking assumptions B–E and then applying theorem 1. In section 5 we show, under the non-negativity constraint on the boundary values b , that $C(x, y; \phi) \geq 0$ at every grid site (x, y) (see the proof of proposition 1, part (a)); thus, the range of the forward map f is the positive orthant $\mathbb{M} = \mathbb{R}_+^m$. For $\mu, \mu' \in \mathbb{M}$, we define the preorder $\mu \lesssim \mu'$ by $\mu_i \leq \mu'_i$, for each $i = 1, \dots, m$, which is the standard partial order in \mathbb{R}^m .

The likelihood $p(\mathbf{y} | f(\phi))$ is clearly a positive and bounded function of ϕ . Continuity is a consequence of proposition 1(b) in section 5. This establishes part (i) of assumption B. For part (ii), we need to find $\mu_0 \in \mathbb{M}$ such that $\mu \mapsto p(\mathbf{y} | \mu)$ is decreasing beyond μ_0 . Take $\mu_0 \in \mathbb{M}$ to be the data vector $\mu_0 = \mathbf{y}$. When $\mu_0 \lesssim \mu \lesssim \mu'$, because $\mu \mapsto p(\mathbf{y} | \mu)$ is a Gaussian density with mean \mathbf{y} , it is easy to see that $p(\mathbf{y} | \mu) \geq p(\mathbf{y} | \mu')$. Furthermore, for any $\mu_1 \in \mathbb{M}$, the set $\{\mu \in \mathbb{M} : \mu \lesssim \mu_1\}$ is compact in \mathbb{R}^d and as $p(\mathbf{y} | \mu)$ is positive and continuous in μ , it follows that $\inf_{\mu \lesssim \mu_1} p(\mathbf{y} | \mu) > 0$, so part (iii) holds as well.

To check assumptions C and D, we use a discrete version of the maximum principle (DMP) for elliptic PDE. In proposition 1(c), we establish that the numerical solution $C(\phi)$ to (5) attains its maximum at a boundary site. By symmetry, the minimum is also attained at a boundary site. For ϕ and ϕ' such that $\phi_D \leq \phi'_D$, by the linearity of the PDE, the difference $C(\phi') - C(\phi)$ is the solution of a discretized Dirichlet problem (5) having all boundary values non-negative. This solution attains its minimum on the boundary, thus, $C(\phi) \leq C(\phi')$, which implies, $f(\phi) \lesssim f(\phi')$. This establishes that the forward map f is monotone along the subvector ϕ_D . To finish checking assumption C, we need to show that $\mu_0 \lesssim f(\phi)$ when $|\phi_D|$ is sufficiently large. In proposition 2, we show that the tracer concentration $C(x, y; \phi)$ at every interior grid site $(x, y) \in \Omega_G \setminus \partial\Omega_G$ can be made arbitrarily large by making $|\phi_D|$ large enough. Condition (7) is essential in establishing this. Thus, $\mathbf{y} = \mu_0 \lesssim f(\phi)$ for sufficiently large $|\phi_D|$. This proves assumption C.

For assumption D, let $g(\phi_D) = |\phi_D|_\infty \cdot \mathbb{1}$, where $\mathbb{1} \in \mathbb{M}$ is a vector of ones and $|\cdot|_\infty$ denotes the sup norm. Then, using the DMP again, $f(\phi) \lesssim g(\phi_D)$ for all $\phi \in \mathbb{R}^d$. It is also clear that $\phi_D \leq \phi'_D$ implies $g(\phi_D) \lesssim g(\phi'_D)$.

The final step is to verify that assumption E holds for GMRF priors, as in (8); this is done in proposition 5 in the Appendix.

