

A Note on Formal Constructions of Sequential Conditional Couplings

by

Gareth O. Roberts¹ and Jeffrey S. Rosenthal²

(November 2012; last revised May 2013.)

Abstract. We discuss a formal mathematical framework for certain coupling constructions via minorisation conditions, which are often used to prove bounds on convergence to stationarity of stochastic processes and MCMC algorithms.

1. Introduction.

Bounds on the convergence of Markov chains and other stochastic processes to stationary distributions has become a very widely studied topic in recent years, motivated largely by applications to Markov chain Monte Carlo (MCMC) algorithms, e.g. [9, 22, 10, 11, 7]. One common method of obtaining such bounds is through *coupling constructions*, see e.g. [21, 19, 16]. Here a second copy of a similar or identical process is constructed, jointly with the original process, and the probability of the two chains becoming (or remaining) equal is then examined and used.

Such coupling constructions are often presented in a somewhat informal and intuitive style, of the form “First construct one random variable as follows, then find a joint distribution for these two other random variables conditional on the first one, then conditionally construct a fourth random variable like this”, etc. We believe this to be acceptable, and to lead to rigorously valid coupling constructions. However, we recently became aware that at least one mathematically-minded reader is not comfortable with such informal descriptions. Thus, the purpose of this note is to provide a more formal, mathematical version of common methods of constructing couplings of pairs of stochastic processes. We emphasise that none of the results presented here are particularly novel – they are simple consequences of the Kolmogorov Extension Theorem, the coupling inequality, and maximal couplings – but we hope that they will help clarify the application of coupling constructions to MCMC algorithms.

¹Department of Statistics, University of Warwick, CV4 7AL, Coventry, U.K. Email: g.o.roberts@lancaster.ac.uk. Supported in part by EPSRC grants EP/20620/01 and EP/S61577/01.

²Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: jeff@math.toronto.edu. Web: <http://probability.ca/jeff/> Supported in part by NSERC of Canada.

Coupling constructions are often used to bound the probability of two Markov chains with identical transition kernels *becoming* equal, thus bounding the total variation distance between them (see e.g. [21, 18, 19]). Some other approaches instead bound the probability of two processes with different probability laws *staying* equal for all times up to time N . Both cases often involve conditional-type coupling constructions which are usually stated informally, but which can be stated formally if desired. Below, for concreteness, we concentrate on the second case, since the recent questions about coupling constructions originated there. However, similar methods can be used to “formalize” the coupling constructions in the first case too – and indeed in any situation in which couplings are constructed informally, one random variable at a time, in terms of various conditional distributions.

2. Statement of Main Result.

Let $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$ be two different stochastic processes, defined possibly on different probability spaces, but taking values in the same Polish measurable state space $(\mathcal{X}, \mathcal{B})$ (e.g. on \mathbf{R}^d with the Borel subsets). Let $F_n = (X_0, X_1, \dots, X_n)$ and $F'_n = (X'_0, X'_1, \dots, X'_n)$ be the two processes’ histories up to time n . For $n \geq 1$, and $A \in \mathcal{B}$, and a state history vector $s^{(n-1)} = (s_0, s_1, \dots, s_{n-1}) \subseteq \mathcal{X}^n$, let $Q_n(A; s^{(n-1)}) = \mathbf{P}[X_n \in A \mid F_{n-1} = s^{(n-1)}]$ and $Q'_n(A; s^{(n-1)}) = \mathbf{P}[X'_n \in A \mid F'_{n-1} = s^{(n-1)}]$ be the regular conditional probability distributions. As a special case, when $n = 0$, let F_{-1} and F'_{-1} and $s^{(-1)}$ each be the empty set, so that $Q_0(A; s^{(-1)}) = \mathbf{P}[X_0 \in A \mid \emptyset = \emptyset] = \mathbf{P}[X_0 \in A]$ and $Q'_0(A; s^{(-1)}) = \mathbf{P}[X'_0 \in A]$.

In terms of these definitions, a formal statement about sequential coupling constructions which attempt to keep the two processes equal, is as follows.

