A NEW METHOD FOR COUPLING RANDOM FIELDS

L.A. BREYER AND G.O. ROBERTS

ABSTRACT. Given a Markov chain, a stochastic flow which simultaneously constructs sample paths started at each possible initial value can be constructed as a composition of random fields. In this paper, we describe a method for coupling flows by modifying an arbitrary field (consistent with the Markov chain of interest) by an independence Metropolis Hastings iteration. We show that the resulting stochastic flow has many desirable coalescence properties, regardless of the form of the original flow.

1. INTRODUCTION

In this note, we describe a novel approach to the problem of coupling random fields, and by extension, Markov processes. The problem setting is as follows: Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, upon which is defined a family $\big(\mathfrak{X}(z) : z \in Z\big)$ of random variables (called a random field), Z being some index set and each variable $\mathfrak{X}(z)$ taking values in some common measurable space $(E, \mathcal{E})$. Under measurability conditions, the marginal distributions of the field are fully described by a kernel $P : Z \times \mathcal{E} \to [0, 1]$, in the sense that

$$(1) \qquad \mathbb{P}\big(\mathfrak{X}(z) \in dy\big) = P(z, dy), \quad z \in Z.$$

Conversely, if we start with a kernel P and take (1) as a hypothesis, we can construct some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family $X(z) : \Omega \to E$ ($z \in Z$) such that (1) holds. Such a probability is often called a coupling, at least when Z consists of two points only. The choices for $\mathbb{P}$ are limitless, and it therefore becomes interesting to ask whether a probability space and a family $\mathfrak{X} = \big(\mathfrak{X}(z) : z \in Z\big)$ can be constructed which additionally satisfy some previously agreed upon condition.

The case $|Z| = 2$. When the index set Z consists of only two (distinct) points, the problem we shall deal with can be phrased as follows: given two marginal distributions $P_1(dy)$ and $P_2(dy)$ on a common space E, find a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a realization (X_1, X_2) such that

$$(2) \qquad \mathbb{P}(X_1 = X_2) > 0, \quad X_1 \sim P_1(\cdot), X_2 \sim P_2(\cdot).$$

This problem can be solved (Lindvall, 1992) if and only if there exists some non-trivial measure $\nu(\cdot)$ on E such that both $P_1 \geq \nu$ and $P_2 \geq \nu$, in which case there exists a *maximal* probability $\mathbb{P}$ satisfying (2) on a canonical probability space Ω. This result has been the basis for a very successful method, originally proposed by Nummelin, for coupling Markov chains (Meyn and Tweedie, 1993).

The case $|Z| > 2$. When the indexing set Z consists of more than two distinct points, we must replace (2) by a more suitable requirement on $\mathbb{P}$. There is no universal definition extending (2), but the following seems natural.

Definition 1. *Let Z be some index set, and consider a kernel $P : Z \times \mathcal{E} \to [0, 1]$ of probability measures on some measurable space $(E, \mathcal{E})$. We say that a family $(\mathfrak{X}(z) : z \in Z)$ of random variables with values in E, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a* finite *(resp.* discrete*) coupling of the field determined by P if both* (1) *holds and*

$$\mathbb{P}\Big(z \mapsto \mathfrak{X}(z) \text{ has a finite [resp. discrete] range in } E\Big) = 1.$$

One aim will be to find conditions on the kernel P such that there exists a finitely (resp. discretely) coupled field. Interest in finite couplings is due to the recent developments in the literature on Perfect Simulation, which was spawned by the seminal work of Propp and Wilson (1995). The proof of existence we offer is constructive, and susceptible to be implemented on a computer.

The construction of Markov chains for simulation purposes is usually accomplished by the use of a stochastive recursive representation of the Markov chain transitions (Borovkov and Foss, 1992). This in turn implies a succession of stochastic fields, mapping at each simulation epoch the current state z of the chain into its subsequent value $\mathfrak{X}(z)$. On composition of these fields, we obtain a stochastic flow describing the simultaneous dynamics of the chain from all possible initial states.

It is of interest (especially in the fast growing area of perfect simulation) to construct this flow in such a way as to force coalescence of chains started from all possible initial states. Unfortunately, many natural stochastic recursive constructions do not lead to this, and it would be desirable to exhibit a general recipe for the modification of an arbitrary stochastic flow (consistent with this particular Markov chain dynamic) in order to guarantee the desired coalescence property. This is the motivation behind the present paper.

Our approach will be to modify the individual component fields of an existing flow, and recombine these new fields into a flow which has improved coalescence properties. We will use a method of modification based on the independence sampler (as commonly used in MCMC). Our first two main results, Theorems 2 and 5 deal with the modified fields themselves, showing that under suitable uniform ergodicity assumptions, the fields, $\mathfrak{X}(z)$, can be constructed so as to have finite image with arbitrarily high probability (Theorem 2) and with probability 1 by a suitable perfect simulation construction (Theorem 5).

Theorem 6 and Theorem 8 deal with the whole flow as constructed from the modified fields, $\mathfrak{X}^{(1)}(z)$, $\mathfrak{X}^{(2)}(z)$, ..., say. Theorem 6 ensures the coupling of chains started at any two arbitrary initial values. Theorem 8 extends this result to an arbitrary collection of starting values. Finally in Proposition 9, we focus on the problem of constructing perfect simulation algorithms for Metropolis Hastings chains. These algorithms are not covered by our previous results, although their specific structure allows specialised results to be proved. Examples include the random walk Metropolis algorithm and the Langevin algorithm.

2. A Markov chain on Field Space

The construction of the finitely (resp. discretely) coupled field will be accomplished by applying ideas originally devised by Propp and Wilson (1995) to a Metropolis-Hastings type chain whose states are represented by candidate fields. The successively produced fields in this chain will have progressively "smaller" ranges until such time as only finite range realizations of the field are produced. This will a.s. occur after a finite number of iterations, uniformly fast in the "seed" field. By the technique of "Coupling From The Past" we generate from this chain a finitely coupled field.

We proceed to set up some notation, and describe the transitions of the above mentioned Markov chain. Let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space, Z a topological space. We shall denote by $\mathfrak{F}(Z; E)$ the space of Borel measurable functions of Z into E. This will be thought of as the state space for a Markov chain $\mathfrak{X}_n$ defined below.

We also suppose given a probability kernel $P(z, dy) = p(z, y)\mu(dy)$ of Z into E, absolutely continuous with respect to μ. The existence of the density $p(z, y)$ here is mainly a matter of convenience, and can be dispensed with at the cost of small changes in the theory below.

For the construction of our Markov chain(s), we shall need an appropriate probability space $(\Omega, \mathcal{F})$, which we now describe. Take $\Omega = \mathfrak{F}(Z; E)$ and $\mathcal{F}$ to be the σ-algebra generated by the function evaluations $z \mapsto \omega(z)$, $\omega \in \Omega$. Define also the set Π of probability measures on $(\Omega, \mathcal{F})$ which prescribe the marginals of the canonical random variable:

$$(3) \quad \Pi = \Big\{ \mathbb{P} \in \mathcal{P}(\Omega) : \mathbb{P}\big(\omega : \omega(z) \in A\big) = P(z, A) \text{ for all } z \in Z, A \in \mathcal{B}(E) \Big\}.$$

Note that Π is nonempty, since we can always take the product measure $\mathbb{P} = \bigotimes_z P(z, \cdot) \in \Pi$. However, in the interests of readability, given a random variable $\mathfrak{X} : \Omega \to \mathfrak{F}(Z; E)$, we shall often write $\mathfrak{X} \in \Pi$ instead of the more cumbersome "the law of $\mathfrak{X}$ belongs to Π".

We are now ready to define our Markov chain. In general, we shall need a family $\Theta_z(x, dy)$ of probability kernels satisfying $\Theta_z(x, \cdot) \gg P(x, \cdot)$ for each $z \in Z$; These will represent "proposals". We suppose here also the existence of a density, $\Theta_z(x, dy) = \theta_z(x, y)\mu(dy)$.

