Illinois Journal of Mathematics
Volume 54, Number 3, Fall 2010, Pages 963–979
S 0019-2082

STOCHASTIC ALTERNATING PROJECTIONS

PERSI DIACONIS, KSHITIJ KHARE AND LAURENT SALOFF-COSTE

ABSTRACT. We show how basic work of Don Burkholder on iterated conditional expectations is intimately connected to a standard tool of scientific computing—Glauber dynamics (also known as the Gibbs sampler). We begin with von Neumann's alternating projection theorem using an example of Burkholder's. We then review Burkholder's theorem. Finally, we introduce Glauber dynamics and show how Burkholder's theorem can be harnessed to prove convergence. In the other direction, we show how classical convergence rates involving the angle between subspaces can be substantially refined in several cases.

1. Introduction

Alternating projection algorithms are basic tools of applied mathematics. They are used for minimization in coordinate descent, to project onto convex sets and (as seen below) as a basic tool for simulation. One mathematical foundation for these algorithms is von Neumann's alternating projection theorem.

THEOREM 1 (von Neumann). *Let P_1, P_2 be the orthogonal projections onto closed subspaces $\mathcal{M}_1$, $\mathcal{M}_2$ of a Hilbert space $\mathcal{H}$. Let P_I be the orthogonal projection onto the intersection $\mathcal{M}_1 \cap \mathcal{M}_2$. If $T = P_2P_1$, then $T^k \to P_I$ as $k \to \infty$. That is, $\|T^k(h) - P_I(h)\| \to 0$ for each $h \in \mathcal{H}$.*

We review the literature on von Neumann's theorem, extensions to several projections and rates of convergence in Section 2. We present there a charming, unpublished illustration shown to us by Don Burkholder.

In the special case that $\mathcal{H} = L_2(\mu)$ for some probability measure μ and $\mathcal{M}_i$ are subspaces of measurable functions with respect to two sub σ-algebras,

Received November 5, 2009; received in final form April 25, 2010.
2010 *Mathematics Subject Classification.* Primary 60J10, 46N30. Secondary 60J27, 65C40.

Burkholder gave a definitive treatment of von Neumann's theorem, providing necessary and suffucient conditions on h for convergence to hold. He showed (roughly) that convergence takes place almost surely if and only if h above is in $L\log(1+L)$, that is, $\int |h|\log(1+|h|)\,d\mu < \infty$. We review this result and some recent progress in Section 3.

In Section 4, we show that Burkholder's theorem (and von Neumann's theorem) are useful in proving basic convergence results for a widely used algorithm—Glauber dynamics. We explain the algorithm and the connection. The rate of convergence in von Neumann's theorem is shown to be related to the probabilists' 'maximal correlation'. The Hilbert space notion of strong convergence is shown to be fairly weak in the probabilistic setting where much more demanding topologies are standard fare. This suggests new problems in both the Hilbert space and measure space settings.

2. Alternating projection algorithms

We begin with a charming demonstration of the theorems on alternating projections shown to us by Don Burkholder. Take a piece of string about two feet long. Attach two paper clips at two arbitrary positions. Call these the "Left" and "Right" paper clips (see Figure 1(a)).

At any stage, proceed as follows: Fold the right end of the string over to touch the left paper clip. Holding this clip (Figure 1(b), top) (and the right end) with the left hand fingers, slide the right clip to the right until it hits the right end of the loop formed (Figure 1(b), bottom).

Unfold the string, fold the left end of the string over to touch the right clip (Figure 1(c), top). Hold it there with the right hand fingers and slide the left-most clip to the left until it hits the left end of the loop formed (Figure 1(c), bottom).

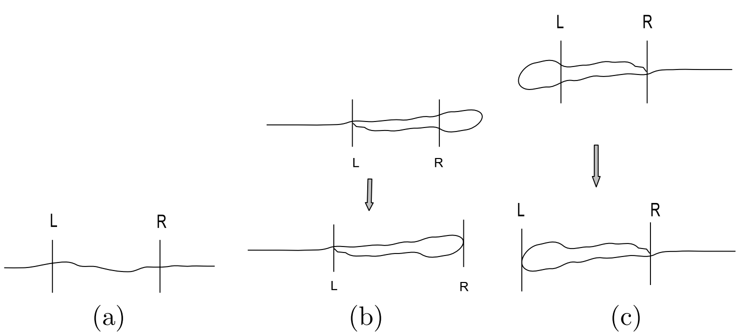

FIGURE 1. Successive stages for Burkholder's alternating projection demonstration.

