The Annals of Statistics
1996, Vol. 24, No. 1, 122–140

CONVERGENCE PROPERTIES OF THE GIBBS SAMPLER FOR PERTURBATIONS OF GAUSSIANS[1]

BY YALI AMIT

University of Chicago

The exact second eigenvalue of the Markov operator of the Gibbs sampler with random sweep strategy for Gaussian densities is calculated. A comparison lemma yields an upper bound on the second eigenvalue for bounded perturbations of Gaussians which is a significant improvement over previous bounds. For two-block Gibbs sampler algorithms with a perturbation of the form $\chi(g_1(x^{(1)}) + g_2(x^{(2)}))$ the derivative of the second eigenvalue of the algorithm is calculated exactly at $\chi = 0$, in terms of expectations of the Hessian matrices of g_1 and g_2.

1. Introduction. Monte-Carlo simulations are widely used today in many statistical applications, most commonly for simulating the posterior distribution in a Bayesian framework. One of the first examples of the use of these methods can be found in Geman and Geman (1984), in the context of image restoration. An extensive bibliography can be found in Geman (1991). Other examples in the context of higher level image analysis can be found in Knoerr (1988) and Grenander, Chow and Keenan (1991). In Besag and Green (1993) and Smith and Roberts (1993) a host of other applications of these methods in the context of hierarchical Bayesian models is described.

In all these examples the question arises as to the rate of convergence of the Markov chain. There is no unified theory to answer this question, especially in the case of continuous densities on R^n, and currently many different approaches are being attempted for the different types of simulation methods. One direction of research uses Doeblin type bounds on bounded sets together with geometric drift conditions or some form of Lyapounov functions to obtain rather general results on exponential convergence for continuous densities on R^n; see Mengersen and Tweedie (1996) for the Metropolis algorithm and Rosenthal (1995) for the Gibbs sampler. These results are then applied to certain special cases and yield bounds on the rate of convergence in variational norm distance. Another approach analyzes a discretized version of the problem using general discrete state space Markov chain techniques; see Frieze, Kannan and Polson (1994).

Finally there are attempts to use spectral analysis to obtain rates of convergence in the L_2 sense, where the Markov chain operator is considered as a bounded operator in the Hilbert space of square integrable functions

Received January 1995; revised July 1995.
[1]Research supported in part by the University of Chicago Block Fund and ARO DAAL03-92-G-0322.
AMS 1991 *subject classifications*. 60J10, 47B38.
Key words and phrases. Second eigenvalue, Markov chains, integral operators.

with respect to the target distribution. The largest eigenvalue of this operator is always 1, the eigenfunctions being the constant functions. The question is then what is the next largest eigenvalue (in absolute value). The difference between the two is often called the *gap*. This approach was initially applied by Goodman and Sokal (1989) to the Gibbs sampler (heat bath algorithm in the physics literature), for Gaussian distributions, and subsequently by Amit (1991) to bounded perturbations of Gaussians. For other treatments of Gaussian distributions, see Amit and Grenander (1991) and Barone and Frigessi (1990).

The elementary operation in the Gibbs sampler is sampling from conditional distributions of some set of coordinates given the values on the complement. Given a partition T_1, T_2 of the set $I_n = \{1, \ldots, n\}$, a typical step will simulate from the conditional distribution of X_{T_1} given $X_{T_2} = x_{T_2}$, where X has distribution ϕ. This distribution is completely described by the conditional expectations $E(f(X)|X_{T_2} = x_{T_2})$ for all functions f in the Hilbert space $L_2(R^n, \phi)$. These conditional expectations are orthogonal projections. Typically the Gibbs sampler sweeps through the coordinates one by one in some fixed order: sequential sweep strategy. Writing $P_i f = E(f(X)|X_{(i)} = x_{(i)})$, where $(i) = I_n - \{i\}$, the entire sweep corresponds to the Markov chain operator $L_{\text{seq}} = P_n P_{n-1} \cdots P_1$, namely, a product of projections. Observe that P_i is the projection onto the space $S_i = \{f \in L_2(R^n, \phi)\colon f(x) = f(x_{(i)})\}$. Alternatively, the choice of coordinates can be done at random: random sweep strategy. In this case the operator corresponding to each step is $L_{\text{ran}} = (1/n)\sum_{i=1}^{n} P_i$, namely, an average of projections.

The family of continuous densities which can be analyzed using spectral analysis and perturbation methods is limited at this point to perturbations of Gaussians. However in some cases where these methods apply, the bounds are much more precise than those derived using the Doeblin techniques. Take for example the bivariate normal model from Rosenthal (1995) with $\mu = 0$. The rate obtained using the minorization techniques is over 0.95. On the other hand, the spectral gap for the Gibbs sampler operator is exactly 0.5 as can be seen from Amit [(1991), Lemma 2], and assuming the initial distribution is a density, the variational norm distance also converges as $C \cdot 0.5^k$.

Perturbed Gaussian models are very common in statistical applications where the data cannot be directly modeled as a Gaussian process or field, rather as a nonlinear function of a Gaussian process. These could be called hidden Gaussian models in analogy with the hidden Markov models which are extensively used in statistics and engineering. We illustrate this with an example from image analysis.

In Amit, Grenander and Piccioni (1991) a Gaussian random vector field describes the prior on displacements of the pixel lattice, but is only observed indirectly through the deformed template image which is produced by the displacements and which is also degraded by additive independent Gaussian noise. If $T(x)$ is the template image, $D(x)$ is the observed image and $U_1(x), U_2(x)$ the random vector field at pixel x, the model assumes that $D(x)$ is normally distributed with variance σ and mean given by $T(x_1 + U_1(x),$

$x_2 + U_2(x)$), that is, the value of the template at the displaced location of x. The posterior distribution will have the form

$$\frac{1}{Z}\exp\Bigg[-\sum_x\Big[(LU_1(x))^2+(LU_2(x))^2\Big] - \frac{1}{\sigma^2}\sum_x\big[T(x_1+U_1(x),x_2+U_2(x))-D(x)\big]^2\Bigg],$$

where L is a finite difference approximation to some differential operator. The quadratic term imposes a Gaussian smoothness prior on the displacement field U. Given an observed image D, inference is then made with respect to the underlying displacement field U. This could also be thought of as a hidden Gaussian model. An underlying Gaussian process drives the system but is not directly observed, rather through some nonlinear transformation. The resulting posterior distribution in these models has the form of a perturbed Gaussian. If the nonlinear transformation has bounded range the perturbation is bounded. These types of bounded perturbations were analyzed in Amit (1991) for the sequential updating scheme.

The goal of this paper is twofold. First the results of Amit (1991) are extended to the random updating scheme. For the Gaussian distribution the projections are analyzed in terms of the Hermite polynomial basis, to yield the exact gap for the Markov chain operator. Then a lower bound on the gap for the perturbed distribution is obtained using a comparison lemma on the gaps of the two operators. These results are similar in flavor to those of Diaconis and Saloff-Coste (1993) for analyzing various types of random walks on groups. Certain cases lend themselves to precise spectral analysis essentially borrowing from the theory of special functions and group representation theory; then other cases are analyzed using general comparison lemmas. The lower bound obtained in the comparison lemma in the random updating scheme is much larger than the one achievable in the comparison lemma for the deterministic updating scheme analyzed in Amit (1991).