To define the Markov chain $(\mathfrak{X}_n)$ we take $\mathfrak{X}_0 \in \Pi$ ("any random variable $\mathfrak{X}_0$ whose law belongs to Π"), and then use induction. Given $\mathfrak{X}_n = \big(\mathfrak{X}_n(z) : z \in Z\big)$, let $\Phi_z \sim \theta_z(\mathfrak{X}_n(z), \cdot)\mu(\cdot)$, and put $\mathfrak{X}_{n+1} = \big(\mathfrak{X}_{n+1}(z) : z \in Z\big)$ where

$$(4) \quad \mathfrak{X}_{n+1}(z) = \begin{cases} \Phi_z & \text{if } p(z, \Phi_z)\theta_z\big(\Phi_z, \mathfrak{X}_n(z)\big) > \xi \cdot p\big(z, \mathfrak{X}_n(z)\big)\theta_z\big(\mathfrak{X}_n(z), \Phi_z\big) \\ \mathfrak{X}_n(z) & \text{otherwise.} \end{cases}$$

Here $\xi \sim U[0, 1]$ independently of $\mathfrak{X}_n$.

If we fix $z \in Z$ in the above, it is immediately seen that the *process* $n \mapsto \mathfrak{X}_n(z)$ is a Metropolis-Hastings chain with stationary distribution $P(z, \cdot)$ and proposal

kernel $\Theta_z(x, dy)$. Since moreover by assumption $\mathfrak{X}_0(z) \sim p(z, \cdot)\mu(\cdot)$, one checks immediately (or see Roberts and Smith, 1993) that

$$(5) \qquad \mathbb{P}\Big(\mathfrak{X}_n(z) \in dy\Big) = p(z, y)\mu(dy), \quad n > 0, \quad i.e.\ \mathfrak{X}_n \in \Pi.$$

However, when viewed as a family indexed by z, the chains $\mathfrak{X}_{\cdot}(z)$ are not independent as z varies throughout Z, since we *reuse the same random number* ξ.

In the examples below, we describe some choices of the kernel density $\theta_z(x, y)$ which guarantee that, for sufficiently large (but finite) n, the realization $z \mapsto \mathfrak{X}_n(z)$ a.s. has finite range. None of the choices we list claim to be "optimal".

Example 1. The simplest example of such a chain is arguably obtained if we take $\theta_z(x, y) \equiv \theta(y)$, a function of y alone. We call this the *Simple Field Coupler*. For this choice, the formula (4) simplifies to

$$(6) \qquad \mathfrak{X}_{n+1}(z) = \begin{cases} \Phi \sim \theta(\cdot)\mu(\cdot) & \text{if } p(z, \Phi)\theta\big(\mathfrak{X}_n(z)\big) > \xi \cdot p\big(z, \mathfrak{X}_n(z)\big)\theta(\Phi) \\ \mathfrak{X}_n(z) & \text{otherwise.} \end{cases}$$

Note first that if $z \mapsto \mathfrak{X}_k(z)$ has finite range, then every subsequent function $z \mapsto \mathfrak{X}_{k+\ell}(z)$ will have finite range also. Thus the set of states in $\mathfrak{F}(Z; E)$ with finite range forms an absorbing set for any chain defined by (6). Let us now show that this chain $\mathfrak{X}_n$ indeed evolves towards a state with finite range. We shall explain in the next section how to generate a finitely coupled field using a perfect simulation technique.

Recall that a subset $C \subset Z$ is called 1-*small* (Roberts and Rosenthal, 1996) for the kernel $P(x, dy)$ on E if there exists a probability density ν and a constant $\epsilon > 0$ such that

$$(7) \qquad \inf_{x \in C} p(x, y)\mu(dy) \geq \epsilon\nu(y)\mu(dy).$$

Theorem 2. *Suppose that Z can be entirely covered by a finite union of 1-small sets $C_1, \ldots, C_N$, and suppose we can choose a probability density θ in such a way that*

$$(8) \qquad \epsilon_i\nu_i(\cdot) \leq \inf_{z \in C_i} p(z, \cdot) \leq \sup_{z \in C_i} p(z, \cdot) \leq \gamma_i\theta(\cdot), \quad i = 1, \ldots, N$$

for positive numbers $\gamma_1, \ldots, \gamma_N$. *If we define a Markov chain* $\mathfrak{X}_n$ *on* Π *by the prescription* (6), *then*

$$\mathbb{P}\Big(\mathfrak{X}_n \ \textit{has finite range eventually}\ \Big) = 1.$$

Proof. Clearly from (6), if z is fixed the process $n \mapsto \mathfrak{X}_n(z)$ is a Metropolis-Hastings Markov chain on E (an Independence Sampler, in fact), with stationary distribution $p(z, \cdot)\mu(\cdot)$. Therefore we certainly have (5), since $\mathfrak{X}_0(z)$ is already stationary. Now for every $z \in Z$, the chain $n \to \mathfrak{X}_n(z)$ must almost certainly move eventually. This will happen the first time from (6) at time n if and only if

$$(9) \qquad\qquad \Big(p(z, \Phi)\theta\big(\mathfrak{X}_n(z)\big)\Big) > \xi \cdot \Big(p\big(z, \mathfrak{X}_n(z)\big)\theta(\Phi)\Big).$$

By (7) and (8), if $z \in C_k$, the event (9) contains the event

$$(10) \qquad\qquad \epsilon_k \nu_k(\Phi) > \gamma_k \xi\ \theta(\Phi),$$

which will occur eventually for some pair (Φ, ξ) since the pairs are independent, by the Borel-Cantelli lemma. When it occurs at time n_k say, *all* Markov chains $\mathfrak{X}_{n_k}(z)$ for $z \in C_k$ will accept the *same* jump to Φ, reducing the range of $z \mapsto \mathfrak{X}_{n_k}(z)$ on C_k to a single point. Note that after this common jump, the range over C_k is likely to break into more than one point, but never more than a finite number at each iteration. This analysis is valid on each of the small sets $C_1, \ldots, C_N$ which cover Z, and another application of the Borel-Cantelli lemma now shows that each of these sets must experience a common jump within a finite time $\tau = \max\{n_1, \ldots, n_N\}$, beyond which the realization $\mathfrak{X}_{\tau+n}$ must always have a finite range. $\square$

It is clear that the choice of proposal Φ will influence the acceptance rate of the algorithm. For any given $z_0 \in Z$, we can judiciously adapt the proposal density $\theta(\cdot)$ to the target density $p(z_0, \cdot)$, but this will be useless if the density $p(z, \cdot)$ varies highly as a function of z. This difficulty can be addressed very simply by implementing an auxiliary variable which can be used to tune the proposal θ. The next example shows one way to do this. It will be called the *Independence Field Coupler*.

The Simple Field Coupler does not usually generate finitely coupled random fields in a fixed, finite number of iterations n say, if the "seed" field $\mathfrak{X}_0$ is not already finitely coupled. To state an analogy with the behaviour of ergodic Markov chains, we make the following definition:

Definition 3. *A random field* $\mathfrak{X} = (\mathfrak{X}(z) : z \in Z)$ *is called finitely coupled within* $\epsilon > 0$ *if*

$$\mathbb{P}\big(z \mapsto \mathfrak{X}(z) \text{ has finite range}\big) \geq 1 - \epsilon.$$

With this definition, we can paraphrase Theorem 2 as follows: For each $\epsilon > 0$, there exists n such that $\mathfrak{X}_n$ is finitely coupled within ϵ. It is *not* true that $\mathfrak{X}_n$ is finitely coupled (within $\epsilon = 0$) for *any* n. Thus, given a realization $\mathfrak{X}_0$ of a seed field, we can never be certain that the realization $\mathfrak{X}_n$ generated by the Simple Field Coupler has finite range. We shall construct a finitely coupled field $\mathfrak{X}$ by a perfect simulation technique in the next section. For practical implementation of this technique, it is not even necessary to be able to generate an initial field $\mathfrak{X}_0$, since its shape over the small set C_k is completely forgotten once the event (10) occurs.

The bounds (8) are not needed to implement the Simple Field Coupler. However, if they are available, we can get an estimate of the speed of convergence to a finitely coupled field within $\epsilon > 0$. By the proof of Theorem 2, the realization $\mathfrak{X}_n$ has finite range as soon as the event (10) has occurred for each $k = 1, \cdots, N$. If we denote by T_k the first hitting time of the set $\{(y, x) : \epsilon_k \nu_k(y) > \gamma_k x \theta(y)\}$ by the IID series $(\Phi^{(n)}, \xi^{(n)})$, we have

$$\mathbb{P}\big(\mathfrak{X}_n \text{ has finite range }\big) \geq \mathbb{P}\big(\max(T_1, \ldots, T_N) \leq n\big),$$

and this is independent of $\mathfrak{X}_0$. This bound can be explicitly computed in particular cases.