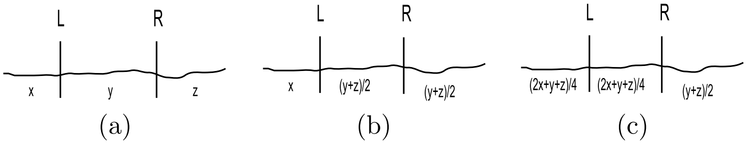

FIGURE 2. Lengths during one iteration of the demonstration.

Unfold the string. These two stages constitute one pass of the algorithm. If this is repeated a few times, the position of the clips will converge to $\frac{1}{3}$ and $\frac{2}{3}$ of the total length.

To make the connection with alternating projections, suppose the clips are originally at distance x and $x+y$ on a string of length $x+y+z$ (Figure 2(a)). Folding the right end over and sliding the right clip results in (Figure 2(b)). Folding the left end over and sliding the left clip results in (Figure 2(c)).

These transformations correspond to two projections.

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \qquad P_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{with } P_2P_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Since it is doubly stochastic with $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]^T$ as a unique stationary vector,

$$(P_2P_1)^n \begin{bmatrix} x \\ y \\ z \end{bmatrix} \to \begin{bmatrix} \frac{x+y+z}{3} \\ \frac{x+y+z}{3} \\ \frac{x+y+z}{3} \end{bmatrix}.$$

If one were to demonstrate this, say by making a prediction for the length of the leftmost clip before the initial placement of the clips, it is desirable to know how many iterations are required for convergence. The eigenvalues and right eigenvectors of P_2P_1 are

$$1, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \frac{1}{4}, \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \qquad 0, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Then

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{x+y+z}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \left(\frac{2}{3}x - \frac{y}{3} - \frac{z}{3}\right) \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + (y-x) \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Thus,

$$(P_2P_1)^n \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{x+y+z}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{4^n}\left(\frac{2}{3}x - \frac{y}{3} - \frac{z}{3}\right) \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}.$$

Given $x+y+z=c$, the error is largest when $x=c, y=z=0$. This corresponds to both clips staring at the right end of the string. The deviation of the leftmost clip from $\frac{c}{3}$ after n steps is $\frac{2c}{3(4^n)}$. For $c=2$ feet, if we want an error of $\frac{1}{10}$ inch, we must take $n \geq 4$. For the historical record, we note that Burkholder demonstrated this to us by repeatedly folding a 3×5 index card instead of a piece of string.

Von Neumann's alternating projection theorem stated above has many extensions and applications. A splendid textbook account appears in [19, Chapter 1], while applications are surveyed in [18]. These trace the history back to Schwarz [34], who used the method to solve the Dirichlet problem on a region given as a union of regions each having a simple-to-solve Dirichlet problem (e.g., a union of disks). Von Neumann's theorem has been widely developed and applied. For alternating minimization procedures, see [5]. For convex optimization using random alternating projections, see [14]. Applications to best approximation are in [6]. Proofs of extensions of the Ergodic theorem are in [9]. A useful survey geared towards projections into the intersections of convex sets is in [4].

One important extension is due to Halperin [23], who shows that the conclusion holds as stated for n subspaces with $T = P_1 P_2 \cdots P_n$. An elegant proof of this using elementary arguments is in [29]. See Kassabov [25] for crucial refinements in the present context.

There is some classical work on the rate of convergence in von Neumann's theorem. If $\mathcal{M}_1$, $\mathcal{M}_2$ are closed subspaces of the Hilbert Space $\mathcal{H}$, let

$$c = \sup\{\langle v_1, v_2\rangle \mid v_i \in \mathcal{M}_i \cap (\mathcal{M}_1 \cap \mathcal{M}_2)^{\perp}, \|v_i\| \leq 1\}. \tag{1}$$

This is the cosine of the angle between $\mathcal{M}_1$ and $\mathcal{M}_2$. If P_i is the projection into $\mathcal{M}_i$ and P_I is the orthogonal projection into $\mathcal{M}_1 \cap \mathcal{M}_2$, Aronszajn [2], proved that

$$\|(P_2 P_1)^n(x) - P_I(x)\| \leq c^{2n-1}\|x\| \quad \text{for all } x \in \mathcal{H}. \tag{2}$$

This result is best possible, and some extensions to more subspaces are available. See [19, p. 220] and Section 5.5 below. To compare the exact convergence rate with (1), (2) above, we note that the cosine of the two subspaces in the Burkholder's paper clip example is $c = \frac{1}{2}$. Section 4.2 below contains illustrations for the Gibbs sampler.