The second component of this paper involves the asymptotic analysis of the gap for small perturbations. When the additive perturbation in the exponent is multiplied by a parameter χ, an explicit formula is derived for the derivative of the gap in χ at $\chi = 0$ (the Gaussian). Here the perturbation is not assumed to be bounded. This calculation does not yield an explicit estimate on the rate of convergence; however, it can indicate whether the gap is expected to increase or decrease.

The two-block Gibbs sampler involves partitioning the coordinates into two disjoint sets T_1, T_2, and successively simulating from the conditional distribution of one set of variables given the other. If the perturbation has the form $\chi(g_1(x_{T_1}) + g_2(x_{T_2}))$, then it is shown in Section 3 that the first derivative of the second eigenvalue of the operator corresponding to the two-block Gibbs sampler algorithm, as a function of χ, at $\chi = 0$, is given in terms of expectations involving the Hessians of g_1 and g_2. In particular if both are

convex, the derivative is negative implying that the gap increases and the convergence rate increases. It is interesting to note that in the two-block case the rate of convergence is precisely the square of the *maximal correlation* between the two sets of coordinates, or variables.

In general the two-block method is no more feasible than direct simulation of the entire distribution. In some cases, however, the coordinates of $x^{(1)} = x_{T_1}$ are mutually independent given the values of $x^{(2)} = x_{T_2}$ and vice versa. Then block updating is equivalent to the ordinary sequential sweep provided all the coordinates in $x^{(1)}$ are updated first and then those of $x^{(2)}$.

Similar calculations are carried out for the random updating scheme, when the perturbation is of the form $\chi \sum_i g(x_i)$; in other words, a fixed function g applied to each coordinate separately. This type of perturbation appears in all the image analysis applications described above.

2. Rates for the random sweep strategy. In Amit (1991) the Gibbs sampler algorithm was analyzed for the sequential sweep strategy, namely, when the coordinates are visited in a fixed order. The same can be done for the random sweep strategy as will be presented in this section.

Let $X_k(x)$ be the kth step of the Markov chain starting at x. For any function $f \in L_2(R^n, \phi)$, the Hilbert space of square integrable functions with respect to ϕ, define

$$Lf(x) = L_{\text{ran}} f(x) = Ef(X_1(x)) = (1/n) \sum_{i=1}^{n} P_i f(x).$$

Then $Ef(X_k(x)) = L^k(x)$, and the question is how fast does $L^k f$ converge to $\phi(f)$, the integral of f with respect to the density ϕ. All Markov chain Monte-Carlo schemes have the property that $\|L\| \leq 1$ and $L\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the constant function 1. In the present setting L is an average of projections and therefore all eigenvalues are between 0 and 1. The question is then what is the second largest eigenvalue of L.

Gaussian case. First let ϕ be the density of a multivariate normal with mean 0 and inverse covariance matrix Q. We assume without loss of generality that Q has all ones on the diagonal. Let K be the symmetric square root of the matrix Q. Since the diagonal elements of Q have been assumed to be all 1, each row of the matrix K is a unit vector. Let $K^{(i)}$ denote the projection in R^n onto the vector space spanned by the ith row of K and let $T^{(i)} = I - K^{(i)}$.

For a multi-index of nonnegative integers $\alpha = (\alpha_1, \ldots, \alpha_n) \in Z_+^n$ and $a \in R^n$, let $a^\alpha = a_1^{\alpha_1} \cdots a_n^{\alpha_n}$ and $\alpha! = \alpha_1! \cdots \alpha_n!$. Let $\mathscr{H}_\alpha(x) = h_{\alpha_1}(x_1) \cdots h_{\alpha_n}(x_n)$, with h_k the kth order Hermite polynomial [see, e.g., Cramér (1946)]. Using standard properties of Hermite polynomials it follows that the family of functions $\hat{\mathscr{H}}_\alpha = (\mathscr{H}_\alpha(Kx))/\sqrt{\alpha!}$, $\alpha \in Z_+^n$, forms an orthonormal basis of

$L_2(R^n, \phi)$. For $a \in R^n$ let f_a denote the generating function parameterized by a given by

$$f_a(x) = \exp\left[\langle a, Kx\rangle - \frac{|a^2|}{2}\right] = \sum_\alpha a^\alpha \frac{\mathscr{H}_\alpha(Kx)}{\alpha!}.$$

A simple computation involving completing the square yields

$$\begin{aligned} P_i(f_a) &= \int f_a(x)\,\phi\big(x_i|x_{(i)}\big)\,dx_i \\ &= \exp\Big[-\tfrac{1}{2}|a|^2 + \tfrac{1}{2}(Ka)_i^2 + \langle Ka, x\rangle - (Ka)_i(Qx)_i\Big]. \end{aligned}$$

Write D_i for the matrix which is all zeros except for 1 at the ith place on the diagonal. Then $D_i^2 = D_i$ and $KD_iK = K^{(i)}$. We obtain

$$P_i(f_a) = \exp\Big[-\tfrac{1}{2}|a|^2 + \tfrac{1}{2}\langle D_iKa, D_iKa\rangle + \langle a, Kx\rangle - \langle D_iKa, D_iQx\rangle\Big].$$

Writing the second term as $|KD_iKa|^2$ and the last term as $\langle a, KD_iKKx\rangle$ yields

$$P_i(f_a) = f_{T^{(i)}a}.$$

Consequently,

$$L(f_a) = \frac{1}{n}\sum_{i=1}^n {}_i f_a = \frac{1}{n}\sum_{i=1}^n f_{T^{(i)}a}$$

for all $a \in R^n$ or

$$L\left(\sum_\alpha a^\alpha \frac{\mathscr{H}_\alpha(Kx)}{\alpha!}\right) = \sum_\alpha \left(\frac{1}{n}\sum_{i=1}^n (T^{(i)}a)^\alpha\right)\frac{\mathscr{H}_\alpha(Kx)}{\alpha!}.$$

The coefficients of a^α must match on both sides of the equality, thus providing an explicit expression of the image of each basis element $\hat{\mathscr{H}}_\alpha$ under L. The coefficient of a^α on the right-hand side is computed as follows.