If Z is only covered by a countable collection of 1-small sets, we can still run the algorithm (6), but the realization $\mathfrak{X}_n$ will generally not have discrete range in a finite time.

We end our description of the properties of the Simple Field Coupler with some comments about the assumptions of Theorem 2. Let us suppose that $Z = E$, so that we may think of the kernel P as the transition function of a Markov chain with state space E.

It is a fairly easy exercise (left to the reader) to show that *coverage of E by a finite number of 1-small sets* and irreducibility implies that the whole set E is

petite, and aperiodicity then gives the whole set E as small. This last statement *implies that P is uniformly ergodic.*

It is therefore of interest to know if the converse holds. Unfortunately, this case is not so straightforward. Indeed, *if P is uniformly ergodic, it does not follow that E can be covered by a finite, even countable collection of 1-small sets.*

As a counterexample, consider the Random Scan Gibbs Sampler with uniform target on $E = [0,1] \times [0,1]$. Here the kernel P is given explicitly by

$$P(x_1, x_2; dy_1 \times dy_2) = \frac{1}{2}\big(\delta_{x_1}(dy_1)dy_2 + \delta_{x_2}(dy_2)dy_1\big).$$

Since for $(x_1, x_2) \neq (x'_1, x'_2) \in [0,1] \times [0,1]$, $P(x_1, x_2; \cdot)$ and $P(x'_1, x'_2; \cdot)$ are mutually singular measures unless $x_1 = x'_1$ or $x_2 = x'_2$, it is impossible to partition E into even a *countable* collection of 1-small sets. However, E is still small (hence P is uniformly ergodic), for in two steps, every part of the state space can be reached:

$$P^2(x_1, x_2; dy_1 \times dy_2) \geq \frac{1}{2}dy_1 \times dy_2 = \epsilon\nu(dy_1 \times dy_2).$$

In this counterexample, P did not have a density with respect to some σ-finite product measure on E. However, the random scan Gibbs sampler on a discrete state space $E = \mathbb{Z}^2$ say does have a joint density $p(x,y)$ and still there is only a countable covering by 1-small sets. Note that with a strictly positive continuous density and a locally compact state space E, there always is at least a *countable* covering by 1-small sets (see Corollary below).

Intuitively, what is needed for a finite covering by 1-small sets is that the chain be able to do *large* jumps in the state space. We offer the following sufficient condition (which implies that the whole state space E is 1-small):

Corollary 4. *Suppose that $P(x, dy)$ has a strictly positive continuous density with respect to some σ-finite reference probability μ on E. If Z is compact, or Z is only locally compact and the limit $p(\infty, y) := \lim_{z\to\infty} p(z, y) > 0$ exists and defines a probability density, then*

$$\mathbb{P}\Big(\mathfrak{X}_n \text{ has finite range in finite time}\Big) = 1.$$

Proof. If Z is locally compact, we add a point at infinity, making it compact. Thus we can always assume that Z is compact. Now by the strict positivity of the density $p(x, y)$ the sets $U_n = \{(z, y) : p(z, y) \in (1/n, 2/n)\}$ entirely cover the

compact set $Z \times Z$. Hence there exists a finite subcovering $U_{n_1}, \ldots, U_{n_k}$ say, and the sets $C_j = \{z : (z,y) \in U_{n_j}\}$ for $j = 1, \ldots, k$ are 1-small and cover Z:

$$P(z, dy) = p(z, y)\mu(dy) \geq \frac{1}{n} 1_{U_{n_j}}(y)\mu(dy)$$
$$= \epsilon_j \nu_j(dy) \text{ if } z \in C_j, \quad j = 1, \ldots, k.$$

$\square$

To illustrate the Corollary, we suppose that P is the transition kernel associated with the random field on $Z = E = \mathbb{R}$ given by

$$\mathfrak{X}_0(z) = z + \sigma^2(z)\pi'(z)/2\pi(z) + \sigma(z)W, \quad W \sim \mathcal{N}(0,1),$$

where we specify π and σ as follows: Let $\pi(x)$ be some function such that $\pi(x) \sim \exp(-\alpha |x|^\beta)$ as $|x| \to \infty$, where α and β are strictly positive real numbers. Taking any sequence $\gamma_s \geq 0$ such that $\gamma_s + \beta > 2$ for all s, let σ satisfy

$$\sigma^2(x)/(-\log\pi(x))^{\gamma_s/\beta} \to 1 \text{ as } |x| \to \infty.$$

It is shown in (Stramer and Tweedie, 1997) that the kernel P associated with the field $\mathfrak{X}_0$ is uniformly ergodic, and it is left to the reader to check that the assumptions of the corollary above are verified. Thus Theorem 2 applies to the "seed" field $\mathfrak{X}_0$.

Example 2. In this example, we present a version of the chain (4) which allows a range of proposal distributions, thereby taking the shape of $p(z, \cdot)$ better into account. Referring back to (4), we suppose given a family $\theta_z(y)$ of proposal distributions. Typical choices of $\theta_z(\cdot)$ would be "close" to $p(z, \cdot)$. We will need a selection mechanism, to decide which distribution θ_{z_0} is to be used at every step of the algorithm. The simplest choice is to take z_0 distributed independently on Z according to some distribution $\eta(\cdot)$, perhaps assigning higher weight to highly variable or otherwise difficult regions of the field. We call *Independence Field Coupler* the following Markov chain $(\mathfrak{X}_n)$ on $\mathfrak{F}(Z; E)$: Given $\mathfrak{X}_n$, draw independently $Y^{(n+1)} := Y \sim \eta(\cdot)$ and set

(11)

$$\mathfrak{X}_{n+1}(z) = \begin{cases} \Phi_Y \sim \theta_Y(\cdot)\mu(\cdot) & \text{if } p(z, \Phi_Y)\theta_Y\big(\mathfrak{X}_n(z)\big) > \xi \cdot p\big(z, \mathfrak{X}_n(z)\big)\theta_Y(\Phi_Y), \\ \mathfrak{X}_n(z) & \text{otherwise.} \end{cases}$$

Note that we can think of the object $\big(Y^{(n)}, \mathfrak{X}_n(z) : z \in Z\big)$ as a new random field, where the index set is now $Z \cup \{z_0\}$ for some isolated point $z_0 \notin Z$, and $\mathfrak{X}_n(z_0) := Y^{(n)}$. With this notation, we can fit the rule (11) in the framework (4) (exercise for the reader).

The auxiliary variables $Y^{(n)}$ are IID, and given $\mathfrak{X}_n \in \Pi$ (inductive hypothesis), we also have

$$
\begin{aligned}
\mathbb{P}\big(\mathfrak{X}_{n+1}(z) \in dx\big) &= \int \mathbb{P}\big(\mathfrak{X}_{n+1}(z) \in dx \,|\, Y = y\big)\eta(dy) \\
&= \int p(z,x)\mu(dx)\eta(dy), \quad \text{using (5)} \\
&= p(z,x)\mu(dx),
\end{aligned}
$$

since if $Y = y$ is given, we just have the previous Simple Field Coupler (6) with $\theta(\cdot) = \theta_y(\cdot)$. Thus clearly $\mathfrak{X}_{n+1} \in \Pi$ if $\mathfrak{X}_n \in \Pi$. We can now repeat the arguments of the previous example, yielding identical theorems for this new Field Coupler if we simply replace (8) with the generalization

$$
(12) \qquad \epsilon_i \nu_i(\cdot) \leq p(z,\cdot) \leq \gamma_i \theta_y(\cdot), \quad z,y \in C_i, \text{ for } i = 1, \ldots, N.
$$

Let us assume this done.

The construction (11), while an improvement over (6), may suffer another defect. Due to the choice of auxiliary variable Y independently of the current state of the field $\mathfrak{X}_n = \big(\mathfrak{X}_n(z) : z \in Z\big)$, we can expect some redundancy in the proposed values $\Phi^{(1)}$, $\Phi^{(2)}$, ... when $Y^{(n)}$ occurs in a location sufficiently close to a previous value $Y^{(k)}$ $(k < n)$, whose associated proposal $\Phi^{(k)}$ was accepted. In our simulations, this appeared to matter only when the choices of θ_z and η were 'bad'. Clearly the scope for further extensions of the Field Coupler $(\mathfrak{X}_n)$ is limitless.