3. Burkholder's theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and consider the Hilbert space $\mathcal{H} = L^2(P)$. Let $\mathcal{A}_1$, $\mathcal{A}_2$ be sub-σ algebras of $\mathcal{F}$ and $\mathcal{M}_1, \mathcal{M}_2$ the closed subspaces of $\mathcal{A}_1, \mathcal{A}_2$ measurable functions respectively in $\mathcal{H}$. For $U \in L^2(P)$, it is well known that the orthogonal projection of U onto $\mathcal{M}_i$ is given by the conditional expectation $\mathbf{E}(U|\mathcal{A}_i)$. Thus let

$$U_0 = U, \qquad U_{2i+1} = \mathbf{E}(U_{2i}|\mathcal{A}_1), \qquad U_{2i+2} = \mathbf{E}(U_{2i+1}|\mathcal{A}_2), \quad 0 \le i < \infty.$$

Burkholder and Chow [13] proved that

$$U_n \to \mathbf{E}(U|\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2), \quad \text{almost surely.} \tag{3}$$

Here $\bar{\mathcal{A}}_1, \bar{\mathcal{A}}_2$ denote the P-completions of the σ-algebra's $\mathcal{A}_1, \mathcal{A}_2$, respectively. The original motivation for studying (3) came from mathematical statistics. In trying to construct 'minimal sufficient subfields' Burkholder [12] found examples of families of probability measures admitting $\mathcal{A}_1$ and $\mathcal{A}_2$ as sufficient subfields with $\mathcal{A}_1 \cap \mathcal{A}_2$ not sufficient. This is still surprising, 50 years later. The limiting result (3) allowed him to get his hands on $\mathbf{E}(U|\mathcal{A}_1 \cap \mathcal{A}_2)$.

The convergence (3) was abstracted and refined by Burkholder [10], [11], [13], Rota [33] with recent results in [7], [16], [17], [35]. Burkholder [11] and Ornstein [31] show that convergence in (3) can fail if only $X \in L^1(P)$. These authors show that a necessary and sufficient condition is that

$$\int |X(\omega)| \log\bigl(1 + |X(\omega)|\bigr) P(d\omega) < \infty. \tag{4}$$

The equivalence (3) $\Leftrightarrow$ (4) is called Burkholder's theorem.

The extension to more than two σ-algebras was open for more than forty years. Consider the case of three σ-algebras $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$. If the iterations are taken in order

$$\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_2, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_2, \mathcal{A}_1, \ldots$$

then the original arguments go through to show almost sure convergence to $\mathbf{E}(U|\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 \cap \bar{\mathcal{A}}_3)$.The straightforward extension of Halperin's theorem, where the iterations are taken in order

$$\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \ldots$$

resisted solution. Convergence was finally proved for $U \in L_2$ by Delyon and Delyon [17], and for $U \in L_p$ by Cohen [16]. Their argument uses a fascinating extension of spectral theory to nonnormal operators. It cries out for a more probabilistic proof.

We close this section with three remarks on Burkholder's theorem. The first remark translates Burkholder's folding example (Section 2) into probability language, the second remark deals with projections which cannot be represented via conditional expectation, and the third remark deals with null sets.

REMARK 1. It is straightforward to give a probabilistic setting for the folded string example at the start of Section 2. Let $\Omega = \{1, 2, 3\}$, $\mathcal{F} = \{$All subsets of $\Omega\}$, $P(\{1\}) = P(\{2\}) = P(\{3\}) = \frac{1}{3}$. Using the notation of the example, let $U \in L^2(P)$ with $U(1) = x, U(2) = y, U(3) = z$. Set

$$\mathcal{A}_1 = \sigma(\{1, 2\}, \{3\}), \qquad \mathcal{A}_2 = \sigma(\{1\}, \{2, 3\}).$$

Then,

$$\mathbf{E}(U|\mathcal{A}_1)(i) = \begin{cases} \frac{x+y}{2} & \text{if } i=1, \\ \frac{x+y}{2} & \text{if } i=2, \\ z & \text{if } i=3. \end{cases} \qquad \mathbf{E}(U|\mathcal{A}_2)(i) = \begin{cases} x & \text{if } i=1, \\ \frac{y+z}{2} & \text{if } i=2, \\ \frac{y+z}{2} & \text{if } i=3. \end{cases}$$

Here $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 = \{\phi, \Omega\}$ and $\mathbf{E}(U|\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2)(i) = \frac{x+y+z}{3}$ for $i = 1,2,3$. The successive foldings amount precisely to iterated expectations.