Let $k = |\alpha| = \alpha_1 + \cdots + \alpha_n$. The term a^α will appear only for those multi-indexes β such that $|\beta| = k$. Let ε^* denote the sequence of k elements from $\{1, \ldots, n\}$ with α_1 1s, $\ldots$, α_n ns. Let δ^* denote the corresponding sequence for β. Any distinct permutation ε of ε^* denotes an order of choosing the factors of a^α from the k factors of $(T^{(i)}a)^\beta$. The coefficient of a^α on the right-hand side is therefore

$$\frac{1}{n}\sum_\varepsilon \sum_{i=1}^n T^{(i)}_{\delta_1^*\varepsilon_1} \cdots T^{(i)}_{\delta_k^*\varepsilon_k},$$

where ε runs over all $k!/\alpha!$ distinct permutations of ε^*. In fact, substituting any permutation δ of δ^* in the above expression will yield the same result,

since Σ_ε loops over all distinct permutations of ε^*. There are $k!/\beta!$ such distinct permutations of δ^* so we can write

$$L\big(\hat{\mathscr{H}}_\alpha(x)\big) = \frac{1}{n} \sum_{\beta:|\beta|=k} \frac{\sqrt{\alpha!}\sqrt{\beta!}}{k!} \left(\sum_{i=1}^{n} \sum_{\varepsilon,\delta} T^{(i)}_{\delta_1\varepsilon_1} \cdots T^{(i)}_{\delta_k\varepsilon_k} \right) \hat{\mathscr{H}}_\beta(x).$$

Consequently the subspace H_k spanned by all $\hat{\mathscr{H}}_\beta$ such that $|\beta| = k$ is invariant under L and the matrix corresponding to the operator L in that subspace is obtained from normalized block sums of entries of the tensor

$$\mathscr{T}_k = \frac{1}{n} \sum_{i=1}^{n} \big(T^{(i)}\big)^{\otimes k},$$

where $(k!)^2/(\alpha!\beta!)$ is the number of entries in the block. This implies that the largest eigenvalue of L in H_k is bounded by the largest eigenvalue of $\mathscr{T}_k$ as shown by the following lemma.

LEMMA 1. *Let A be a positive definite $r \times r$ matrix. Let $S_1, \ldots, S_q$ be a partition of $\{1, \ldots, r\}$ and let $|S_k|$ denote the cardinality of S_k. Define B to be the $q \times q$ matrix defined as*

$$B_{kl} = \frac{1}{\sqrt{|S_k||S_l|}} \sum_{\substack{i \in S_k \\ j \in S_l}} A_{ij}.$$

Then the maximum eigenvalue of B is less than the maximum eigenvalue of A.

PROOF. The largest eigenvalue of B is given as

$$\begin{aligned} \sup_{\substack{|v|=1 \\ v \in R^q}} v^t B v &= \sup_{\substack{|v|=1 \\ v \in R^q}} \sum_{k,l} v_k B_{kl} v_l \\ &= \sup_{\substack{|v|=1 \\ v \in R^q}} \sum_{k,l} \sum_{\substack{i \in S_k \\ j \in S_l}} \frac{v_k}{\sqrt{|S_k|}} A_{ij} \frac{v_l}{\sqrt{|S_l|}} \\ &= \sup_{|u| \in F^r} u^t A u, \end{aligned}$$

where F^r is the subset of unit vectors in R^r obtained from unit vectors in R^q by replicating the kth coordinate $|S_k|$ times multiplied by the factor $1/\sqrt{|S_k|}$. The last term in the equality above is bounded by the largest eigenvalue of A. □

The spaces H_k are mutually orthogonal. The space H_0 is simply the space of constant functions and corresponds to the eigenvalue 1. The space H_1 is

the n dimensional space of linear functions, and the matrix of the restriction of L to H_1 is precisely

$$L|_{H_1} = \mathscr{T}_1 = \frac{1}{n}\sum_{i=1}^{n} T^{(i)} = \frac{1}{n}\sum_{i=1}^{n} I - K^{(i)} = I - \frac{1}{n}\sum_{i=1}^{n} K^{(i)}$$
$$= I - \frac{1}{n}\sum_{i=1}^{n} KD_iK = I - \frac{1}{n}Q.$$

This implies that the largest eigenvalue in H_1 is $1 - (1/n)\lambda_m$, where λ_m is the smallest eigenvalue of Q.

Observe that

$$\mathscr{T}_l = \mathscr{T}_{l-1} \otimes I - \frac{1}{n}\sum_{i=1}^{n} (I - K^{(i)})^{\otimes l-1} \otimes K^{(i)}.$$

The second term is an average of tensor products of projections and is therefore a nonnegative operator. Since $\mathscr{T}_{l-1} \otimes I$ has the same eigenvalues as $\mathscr{T}_{l-1}$, the largest eigenvalue of $\mathscr{T}_l$ is less than or equal that of $\mathscr{T}_{l-1}$. Thus the largest eigenvalue of L which is less than 1 is obtained in H_1 and is therefore the largest eigenvalue of $I - Q/n$, which is $1 - \lambda_m/n$, where λ_m is the smallest eigenvalue of Q. The following theorem summarizes the foregoing text.

THEOREM 1. *The second largest eigenvalue of the operator L corresponding to the random updating Gibbs sampler for the Gaussian distribution ϕ is $1 - (1/n)\lambda_m$ or $\mathrm{GAP}(L) = \lambda_m/n$, where λ_m is the smallest eigenvalue of the inverse covariance matrix. The corresponding normalized eigenfunction is $B(x) = \sqrt{\lambda_m}\langle b, x\rangle$ where b is an eigenvector of Q corresponding to λ_m.*

Perturbations. Let ϕ be any density for which the gap of the random updating Gibbs sampler operator L is known. Consider now a perturbed density $\tilde{\phi} = (\rho/Z)\phi$, where ρ is a function such that $0 < c < \rho(x) \le C$. Let $S = L_2(R^n, \phi)$ and $\tilde{S} = L_2(R^n, \tilde{\phi})$. Similarly define S_i, $i = 1, \ldots, n$, as above and $\tilde{S}_i$ the corresponding subspaces of $\tilde{S}$, with $\tilde{P}_i$ the corresponding projections. Since ρ is bounded away from zero and infinity, the elements in S and $\tilde{S}$ are the same; similarly for S_i and $\tilde{S}_i$ for $i = 1, \ldots, n$.

LEMMA 2 (Comparison lemma). *Let $L, \tilde{L}$ denote the operators in $S, \tilde{S}$, respectively, of the Gibbs sampler with random updating scheme for the densities $\phi, \tilde{\phi}$. Then* $\mathrm{GAP}(\tilde{L}) \ge (c/C)\mathrm{GAP}(L)$.

PROOF. Observe that

$$\mathrm{GAP}(L) = \inf_{\substack{f \in S:\ \phi(f) = 0 \\ |f|_S = 1}} \langle f, (I - L)f\rangle_S,$$

and similarly for GAP($\tilde{L}$). The proof consists of a simple sequence of inequalities.