3. Generating finitely coupled fields

The Field Couplers of the previous section are straightforward to implement, and computer experiments suggest that a small number of iterations suffice to obtain a finite range. Here we mean of course that the Markov chain $\mathfrak{X}_n$ must be simulated for a small but fixed (non-random) number of iterations, starting from an uncoupled version of the field. However, for any n, the realized field $\mathfrak{X}_n$ will only ever be finitely coupled within some $\epsilon > 0$.

For application to simulation problems, it is desirable to implement a procedure which will output *with certainty* a finitely coupled version of the field of interest. Here, we describe how this can be accomplished in general, while in the next section, some concrete Markov Chain Monte Carlo examples will be discussed.

The difficulty encountered when trying to output a finitely coupled field using the Markov chain $\mathfrak{X}_n$ is that, even though $\mathfrak{X}_0 \in \Pi$ and $\mathfrak{X}_n \in \Pi$ for all deterministic times $n \geq 0$ (recall that Π was defined in (3)), if we stop the chain at some random time τ by observing the field $(\mathfrak{X}_n(z) : z \in Z)$ and deciding it has fully coupled, the result $\mathfrak{X}_\tau$ no longer belongs to Π in general. To illustrate, consider the following naive procedure.

Faulty reasoning. *Suppose that Z is compact and that $p(z,y)$ and $\theta_z(y)$ are continuous. It is often natural to start with an initial field $\mathfrak{X}_0$ such that $z \mapsto \mathfrak{X}_0(z)$ is a.s. continuous. Consider a fixed $z \in Z$. If T_z denotes the first time that the chain $\mathfrak{X}_n(z)$ jumps (this happens when (9) occurs), it will also be the case that $\mathfrak{X}_n(z')$ must jump for all z' in a sufficiently small open neighborhood U of z. Now 'freeze' this neighborhood, and continue coupling only the field $\big(\mathfrak{X}(z) : z \in Z \backslash U\big)$. Very quickly we end up with a function $\mathfrak{X}$ on Z having finite range, but does this belong to Π? For fixed z, the variable $\mathfrak{X}_{T_z}(z)$ is the first transition of the jump chain associated with $\mathfrak{X}_{\cdot}(z)$, and this is well known to have stationary distribution $(1 - r_z(\cdot))p(z, \cdot)\mu(\cdot)$, where $r_z(x) = \mathbb{P}\big(\mathfrak{X}_1(z) = x \,|\, \mathfrak{X}_0(z) = x\big)$. Since we began with $\mathfrak{X}_0(z) \sim p(z, \cdot)\mu(\cdot)$, an easy calculation shows that we must have*

$$\mathbb{P}\Big(\mathfrak{X}_{T_z}(z) \in dy\Big) = \int p(z,x)\big(1 - r_z(x)\big)^{-1}\mathbb{P}\Big(\mathfrak{X}_1(z) \in dy \,|\, \mathfrak{X}_0(z) = x\Big)\mu(dx).$$

In general, stopping times τ for which $\mathfrak{X}_\tau(z) \sim P(z, \cdot)$ are hard to find. If both $\mathfrak{X}_\tau \in \Pi$ and $\mathfrak{X}_\tau$ is independent of τ, then τ is known as a *strong stationary time*.

There is another method to generate a finitely coupled field in finite time, by going backwards in time rather than forward. The details will occupy the rest of this section. Throughout, we hypothesize that Z is covered by a finite number of 1-small sets $C_1, \ldots, C_N$, and that (12) holds. We now briefly remind the reader of the Propp and Wilson formalism for *perfect simulation* by Markov chains, duly adapted to our requirements.

The equations (6) and (11) can be viewed as implicitly defining a sequence of independent random functions $G_n : \mathfrak{F}(Z; E) \times \Omega \to \mathfrak{F}(Z; E)$ so that $\mathfrak{X}_{n+1}(\omega) = G_n(\mathfrak{X}_n(\omega), \omega)$. If function composition occurs in the first variable, then we have equivalently

$$\mathfrak{X}_n(\omega) = \big(G_n \circ G_{n-1} \circ \cdots \circ G_1\big)(\mathfrak{X}_0(\omega), \omega).$$

It is convenient to extend the family (G_n) for values $n = 0, -1, -2, -3, \ldots$ so that we may think of the chain $\mathfrak{X}_n$ as evolving from the distant past.

By (5), it is clear that for any $\mathfrak{u} \in \mathfrak{F}(Z; E)$,

(13) $$\big(G_n \circ G_{n-1} \circ \cdots \circ G_1\big)(\mathfrak{u}) \in \Pi \quad \text{if } \mathfrak{u} \in \Pi.$$

But the maps G_n are exchangeable, so if we write $\mathfrak{Y}_n(\mathfrak{u}) = G_0 \circ G_{-1} \circ \cdots \circ G_{-n}(\mathfrak{u})$ for $\mathfrak{u} \in \Pi$, then by (13) we have automatically

(14) $$\lim_{n \to \infty} \mathfrak{Y}_n(\mathfrak{u}) \in \Pi, \quad \text{if } \mathfrak{u} \in \Pi \text{ and the limit exists.}$$

Suppose moreover that we can prove (as we do in Theorem 5) that the random time

$$T = \inf\big\{n : \mathfrak{Y}_n(\cdot) \text{ is constant on } \Pi\big\}$$

is a.s. finite. In that case, we have

(15) $$\mathfrak{Y}_n(\cdot) = \mathfrak{Y}_T \circ G_{T-1} \circ \cdots \circ G_{-n}(\cdot) = \mathfrak{Y}_T(\cdot) \quad \text{on } \{T \le n\}.$$

This fact, first noted by Propp and Wilson (1996) in a different context, forms the basis for their perfect simulation method. In our setting, it suffices now to construct a random time $T' \ge T$ for which we can conveniently test if $\{T' \le n\}$ has occurred. By (14) and (15) we must then have $\mathfrak{X}_0 \equiv \mathfrak{Y}_n(\cdot) \in \Pi$, a perfect specimen of a finitely coupled field. We shall do this now for the algorithm (11), as (6) can be viewed as the special case when $\theta_y(\cdot) \equiv \theta(\cdot)$ for all $y \in Z$.

Intuitively, we shall think of the field $\mathfrak{X}_n$ defined by (11) as a random point pattern on $Z \times E$ (the graph of the map $z \mapsto \mathfrak{X}_n(z)$). One iteration of the Independence Field Coupler (11) on the field $\mathfrak{X}_n$ consists in the deletion of a section of the point pattern, replacing it by a new pattern. If we consider two initially distinct point patterns $\mathfrak{u}$ and $\mathfrak{u}'$, then for any ℓ, the patterns $G_\ell(\mathfrak{u})$ and $G_\ell(\mathfrak{u}')$ will agree at all points that were newly added, provided these had been accepted by

both patterns. Thus we can expect that for all sufficiently large n, all the points making up the patterns $\mathfrak{u}$ and $\mathfrak{u}'$ originally have disappeared from the patterns $G_0 \circ G_{-1} \circ \cdots \circ G_{-n}(\mathfrak{u})$ and $G_0 \circ G_{-1} \circ \cdots \circ G_{-n}(\mathfrak{u}')$, which will therefore agree: $\mathfrak{Y}_n(\mathfrak{u}) = \mathfrak{Y}_n(\mathfrak{u}')$.

We now construct the stopping time $T' \geq T$. We assume that (12) holds. To generate a finitely coupled field,

---

(FC-1): Generate IID sequences $(\xi^{(0)}, Y^{(0)}, \Phi^{(0)}), (\xi^{(1)}, Y^{(1)}, \Phi^{(1)}), \ldots$ where

(16) $\qquad \xi^{(k)} \sim U[0,1], \quad Y^{(k)} \sim \eta(\cdot), \Phi^{(k)} \sim \theta_{Y^{(k)}}(\cdot)\mu(\cdot), \quad k = 0, 1, 2, 3, \ldots$

(FC-2): Let $T'_k = \min\big\{s \geq 0 : \epsilon_k \nu_k(\Phi^{(s)})/\theta_{Y^{(s)}}(\Phi^{(s)}) > \gamma_k \xi^{(s)}\big\}$ and set $T' = \max\{T'_k : k = 1, \ldots, N\}$.