REMARK 2. Not all projections into closed subspaces can be realized by computing conditional expectations. For example, consider $\mathcal{H} = L^2((-\pi,\pi), \frac{dx}{2\pi})$. The subspace $\mathcal{M}_1$ of functions which vanish (almost surely) on a subset of $(-\pi,\pi)$ is closed, but not the range of conditional expectations given a sub-σ-algebra (the constant functions are not in $\mathcal{M}_1$). The subspace $\mathcal{M}_2$ of functions with vanishing negative Fourier coefficients is a closed subspace containing the constants, but not the range of conditional expectation given a sub-σ-algebra (the projection of a positive function in $\mathcal{M}_2$ need not be positive). Alternating projections into these subspaces are a key ingredient of the classical work of Landau, Logan, Pollack and Slepian on band limited functions. See [26] for a readable overview. An elegant necessary and sufficient condition for a subspace of $L^2(P)$ to be the range of a conditional expectation operator (for some sub-σ-algebra) is given in Neveu [30, Exercise IV.3.1, p. 123]. The subspace V must be closed, contain the constants and if f is in V, then $\max(f,0)$ must be in V.

EXAMPLE 1. For a finite space $\mathcal{X} = \{1,2,\dots,n\}$, let $\theta_i > 0, \sum_{i=1}^n \theta_i = 1$. A σ-field $\mathcal{A}$ on $\mathcal{X}$ is specified by a partition $\{A_1, A_2, \dots, A_k\}, A_i \neq \emptyset, A_i \cap A_j = \emptyset, \bigcup_{i=1}^k A_i = \mathcal{X}$. Let $P_{\mathcal{A}}$ be the projection from $L^2(\mathcal{X}, \underline{\theta})$ to $L^2(\mathcal{X}, \mathcal{A}, \underline{\theta})$. The matrix of $P_{\mathcal{A}}$ has (i,j) entry $\bar{\theta}_j \delta_{\mathcal{A}}(i,j)$, $1 \le i,j \le n$, where $\delta_{\mathcal{A}}(i,j)$ is one or zero as i and j are in the same block A_l and $\bar{\theta}_j = \frac{\theta_j}{\sum_{i \in A_l} \theta_i}$. It follows that the number of subspaces of $L^2(\mathcal{X}, \underline{\theta})$ which are the range of a conditional expectation operator is the Bell number $B(n)$ (that is, the number of set partitions on $\{1,2,\dots,n\}$). There are a continuum of other subspaces.

REMARK 3. The projection on the right side of (3) is onto the $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2$ measurable functions. Without the completions, convergence may fail. We find the following examples instructive.

EXAMPLE 2. Let $(\Omega, \mathcal{F})$ be the Borel unit square, with $\mathcal{A}_i$ as in Example 2, with P the uniform probability on the diagonal Δ. Take $\mathcal{A}_i = \sigma(X_i)$ with $X_i(\omega) = \omega_i$ for $\omega = (\omega_1, \omega_2)$. Then $\mathcal{A}_1 \cap \mathcal{A}_2 = \{\phi, \Omega\}$, and $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 = \mathcal{F}$. Here $U_{2i+1}(\omega_1, \omega_2) = U(\omega_1, \omega_1)$ and $U_{2i+2}(\omega_1, \omega_2) = U(\omega_2, \omega_2)$. The iterations do not converge.

EXAMPLE 3. Let Ω be the Borel unit square. Let P be the uniform distribution supported on the upper left and lower right quarter squares. Now P

has a density $f(x,y)$ with respect to Lebesgue measure on Ω, but $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2$ contains the 4 quarter squares. Thus, the iterations do not converge to constant functions.

A crucial step in connecting the Burkholder–Chow result (3) and other results below to the Gibbs sampler is understanding when $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 = \overline{\mathcal{A}_1 \cap \mathcal{A}_2}$. After all, if $\Omega = \Omega_1 \times \Omega_2$ is a product space and $\mathcal{A}_i$ is the σ-algebra generated by the projection on the ith coordinate, $\mathcal{A}_1 \cap \mathcal{A}_2$ is the trivial σ-algebra and (3) then says U_n converges to $\mathbf{E}(U)$ if $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 = \overline{\mathcal{A}_1 \cap \mathcal{A}_2}$. An elegant necessary and sufficient condition has been developed in response to questions raised by an early version of the present paper by Patrizia Berti, Luca Pratelli and Pietro Rigo [7], [8]. Here is one version of their result.