Theorem 1. *Let N be a non-negative integer. Suppose that for each $0 \leq n \leq N$, there is $a_n \geq 0$ such that for each state history vector $s^{(n-1)} = (s_0, s_1, \dots, s_{n-1}) \subseteq \mathcal{X}^n$, either:*

- (i) $\sup_{A \in \mathcal{B}} |Q_n(A; s^{(n-1)}) - Q'_n(A; s^{(n-1)})| \leq a_n$; or
- (ii) *there are random variables W and W' , defined jointly on some probability measure space, each taking values in \mathcal{X} , which are measurable functions of $s^{(n-1)}$ (i.e., such that if $W = W(s^{(n-1)})$ and $W' = W'(s^{(n-1)})$, then for each $A \in \mathcal{B}$, the subsets $\{s^{(n-1)} : W(s^{(n-1)}) \in A\}$ and $\{s^{(n-1)} : W'(s^{(n-1)}) \in A\}$ are measurable subsets of \mathcal{X}^n), such that $\mathbf{P}[W \in A] = Q_n(A; s^{(n-1)})$ and $\mathbf{P}[W' \in A] = Q'_n(A; s^{(n-1)})$ for all $A \in \mathcal{B}$, and $\mathbf{P}[W = W'] \geq 1 - a_n$; or*
- (iii) *there is a probability measure $\nu(\cdot)$ on $(\mathcal{X}, \mathcal{B})$ which is a measurable function of $s^{(n-1)}$ (i.e., such that if $\nu(\cdot) = \nu_{s^{(n-1)}}(\cdot)$, then for each $A \in \mathcal{B}$, the mapping $s^{(n-1)} \mapsto \nu_{s^{(n-1)}}(A)$ is a measurable function of $s^{(n-1)} \in \mathcal{X}^n$), such that $Q_n(A; s^{(n-1)}) \geq (1 - a_n) \nu(A)$ and $Q'_n(A; s^{(n-1)}) \geq (1 - a_n) \nu(A)$ for all $A \in \mathcal{B}$.*

Then there exist random variables $\{\tilde{X}_n, \tilde{X}'_n\}_{n=0}^N$ defined jointly on some probability measure space, such that $\mathcal{L}(\tilde{X}_0, \dots, \tilde{X}_N) = \mathcal{L}(X_0, \dots, X_N)$, and $\mathcal{L}(\tilde{X}'_0, \dots, \tilde{X}'_N) = \mathcal{L}(X'_0, \dots, X'_N)$, and furthermore

$$\mathbf{P}(\tilde{X}_i = \tilde{X}'_i \text{ for } 0 \leq i \leq N) \geq 1 - a_0 - a_1 - a_2 - \dots - a_N. \quad (*)$$

3. Background Tools.

In this section we collect a few standard results that will be used to prove the above Theorem.

Proposition 2. *Let ρ and σ be two probability measures on $(\mathcal{X}, \mathcal{B})$, and let $\epsilon \geq 0$. Then the following are equivalent:*

- (i) $\sup_{A \in \mathcal{B}} |\rho(A) - \sigma(A)| \leq \epsilon$.
- (ii) *there are jointly-defined random variables Y and Z taking values on $(\mathcal{X}, \mathcal{B})$, such that $\mathbf{P}(Y \in A) = \rho(A)$, and $\mathbf{P}(Z \in A) = \sigma(A)$ for all $A \in \mathcal{B}$, and $\mathbf{P}[Y = Z] \geq 1 - \epsilon$.*
- (iii) *there is a probability measure $\nu(\cdot)$ on $(\mathcal{X}, \mathcal{B})$ such that $\rho(A) \geq (1 - \epsilon)\nu(A)$ and $\sigma(A) \geq (1 - \epsilon)\nu(A)$ for all $A \in \mathcal{B}$.*

Proof. That (ii) implies (i) is the standard *coupling inequality*; see e.g. [21], or equation (13) of [16].