For any f such that $\tilde{\phi}(f) = 0$ and $|f|_{\tilde{S}} = 1$,

$$\begin{aligned}
\langle f, (I - \tilde{L}) f \rangle_{\tilde{S}} &= \frac{1}{n} \sum_{i=1}^{n} \langle f, (I - \tilde{P}_i) f \rangle_{\tilde{S}} \\
&= \frac{1}{n} \sum_{i=1}^{n} \int \big(f(x) - \tilde{P}_i f(x)\big)^2 \tilde{\phi}(x)\, dx \\
&\geq \frac{c}{Z} \frac{1}{n} \sum_{i=1}^{n} \int \big(f(x) - \tilde{P}_i f(x)\big)^2 \phi(x)\, dx \\
&\geq \frac{c}{Z} \frac{1}{n} \sum_{i=1}^{n} \int \big(f(x) - P_i f(x)\big)^2 \phi(x)\, dx \\
&= \frac{c}{Z} \langle (f - \phi(f)), (I - L)(f - \phi(f)) \rangle_S \\
&\geq \frac{c}{Z} \operatorname{GAP}(L) \int (f(x) - \phi(f))^2 \phi(x)\, dx \\
&\geq \frac{c}{C} \operatorname{GAP}(L) \int (f(x) - \phi(f))^2 \tilde{\phi}(x)\, dx \\
&\geq \frac{c}{C} \operatorname{GAP}(L) \int \big(f(x) - \tilde{\phi}(f)\big)^2 \tilde{\phi}(x)\, dx \\
&= \frac{c}{C} \operatorname{GAP}(L).
\end{aligned}$$

The second inequality follows from the fact that P_i is an orthogonal projection, in S, onto S_i, and $\tilde{P}_i f \in S_i$. The third inequality follows from the definition of the gap. □

COROLLARY A. *If ϕ is a Gaussian density as in Theorem* 1, *then*

$$\operatorname{GAP}(\tilde{L}) \geq \frac{c}{C} \frac{\lambda_m}{n}. \qquad \square$$

For the sequential updating scheme, Theorem 3 of Amit (1991) states that for $|f|_{\tilde{S}} = 1$

$$|\tilde{L}_{\text{seq}} f|_{\tilde{S}} \leq \big(1 - (c/C)^{n-1} \det(Q)\big) \sim \exp\big[-(c/C)^{n-1} \det(Q)\big]$$

for large n. This bound refers to a sweep through all n coordinates and so should be compared to n steps of the random updating strategy, namely,

$$|\tilde{L}^n_{\text{ran}} f|_{\tilde{S}} \leq \left(1 - \frac{(c/C)\lambda_m}{n}\right)^n \sim \exp\left[-\left(\frac{c}{C}\right)\lambda_m\right].$$

Since the diagonal elements of Q are all 1, the determinant is bounded by 1 (the geometric mean of the eigenvalues is bounded by the arithmetic mean). Hence the presence of an exponentially small term in the exponent of the first bound implies that in general it will be much larger than the second bound. This is not to say that the random updating scheme converges faster, merely that the obtainable bounds are better.

3. Derivative of the gap for a two-block Gibbs sampler. Let ϕ be a positive continuous density. In the two-block Gibbs sampler the coordinates are divided into two subsets $x^{(1)}$ and $x^{(2)}$. The conditional distribution of $x^{(2)}$ is simulated given the values of $x^{(1)}$ and then vice versa. For simplicity assume $x^{(1)} = (x_1, \ldots, x_l)$ and $x^{(2)} = (x_{l+1}, \ldots, x_n)$, and let $S_1, S_2 \subset L_2(\phi)$ be the subspaces of functions depending on $x^{(1)}$ and $x^{(2)}$, respectively. The operator corresponding to the two-block Gibbs sampler is simply the product $L_\phi = P_1 P_2$ of the two projections on the two spaces. Restricted to S_1, the operator L_ϕ has the kernel

$$K_\phi(x^{(1)}, y^{(1)}) = \int \phi(y^{(1)}|y^{(2)})\phi(y^{(2)}|x^{(1)})\, dy^{(2)}. \tag{1}$$

The second eigenvalue of this operator is precisely the cosine squared of the angle between these two spaces or alternatively the square of the maximal correlation between the two sets of coordinates:

$$\begin{aligned}\Gamma &= \sup\{\langle f_1, P_1 P_2 f_1\rangle;\ |f_1| = 1,\ \phi(f_1) = 0,\ f_1 \in S_1\}\\ &= \sup\{\langle f_1, f_2\rangle^2;\ |f_1| = |f_2| = 1,\ \phi(f_1) = \phi(f_2) = 0,\ f_1 \in S_1,\ f_2 \in S_2\}\\ &= \cos^2(S_1, S_2) = \sigma^2,\end{aligned}$$

where Γ is the second eigenvalue and σ is the maximal correlation coefficient.

The gap for the Gaussian, $\chi = 0$. Let $\phi(x) = \phi(x^{(1)}, x^{(2)})$ be the density of a Gaussian with mean zero and inverse covariance

$$Q = \begin{pmatrix} I_1 & C \\ C^t & I_2 \end{pmatrix},$$

where I_1 is an $l \times l$ identity matrix and I_2 is a $(n-l)\times(n-l)$ identity matrix. The perturbation in the exponent is assumed to have the form $\chi(g_1(x^{(1)}) + g_2(x^2))$, where g_1 and g_2 are bounded from below, so that

$$\phi_\chi(x^{(1)}, x^{(2)}) = \frac{1}{c}\exp\big[-\chi\big(g_1(x^{(1)}) + g_2(x^{(2)})\big)\big]\phi(x^{(1)}, x^{(2)}). \tag{2}$$

The marginals of the Gaussian density ϕ on $x^{(1)}$ and $x^{(2)}$ are denoted $\phi(x^{(1)})$, and $\phi(x^{(2)})$, respectively. The inverse of Q is given by

$$Q^{-1} = \begin{pmatrix} (I_1 - CC^t)^{-1} & -C(I_2 - C^tC)^{-1} \\ -C^t(I_1 - CC^t)^{-1} & (I_2 - C^tC)^{-1} \end{pmatrix} = \begin{pmatrix} F_1 & -CF_2 \\ -C^tF_1 & F_2 \end{pmatrix}, \tag{3}$$

where $F_1 = (I_1 - CC^t)^{-1}$ and $F_2 = (I_2 - C^tC)^{-1}$. For notational convenience we also write $D_1 = CC^t$ and $D_2 = C^tC$.

Applying (1) we get

$$R_0(x^{(1)}, y^{(1)}) \equiv K_\phi(x^{(1)}, y^{(1)}) = \frac{1}{Z} \cdot \exp\left[-\frac{1}{2} \left\langle y^{(1)} - D_1 x^{(1)}, (I_1 + D_1)^{-1} \left(y^{(1)} - D_1 x^{(1)} \right) \right\rangle \right]$$

for some normalizing constant Z. In Amit (1991) it is shown that the second eigenvalue $\Gamma_2(0)$ of the block Gibbs sampler algorithm for the Gaussian density ϕ is precisely the largest eigenvalue λ_M of D_1. The corresponding eigenfunctions are linear in $x^{(1)}$, namely, of the form $\langle v, x^{(1)} \rangle$, where $D_1 v = \lambda_M v$. Assuming λ_M is a simple eigenvalue of D_1 with corresponding eigenvector v, then it is also a simple eigenvalue of the block Gibbs sampler operator. Since v is also an eigenvector of F_1, which is the covariance matrix of the marginal of ϕ on the first l coordinates, normalizing in $L_2(R^l, \phi(x^{(1)}))$ we get the normalized eigenfunction $B(x^{(1)}) = (1 - \lambda_M)^{1/2} \langle v, x^{(1)} \rangle$.