THEOREM 2 (Berti, Pratelli, Rigo). *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{F}$ *sub* σ*-fields. Let* $\mathcal{N} = \{F \in \mathcal{F} : P(F) \in \{0,1\}\}, \bar{\mathcal{G}} = \sigma(\mathcal{G} \cup \mathcal{N})$ *for any subclass* $\mathcal{G} \subseteq \mathcal{F}$. *In order that* $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 = \overline{\mathcal{A}_1 \cap \mathcal{A}_2}$ *it is necessary and sufficient that*

$$A_1 \in \mathcal{A}_1, \qquad A_2 \in \mathcal{A}_2 \quad \textit{and} \quad P(A_1 \cap A_2) = P(A_1^c \cap A_2^c) = 0$$

implies $P(A_1 \triangle B) = 0$ *or* $P(A_2 \triangle B) = 0$ *for some* $B \in \mathcal{A}_1 \cap \mathcal{A}_2$.

They give many useful corollaries, some of which are down to earth and useful for the Gibbs sampling application. See Section 4.

4. Glauber dynamics

Glauber dynamics, also known as the Gibbs sampler or the Heat–Bath algorithm, is an important tool of scientific computing. It gives a way of sampling from an intractable high dimensional probability density (perhaps known up to a normalizing constant) by a sequence of one-dimensional updates. We will not attempt to review the extensive literature on the Gibbs Sampler here. See [15] for a gentle introduction, [27] for a textbook treatment and [24], [32] for the literature on rates of convergence. A more detailed review is in [20, Section 2] and [21]. For background on Markov chains and Monte Carlo methods see [3], [24], [28].

In the present paper, we focus on two-component Gibbs Samplers. Thus, let $f(x,y)$ be a probability density with respect to the σ-finite measure $\mu \times \nu$ on $\mathcal{X} \times \mathcal{Y}$. This has marginal densities

$$m_1(x) = \int f(x,y)\nu(dy), \qquad m_2(y) = \int f(x,y)\mu(dx).$$

For simplicity, suppose $0 < m_1(x), m_2(y) < \infty$ for all x, y. Then, the conditional densities

$$f(x|y) = \frac{f(x,y)}{m_2(y)}, \qquad f(y|x) = \frac{f(x,y)}{m_1(x)}$$

are well defined. The Gibbs Sampler is a Markov chain which may be described by

- From (x,y) choose y' from $f(y'|x)$ and then x' from $f(x'|y')$.

This gives a kernel (w.r.t. $\mu \times \nu$):

$$k(x,y;x',y') = f(y'|x)f(x'|y').$$

Let $k^n(x,y;x',y') = \int k^{n-1}(x,y;w,z)k(w,z;x',y')\mu(dw)\nu(dz)$, with K^n the associated operator on $L^2(f)$. There is an obvious extension to higher dimensional problems. We stick to the two dimensional case for most of this section.

Here is a concrete example used throughout this section. Let

$$\mathcal{X} = \{0,1,2,3,\ldots\}, \qquad \mathcal{Y} = (0,\infty),$$

$\mu(dx) =$ Counting measure, and $\nu(dy) =$ Lebesgue measure. Set

$$f(x,y) = \frac{e^{-2y}y^x}{x!}.$$

This example is natural in a Bayesian Statistics setting where

$$f(x|y) = \frac{e^{-y}y^x}{x!}$$

is the Poisson distribution with parameter y. If e^{-y} is taken as the prior density of y, the joint density is $f(x,y)$. The conditional density

$$f(y|x) = \frac{2^{x+1}e^{-y}y^x}{x!}$$

is the $\Gamma\bigl(x+1,\frac{1}{2}\bigr)$ density.

The successive steps in Glauber dynamics amount to iterated projections in the Hilbert space $L^2(f)$. This allows the theorems of von Neumann and Burkholder (and their extensions) to be brought in. The connection with von Neumann's theorem has been used by Amit [1] to give rates of convergence for special cases. These ideas are extended below. In Section 4.1, we make a more explicit connection between projections and the steps of Glauber dynamics. Section 4.2 pulls together some (known) results linking angles between subspaces, maximal correlation and spectral gaps. Section 4.3 makes explicit the strengths and weaknesses of the rate from von Neumann's theorem. Section 4.4 explains how Burkholder's theorem coupled with a result of Berti–Pratelli–Rigo [7] translates into a natural ergodic theorem when specialized to Glauber dynamics.

4.1. Gibbs sampling as alternating projections. We now show that the Gibbs sampler can be regarded as an alternating projection algorithm. This connection was first exploited by Amit [1]. He used an extension of Halperin's bound to get rates of convergence in a weak L^2 distance for Gaussian variables and then used the comparison approach of [22] to treat perturbations. As explained below, these techniques give bounds that can be "off" by orders of magnitude, in natural problems that are far from Gaussian.