3.1. Remark

The parameterization needed for the uniform Péclet condition (7) can be arranged in various ways. The most natural way is to allow v_j and k_j to be unbounded functions of ϕ while controlling their relative magnitude via a common-scale parameter. More specifically, let h be a positive and continuous real function, bounded away from zero and bounded above, and let

$$k_j(\phi) = \left(\kappa_0 + |\phi^{\text{scale}}| \right) h(\phi_j^{\text{diff}}),$$

where $\kappa_0 > 0$ is fixed and $j = 1, 2$. Here the scale parameter ϕ^{scale} is shared between the diffusion coefficients, whereas ϕ_j^{diff} is specific to k_j , and both are regarded as components of $\phi \in \mathbb{R}^d$. Clearly $\inf_\phi k_j(\phi) > 0$ and $k_1(\phi) \asymp k_2(\phi)$. Similarly, let g be a bounded and continuous real function, and express the velocities v_j (using more components of ϕ) as:

$$v_j(x, y; \phi) = |\phi^{\text{scale}}| g(\phi_{j,x,y}^{\text{vel}}),$$

where (x, y) ranges over locations on the grid Ω_G . Plausible values of κ_0 and bounds on the functions h, g are readily obtained by consulting the oceanography literature (as discussed in McKeague *et al.*, 2005 and Herbei *et al.*, 2008). To check the first part of (7), note that

$$\frac{|v_j(x, y; \phi)|}{k_j(\phi)} = \frac{|\phi^{\text{scale}}|}{\kappa_0 + |\phi^{\text{scale}}|} \frac{|g(\phi_{j,x,y}^{\text{vel}})|}{h(\phi_j^{\text{diff}})},$$

which is uniformly bounded above by some constant M . By refining the mesh grid sizes, we can then arrange for $\sup_{\phi} \Delta_j |v_j|/k_j \leq \Delta_j M < 2$, as required. GMRF priors can be used for $\{\phi_{j,x,y}^{\text{vel}} : (x, y) \in \Omega_G\}$, along with independent Gaussian priors for ϕ^{shared} and ϕ_j^{diff} , $j = 1, 2$.

4. Discussion

In recent years, Bayesian/MCMC techniques have become feasible approaches in numerous scientific applications involving ill-posed inverse problems. However, assessing whether a particular chain has converged is still a challenge, so it is useful to develop sufficient conditions for fast (geometric) rates of convergence that apply to wide classes of MCMC samplers arising in such applications. On the other hand, it can be difficult to know the number of iterations required because a large unknown constant may offset the geometric rate; MCMC methods still need to be used with caution, particularly when the state space is high-dimensional.

In many problems, some functional $B(\phi)$ (such as the coefficients of (5)) may be of greater interest than ϕ itself. Under mild assumptions on $B(\phi)$ (e.g. a finite $2 + \epsilon$ moment for some $\epsilon > 0$, as in Tierney, 1994), a central limit theorem is available for geometrically ergodic chains: given successive iterations $\phi^1, \phi^2, \dots, \phi^n$, the sample mean $n^{-1} \sum_{i=1}^n B(\phi^i)$ converges weakly at \sqrt{n} -rate to a Gaussian distribution, a result that can be used in selecting the posterior mean as a representative value for $B(\phi)$ and to find Monte Carlo standard errors.

Our ergodic result is applicable to other interesting ill-posed inverse problems beyond the oceanography example. Calvetti *et al.* (2006) studied the estimation of mitochondrial oxygen consumption during muscle activities based on discrete noisy observation of the oxygen concentration on the muscle surface. They used an optimization approach, finding the maximum a posteriori (MAP) estimate of parameters in a system of diffusion-consumption partial and ordinary differential equations with known initial and boundary values. The parameters in ϕ represent fluxes of oxygen consumption at resting state and during muscle stimulation. The forward map $f(\phi)$ is the time-varying oxygen concentration in the chamber used in the experiment, and the data y are noisy measurements of $f(\phi)$ at regularly spaced time points. The forward map in this case is bounded: $0 \leq f(\phi) \leq C_{\text{ch},0}$, where $C_{\text{ch},0}$ is the (known) initial oxygen concentration in the chamber, so all the parameters can be treated as non-dominant and assumptions C and D hold automatically. If $C_{\text{ch},0}$ is not known exactly, but measured with observation error, then it plays the role of a dominant parameter (suitably parameterized along the lines of section 3.1) and assumptions C and D come into play. Under the usual Gaussian observation model, assumption B holds with the preorder induced by the Euclidean norm. A suitable prior on the parameters, satisfying assumption E, then guarantees that the RSM sampler is V-uniformly ergodic. This result is useful in finding a ‘representative’ value of ϕ , the most popular choice being the posterior mean, which can be significantly different from the MAP estimate.