The derivative at $\chi = 0$. For the perturbed density let K_{ϕ_χ} denote the kernel of L_{ϕ_χ} restricted to the subspace $S_1^\chi = L_2(R^l, \phi_\chi(x^{(1)}))$, where

$$\phi_\chi(x^{(1)}) = \frac{\exp\left[-\chi g_1(x^{(1)}) \right] \phi(x^{(1)})}{c} \int \exp\left[-\chi g_2(x^{(2)}) \right] \phi(x^{(2)}|x^{(1)})\, dx^{(2)}.$$

In order to fix everything in the same space we use the unitary map U of $S_1^\chi = L_2(R^l, \phi_\chi(x^{(1)}))$ onto $S_1 = L_2(R^l, \phi(x^{(1)}))$ defined by

$$Uf(x^{(1)}) = \frac{\rho_\chi^{1/2}(x^{(1)})}{c^{1/2}} \left[\int \rho_\chi(x^{(2)}) \phi(x^{(2)}|x^{(1)})\, dx^{(2)} \right]^{1/2} f(x^{(1)}),$$

where $\rho_\chi(x^{(i)}) = \exp(-\chi g_i(x^{(i)}))$. The similarity transformation obtained through U yields $R_\chi = UL_{\phi_\chi}U^{-1}$ on S_1 with kernel

$$R_\chi(x^{(1)}, y^{(1)}) = \frac{\rho_\chi^{1/2}(x^{(1)})\, \rho_\chi^{1/2}(y^{(1)})}{\left(\int \rho_\chi(u^{(2)}) \phi(u^{(2)}|x^{(1)})\, du^{(2)} \int \rho_\chi(v^{(2)}) \phi(v^{(2)}|y^{(1)})\, dv^{(2)} \right)^{1/2}} \times \int \frac{\rho_\chi(y^{(2)})}{\int \rho_\chi(w^{(1)}) \phi(w^{(1)}|y^{(2)})\, dw^{(1)}} \phi(y^{(1)}|y^{(2)}) \phi(y^{(2)}|x^{(1)})\, dy^{(2)}.$$

Let $\Gamma_2(\chi)$ denote the second eigenvalue of the Gibbs sampler for the density ϕ_χ, which is also the second eigenvalue of the symmetric operator R_χ on H_1. By the discussion above, $\langle B, R_\chi B\rangle = \lambda_M$ at $\chi = 0$. The derivative of $\langle B, R_\chi B\rangle$ in χ at $\chi = 0$ yields $\Gamma_2'(0)$, under the appropriate stability conditions which are addressed in the Appendix. [See Kato (1976), Theorem 2.6, page 445.] In other words,

$$\Gamma_2'(0) = \int B(x^{(1)})\left(\int B(y^{(1)})\frac{\partial R_\chi(x^{(1)}, y^{(1)})}{\partial\chi}\bigg|_0 dy_1\right)\phi(x^{(1)})\,dx^{(1)}. \tag{4}$$

Now

$$\frac{\partial R_\chi(x^{(1)}, y^{(1)})}{\partial\chi}\bigg|_0 = \int\left(\frac{1}{2}\big(q(x^{(1)}) + q(y^{(1)})\big) + r(y^{(2)})\right) \times\phi(y^{(1)}|y^{(2)})\phi(y^{(2)}|x^{(1)})\,dy^{(2)},$$

where

$$q(x^{(1)}) = -g_1(x^{(1)}) + \int g_2(u^{(2)})\phi(u^{(2)}|x^{(1)})\,du^{(2)}$$

and

$$r(y^{(2)}) = -g_2(y^{(2)}) + \int g_1(w^{(1)})\phi(w^{(1)}|y^{(2)})\,dw^{(1)}.$$

Substituting this in (4) we get

$$\begin{aligned}\Gamma_2'(0) = (1-\lambda_M)\sum_{a,b=1}^{l}\int v_a v_b x_a^{(1)} y_b^{(1)}\big[-\tfrac{1}{2}\big(g_1(x^{(1)}) + g_1(y^{(1)})\big) \\ + \tfrac{1}{2}\big(g_2(u^{(2)}) + g_2(v^{(2)})\big) - g_2(y^{(2)}) + g_1(w^{(1)})\big] \\ \times\Phi(u^{(2)}, v^{(2)}, w^{(1)}, x^{(1)}, y^{(1)}, y^{(2)})\,du^{(2)}\,dv^{(2)}\,dw^{(1)}\,dx^{(1)}\,dy^{(1)}\,dy^{(2)},\end{aligned} \tag{5}$$

where

$$\Phi(u^{(2)}, v^{(2)}, w^{(1)}, x^{(1)}, y^{(1)}, y^{(2)}) = \phi(u^{(2)}|x^{(1)})\phi(v^{(2)}|y^{(1)})\phi(w^{(1)}|y^{(2)}) \times\phi(y^{(1)}|y^{(2)})\phi(y^{(2)}|x^{(1)})\phi(x^{(1)}).$$

Observe that Φ is the density of a normal distribution in all the variables with inverse covariance matrix written in block form

$$\mathscr{C} = \begin{pmatrix} I_2 & 0 & 0 & C^t & 0 & 0 \\ 0 & I_2 & 0 & 0 & C^t & 0 \\ 0 & 0 & I_2 & 0 & 0 & C \\ C & 0 & 0 & I_1 + D_1 & 0 & C \\ 0 & C & 0 & 0 & I_1 + D_1 & C \\ 0 & 0 & C^t & C^t & C^t & I_2 + 2D_2 \end{pmatrix},$$

and covariance matrix

$$\mathscr{C}^{-1} = \begin{pmatrix} F_2 & D_2^2F_2 & -C^tD_1F_1 & -C^tF_1 & -C^tD_1F_1 & D_2F_2 \\ D_2^2F_2 & F_2 & -C^tD_1F_1 & -C^tD_1F_1 & -C^tF_1 & D_2F_2 \\ -CD_2F_2 & -CD_2F_2 & F_1 & D_1F_1 & D_1F_1 & -CF_2 \\ -CF_2 & -CD_2F_2 & D_1F_1 & F_1 & D_1F_1 & -CF_2 \\ -CD_2F_2 & -CF_2 & D_1F_1 & D_1F_1 & F_1 & -CF_2 \\ D_2F_2 & D_2F_2 & -C^tF_1 & -C^tF_1 & -C^tF_1 & F_2 \end{pmatrix}.$$

The calculation of $\Gamma_2'(0)$ as given in (5) reduces to calculating moments of Gaussians.

For any d dimensional multivariate Gaussian Z and any polynomial g on R^r define

$$\|g\|^{(2)} = E\left(|g(Z)|^2 + \sum_{a,b} |\partial_{ab}g(Z)|^2\right),$$

where $\partial_{ab}g$ are the second derivatives of g. Let $\mathscr{H}^{(2)}(Z)$, $i = 1, 2$, denote the closure of the space of polynomial functions in $\|\cdot\|^{(2)}$. This space contains all twice continuously differentiable functions with finite $\|\cdot\|^{(2)}$ norm.