(FC-3): For fixed n, define the random map G_{-n} from $(\xi^{(n)}, Y^{(n)}, \Phi^{(n)})$ by: if $\mathfrak{u} = (\mathfrak{u}(z) : z \in Z)$ is a field,

$$
\big(G_{-n}(\mathfrak{u})\big)(z) = \begin{cases} \Phi^{(n)} & \text{if } p\big(z, \Phi^{(n)}\big)\theta_{Y^{(n)}}\big(\mathfrak{u}(z)\big) > \xi^{(n)} p\big(z, \mathfrak{u}(z)\big)\theta_{Y^{(n)}}\big(\Phi^{(n)}\big) \\ \mathfrak{u}(z) & \text{otherwise.} \end{cases}
$$

Note immediately that $\big(G_{-T'_k}(\mathfrak{u})\big)(z)$ is independent of $\mathfrak{u}$ for $z \in C_k$, and hence so is $\big(\mathfrak{Y}_n(\mathfrak{u})\big)(z)$ for all $n \geq T'_k$. Consequently, $\mathfrak{Y}_{T'}(\mathfrak{u})$ is a random field which is independent of $\mathfrak{u}$. Moreover, $\mathbb{P}(T' < \infty) = 1$ as follows from Theorem 5 below.

(FC-4): If $\mathfrak{u} \in \Pi$, then $\mathfrak{X} = G_0 \circ G_{-1} \circ \cdots \circ G_{-T'}(\mathfrak{u})$ is finitely coupled and belongs to Π. We output $\mathfrak{X}$.

To summarize, we have the following result:

Theorem 5. *Let Z be coverable by a finite number of 1-small sets $C_1, \ldots, C_N$ such that (12) holds. There exists a finitely coupled field $\mathfrak{X}$ corresponding to the kernel P which a.s. can be constructed in a finite number of iterations of the algorithm (FC-1, ... ,FC-4).*

Proof. All we need to check is that $\mathbb{P}\big(T' < \infty\big) = 1$. Now each T'_k defined in (FC-2) is binomially distributed with strictly positive success probability, from which the claim follows. $\qquad \square$

At each iteration, we can test very simply whether T' has occurred or not. The number $T' - T$ of iterations "wasted" will depend upon the overall quality of the estimates (12). We shall describe some of our findings in the next section.

Note also that the total number of values in the range of $\mathfrak{X}$ is bounded by T (hence by T', which has geometric tails).

The algorithm (FC-1, ... ,FC-4) given above is not *interruptible*. Thus it is possible that a particular run to produce a finitely coupled field $\mathfrak{X}$ can take a large number of computations. If we stop the run when a fixed number n_0 of iterations has been reached, and begin afresh, we are biasing the output $\mathfrak{X}$ in favour of fields which require at most n_0 iterations to generate. Since n_0 is also the maximum number of distinct values contained in the range of $z \mapsto \mathfrak{X}(z)$, we are effectively biasing in favour of a bounded range. While this is a potential problem, the stopping time T' will often be very small unless the estimates (12) are poor.

Fill (1997) proposed another method for perfect simulation of a target probability density, based on rejection sampling. This has the advantage of *being* interruptible. However, the chain $(\mathfrak{X}_n, Y^{(n)})$ is not reversible on Π (since the set of finitely coupled fields is absorbing) and a modification of Fill's method for Field Coupling appears more complicated.

4. Flows and Markov chains: coupling two chains

For the remainder of this paper, we shall take $Z = E$. We are given a kernel $P(x, dy)$ on E representing the transition function of some Markov chain. For any sequence of random fields $\mathfrak{X}^{(1)}, \mathfrak{X}^{(2)}, \cdots \in \Pi$, we can define a corresponding *flow* $F_{s,t} : E \to E$,

$$(17) \qquad F_{s,t}(x) = \mathfrak{X}^{(t-1)} \circ \cdots \circ \mathfrak{X}^{(s)}(x), \quad s < t, \quad x \in E.$$

Then if we set $X_t(x) = F_{0,t}(x)$, we have simulateously realized all Markov chains on E with transition function P and deterministic initial condition. It is common in simulation studies to take the random fields in the form $\mathfrak{X}^{(k)}(z) = f(z, \zeta_k)$, where f is some measurable function and (ζ_k) is an IID sequence. This has been called a Stochastic Recursive Sequence (SRS) construction (Borovkov and Foss, 1992). It should be noted however that it is by no means necessary to take the *same* function f for each k, so long as the resulting field $\mathfrak{X}^{(k)}$ always belongs to Π.

We shall now discuss briefly an application of the Field Couplers to the problem of coupling two such Markov chains, $X_t(x_1)$ and $X_t(x_2)$ say, both with the same transition kernel P.

Recall first that a *coupling time* here is a random time $T \geq 0$ such that $X_T(x_1) = X_T(x_2)$. Since both chains are built from the same flow, this also means that $X_{T+t}(x_1) = X_{T+t}(x_2)$ for all $t > 0$. The coupling time is called *successful* if $\mathbb{P}(T < \infty) = 1$. The existence of a successful coupling time is by no means guaranteed, and obvioulsy depends on the choice of flow $F_{s,t}$.

Some flows do not have successful coupling times, even for uniformly ergodic kernels P. As an example, consider the random walk on the unit circle $S^1 = \{e^{i\theta} : \theta \in \mathbb{R}\}$, given by the flow

$$F_{t,t+1}\big(\exp(ix)\big) = \exp i(x + W_t), \quad W_t \sim \mathcal{N}(0,1).$$

The Markov chain $X_t(e^{ix})$ is aperiodic and uniformly ergodic, with stationary distribution given by the Lebesgue measure on S^1. From the description of $F_{t,t+1}$ it is also clear that two chains $X_t\big(\exp(ix_1)\big)$ and $X_t\big(\exp(ix_2)\big)$ evolve in parallel, and can never couple unless $x_1 = x_2(\mod 2\pi)$.

The above example does not preclude the existence of some alternative flow $\widetilde{F}_{t,t+1}$ which will allow a successful coupling to take place (indeed, the uniform ergodicity guarantees this existence). Now given *any* flow of the type in (17), we claim that the Field Couplers can generate such a new flow $\widetilde{F}_{s,t}$, for which the coupling of the associated chains $\widetilde{X}_t(x_1)$ and $\widetilde{X}_t(x_2)$ is successful.

Indeed, suppose that for each s, we iterate the Simple Field Coupler n times with seed $\mathfrak{X}^{(s)}$, obtaining a new field $\mathfrak{X}_n^{(s)} \in \Pi$. We define the flow

(18) $$\widetilde{F}_{s,t}(z) = \mathfrak{X}_n^{(t-1)} \circ \cdots \circ \mathfrak{X}_n^{(s)}(z),$$

and corresponding Markov chains $\widetilde{X}_t(z) = \widetilde{F}_{0,t}(z)$. We shall consider fields $\mathfrak{X}_n^{(s)}$ generated by the Simple Field Coupler only here, for simplicity and because the Independence Field Coupler can be treated in a very similar way.

The following result shows that if the proposal density θ has heavy enough tails, then coupling of $\widetilde{X}_t(x_1)$ and $\widetilde{X}_t(x_2)$ is always successful for uniformly ergodic kernels P. This will be used in the next section.

Note that the coupling method proposed here does not require any explicit computations of minorization conditions, merely the knowledge that they exist. This is unlike the classical coupling method based on splitting, whereby two chains both have to enter an explicitly defined small set C simultaneously, right before coupling.

The proof below also shows that if P is only positive Harris recurrent, then coupling is successful with strictly positive probability.