5. Checking assumptions

Various properties of the finite difference scheme needed for the proof of theorem 2 are collected in this section.

For each interior grid site (x, y) , the gradients in (5) are approximated using the centred finite differences

$$\begin{aligned} \frac{\partial^2 C(x, y)}{\partial x^2} &\approx \frac{1}{\Delta_1^2} [C(x + \Delta_1, y) - 2C(x, y) + C(x - \Delta_1, y)], \\ \frac{\partial^2 C(x, y)}{\partial y^2} &\approx \frac{1}{\Delta_2^2} [C(x, y + \Delta_2) - 2C(x, y) + C(x, y - \Delta_2)], \\ \frac{\partial C(x, y)}{\partial x} &\approx \frac{1}{2\Delta_1} [C(x + \Delta_1, y) - C(x - \Delta_1, y)], \\ \frac{\partial C(x, y)}{\partial y} &\approx \frac{1}{2\Delta_2} [C(x, y + \Delta_2) - C(x, y - \Delta_2)]. \end{aligned}$$

Based on these approximations, the discrete version of the Dirichlet problem (5) is given by the following system of linear equations:

$$\left\{ \begin{aligned} &\left(\lambda + \frac{2k_1}{\Delta_1^2} + \frac{2k_2}{\Delta_2^2} \right) C(x, y) + \left(-\frac{k_1}{\Delta_1^2} + \frac{v_1(x, y)}{2\Delta_1} \right) C(x + \Delta_1, y) \\ &+ \left(-\frac{k_1}{\Delta_1^2} - \frac{v_1(x, y)}{2\Delta_1} \right) C(x - \Delta_1, y) + \left(-\frac{k_2}{\Delta_2^2} + \frac{v_2(x, y)}{2\Delta_2} \right) C(x, y + \Delta_2) \\ &+ \left(-\frac{k_2}{\Delta_2^2} - \frac{v_2(x, y)}{2\Delta_2} \right) C(x, y - \Delta_2) = 0 \quad \text{for } (x, y) \in \Omega_G \setminus \partial\Omega_G, \\ &C(x, y) = b(x, y) \quad \text{for } (x, y) \in \partial\Omega_G, \end{aligned} \right.$$

which we write in matrix form as:

$$A(\phi)C(\phi) = b^*(\phi). \tag{10}$$

Here $A(\phi)$ is a $|\Omega_G| \times |\Omega_G|$ -matrix, $C(\phi)$ is the $|\Omega_G|$ -dimensional column vector representing the tracer concentrations over the grid and $b^*(\phi)$ is a column vector with zero entries at interior grid points and $b(x, y)$ on the boundary. The system of equations (10) has a unique solution $C(\phi) = A(\phi)^{-1}b^*(\phi)$ provided that $A(\phi)$ is invertible, and we now show that this is the case under our assumptions for theorem 2.

Proposition 1

Suppose the Péclet condition (6) holds and $b, v_1, v_2, k_1, k_2, \lambda$ are continuous in $\phi \in \mathbb{R}^d$ and satisfy $\lambda > 0, b \geq 0$. Then

- (a) $A(\phi)$ is an M -matrix;
- (b) $C(\phi)$ is continuous in $\phi \in \mathbb{R}^d$;
- (c) $C(\phi)$ satisfies the (weak) DMP:

$$\max_{(x,y) \in \Omega_G} C(x, y; \phi) = \max_{(x,y) \in \partial\Omega_G} b(x, y; \phi). \tag{11}$$

Part (c) of this proposition is a DMP that is well known for elliptical PDE. This result holds under greater generality than we need (see, e.g. Ciarlet, 1970; Al-Mahameed, 2005), but the proof is easy in our case and is included for completeness.