LEMMA 3. *Let the random variables $(X, Y, Z_1, \ldots, Z_r)$, be multivariate Gaussian with mean zero and let $Z = (Z_1, \ldots, Z_r)$. For any $g \in \mathscr{H}^{(2)}(Z)$,*

$$EXYg(Z) = EXY \cdot Eg(Z) + \sum_{i,j=1}^{r} EXZ_i \cdot EYZ_j \cdot E\big(\partial_{ij}g(Z)\big). \tag{6}$$

PROOF. First let $g(z_1, \ldots, z_r) = z_1^{n_1} \cdots z_r^{n_r}$. Since the only nonzero cumulants of Gaussians are second order, moments of multivariate Gaussians can be expressed in terms of a sum of products of covariances, where the sum extends over all partitions into pairs, of the variables, and a variable is counted as many times as the corresponding power. See, for example, Ibragimov and Rozanov [(1978), Chapter I, (5.10)] or McCullagh [(1987), Section 2.3]. We obtain

$$E(XYg(Z)) = E(XY)\sum + 2\sum_{i<j} n_i n_j E(XZ_i)E(YZ_j)\sum_{ij} + \sum_{i=1}^{r} n_i(n_i - 1)E(XZ_i)E(YZ_i)\sum_{ii},$$

where $\sum$ is the sum of products of covariances over all partitions of $g(Z)$ into pairs and $\sum_{ij}$ is the sum over all partitions into pairs of $g(Z)$ with one i-variable omitted and one j-variable omitted, which are the same no matter which ones are omitted. Finally $\sum_{ii}$ is the sum over all partitions into pairs of $g(Z)$ with two i-variables omitted, which again are the same no matter which ones are omitted.

This can be rewritten as

$$E(XYg(Z)) = E(XY)Eg(Z) + 2\sum_{i<j} E(XZ_i)E(YZ_j)E\big(\partial_{ij}g(Z)\big) + \sum_{i=1}^{r} E(XZ_i)E(YZ_i)E\big(\partial_{ii}g(Z)\big),$$

which leads to the required result for monomials.

From linearity the result applies to any polynomial. By the density of the polynomials in $\mathscr{H}^{(2)}$, the result follows for all functions in this space. □

REMARK. The same result holds if X or Y are equal to Z_i for some $i = 1, \ldots, r$.

We are now in position to calculate $\Gamma_2'(0)$. Let ϕ be a Gaussian with inverse covariance matrix

$$Q = \begin{pmatrix} I_1 & C \\ C^t & I_2 \end{pmatrix}.$$

with I_1 an $l \times l$ identity matrix and I_2 an $(n-l) \times (n-l)$ identity matrix.

THEOREM 2. *If CC^t has a simple largest eigenvalue λ_M with normalized eigenvector v and if*

$$G = \begin{pmatrix} I_1 - 2CC^t & 0 & C \\ 0 & I_1 - 2CC^t & C \\ C^t & C^t & I_2 - 3C^tC \end{pmatrix}$$

is positive definite, then for any two functions $g_1, g_2 \in \mathscr{H}^{(2)}$ that are bounded from below, the derivative at $\chi = 0$ of the second eigenvalue of L_{ϕ_χ} is given by

$$\Gamma_2'(0) = -\lambda_M \int v^t R_1(x^{(1)}) v \phi(x^{(1)})\, dx^{(1)} - \int v^t C R_2(x^{(2)}) C^t v \phi(x^{(2)})\, dx^{(2)}, \tag{7}$$

where R_1, R_2 are the matrices of second derivatives of g_1, g_2, respectively.

PROOF. The stability of the perturbation, as required by Kato [(1976), Theorem 2.6, page 445], under the assumption that G is positive definite is proved in the Appendix. Let $U^{(2)}$, $V^{(2)}$, $W^{(1)}$, $X^{(1)}$, $Y^{(1)}$ and $Y^{(2)}$ be Gaussian

random vectors jointly distributed according to Φ. Apply Lemma 3 to (5) to get

$$\begin{aligned}
\Gamma_2'(0) &= (1-\lambda_M)\sum_{a,b=1}^{l}\int v_a v_b x_a^{(1)} y_b^{(1)}\big[-g_1(x^{(1)}) \\
&\qquad +g_1(w^{(1)})+g_2(u^{(2)})-g_2(y^{(2)})\big] \\
&\quad \times \Phi\big(u^{(2)},v^{(2)},w^{(1)},x^{(1)},y^{(1)},y^{(2)}\big)\,du^{(2)}\,dv^{(2)}\,dw^{(1)}\,dx^{(1)}\,dy^{(1)}\,dy^{(2)} \\
&= v^t\Big[-E\Big[X^{(1)}(X^{(1)})^t\Big]E\big[R_1(X^{(1)})\big]E\Big[X^{(1)}(Y^{(1)})^t\Big] \\
&\quad +E\Big[X^{(1)}(W^{(1)})^t\Big]E\big[R_1(W^{(1)})\big]E\Big[W^{(1)}(Y^{(1)})^t\Big] \\
&\quad -E\Big[X^{(1)}(Y^{(2)})^t\Big]E\big[R_2(Y^{(2)})\big]E\Big[Y^{(2)}(Y^{(1)})^t\Big] \\
&\quad +E\Big[X^{(1)}(U^{(2)})^t\Big]E\big[R_2(U^{(2)})\big]E\Big[U^{(2)}(Y^{(1)})^t\Big]\Big]v. \qquad (8)
\end{aligned}$$

The first equality follows from the symmetry between $x^{(1)}$ and $y^{(1)}$ and the symmetry between $u^{(2)}$ and $v^{(2)}$. In the second equality we have used the fact that the first terms coming from the right-hand side of (6) cancel because the covariance of $X^{(1)}$ is the same as that of $W^{(1)}$ and similarly for $U^{(2)}$ and $Y^{(2)}$. Substituting the covariances in (8) and using the fact that $CF_2 = F_1C$ and $F_1v = (1/(1-\lambda_M))v$, $D_1v = \lambda_M v$, we obtain

$$\begin{aligned}
\Gamma_2'(0) &= (1-\lambda_M)\big[-v^tE\big(R_1(X^{(1)})D_1F_1v - v^tCF_2E\big(R_2(Y^{(2)})\big)C^tv\big] \\
&= -\lambda_M v^tE\big(R_1(X^{(1)})\big)v - v^tCE\big(R_2(Y^{(2)})\big)C^tv \qquad (9) \\
&= -\lambda_M\int v^tR_1(x^{(1)})v\phi(x^{(1)})\,dx^{(1)} - \int v^tCR_2(x^{(2)})C^tv\phi(x^{(2)})\,dx^{(2)}.
\end{aligned}$$

□

REMARK. If Q has the more general form

$$Q = \begin{pmatrix} A & C \\ C^t & B \end{pmatrix},$$

the same type of result holds. Set $\tilde{C} = A^{-1/2}CB^{-1/2}$ and let $\tilde{\phi}$ be the Gaussian density with inverse covariance

$$\tilde{Q} = \begin{pmatrix} I_1 & \tilde{C} \\ \tilde{C}^t & I_2 \end{pmatrix}.$$

For $f \in L_2(R^n,\phi)$ define $Uf(x^{(1)},x^{(2)}) = f(A^{-1/2}x^{(1)}, B^{-1/2}x^{(2)})$. It is easily verified that U is a unitary map from $L_2(R^n,\phi)$ onto $L_2(R^n,\tilde{\phi})$ which maps H_1 onto $\tilde{H}_1$ and H_2 onto $\tilde{H}_2$. Thus the spectral properties of the two-block Gibbs sampler algorithm for ϕ are the same as those for $\tilde{\phi}$, which is of the form assumed in Theorem 2.