Let $(\mathcal{X}, \mu(dx)), (\mathcal{Y}, \nu(dy))$ be σ-finite measure spaces. Let $f(x,y)$ be a probability density with respect to $\mu \times \nu$. This determines a Hilbert space $L^2(f)$. If $X(x,y) = x, Y(x,y) = y$ are the coordinate projections and $\mathcal{A}_1 = \sigma(Y), \mathcal{A}_2 = \sigma(X)$ the associated σ-algebras, let the marginals be

$$m_X(x) = \int f(x,y)\nu(dy), \qquad m_Y(y) = \int f(x,y)\mu(dx).$$

Let

$$\mathcal{M}_1 = L^2(\sigma(Y), f) \simeq L^2(m_Y), \qquad \mathcal{M}_2 = L^2(\sigma(X), f) \simeq L^2(m_X).$$

These are closed subspaces of $L^2(f)$ and the orthogonal projections onto $\mathcal{M}_1, \mathcal{M}_2$ are realized by the conditional expectations $\mathbf{E}(\cdot|\mathcal{A}_1), \mathbf{E}(\cdot|\mathcal{A}_2)$ respectively. See [30, Proposition IV.3.1, p. 122].

Consider the Gibbs Sampling Markov chain introduced at the beginning of this section. This has transition density

$$k(x,y;x',y') = f(x'|y')f(y'|x). \tag{5}$$

We provide one of our main results.

THEOREM 3. *Let K be the operator on $L^2(f)$ associated to the Gibbs sampling kernel (5). Let P_1, P_2 be orthogonal projections onto the subspaces $\mathcal{M}_1, \mathcal{M}_2$ defined above. Then,*

$$K = P_2P_1 \quad \text{and so } K^n = (P_2P_1)^n \text{ for } n = 0, 1, 2, \ldots.$$

Proof. Using (5), for $U \in L^2(f)$,

$$K(U)(x,y) = \int_{\mathcal{X}}\int_{\mathcal{Y}} U(x',y')f(y'|x)f(x'|y')\nu(dy')\mu(dx'),$$

and,

$$\begin{aligned} P_2P_1(U)(x,y) &= E[P_1(U)|X = x] \\ &= E[E[U|\sigma(Y)]|X = x] \\ &= \int_{\mathcal{Y}}\left\{\int_{\mathcal{X}} U(x',y')f(x'|y')\mu(dx')\right\}f(y'|x)\nu(dy') \\ &= \int_{\mathcal{Y}}\int_{\mathcal{X}} U(x',y')f(x'|y')f(y'|x)\mu(dx')\nu(dy') \\ &= \int_{\mathcal{X}}\int_{\mathcal{Y}} U(x',y')f(y'|x)f(x'|y')\nu(dy')\mu(dx'). \end{aligned}$$

This proves the result for $n=1$. The result for general n now follows by induction. $\square$

The argument can be generalized to higher dimensions. Let $(\mathcal{X}_i, \mu_i(dx_i))$, be σ-finite measure spaces for $i=1,2,\dots,k$. Let $f(x_1,x_2,\dots,x_k)$ be a probability density with respect to $\prod_{i=1}^k \mu_i$. In what follows, we will write $\prod_{i=1}^k dx_i$ for the σ-finite dominating measure $\prod_{i=1}^k \mu_i(dx_i)$. Let $X_i(x_1,x_2,\dots,x_k)=x_i, i=1,2,\dots,k$, be the coordinate projections. Define the σ-algebras

$$\mathcal{A}_j = \sigma(X_1,X_2,\dots,X_{j-1},X_{j+1},\dots,X_k), \quad 1 \le j \le k,$$

and the corresponding Hilbert spaces

$$\mathcal{M}_j = L^2(\mathcal{A}_j, f), \quad 1 \le j \le k.$$

The orthogonal projection P_j onto $\mathcal{M}_j$ is realized by the conditional expectation $\mathbf{E}(\cdot|\mathcal{A}_j), 1 \le j \le k$. Consider the Gibbs sampling algorithm with transition density

$$k(x_1,x_2,\dots,x_k;x_1',x_2',\dots,x_k') = \prod_{i=1}^k f(x_i'|x_1,x_2,\dots,x_{i-1},x_{i+1}',\dots,x_k'). \tag{6}$$

THEOREM 4. *Let K be the operator on $L^2(f)$ associated to the Gibbs sampling chain* (6). *Then*,

$$K = P_k P_{k-1} \cdots P_1 \quad \textit{and so } K^n = (P_k P_{k-1} \cdots P_1)^n \textit{ for all } 0 \le n < \infty.$$