Proof. For (a), the conditions $2k_1 > |v_1|\Delta_1$ and $2k_2 > |v_2|\Delta_2$ imply that $\kappa_1, \kappa_2 \geq 0$. Poole & Boullion (1974), in their theorem 2.1, give several equivalent definitions of M -matrices. We will check condition (2f), that is, there exists a vector $\mathbf{x} \in \mathbb{R}_+^{|\Omega_G|}$ such that $A(\phi)\mathbf{x} \in \mathbb{R}_+^{|\Omega_G|}$. Indeed, the diagonal entries of $A(\phi) = (a_{ij})$, namely

$$a_{ll} = \begin{cases} \lambda + \frac{2k_1}{\Delta_1^2} + \frac{2k_2}{\Delta_2^2} & \text{if } l \text{ indexes an interior site,} \\ 1 & \text{if } l \text{ indexes a boundary site,} \end{cases}$$

are positive. Also, for the off-diagonal entries, $a_{ll'} \leq 0$ for $l \neq l'$, and

$$a_{ll} = |a_{ll}| > \sum_{l' \neq l} |a_{ll'}|.$$

Let $\mathbf{x} = (1, 1, \dots, 1)^T \in \mathbb{R}_+^{|\Omega_G|}$, and note that

$$(A(\phi)\mathbf{x})_l = \sum_{l'=1}^{|\Omega_G|} a_{ll'} = \begin{cases} \lambda & \text{if } l \text{ indexes an interior site,} \\ 1 & \text{if } l \text{ indexes a boundary site,} \end{cases}$$

so $A(\phi)\mathbf{x} \in \mathbb{R}_+^{|\Omega_G|}$. Thus, condition (2f) of theorem 2.1 in Poole & Boullion (1974) is satisfied, which implies that $A(\phi)$ is an M -matrix. As a consequence, condition (2e) of the same theorem ensures that $A(\phi)^{-1}$ exists and that $A(\phi)^{-1} \geq 0$ (element-wise). This also implies that the solution of (10) satisfies $C(\phi) = A(\phi)^{-1}b^*(\phi) \geq 0$ as we assume that the boundary values are non-negative. This was a claim made at the start of the proof of theorem 2.

For (b), as $b^*(\phi)$ is continuous in ϕ , it suffices to show that $A(\phi)^{-1}$ is continuous in ϕ with respect to the matrix norm $\|\cdot\|$. This follows from the continuity of $A(\phi)$ with respect to the matrix norm and the inequality

$$\|A(\phi^1)^{-1} - A(\phi^0)^{-1}\| \leq \frac{\|A(\phi^0)^{-1}\|^2 \|A(\phi^1) - A(\phi^0)\|}{1 - \|A(\phi^0)^{-1}\| \|A(\phi^1) - A(\phi^0)\|}$$

for all $\phi^0, \phi^1 \in \mathbb{R}^d$ such that $\|A(\phi^0)^{-1}\| \|A(\phi^1) - A(\phi^0)\| < 1$ (Gandy & Jensen, 2005, lemma 7).

For (c), we argue by contradiction. Suppose there is an interior grid site (x_0, y_0) , where $C(\phi)$ attains its maximum over Ω_G . Also suppose that we are in the non-trivial case in which at least one boundary value is positive; otherwise $C(\phi) = 0$ and (11) holds. Using the finite difference scheme,

$$\begin{aligned} C(x_0, y_0) \left(\lambda + \frac{2k_1}{\Delta_1^2} + \frac{2k_2}{\Delta_2^2} \right) &= \left(\frac{k_1}{\Delta_1^2} - \frac{v_1(x_0, y_0)}{2\Delta_1} \right) C(x_0 + \Delta_1, y_0) \\ &\quad + \left(\frac{k_1}{\Delta_1^2} + \frac{v_1(x_0, y_0)}{2\Delta_1} \right) C(x_0 - \Delta_1, y_0) \\ &\quad + \left(\frac{k_2}{\Delta_2^2} - \frac{v_2(x_0, y_0)}{2\Delta_2} \right) C(x_0, y_0 + \Delta_2) \\ &\quad + \left(\frac{k_2}{\Delta_2^2} + \frac{v_2(x_0, y_0)}{2\Delta_2} \right) C(x_0, y_0 - \Delta_2) \\ &\leq \left(\frac{2k_1}{\Delta_1^2} + \frac{2k_2}{\Delta_2^2} \right) C(x_0, y_0). \end{aligned}$$

It follows that $\lambda C(x_0, y_0) \leq 0$ and thus $C(x, y) \leq C(x_0, y_0) \leq 0$ at all sites (x, y) . When (x, y) is a boundary site, this contradicts the assumption that at least one boundary value is positive.