COROLLARY 1. *If g_1 and g_2 are convex, R_1 and R_2 are positive definite so that $\Gamma_2'(0) < 0$ and the speed of convergence increases for small χ.*

REMARK. A general convex perturbation of the form $g(x) = g(x^{(1)}, x^{(2)})$, which is not of the form $g_1(x^{(1)})g_2(x^{(2)})$, will not necessarily increase the speed of convergence. For example, take

$$g(x) = x^t \begin{pmatrix} I_1 & (1+\alpha)C \\ (1+\alpha)C^t & I_2 \end{pmatrix} x.$$

Then ϕ_χ would be Gaussian with covariance

$$\begin{pmatrix} (1+\chi)I_1 & (1+\chi+\chi\alpha)C \\ (1+\chi+\chi\alpha)C^t & (1+\chi)I_2 \end{pmatrix}.$$

The two-block Gibbs sampler algorithm for this density would have a second eigenvalue given by the spectral radius of

$$\frac{(1+\chi+\chi\alpha)^2}{(1+\chi)^2} CC^t,$$

which is greater than the spectral radius of CC^t. Therefore this convex perturbation decreases the speed of convergence.

COROLLARY 2. *If all the one-dimensional marginals of ϕ are the same and $g_i(x^i) = \sum_a g(x_a^i)$, $i = 1, 2$, for some univariate function g, then*

$$\Gamma_2'(0) = -2\lambda_M \int g''(x_1)\phi(x_1)\,dx_1.$$

PROOF. Under the assumptions, R_i are both diagonal and (7) becomes

$$\begin{aligned} \Gamma_2'(0) &= \left(-\lambda_M |v|^2 - |C^t v|^2\right) \int g''(x_1)\phi(x_1)\,dx_1 \\ &= -2\lambda_M \int g''(x_1)\phi(x_1)\,dx_1. \end{aligned}$$

□

COROLLARY 3. *If C is square and circulant, then the two marginals $\phi(x^{(1)})$ and $\phi(x^{(2)})$ are the same, and if $g_1 = g_2$, then*

$$\Gamma_2'(0) = -2\lambda_M \int v^t R_1(x^{(1)}) v \phi(x^{(1)})\,dx^{(1)}.$$

Random sweep strategy. A similar calculation can be carried out for the random sweep strategy. Again it is assumed that the diagonal entries of the inverse covariance matrix are all 1, and that the perturbation is of the form $\sum_{i=1}^n g(x_i)$. Also as in Corollary 2 it is assumed that all the one-dimensional marginals are the same. Since the details are very similar to those above, we

only provide a brief outline. The integral kernel for each of the n projections in $L_2(R^n, \phi)$ is simply the conditional density

$$\phi_\chi^{(i)}\big(y_i|x_{(i)}\big) = \frac{\exp\big(-\chi\big(g(x_i) + g(y_i)\big)\big)\phi\big(y_i|x_{(i)}\big)}{\int \exp\big(-\chi g(t_i)\big)\phi\big(t_i|x_{(i)}\big)}.$$

Using the unitary transformation from $L_2(R^n, \phi)$ to $L_2(R^n, \phi_\chi)$ given by $Uf(x) = \exp[\frac{1}{2}\Sigma_i g(x_i)]$, we obtain the corresponding integral kernel in $L^2(R^n, \phi)$ given by

$$R_\chi^{(i)}(x; y) = R_\chi^{(i)}(x_{(i)}; y_i) = \frac{\exp\big(-\frac{1}{2}\chi\big(g(x_i) + g(y_i)\big)\big)\phi\big(y_i|x_{(i)}\big)}{\int \exp\big(-\chi g(t_i)\big)\phi\big(t_i|x_{(i)}\big)}\, dt_i,$$

and the derivative in χ at 0 of $R_\chi^{(i)}$ is

$$\tilde{R}^{(i)}(x; y) = \int\big(-\tfrac{1}{2}\big[g(x_i) + g(y_i)\big] + g(t_i)\big)\phi\big(y_i|x_{(i)}\big)\phi\big(t_i|x_{(i)}\big)\, dt_i. \tag{10}$$

From Theorem 1 the eigenfunction $B(x)$ corresponding to the second largest eigenvalue is $\sqrt{\lambda_m}\langle b, x\rangle$, where λ_m is the smallest eigenvalue of Q. As in Section 3, the derivative of the eigenvalue is given by

$$\int B(x)\left(\int \tilde{R}(x; y)B(y)\, dy\right)\phi(x)\, dx,$$

where $\tilde{R}(x, y) = (1/n)\sum_{i=1}^n \tilde{R}^{(i)}(x; y)$. Writing this integral explicitly leads to various terms of the form addressed in Lemma 3. The calculation is lengthy but straightforward and yields

$$\begin{aligned}&\int B(x)\int \tilde{R}(x; y)B(y)\, dy\, \phi(x)\, dx\\ &\quad = -\frac{1}{n}(1 - \lambda_m)\int g''(x_1)\phi(x_1)\, dx_1.\end{aligned} \tag{11}$$

The $1/n$ factor disappears if one considers n steps at a time. Observe that the derivative for the two-block Gibbs sampler given in Corollary 2 is in terms of λ_M, which is the largest eigenvalue of $CC^t = D_1$, whereas λ_m in (11) is the smallest eigenvalue of the full inverse covariance matrix Q.

APPENDIX

Stability of the perturbation. Following are sufficient conditions under which R_χ is a stable perturbation.

LEMMA A1. *If*

$$G = \begin{pmatrix} I_1 - 2D_1 & 0 & C \\ 0 & I_1 - 2D_1 & C \\ C^t & C^t & I_2 - 3D_2 \end{pmatrix}$$

is positive definite, then $\|R_\chi - R_0\| \to 0$ *as* $\chi \to 0$.