Theorem 6. *Let $F_{s,t}$ be an arbitrary flow for P, and suppose that the target density $p(x,y)$ and proposal density $\theta(y)$ satisfy $p(z,y) < \gamma(z)\theta(y)$ for some function $\gamma(z)$ which is bounded on a 1-small set C. If P is aperiodic and uniformly ergodic, then*

$$\mathbb{P}\big(\widetilde{X}_t(x_1) = \widetilde{X}_t(x_2) \text{ eventually }\big) = 1 \text{ for all } x_1, x_2 \in E.$$

Proof. Since $F_{s,t}$ is arbitrary, we shall assume that (18) holds with $n = 1$. Recall that we denote by $X_s(x)$ the MC started from x and constructed from the original flow $F_{s,t}$, and by $\widetilde{X}_s(x)$ the MC started from x constructed from the new flow $\widetilde{F}_{s,t}$ obtained after iterating the Simple Field Coupler. We shall analyse the law of

$$\widetilde{T}(x_1, x_2) = \min\{s \geq 0 : \widetilde{X}_s(x_1) = \widetilde{X}_s(x_2)\}.$$

Since P is uniformly ergodic, there exist m and $\alpha > 0$ such that $\mathbb{P}(X_m(x) \in C) \geq \alpha$ holds for all $x \in E$. Denote by $H(y)$ the first exit time of the chain $(\Phi^{(s-1)}, \Phi^{(s)}, \xi^{(s)})$, started initially at $(y, \Phi^{(1)}, \xi^{(1)})$, from the set $\{(u,v,w) : p(u,v) > w\gamma(u)\theta(v)\}$. If $H(y) > m$, then because the proposal $\theta(\cdot)$ dominates $p(z, \cdot)$, we must have

$$p(y, \Phi^{(1)}) > \gamma(y)\xi^{(1)}\theta(\Phi^{(1)}),$$

$$p(\Phi^{(1)}, \Phi^{(2)}) > \gamma(\Phi^{(1)})\xi^{(2)}\theta(\Phi^{(2)}),$$

$$\vdots$$

$$p(\Phi^{(m-1)}, \Phi^{(m)}) > \gamma(\Phi^{(m-1)})\xi^{(m)}\theta(\Phi^{(m)}),$$

and consequently $\widetilde{X}_1(y) = \Phi^{(1)}$, $\widetilde{X}_2(y) = \Phi^{(2)}$, ... ,$\widetilde{X}_m(y) = \Phi^{(m)}$. In particular, either $\widetilde{X}_m(x) = \widetilde{X}_m(y) = \Phi^{(m)}$ (when $\widetilde{T}(x,y) \leq m$ holds), or else $\widetilde{X}_s(x) \neq \Phi^{(s)}$ for all $s \leq m$ (when $\widetilde{T}(x,y) > m$ holds), and then $\widetilde{X}_s(x) = X_s(x)$.

Now consider the event $B_{m+1} = \Big\{\epsilon\nu\big(\Phi^{(m+1)}\big) > \max_{z \in C} \gamma(z)\xi^{(m+1)}\theta\big(\Phi^{(m+1)}\big)\Big\}$. When this holds, because the proposal $\theta(\cdot)$ dominates $p(z, \cdot)$, all chains which satisfy

$\widetilde{X}_m(x) \in C$ must couple,i.e. $\widetilde{X}_{m+1}(x) = \Phi^{(m+1)}$. Consequently,

$$\mathbb{P}\big(\widetilde{T}(x,y) \le m + 1\big) \ge \mathbb{P}\big(X_m(x) \in C, \widetilde{X}_m(y) \in C, H(y) > m, B_{m+1}\big)$$
$$= \mathbb{P}\big(X_m(x) \in C\big)\mathbb{P}\big(\Phi^{(m)} \in C, H(y) > m\big)\mathbb{P}(B_{m+1}).$$

Let A be a set of strictly positive stationary measure and $\beta > 0$ a number such that $\inf_{y \in A} \mathbb{P}(\Phi^{(m)} \in C, H(y) > m) > \beta$ holds. Such a set must exist, for otherwise we would have $\widetilde{F}_{s,s+1} \equiv F_{s,s+1}$ a.s. We estimate

$$\inf_{y \in A} \mathbb{P}\big(\widetilde{T}(x,y) \le m + 1\big) \ge \alpha\beta\mathbb{P}(B_{m+1}) = \delta > 0,$$

and note that this is independent of x. To finish the proof, observe that if we begin with two chains $\widetilde{X}_s(x_1)$ and $\widetilde{X}_s(x_2)$ such that $x_1 \ne x_2$, then because $\widetilde{X}_s(x_2)$ is positive recurrent, it must enter the set A infinitely often. Each time this occurs, regardless of the current location x of the chain $\widetilde{X}_s(x_1)$, there exists at least δ probability that the two chains couple within $m + 1$ steps. The Borel-Cantelli lemma therefore gives $\mathbb{P}(\widetilde{T}(x_1, x_2) < \infty) = 1$. □

In case the kernel density $p(x, y)$ and proposal $\theta(y)$ are continuous, on a locally compact state space E, a sufficient condition for the existence of $\gamma(z)$ is that $\lim_{y \to \infty} p(z, y)/\theta(y) = 0$ for all z.

Theorem 6 is also valid for the Independence Field Coupler (defined in Section 2, Example 2). In this case, since there is a range of proposal densities $\theta_z(\cdot)$ to choose from, the hypothesis can be weakened as follows: There exists some function $\gamma(z, w)$ which is bounded on a set $C \times D$ such that $p(z, y) < \gamma(z, w)\theta_w(y)$, with C small and D of positive η-measure.

5. Coupling many Markov chains

In this section, we are interested in the question whether the whole collection $X_t(x) : x \in E$ of Markov chains can be coupled in a common finite time. We already know from Theorem 6 that any two given chains $X_t(x_1)$ and $X_t(x_2)$ couple successfully in a time $T(x_1, x_2) < \infty$ a.s. , but the possibility remains that $\sup_{x_1, x_2} T(x_1, x_2) = \infty$. Indeed, it was shown by Foss and Tweedie (1997) that there exists *some* flow $F_{s,t}$ corresponding to P such that $\sup_{x_1, x_2} T(x_1, x_2) < \infty$ if and only if P is *uniformly ergodic*. Thus we must here too restrict ourselves to

such kernels P. We would like to investigate the case that $F_{s,t}$ is fixed to be of the form (18) with $n \geq 1$.

Definition 7. *Let $F_{s,t}$ be a flow for the transition kernel P on the state space E. We say that $F_{s,t}$ collapses in finite time if*

$$\mathbb{P}\big(x \mapsto F_{s,t}(x) \text{ is constant for all large values of } t\big) = 1.$$

We say that $F_{s,t}$ thins at time $r > s$ if

$$\left|\{F_{s,r+1}(x) : x \in E\}\right| < \left|\{F_{s,r}(x) : x \in E\}\right|.$$

As a direct consequence of the previous sections, we can state

Theorem 8. *Let P be uniformly ergodic with a density $p(x,y)$ which satisfies (8) for some collection $C_1, \ldots, C_N$ of 1-small sets covering E. If $F_{s,t}$ is any random flow with one-step transition probability P, then the flow $\widetilde{F}_{0,t}$ defined by (18) using the Simple Field Coupler with $n \geq N$ (where $\mathfrak{X}^{(t)} = F_{t,t+1}$) collapses in finite time.*

Proof. Applying the Simple Field Coupler $n \geq N$ times ensures that the random field $\mathfrak{X}_n^{(s)}$ is finitely coupled within some $\epsilon > 0$. Consequently, the Borel-Cantelli lemma implies that the field $\widetilde{F}_{0,t_0}$ has finite range a.s. , for some finite t_0. Now using Theorem 6 ensures that the flow $\widetilde{F}_{0,t_0+t}$ thins repeatedly until collapse. $\square$

Variations on the above theme are possible. For example, we can generate (perfectly) finitely coupled fields $\mathfrak{X}^{(s)}$ and use these to construct $\widetilde{F}_{s,t}$. This is discussed below for Metropolis-Hastings chains.

In principle, we can now collapse the flows of any uniformly ergodic Markov chains, provided the transition kernel P has a density. We give two examples.

Example: Random Walk. Let $P(z, dy) = q(y - z)dy$, where $q(\cdot)$ is a continuous, not necessarily symmetric density on $\mathbb{R}^d$. A random field belonging to Π is easily found, namely

$$\mathfrak{X}(z) = z + W, \quad \text{where } W \sim q(y)dy.$$

We propose to find a finitely coupled version of $\widetilde{\widetilde{\mathfrak{X}}}$ on a compact subset $Z \subset \mathbb{R}^d$. We cover Z by a finite number of closed balls $C_i = B(z, r_i)$ say, $i = 1, \ldots, N$, and such that $\epsilon_i \nu_i(y) = \inf_{z \in C_i} q(y - z)$ is not identically zero. Let g be the (continuous)

density of a distribution on $\mathbb{R}^d$ with heavier tails than q. We will need to assume that $q(\cdot) \leq Cg(\cdot)$ for some constant C. Then put

$$\theta_z(y) \equiv \theta_i(y) := g(y - z_i) \quad \text{if } z \in C_i = B(z_i, r_i),$$

and calculate the constants $\gamma_i = \max_{z \in C_i, y \in \mathbb{R}^d} q(y - z)/g(y - z_i)$. Finally, take $Y \sim \eta(\cdot)$, where η is the uniform distribution on the set of points $\{z_1, \ldots, z_N\}$. Now applying (FC-1, $\ldots$, FC-4), we obtain a new field $\widetilde{\mathfrak{X}}(z) \sim q(y - z)dy$ which is a.s. finitely coupled on Z.