Proposition 2

Suppose the uniform Péclet condition (7) holds and $\lambda > 0, b(\phi) \geq 0$. If $|b(\phi^n)| \rightarrow \infty$ then $C(x, y; \phi^n) \rightarrow \infty$ for any interior site (x, y) .

Proof. Given $|b(\phi^n)| \rightarrow \infty$, there exists a boundary site (x_0, y_0) such that $C(x_0, y_0; \phi^n) = b(x_0, y_0; \phi^n) \rightarrow \infty$. Let (x_1, y_1) be an interior site neighbouring (x_0, y_0) , and consider the case $x_1 = x_0 + \Delta_1$ and $y_1 = y_0$ (the same argument applies to all such interior sites). From the finite difference scheme, using $C \geq 0$ (see the proof of proposition 1(a)), it follows that

$$C(x_1, y_1) \left(\lambda + \frac{2k_1}{\Delta_1^2} + \frac{2k_2}{\Delta_2^2} \right) \geq \left(\frac{k_1}{\Delta_1^2} + \frac{v_1(x_1, y_1)}{2\Delta_1} \right) C(x_0, y_0),$$

where we have suppressed the dependence on ϕ^n . Dividing through by k_1/Δ_1^2 , we then have

$$\begin{aligned} C(x_1, y_1) \left(\frac{\lambda \Delta_1^2}{k_1} + 2 + \frac{k_2}{k_1} \frac{2\Delta_1^2}{\Delta_2^2} \right) &\geq \left(1 + \frac{\Delta_1 v_1(x_1, y_1)}{2k_1} \right) C(x_0, y_0) \\ &\geq \left(1 - \frac{\Delta_1 |v_1(x_1, y_1)|}{2k_1} \right) C(x_0, y_0). \end{aligned}$$

From the uniform Péclet condition (7) we deduce that the expression $(1 - \Delta_1 |v_1(x_1, y_1)|)/(2k_1)$, as a function of ϕ , is bounded away from zero, and each summand in $(\frac{\lambda \Delta_1^2}{k_1} + 2 + \frac{k_2}{k_1} \frac{2\Delta_1^2}{\Delta_2^2})$ is bounded, as a function of ϕ . Thus, the right-hand side above tends to infinity as $n \rightarrow \infty$. It follows that $C(x_1, y_1) \rightarrow \infty$. This argument can be repeated for all the interior neighbours of (x_1, y_1) and after a finite number of steps we obtain that $C(x, y) \rightarrow \infty$ for any interior site (x, y) .

Acknowledgements

Ian McKeague's research was supported in part by NSF Grant DMS-0806088.

References

- Abramovich, F. & Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85**, 115–129.
- Al-Mahameed, M. M. (2005). Weak discrete maximum principles. *Arch. Math.* **41**, 167–173.
- Andersen, K. E., Brooks, S. P. & Hansen, M. B. (2001). A Bayesian approach to crack detection in electrically conducting media. *Inverse Problems* **17**, 121–136.
- Calvetti, D., Dash, R. K., Somersalo, E. & Cabrera, M. (2006). Local regularization method applied to estimating oxygen consumption during muscle activities. *Inverse Problems* **22**, 229–243.
- Ciarlet, P. G. (1970). Discrete maximum principles for finite difference operators. *Aequationes Math.* **4**, 336–352.
- Fort, G., Moulines, E., Roberts, G. O. & Rosenthal, J. S. (2003). On the geometric ergodicity of hybrid samplers. *J. Appl. Probab.* **40**, 123–146.
- Gandy, A. & Jensen, U. (2005). On goodness-of-fit tests for Aalen's additive risk model. *Scand. J. Statist.* **32**, 425–445.
- Haario, H., Laine, M., Lehtinen, M., Saksman, E. & Tamminen, J. (2004). Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. R. Stat. Soc. Ser. B Statist. Methodol.* **66**, 591–607.
- Hadamard, J. (1923). *Lectures on Cauchy's problem in linear partial differential equations*. Yale University Press, New Haven.
- Herbei, R., McKeague, I. W. & Speer, K. (2008). Gyres and jets: inversion of tracer data for ocean circulation structure. *J. Phys. Oceanogr.* **38**, 1180–1202.
- Jarner, S. & Hansen, E. (2000). Geometric ergodicity of the Metropolis algorithm. *Stochastic Process. Appl.* **85**, 341–361.
- Kirsch, A. (1996). *An introduction to the mathematical theory of inverse problems*. Springer-Verlag, New York.
- Mair, B. A. & Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.* **56**, 1424–1444.
- McKeague, I. W., Nicholls, G. K., Speer, K. G. & Herbei, R. (2005). Statistical inversion of South Atlantic circulation in an abyssal neutral density layer. *J. Mar. Res.* **63**, 683–704.