That (i) implies (ii) is a well-known property of couplings, corresponding to the existence of maximal couplings (e.g. [13]; Proposition 3(g) of [16]). Indeed, let $\eta(\cdot) = \rho(\cdot) + \sigma(\cdot)$ be a joint dominating measure, with corresponding Radon-Nikodym derivatives $g = \frac{d\rho}{d\eta}$ and $h = \frac{d\sigma}{d\eta}$, and let $m = \min(g, h)$. Then, let $a = \int_{\mathcal{X}} m d\eta$, $b = \int_{\mathcal{X}} (g - m) d\eta$, and $c = \int_{\mathcal{X}} (h - m) d\eta$. The statement is trivial if any of a, b, c are zero, so assume they are all positive. Then jointly construct independent random variables R, U, V, I such that R has density m/a , U has density $(g - m)/b$, V has density $(h - m)/c$, and $\mathbf{P}[I = 1] = a$ and $\mathbf{P}[I = 0] = 1 - a$. Finally, let $Y = Z = R$ if $I = 1$, and let $Y = U$ and $Z = V$ if $I = 0$. Then $Y \sim \rho(\cdot)$ and $Z \sim \sigma(\cdot)$, and U and V have disjoint support so $\mathbf{P}[U = V] = 0$. Hence, $\mathbf{P}[Y = Z] = \mathbf{P}[I = 1] = a$. But it is easily seen that $a = 1 - \sup_{A \in \mathcal{B}} |\rho(A) - \sigma(A)|$ (see e.g. [16], Proposition 3(f)), which gives the result.

That (iii) implies (i) follows by setting $\alpha(A) = \epsilon^{-1}[\rho(A) - (1 - \epsilon)\nu(A)]$ and $\beta(A) = \epsilon^{-1}[\sigma(A) - (1 - \epsilon)\nu(A)]$, so that $\alpha(\cdot)$ and $\beta(\cdot)$ are probability measures on $(\mathcal{X}, \mathcal{B})$, and $\rho(\cdot) = \epsilon\alpha(\cdot) + (1 - \epsilon)\nu(\cdot)$ and $\sigma(\cdot) = \epsilon\beta(\cdot) + (1 - \epsilon)\nu(\cdot)$, whence

$$|\rho(A) - \sigma(A)| = |\epsilon\alpha(A) + (1 - \epsilon)\nu(A) - \epsilon\beta(A) - (1 - \epsilon)\nu(A)| = \epsilon|\alpha(A) - \beta(A)| \leq \epsilon.$$

Finally, that (ii) implies (iii) follows by setting $\nu(A) = \frac{\mathbf{P}[Y \in A, Y=Z]}{\mathbf{P}[Y=Z]}$. ■

Corollary 3. *Conditions (i), (ii), and (iii) of Theorem 1 are all equivalent.*

Proof. This is essentially the content of Proposition 2, except that we have to check the measurable dependence on $s^{(n-1)}$. However, using the explicit constructions in the proof of Proposition 2 that (i) implies (ii), and that (ii) implies (iii), it is easily seen that the measurable dependence conditions can all be preserved when establishing these equivalences. In particular, when moving from (i) to (ii), the processes $\{X_n\}$ and $\{X'_n\}$ being jointly defined imply that the probabilities $Q_n(\cdot; s^{(n-1)})$ and $Q'_n(\cdot; s^{(n-1)})$ respectively are measurable functions of $s^{(n-1)}$; it then follows that the distributions of R , U , V , and I are themselves measurable functions of $s^{(n-1)}$, and hence so are the distributions of Y and Z , as required. ■

Proposition 4. *Given a probability measure $\mu(\cdot)$ on $(\mathcal{X}, \mathcal{B})$, and a family of probability measures $\nu_y(\cdot)$ on $(\mathcal{X}, \mathcal{B})$ for all $y \in \mathcal{X}$ such that the mapping $y \mapsto \nu_y(B)$ is measurable for all $B \in \mathcal{B}$, then there exist random variables Y and Z defined jointly on some probability measure space, such that $\mathbf{P}[Y \in A] = \mu(A)$ and $\mathbf{P}[Y \in A, Z \in B] = \int_{y \in A} \nu_y(B) \mu(dy)$ for all $A, B \in \mathcal{B}$.*

Proof. This follows directly from the Kolmogorov Extension Theorem (see e.g. Theorem 36.1 of [6]), upon defining the finite-dimensional distributions $\{\rho_{i_1, \dots, i_k}\}_{\substack{1 \leq k \leq 2 \\ i_j \in \{1, 2\}}}$ by

$$\rho_1(A) = \mu(A), \quad \rho_2(B) = \int_{\mathcal{X}} \nu_y(B) \mu(dy), \quad \rho_{1,2}(A \times B) = \rho_{2,1}(B \times A) = \int_{y \in A} \nu_y(B) \mu(dy).$$
■

4. Proof of Theorem 1.

We now prove Theorem 1. By Corollary 3, conditions (i) and (ii) and (iii) of Theorem 1 are all equivalent, so we can assume that they all hold, and in particular that condition (ii) always holds.