Proof. Let $\underline{x} = (x_1,x_2,\dots,x_n)$ and $\underline{x'} = (x_1',x_2',\dots,x_n')$. Using (6), for $U \in L^2(f)$, we have,

$$\begin{aligned} &K(U)(\underline{x}) \\ &\quad = \int_{\mathcal{X}_1}\int_{\mathcal{X}_2}\cdots\int_{\mathcal{X}_k} U(\underline{x'}) \prod_{i=1}^k f(x_i'|x_1,x_2,\dots,x_{i-1},x_{i+1}',\dots,x_k') \prod_{i=1}^k dx_i'. \end{aligned}$$

Define $U_0 = U$ and

$$U_{ki+j} = \mathbf{E}\big(U_{ki+(j-1)}|\mathcal{A}_j\big), \quad j=1,2,\dots,k, i=0,1,2,\dots.$$

Then,

$$\begin{aligned} &P_k P_{k-1}\cdots P_1(U)(x_1,x_2,\dots,x_k) \\ &\quad = U_k(x_1,x_2,\dots,x_{k-1}) \\ &\quad = \mathbf{E}[U_{k-1}|\mathcal{A}_k](x_1,x_2,\dots,x_{k-1}) \\ &\quad = \int_{\mathcal{X}_k} U_{k-1}(x_1,x_2,x_3,\dots,x_{k-2},x_k') f(x_k'|x_1,x_2,\dots,x_{k-1})\,dx_k' \\ &\quad = \int_{\mathcal{X}_k}\int_{\mathcal{X}_{k-1}} U_{k-2}(x_1,\dots,x_{k-2},x_{k-1}',x_k') f(x_{k-1}'|x_1,\dots,x_{k-2},x_k') \\ &\qquad \times f(x_k'|x_1,x_2,\dots,x_{k-1})\,dx_{k-1}'\,dx_k'. \end{aligned}$$

The last statement uses $U_{k-1} = \mathbf{E}[U_{k-2}|\mathcal{A}_{k-1}]$. Continuing like this we get,

$$P_k P_{k-1} \cdots P_1(U)$$
$$= \int_{\mathcal{X}_k} \int_{\mathcal{X}_{k-1}} \cdots \int_{\mathcal{X}_1} U(\underline{x'}) \prod_{i=1}^{k} f(x'_i | x_1, x_2, \ldots, x_{i-1}, x'_{i+1}, \ldots, x'_k) \prod_{i=1}^{k} dx'_i.$$

This proves the result for $n = 1$. The result for general n now follows by induction. □

4.2. Angles between subspaces, maximal correlation and spectral gaps. The results of this section are known, or part of the folklore. We use them below and present them in a unified fashion for the readers convenience. With notation as in Section 4.1, the bound in Aronszajn's estimate of the convergence of $(P_2P_1)^n$ to P_I involves the cosine of angle between $\mathcal{M}_1$ and $\mathcal{M}_2$:

$$c = \sup\{\langle v_1, v_2\rangle \mid v_i \in \mathcal{M}_i \cap (\mathcal{M}_1 \cap \mathcal{M}_2)^{\perp}, \|v_i\| \le 1, i = 1, 2\}.$$

Recall that in the setup of Section 4.1, $P_i = \mathbf{E}(\cdot|\mathcal{A}_i), i = 1, 2$. The maximal correlation between $\mathcal{A}_1$ and $\mathcal{A}_2$ is defined by

$$\gamma(\mathcal{A}_1, \mathcal{A}_2) = \sup\{\mathbf{E}(X_1X_2) \mid X_i \in \mathcal{M}_i, \mathbf{E}(X_i) = 0, \mathbf{E}(X_i^2) \le 1, i = 1, 2\}.$$

Finally, consider the operator $Q = P_1P_2P_1 : L^2(\sigma(X), f) \to L^2(\sigma(X), f)$. This is a self-adjoint contraction. Let β_1 be the second largest eigenvalue of Q; β_1 may be taken as the maximum of the support of the spectral measure of $I - Q$, if eigenvalues do not exist. The classical minimax characterization of eigenvalues shows that

$$\beta_1 = \sup\{\langle Qg, g\rangle \mid g \in \mathcal{M}_1, \mathbf{E}(g) = 0, \mathbf{E}(g^2) = 1\}.$$

THEOREM 5. *Under the assumption $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 = \{\phi, \Omega\}$, with notation as above,*

$$c = \gamma(\mathcal{A}_1, \mathcal{A}_2) = \sqrt{\beta_1}.$$

Proof. Note that if $\bar{\mathcal{A}}_1 \cap \bar{\mathcal{A}}_2 = \{\phi, \Omega\}$, then $\mathcal{M}_1 \cap \mathcal{M}_2$ consists of constant functions. Hence, $(\mathcal{M}_1 \cap \mathcal{M}_2)^{\perp}$ consists of mean zero functions. Since $P_i = \mathbf{E}(\cdot|\mathcal{A}_i)$ it follows that $\gamma(\mathcal{A}_1, \mathcal{A}_2) = c$.