O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statist. Sci.* **1**, 502–527.
 Poole, G. & Boullion, T. (1974). A survey on M -matrices. *SIAM Rev.* **16**, 419–427.
 Rue, H. & Held, L. (2005). *Gaussian Markov random fields*. Chapman & Hall, London.
 Tierney, L. (1994). Markov chain Monte Carlo for exploring posterior distributions. *Ann. Statist.* **22**, 1701–1728.

Received May 2008, in final form March 2009

Radu Herbei, Department of Statistics, Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, USA. E-mail: herbei@stat.osu.edu

Appendix

In this section, we show that condition (1) of assumption A is satisfied by the types of GMRF used in section 3 to define prior distributions for the parameters in ϕ that determine the boundary condition b and the velocities v_j . For simplicity, we just consider GMRF on a one-dimensional grid $\{1, \dots, d\}$, but the same proof applies to multi-dimensional grids.

Proposition 3

Condition (1) of assumption A holds for the unnormalized density $p = \pi$ given by

$$\pi(\phi) \propto \exp \left\{ -\delta_1 \sum_{i \sim j} (\phi_i - \phi_j)^2 - \delta_2 \sum_{i=1}^d \phi_i^2 \right\}, \quad \phi \in \mathbb{R}^d, \tag{12}$$

where $i \sim j$ iff $|i - j| = 1$, and $\delta_1, \delta_2 > 0$.

Proof. Suppose $|\phi^n| \rightarrow \infty$. Use lemma 1 to select a subsequence $\tilde{\phi}^n$ and $l \in \{1, \dots, d\}$ such that $|\tilde{\phi}_l^n| \rightarrow \infty$ and

$$\text{sign}(\tilde{\phi}_l^n)(\tilde{\phi}_l^n - \tilde{\phi}_k^n) \rightarrow L_k \in [0, \infty] \quad \text{for all } k \neq l. \tag{13}$$

Let $y > 0$. Without loss of generality, assume $\tilde{\phi}_l^n > 0$ for each n . First, we consider the second limit in (1). Note that

$$\begin{aligned} \frac{\pi(\tilde{\phi}^n + ye_l)}{\pi(\tilde{\phi}^n)} &= \frac{\exp \left\{ -\delta_1 \left[(\tilde{\phi}_l^n + y - \tilde{\phi}_{l-1}^n)^2 + (\tilde{\phi}_l^n + y - \tilde{\phi}_{l+1}^n)^2 \right] - \delta_2 (\tilde{\phi}_l^n + y)^2 \right\}}{\exp \left\{ -\delta_1 \left[(\tilde{\phi}_l^n - \tilde{\phi}_{l-1}^n)^2 + (\tilde{\phi}_l^n - \tilde{\phi}_{l+1}^n)^2 \right] - \delta_2 (\tilde{\phi}_l^n)^2 \right\}} \\ &= \exp \left\{ -\delta_1 \left[2y^2 + 2y(\tilde{\phi}_l^n - \tilde{\phi}_{l-1}^n) + 2y(\tilde{\phi}_l^n - \tilde{\phi}_{l+1}^n) \right] \right\} \exp \left\{ -2\delta_2 \tilde{\phi}_l^n - \delta_2 y^2 \right\}, \end{aligned}$$

where the first equality follows after canceling all the terms in (12) not involving $\tilde{\phi}_l^n$. The second exponential term above converges to zero as $\tilde{\phi}_l^n \rightarrow \infty$. The first exponential term is bounded because of (13). Thus,

$$\lim_{n \rightarrow \infty} \frac{\pi(\tilde{\phi}^n + ye_l)}{\pi(\tilde{\phi}^n)} = 0.$$