Note that we have not assumed that $q(\cdot)$ is symmetric. For symmetric and unimodal increment distributions, an alternative field coupling method for $\mathfrak{X}$ exists, proposed by Murdoch and Green (1998). Their method, called the *Bisection Coupler*, uses translations and reflections to construct a finitely coupled field, and appears to require fewer computations in general. However, while the number of points in the range of the resulting field is almost surely finite, they show that it has infinite expectation. In contrast, for the field coupler presented here, the range of the field $\widetilde{\mathfrak{X}}$ never contains more points than the number of iterations T' required to generate it, which as a random number has geometric tails.

Example: Time-Discretized Diffusion. On the interval $Z = [0, 1]$, consider the Langevin random field

(19) $$\mathfrak{X}(z) = z + \sigma W + \sigma^2 \pi'(z)/2\pi(z),$$

where $\pi(z)$ is a strictly positive differentiable function on Z and W is a standard Gaussian random variable. Let us write

(20) $$\widetilde{p}(z, y) = \exp\left(-\left|y - z - \sigma^2 \pi'(z)/2\pi(z)\right|^2 / 2\sigma^2\right)$$

for the unnormalized density of $\mathfrak{X}(z)$. None of our calculations requires the normalization constant. For $i = 1, \ldots, N$, we let $C_i = [(i - 1)/N, i/N]$ and calculate $a_i = \inf_{z \in C_i}\left(z + \sigma^2 \pi'(z)/2\pi(z)\right)$ and $b_i = \sup_{z \in C_i}\left(z + \sigma^2 \pi'(z)/2\pi(z)\right)$. In other words, the interval $[a_i, b_i]$ must contain the means of the distributions $q(z, \cdot)$ for all $z \in C_i$. For the proposal $\Theta_z(x, \cdot)$, let us take (independently of x) the density

proportional to

$$\theta_i(y) = \begin{cases} \exp\left[-\left|y - a_i\right|^2 / 2\sigma^2\right] & \text{if } y \le a_i, \\ 1 & \text{if } y \in [a_i, b_i], \\ \exp\left[-\left|y - b_i\right|^2 / 2\sigma^2\right] & \text{if } y \ge b_i. \end{cases}$$

It is immediately seen that $\widetilde{p}(z, \cdot) \le \theta_i(\cdot)$ for all $z \in C_i$, and this inequality is sharp. Also, take $\epsilon_i = 1$ and

$$\nu_i(y) = \begin{cases} \exp\left(-\left|y - b_i\right|^2 / 2\sigma^2\right) & \text{if } y \le (a_i + b_i)/2 \\ \exp\left(-\left|y - a_i\right|^2 / 2\sigma^2\right) & \text{if } y \ge (a_i + b_i)/2 \end{cases}$$

which completes the series of estimates (12). We choose Y uniformly in $\{1, \ldots, N\}$, or according to some other distribution adapted to the gradient $\pi'(z)/\pi(z)$ if desired. Since

$$\nu_i(y)/\theta_i(y) = \begin{cases} \exp -\frac{1}{2\sigma^2}\left(a_i^2 - b_i^2 + 2y(b_i - a_i)\right) & \text{if } y \le a_i \\ \nu_i(y) & \text{if } y \in [a_i, b_i] \\ \exp -\frac{1}{2\sigma^2}\left(b_i^2 - a_i^2 + 2y(a_i - b_i)\right) & \text{if } y \ge b_i, \end{cases}$$

the probability $v_i := \mathbb{P}\big(Y \text{ falls in } C_i \text{ and } \nu_i(\Phi)/\theta_i(\Phi) > \xi\big)$ is easy to calculate, and then we can write down explicitly the distribution of T', the time until termination of the field coupling procedure (FC-1, $\ldots$,FC-4). This could then be optimized in N, the total size of the covering $C_1, \ldots, C_N$.

A similar analysis can be performed for any random field on $\mathbb{R}^d$ of the form

$$\mathfrak{X}(z) = \left(\boldsymbol{b}_i(z) + \sum_{j=1}^d \boldsymbol{\sigma}_{ij}(z)W_j, \; i = 1, \ldots, d\right),$$

where $\boldsymbol{b}(z)$ is a vector field on $\mathbb{R}^d$, $\boldsymbol{\sigma}$ is a matrix-valued field on $\mathbb{R}^d$ and $(W_1, \ldots, W_d)$ is a standard d-dimensional Gaussian random variable. Such fields arise in the recent work by Stramer and Tweedie (1997), as examples of natural uniformly ergodic Metropolis-Hastings proposals on $\mathbb{R}^d$.

Metropolis-Hastings type chains are a class of Markov chains whose transition kernels do not possess a density (unless the state space is countable). Their general form is (cf. Roberts and Smith, 1993)

$$(21) \qquad P(x, dy) = q(x, y)\left\{1 \wedge \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right\}\mu(dy) + r(x)\delta_x(dy),$$

where $r(x)$ is chosen so that $P(x, E) = 1$ and $q(x, y)\mu(dy)$ is the transition density of some arbitrary *proposal* chain. While it is possible to modify the Field Couplers to work on transition kernels of this form (this requires the proposal $\Theta_z(x, \cdot) \gg P(x, \cdot)$), it is usually very impractical to evaluate the rejection probability $r(x)$, and this renders this method inapplicable.

Fortunately however, it is possible to proceed otherwise, as the proposal kernel $q(x, y)\mu(dy)$ is often analytically tractable. Common examples of such kernels were given as examples above, and the Langevin proposal field was explicitly analysed on a simple region.

Suppose now that we have constructed a random field $\mathfrak{Q}(z) \sim q(z, \cdot)\mu(\cdot)$. We shall have $\mathfrak{X}(z) \sim P(z, \cdot)$ with P given by (21) as soon as we set

$$(22) \qquad \mathfrak{X}(z) = \begin{cases} \mathfrak{Q}(z) & \text{if } \pi\big(\mathfrak{Q}(z)\big)q\big(\mathfrak{Q}(z), z\big) > \xi \cdot \pi(z)q\big(z, \mathfrak{Q}(z)\big), \\ z & \text{otherwise,} \end{cases}$$

where $\xi \sim U[0, 1]$. We may view the Metropolis-Hastings accept/reject step as an *operator* on fields, yielding $\mathfrak{X}$ when applied to $\mathfrak{Q}$. As such, it has some useful properties.

Suppose we generate an IID sequence of random fields $\mathfrak{Q}^{(1)}, \mathfrak{Q}^{(2)}, \ldots$, and then construct $\mathfrak{X}^{(1)}, \mathfrak{X}^{(2)}, \ldots$ by (22), using these to generate a flow $F_{s,t}(z) = \mathfrak{X}^{(t-1)} \circ \cdots \circ \mathfrak{X}^{(s)}(z)$ for the Markov chain with transition kernel P given by (21). The random ("rejection") sets $\{z : \mathfrak{X}^{(k)}(z) = z\} \subset Z$ are not usually empty, at least with high probability.

Proposition 9. *Suppose that $\mathfrak{Q}^{(1)}, \mathfrak{Q}^{(2)}, \ldots$ are finitely coupled proposal fields and define $\mathfrak{X}^{(1)}, \mathfrak{X}^{(2)}, \ldots$ as a Metropolis Hastings proposal, by (22). Then the random sets*

$$\{z : F_{0,t}(z) = z\} = \{F_t \text{ has finite range}\}^c$$

are strictly decreasing as functions of t. If the state space E is irreducible and coverable by a finite number of 1-small sets, then in fact $\{z : F_t(z) = z\} \downarrow \emptyset$ and the flow collapses in a finite time.