We shall prove the theorem by induction on N . The case $N = 0$ is immediate, since in that case the conclusion is equivalent to condition (ii). We assume now that the theorem has been proved for some non-negative integer $N - 1$, and proceed to prove that it also holds for N .

By our induction assumption, (*) holds for $N - 1$, i.e. there are random variables $\{\tilde{X}_n, \tilde{X}'_n\}_{n=0}^{N-1}$ with $\mathcal{L}(\tilde{X}_0, \dots, \tilde{X}_{N-1}) = \mathcal{L}(X_0, \dots, X_{N-1})$, and $\mathcal{L}(\tilde{X}'_0, \dots, \tilde{X}'_{N-1}) = \mathcal{L}(X'_0, \dots, X'_{N-1})$, and furthermore

$$\mathbf{P}(\tilde{X}_i = \tilde{X}'_i \text{ for all } 0 \leq i \leq N - 1) \geq 1 - a_0 - a_1 - a_2 - \dots - a_{N-1}.$$

Then, by condition (ii), for each state history vector $s^{(N-1)}$ we can find $W = W(s^{(N-1)})$ and $W' = W'(s^{(N-1)})$ such that $\mathbf{P}[W \in A] = Q_N(A; s^{(N-1)})$ and $\mathbf{P}[W' \in A] = Q'_N(A; s^{(N-1)})$ for all $A \in \mathcal{B}$, and $\mathbf{P}[W = W'] \geq 1 - a_N$.

We next apply Proposition 4, with Y taking the role of $(X_0, \dots, X_{N-1}, X'_0, \dots, X'_{N-1}) \in (\mathcal{X}^N)^2$, and Z taking the role of $(X_N, X'_N) \in \mathcal{X}^2$. Specifically, we set

$$\mu(\cdot) = \mathcal{L}(\tilde{X}_0, \dots, \tilde{X}_{N-1}, \tilde{X}'_0, \dots, \tilde{X}'_{N-1}),$$

and for $s, t \in \mathcal{X}^n$ we let $\nu_{(s,s)}(\cdot) = \mathcal{L}(W(s), W'(s))$, with $\nu_{(s,t)}(\cdot)$ defined arbitrarily for $s \neq t$ (say, $\nu_{(s,t)}(\cdot) = \delta_{(x_0, x_0)}(\cdot)$ for some fixed $x_0 \in \mathcal{X}$). Proposition 4 then ensures that there are jointly-defined random variables $\check{X}_0, \dots, \check{X}_N, \check{X}'_0, \dots, \check{X}'_N$ such that $\mathcal{L}(\check{X}_0, \dots, \check{X}_N) = \mathcal{L}(X_0, \dots, X_{N-1}, W)$, and $\mathcal{L}(\check{X}'_0, \dots, \check{X}'_N) = \mathcal{L}(X'_0, \dots, X'_{N-1}, W')$, and

$$\mathcal{L}(\check{X}'_0, \dots, \check{X}'_N | X_0, \dots, X_{N-1}, W) = \mathcal{L}(X'_0, \dots, X'_N | X_0, \dots, X_{N-1}, W).$$

In particular,

$$\begin{aligned} \mathbf{P}(\check{X}_i \neq \check{X}'_i \text{ for some } 0 \leq i \leq N) &\leq \mathbf{P}(\check{X}_i \neq \check{X}'_i \text{ for some } 0 \leq i \leq N - 1) + \mathbf{P}(\check{X}_N \neq \check{X}'_N) \\ &= \mathbf{P}(\tilde{X}_i \neq \tilde{X}'_i \text{ for some } 0 \leq i \leq N - 1) + \mathbf{P}(W \neq W') \\ &\leq (a_0 + a_1 + \dots + a_{N-1}) + a_N. \end{aligned}$$