PROOF. Since $\phi(y^{(1)}|y^{(2)})\phi(y^{(2)}|x^{(1)})$ is a density in $(y^{(1)}, y^{(2)})$ for each x, using the Cauchy–Schwarz inequality we have

$$
\begin{aligned}
&|R_\chi f - R_0 f|^2 \\
&\quad = \int\left[\int\big(R_\chi(x^{(1)}, y^{(1)}) - R_0(x^{(1)}, y^{(1)})\big)f(y^{(1)})\,dy^{(1)}\right]^2 \phi(x^{(1)})\,dx^{(1)} \\
&\quad = \int\left[\iint\big(T_\xi(x^{(1)}, y^{(1)}, y^{(2)}) - 1\big)f(y^{(1)})\phi(y^{(1)}|y^{(2)})\right. \\
&\qquad\qquad \left.\times\phi(y^{(2)}|x^{(1)})\,dy^{(1)}\,dy^{(2)}\right]^2 \phi(x^{(1)})\,dx^{(1)} \qquad (12)\\
&\quad \le |f|^2 \cdot \iiint\big(T_\chi(x^{(1)}, y^{(1)}, y^{(2)}) - 1\big)^2 \phi(y^{(1)}|y^{(2)}) \\
&\qquad \times\phi(y^{(2)}|x^{(1)})\,dy^{(1)}\,dy^{(2)}\,\phi(x^{(1)})\,dx^{(1)},
\end{aligned}
$$

where

$$
\begin{aligned}
&T_\chi(x^{(1)}, y^{(1)}, y^{(2)}) \\
&\quad = \rho_\chi^{1/2}(x^{(1)})\rho_\chi^{1/2}(y^{(1)})\rho_\chi(y^{(2)}) \\
&\qquad \times\left[\left(\int\rho_\chi(u^{(2)})\phi(u^{(2)}|x^{(1)})\,du^{(2)}\int\rho_\chi(v^{(2)})\phi(v^{(2)}|y^{(1)})\,dv^{(2)}\right)^{1/2}\right. \\
&\qquad\qquad \left.\times\int\rho_\chi(w^{(1)})\phi(w^{(1)}|y^{(2)})\,dw^{(1)}\right]^{-1}.
\end{aligned}
$$

Since $T_\chi(x^{(1)}, y^{(1)}, y^{(2)}) \to 1$ as $\chi \to 0$, the result will follow from a uniform integrability argument. Let

$$
s_\chi = \min\left[\frac{1}{\sqrt{2\pi}^l}\int\rho_\chi(t^{(2)})\exp(-|t^{(2)}|^2)\,dt^{(2)}, \frac{1}{\sqrt{2\pi}^{n-l}}\int\rho_\chi(t^{(1)})\exp(-|t^{(1)}|^2)\,dt\right].
$$

Since

$$
\begin{aligned}
\phi(t^{(2)}|t^{(1)}) &= \frac{1}{\sqrt{2\pi}^l}\exp\left[-\frac{1}{2}|t^{(2)} + C^t t^{(1)}|^2\right] \\
&\ge \frac{1}{\sqrt{2\pi}^l}\exp\big[-|t^{(2)}|^2\big]\exp\big[-|C^t t^{(1)}|^2\big],
\end{aligned}
$$

we have

$$
\int\rho_\chi(t^{(2)})\phi(t^{(2)}|t^{(1)})\,dt^{(2)} \ge s_\chi\exp(-|C^t t^{(1)}|^2) = s_\chi\exp(\langle -t^{(1)}, D_1 t^{(1)}\rangle)
$$

and similarly

$$
\int\rho_\chi(t^{(1)})\phi(t^{(1)}|t^{(2)})\,dt^{(1)} \ge s_\chi\exp(-|Ct^{(2)}|^2) = s_\chi\exp(\langle -t^{(2)}, D_2 t^{(2)}\rangle).
$$

Consequently,

$$\left|T_\chi\left(x^{(1)}, y^{(1)}, y^{(2)}\right)\right| \leq s^{-3} \exp\left(\tfrac{1}{2}\left[\langle x^{(1)}, D_1 x^{(1)}\rangle + \langle y^{(1)}, D_1 y^{(1)}\rangle + \langle y^{(2)}, 2D_2 y^{(2)}\rangle\right]\right),$$

where s is a lower bound on s_χ for χ in a neighborhood of 0. Hence the upper bound is independent of χ and all that remains is to prove that it is square integrable with respect to

$$\phi\left(y^{(1)}|y^{(2)}\right)\phi\left(y^{(2)}|x^{(1)}\right)\phi\left(x^{(1)}\right).$$

However, by writing out these densities together with the integrand and aggregating everything in the exponent, we get

$$\text{const} \int \exp\left[-\tfrac{1}{2}\left(x^{(1)}, y^{(1)}, y^{(2)}\right) \times \begin{pmatrix} I_1 - 2D_1 & 0 & C \\ 0 & I_1 - 2D_1 & C \\ C^t & C^t & I_2 - 3D_2 \end{pmatrix} \begin{pmatrix} x^{(1)} \\ y^{(1)} \\ y^{(2)} \end{pmatrix} \right] dx^{(1)}\, dy^{(1)}\, dy^{(2)},$$

which is finite if G is positive definite. □

Acknowledgment. I would like to thank Alan Sokal for his encouragement and for so many useful suggestions.

REFERENCES

AMIT, Y. (1991). On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. *J. Multivariate Anal.* **38** 82–100.

AMIT, Y. and GRENANDER, U. (1991). Comparing sweep strategies for stochastic relaxation. *J. Multivariate Anal.* **37** 197–222.

AMIT, Y., GRENANDER, U. and PICCIONI, M. (1991). Structural image restoration through deformable templates. *J. Amer. Statist. Assoc.* **86** 376–387.

BARONE, P. and FRIGESSI, A. (1990). Improving stochastic relaxation for Gaussian random fields. *Probab. Eng. Inform. Sci.* **4** 369–389.

BESAG, J. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc. Ser. B* **55** 25–37.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

DIACONIS, P. and SALOFF-COSTE, L. (1993). Comparison techniques for random walk on finite groups. *Ann. Probab.* **21** 2131–2156.

FRIEZE, A., KANNAN, R. and POLSON, N. G. (1994). Sampling from log-concave distributions. *Ann. Appl. Probab.* **4** 812–837. [Correction note (1994) **4** 1255.]

GEMAN, D. (1991). *Random Fields and Inverse Problems in Imaging. Lecture Notes in Math.* **1427**. Springer, Berlin.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation. Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721–741.

GOODMAN, J. and SOKAL, A. D. (1989). Multigrid Monte-Carlo method. Conceptual foundations. *Phys. Rev. D* **40** 2035–2071.

GRENANDER, U., CHOW, Y. and KEENAN, D. (1991). *HANDS: A Pattern Theoretic Study of Biological Shapes*. Springer, New York.

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Processes*. Springer, New York.

KATO, T. (1976). *Perturbation Theory for Linear Operators*, 2nd ed. Springer, New York.

KNOERR, A. P. (1988). Global models of natural boundaries: Theory and applications. Ph.D. Thesis, Div. of Applied Mathematics, Brown Univ.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall, London.

MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121.

ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** 558–566.

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 3–23.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5764 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637-1514