The first limit in (1) is analysed similarly:

$$\begin{aligned} \frac{\pi(\tilde{\phi}^n)}{\pi(\tilde{\phi}^n - ye_l)} &= \frac{\exp \left\{ -\delta_1 \left[(\tilde{\phi}_l^n - \tilde{\phi}_{l-1}^n)^2 + (\tilde{\phi}_l^n - \tilde{\phi}_{l+1}^n)^2 \right] - \delta_2 (\tilde{\phi}_l^n)^2 \right\}}{\exp \left\{ -\delta_1 \left[(\tilde{\phi}_l^n - y - \tilde{\phi}_{l-1}^n)^2 + (\tilde{\phi}_l^n - y - \tilde{\phi}_{l+1}^n)^2 \right] - \delta_2 (\tilde{\phi}_l^n - y)^2 \right\}} \\ &= \exp \left\{ -\delta_1 \left[-2y^2 + 2y(\tilde{\phi}_l^n - \tilde{\phi}_{l-1}^n) + 2y(\tilde{\phi}_l^n - \tilde{\phi}_{l+1}^n) \right] \right\} \exp \left\{ -2\delta_2 \tilde{\phi}_l^n - \delta_2 y^2 \right\}. \end{aligned}$$

Again the second exponential term converges to zero while the first one is bounded because of (13). Thus,

$$\lim_{n \rightarrow \infty} \frac{\pi(\tilde{\phi}^n)}{\pi(\tilde{\phi}^n - ye_l)} = 0.$$

We have shown that condition (1) holds with $p = \pi$ and any $y \geq 0$.

The choice of δ and Δ in assumption A only relate to the choice of increment densities q_i . The condition $\zeta = \min_{1 \leq i \leq d} \int_{\delta}^{\Delta} q_i(y) dy > 0$ is clearly satisfied when each q_i is a uniform density over a symmetric interval centred at zero.

Lemma 1

Let $\{x^n\}$ be a sequence in \mathbb{R}^d such that $|x^n| \rightarrow \infty$. Then there exists $l \in \{1, 2, \dots, d\}$ and a subsequence $\{\tilde{x}^n\}$ of $\{x^n\}$ such that for each $k = 1, \dots, d, k \neq l$,

$$\text{sign}(\tilde{x}_l^n)(\tilde{x}_l^n - \tilde{x}_k^n) \rightarrow L_k \in [0, \infty]. \tag{14}$$

Proof. Because $|x^n| \rightarrow \infty$, there exists l such that $|x_l^n| \rightarrow \infty$. Without loss of generality we can assume that $|x_k^n| \rightarrow \infty$ for all k , because (14) holds trivially when x_k^n is bounded. Moreover, we only need to consider the case that $x_k^n \rightarrow \infty$ for all k . The claim is proved by induction over d .

First set $d = 2$. Suppose $|(x_1^n, x_2^n)| \rightarrow \infty, x_1^n \rightarrow \infty$ and $x_2^n \rightarrow \infty$. Consider $y^n = x_1^n - x_2^n$. There exists a subsequence \tilde{y}^n of y^n such that $\tilde{y}^n \rightarrow L_1 \in [-\infty, \infty]$. If $L_1 < 0$, select $l = 2$, then $\tilde{x}_l^n - \tilde{x}_k^n \rightarrow L_k \in [0, \infty]$ for all $k \neq l$. If $L_1 \geq 0$ select $l = 1$, then $\tilde{x}_l^n - \tilde{x}_k^n \rightarrow L_k \in [0, \infty]$ for all $k \neq l$. Thus, the claim is true for $d = 2$.

Now assume the claim is true for $d = m$. Let $d = m + 1$ and consider a sequence $\{x^n\} \subset \mathbb{R}^d$ such that $x_k^n \rightarrow \infty$ for each $k = 1, \dots, m + 1$. Apply the claim to the first m components (x_1^n, \dots, x_m^n) to select $j \in \{1, \dots, m\}$ and a subsequence \hat{x}^n such that

$$\hat{x}_j^n - \hat{x}_k^n \rightarrow L_k^m \in [0, \infty], \quad \text{for } k = 1, \dots, m, \quad k \neq j. \tag{15}$$

Consider $(\hat{x}_j^n, \hat{x}_{m+1}^n)$. Use the claim in the case $d = 2$ to select $l \in \{j, m + 1\}$ such that (14) holds.