This establishes equation (*) for N (with the \tilde{X} now replaced by \check{X}). Therefore, this completes the induction step, and hence the proof of Theorem 1. ■

5. Auxiliary randomness.

Sometimes the transition probabilities for the chain $\{X_n\}$ depend not just on its own previous states, but also on additional auxiliary random variables. This is particularly the case for *adaptive MCMC* algorithms, a recently widely-studied subject (see e.g. [14, 2, 4, 17, 1, 12, 5, 20, 3, 8, 15]). For such algorithms, sometimes the coupling only holds under

certain conditions about the auxiliary randomness. To handle this, a slight extension of our Theorem 1 is required.

To set it up, again let $\{X_n\}_{n=0}^\infty$ and $\{X'_n\}_{n=0}^\infty$ be two different stochastic processes, defined possibly on separate probability spaces, taking values in the same measurable state space $(\mathcal{X}, \mathcal{B})$, and again let $F_n = (X_0, X_1, \dots, X_n)$ and $F'_n = (X'_0, X'_1, \dots, X'_n)$. Let $\{\Gamma_n\}$ be a sequence of auxiliary random variables, taking values in some space $(\mathcal{Y}, \mathcal{G})$, such that the law of X_{n+1} depends not only on X_0, \dots, X_n , but also on Γ_n in some way (but assume for simplicity that there is no auxiliary randomness in the law of X'_{n+1}). For $n \geq 1$, let $Q_n(A; s^{(n-1)}, \gamma_n) = \mathbf{P}[X_n \in A | F_{n-1} = s^{(n-1)}, \Gamma_n = \gamma_n]$, and $Q'_n(A; s^{(n-1)}) = \mathbf{P}[X'_n \in A | F'_{n-1} = s^{(n-1)}]$. Again let F_{-1} and F'_{-1} and $s^{(-1)}$ and Γ_{-1} and γ_{-1} be the empty set. Finally, let $H_n \in \mathcal{G}$ be some event (corresponding, intuitively, to “good” values of the Γ_n).

Theorem 5. *Let N be a non-negative integer. Suppose that for each $0 \leq n \leq N$, there is $a_n \geq 0$ such that for each history $s^{(n-1)}$, and each $\gamma_n \in H_n$, either:*

- (i) $\sup_{A \in \mathcal{B}} |Q_n(A; s^{(n-1)}, \gamma_n) - Q'_n(A; s^{(n-1)})| \leq a_n$; or
- (ii) *there are jointly-defined random variables Y and Y' taking values in $(\mathcal{X}, \mathcal{B})$, such that $\mathbf{P}[Y \in A] = Q_n(A; s^{(n-1)}, \gamma_n)$ and $\mathbf{P}[Y' \in A] = Q'_n(A; s^{(n-1)})$ for all $A \in \mathcal{B}$, and $\mathbf{P}[Y = Y'] \geq 1 - a_n$; or*
- (iii) *there is a probability measure $\nu(\cdot)$ on $(\mathcal{X}, \mathcal{B})$ with $Q_n(A; s^{(n-1)}, \gamma_n) \geq (1 - a_n) \nu(A)$ and $Q'_n(A; s^{(n-1)}) \geq (1 - a_n) \nu(A)$ for all $A \in \mathcal{B}$.*

Suppose also that for each $0 \leq n \leq N$, there is $b_n \geq 0$ such that for each history $s^{(n)}$, and each $\gamma_n \in H_n$,

$$\mathbf{P}(\Gamma_{n+1} \in H_{n+1} | F_n = s^{(n)}, \Gamma_n = \gamma_n) \geq 1 - b_n.$$

Then there exist random variables $\{\tilde{X}_n, \tilde{X}'_n\}_{n=0}^N$ defined jointly on some probability space, such that $\mathcal{L}(\tilde{X}_0, \dots, \tilde{X}_N) = \mathcal{L}(X_0, \dots, X_N)$, and $\mathcal{L}(\tilde{X}'_0, \dots, \tilde{X}'_N) = \mathcal{L}(X'_0, \dots, X'_N)$, and furthermore

$$\mathbf{P}(\tilde{X}_i = \tilde{X}'_i \text{ for } 0 \leq i \leq N) \geq 1 - a_0 - a_1 - a_2 - \dots - a_N - b_0 - b_1 - b_2 - \dots - b_N.$$