Proof. Since $\{z : F_{0,t}(z) = z\} = \{z : \mathfrak{X}^{(0)}(z) = z\} \cap \cdots \cap \{z : \mathfrak{X}^{(t-1)}(z) = z\}$, the sets are clearly decreasing. Moreover, for each $z \in E$, the Markov chain $\mathfrak{X}^{(k)}(z)$

must move eventually, and this occurs simultaneously for all z within a small set in at most geometrically distributed time. Since E is coverable by a finite number of such sets, we get $\{z : F_t(z) = z\} = \emptyset$ in finite time. After that, the map $F_t(\cdot)$ has only a finite range, and thus by Theorem 6 we get collapse eventually. $\square$

The result in Proposition 9 is useful for coupling procedures due to the identity $\{z : F_t(z) = z\}^c = \{F_t \text{ has finite range}\}$. This set must be monitored if we wish to detect whether $F_{0,t}(\cdot)$ has collapsed.

Note that if we do *not* assume that $\mathfrak{Q}^{(k)}$ is finitely coupled (only to within ϵ say), the monotonicity of the sets $\{z : F_t(z) = z\}$ is preserved, but the sets $\{F_t \text{ has finite range}\}$ are no longer their complements. Thus they can grow and shrink in complicated ways, which can make them untrackable for practical purposes.

6. Towards Perfect Simulation with Hastings-Metropolis chains

We have seen in the last section that if P defined by (21) is uniformly ergodic, then we can construct *numerically* a finite time coupling of all Markov chains $X_t(x)$ with transition kernel P and deterministic initial condition $X_0(x) = x \in E$. The exact dynamics of the coupling are no longer tractable, since in fact they are random. Instead, we gain an automatic method of coupling chains, which does not depend on our ability to find clever ways to couple the flows *analytically*.

For uniformly ergodic Metropolis-Hastings chains, we describe now how to perfectly simulate a given target distribution, using the Field Couplers described in this paper. A large class of uniformly ergodic M-H chains compatible with this method can be found in Stramer and Tweedie (1998).

(PMH-1): Choose a covering of E by small sets $C_1, \ldots, C_N$ and calculate the estimates (12).

(PMH-2): Using the steps (FC-1, ... ,FC-4), generate finitely coupled fields $(\mathfrak{Q}^{(k)} : k = -1, -2, \ldots)$. For every $\mathfrak{Q}_k$, keep track of the (finite) range of possible values over each C_j, $j = 1, \ldots, N$.

(PMH-3): Construct $\mathfrak{X}^{(k)}$ from $\mathfrak{Q}^{(k)}$ using formula (22) (with $\xi = \xi^{(k)}$), and for each C_j, given the range of values $\mathfrak{Q}^{(k)}(C_j) = \{q_1, \ldots, q_m\}$

```
say, test whether
```

(23) $\pi(q_r)q(q_r,z) > \xi^{(k)}\pi(z)q(z,q_r)$ `for all` $z \in C_j$, $\quad r = 1,\ldots,m.$

```
Let Bₖ denote the union of all sets Cⱼ for which the test (23)

fails.  Clearly {z : 𝔛⁽ᵏ⁾(z) = z} ⊂ Bₖ.
```
(PMH-4): `Let` $k < 0$ `be given.  Check that`

$$B_k \cap B_{k+1} \cap \cdots \cap B_1 = \emptyset.$$

```
If this holds, then typically there will be k < ℓ < 0 such that
```
$B_k \cap B_{k+1} \cap \cdots \cap B_\ell = \emptyset.$ `For each` $k \le n \le \ell$, `let` $\{q_1,\ldots,q_r\} = \mathfrak{Q}_n(E)$
```
denote the (finite) range of 𝔔⁽ⁿ⁾.  Compute the set
```

$$H_n = \left\{\mathfrak{X}^{(1)} \circ \cdots \circ \mathfrak{X}^{(n+1)}(q_i) : i = 1,\ldots,q_r\right\}.$$

```
If  Hₖ = Hₖ₊₁ = ⋯ = Hₗ = {p} say, then the value of p is a

perfect draw from π.  Stop.  Otherwise, try (PMH-4) with an

earlier value of k.
```

We do not discuss ways of implementing the procedure (PMH-1, ... ,PMH-4) efficiently, although it is clearly possible to exploit various recurrence relations.

Nothing more needs to be said of steps (PMH-1) and (PMH-2). For step (PMH-3), it is necessary to perform some analysis, determining a lower bound

(24) $$\inf_{z \in C_i} \pi(y)q(y,z)/\pi(z)q(z,y) \ge b_i(y), \quad i = 1,\ldots,N$$

so that we can perform the test (23) conveniently. Typically, this can be done by finding bounds $\pi(y)/q(z,y) \ge b_i'(y)$ and $\pi(z)/q(y,z) \le b_i''(y)$, taking $b_i(y) = b_i'(y)/b_i''(y)$.

The step (PMH-4) can be performed efficiently if we construct tables of the correspondences $C_j \mapsto \mathfrak{X}_1 \circ \cdots \circ \mathfrak{X}_n(C_j)$.

Example: Langevin Algorithm. We consider again the Langevin random field (19) on $Z = [0,1]$, this time considering it as a proposal field for the Metropolis-Hastings algorithm with target distribution π. Hence, up to normalization, we have $q(y,z) = \widetilde{p}(y,z)$, the latter being defined by (20). To implement the algorithm (PMH-1, ... ,PMH-4), we need only compute the bounds (24). Recall that we took

$C_i = [(i-1)/N, i/N]$. Let $f(z) = z + \sigma^2 \pi'(z)/2\pi(z)$ denote the mean of the density $q(z, \cdot)$. We have $f(z) \in [a_i, b_i]$ for all $z \in C_i$, consequently

$$\pi(y)/q(z,y) = \pi(y)\exp\left(|z - f(y)|^2/2\sigma^2\right)$$

$$\geq \begin{cases} \pi(y) & \text{if } \frac{i-1}{N} \leq f(y) \leq \frac{i}{N} \\ \pi(y)\min_{j=i,i-1}\exp\left(|j/N - f(y)|^2/2\sigma^2\right) & \text{otherwise,} \end{cases}$$

$$\pi(z)/q(y,z) = \pi(z)\exp\left(|y - f(z)|^2/2\sigma^2\right)$$

$$\leq \sup_{z \in C_i} \pi(z)\exp\left((|y - a_i|^2 \vee |y - b_i|^2)/2\sigma^2\right).$$

Note that it does not matter whether we know the normalizing constant of π or not, since we are now taking the ratio of these two quantities to get $b_i(y)$. In particular, these calculations can be done on any conveniently chosen multiple of π. Setting $\widehat{\pi}_i(y) = \pi(y)/\sup_{z \in C_i} \pi(z)$, we find

$$b_i(y) = \begin{cases} \widehat{\pi}_i(y)\exp\left(-|y - a_i|^2 \vee |y - b_i|^2\right)) & \text{if } \frac{i-1}{N} \leq f(y) \leq \frac{i}{N} \\ \widehat{\pi}_i(y)\exp\left(\min_{j=i,i-1}|j/N - f(y)|^2/2\sigma^2 \right. \\ \qquad \left. -|y - a_i| \vee |y - b_i|^2\right) & \text{otherwise.} \end{cases}$$

7. Acknowledgement

References

[1] Borovkov, A.A. and Foss, S.G. (1992) Stochastically recursive sequences and their generalizations. *Siberian Adv. Math.* **2**, 16-81.

[2] Fill, J.A. (1997) An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.* **7**.

[3] Foss, S. and Tweedie, R. L. (1998) Perfect simulation and backward coupling. Preprint.

[4] Green, P.J. and Murdoch, D.J. (1998) Exact sampling for Bayesian inference: towards general purpose algorithms. To appear in *Proceedings of the 6th Valencia meeting on Bayesian Statistics.*

[5] Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability.* Springer-Verlag, London.

[6] Propp, J. and Wilson, D.B. (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252.

[7] Roberts, G.O. and Rosenthal, J.S. (1996) Quantitative bounds for convergence rates of continuous time Markov processes, *Electronic Journal of Probability*, **1**.

[8] Roberts, G.O. and Smith, A.F.M. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *J. Roy. Stat. Soc. B* **55**, 1, 3-23.

[9] Stramer, O. and Tweedie, R.L. (1997) Self-targeting candidates for Metropolis-Hastings algorithms. Preprint.

DIPARTIMENTO DI MATEMATICA, UNIVERSITÀ DEGLI STUDI ROMA TRE, LARGO S. LEONARDO MURIALDO 1, 00146 ROMA, ITALY

E-mail address: `breyer@mat.uniroma3.it`

DEPARTMENT OF MATHEMATICS AND STATISTICS, LANCASTER UNIVERSITY LA1 4YF

E-mail address: `g.o.roberts@lancaster.ac.uk`