Proof. The proof of Theorem 5 is a simple and direct generalisation of the proof of Theorem 1. In particular, when applying Proposition 4, we now let Y take the role of $(X_0, \dots, X_{N-1}, X'_0, \dots, X'_{N-1}, \Gamma_{N-1}) \in (\mathcal{X}^N)^2 \times \mathcal{Y}$, and let Z take the role of $(X_N, X'_N, \Gamma_N) \in \mathcal{X}^2 \times \mathcal{Y}$. That is, we also condition on the previous auxiliary randomness parameter Γ_{N-1} , and we also construct the new auxiliary randomness parameter Γ_N . The remainder of the proof is virtually identical to that of Theorem 1, so we omit the details. ■

Remark. In particular, Theorem 5 provides a mathematically precise framework for the coupling construction used in the proof of Theorem 5 of [17].

Acknowledgements. We thank Faming Liang for bringing this matter to our attention, thank Christian Robert for very helpful assistance, and thank the anonymous referee for a very careful reading.

References

- [1] C. Andrieu and Y.F. Atchadé (2007), On the efficiency of adaptive MCMC algorithms. *Elec. Comm. Prob.* **12**, 336–349.
- [2] C. Andrieu and E. Moulines (2003), On the ergodicity properties of some adaptive Markov Chain Monte Carlo algorithms. *Ann. Appl. Prob.* **16**, 1462–1505.
- [3] Y.F. Atchadé, and G. Fort (2010). Limit theorems for some adaptive mcmc algorithms with sub-geometric kernels. *Bernoulli* **16**, 116-154.
- [4] Y.F. Atchadé and J.S. Rosenthal (2005), On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11**, 815–828.
- [5] Y. Bai, G.O. Roberts, and J.S. Rosenthal (2011), On the containment condition for adaptive Markov chain Monte Carlo algorithms. *Adv. Appl. Stat.* **21**, 1–54.
- [6] P. Billingsley (1995), *Probability and Measure*, 3rd ed. John Wiley & Sons, New York.
- [7] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng (eds.) (2011), *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton, Florida.
- [8] G. Fort, E. Moulines, and P. Priouret (2012). Convergence of adaptive and interacting Markov chain Monte Carlo algorithms. *Ann. Stat.* **39**, 3262–3289.
- [9] A.E. Gelfand and A.F.M. Smith (1990), Sampling based approaches to calculating marginal densities. *J. Amer. Stat. Assoc.* **85**, 398–409.
- [10] C. Geyer (1992), Practical Markov chain Monte Carlo. *Stat. Sci.*, Vol. **7**, No. **4**, 473-483.
- [11] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, ed. (1996), *Markov chain Monte Carlo in practice*. Chapman and Hall, London.

- [12] P. Giordani and R. Kohn (2010), Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals. *J. Comp. Graph. Stat.* **19(2)**, 243–259.
- [13] D. Griffeath (1975), A Maximal Coupling for Markov Chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **31(2)**, 95–106.
- [14] H. Haario, E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm. *Bernoulli* **7**, 223–242.
- [15] L. Holden, R. Hauge, and M. Holden (2009), Adaptive independent Metropolis-Hastings. *Ann. Appl. Prob.* **19**, 395–413.
- [16] G.O. Roberts and J.S. Rosenthal (2004), General state space Markov chains and MCMC algorithms. *Prob. Surv.* **1**, 20–71.
- [17] G.O. Roberts and J.S. Rosenthal (2007), Coupling and ergodicity of adaptive MCMC. *J. Appl. Prob.* **44**, 458–475.
- [18] J.S. Rosenthal (1995), Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Stat. Assoc.* **90**, 558–566.
- [19] J.S. Rosenthal (2002), Quantitative convergence rates of Markov chains: A simple account. *Elec. Comm. Prob.* **7**, No. 13, 123–128.
- [20] E. Saksman and M. Vihola (2010), On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Prob.* **20**, 2178–2203.
- [21] H. Thorisson (2000), *Coupling, Stationarity, and Regeneration*. Springer, New York.
- [22] L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). *Ann. Stat.* **22**, 1701–1762.