By definition of the orthogonal projection P_1, for any $h \in \mathcal{H}$, we have

$$\langle P_1h, h\rangle^{1/2} = \|P_1h\| = \max\{\langle g, h\rangle\} \mid g \in \mathcal{M}_1, \|g\| \le 1\}. \tag{7}$$

Indeed, for any $g \in \mathcal{M}_1$, $h - P_1h$ is orthogonal to g and thus $\langle g, h\rangle = \langle g, P_1h\rangle$. It follows that

$$\max\{\langle g, h\rangle\} \mid g \in \mathcal{M}_1, \|g\| \le 1\} = \|P_1h\|.$$

The first equality in (7) follows from the fact that P_1 is self-adjoint and is a projection. For any $v_2 \in \mathcal{M}_2 \cap (\mathcal{M}_1 \cap \mathcal{M}_2)^{\perp}$, we have

$$\sup\{\langle v_1, v_2\rangle \mid v_1 \in \mathcal{M}_1 \cap (\mathcal{M}_1 \cap \mathcal{M}_2)^{\perp}, \|v_1\| \le 1\} = \langle P_1v_2, v_2\rangle^{1/2}$$

and it follows that

$$c = \sup\{\langle P_1 v_2, v_2\rangle^{1/2} \mid v_2 \in \mathcal{M}_2 \cap (\mathcal{M}_1 \cap \mathcal{M}_2)^\perp, \|v_2\| \le 1\}. \tag{8}$$

This can also be understood as saying that the norm of P_1 as an operator from $\mathcal{M}_2 \cap (\mathcal{M}_1 \cap \mathcal{M}_2)^\perp$ to $\mathcal{M}_1$ is bounded by c. By symmetry, c is also a bound on the norm of P_2 acting from $\mathcal{M}_1 \cap (\mathcal{M}_1 \cap \mathcal{M}_2)^\perp$ to $\mathcal{M}_2$.

In the Gibbs sampling setup, the operator $Q = P_2P_1P_2 : L^2(\sigma(X), f) \to L^2(\sigma(X), f)$ is evidently self-adjoint and corresponds to the marginal x-chain of the Gibbs sampling markov chain. Its norm

$$\|Q\|_0 = \|Q\|_{L_0^2(\sigma(X),f)\to L_0^2(\sigma(X),f)}$$

on

$$L_0^2(\sigma(X), f) = \{g \in L^2(\sigma(X), f) \mid \mathbf{E}(g) = 0\}$$

can be computed (by self-adjointness) as

$$\|Q\|_0 = \sup\{|\langle Qu, u\rangle| \mid u \in L_0^2(\sigma(X), f); \mathbf{E}(u^2) \le 1\}.$$

Observe further that since P_2 is self-adjoint, for $u \in \mathcal{M}_2$,

$$\langle Qu, u\rangle = \langle P_2P_1P_2u, u\rangle = \langle P_1u, u\rangle.$$

Hence, using (8), we have established that

$$\|Q\|_0 = c^2. \tag{9}$$

Note that by the definition of β_1, we get,

$$\begin{aligned}\beta_1 &= \sup\{\langle Qg, g\rangle \mid g \in L^2(\sigma(X), f), \mathbf{E}(g) = 0, \mathbf{E}(g^2) = 1\}\\ &= \sup\{\langle P_1g, g\rangle \mid g \in L^2(\sigma(X), f), \mathbf{E}(g) = 0, \mathbf{E}(g^2) = 1\}\\ &= c^2.\end{aligned}$$

Hence, the result is proved. □

4.3. Convergence theorems. With notation as above, suppose for the moment that $f(x, y) > 0$ for all x, y. Then, for all $U \in L^2(f)$, with $\bar{U} = \int_{\mathcal{X}\times\mathcal{Y}} U(x, y)f(x, y)\mu(dx)\nu(dy)$, von Neumann's theorem shows that

$$\int \left(K^nU(x, y) - \bar{U}\right)^2 f(x, y)\mu(dx)\nu(dy) \to 0 \quad \text{as } n \to \infty. \tag{10}$$

The examples in the remark at the end of Section 3 show that some conditions on f are needed to ensure convergence. This is the subject of careful study by Berti–Pratelli–Rigo. Their results show that it is sufficient to assume that the support of f is open and connected (assuming that $\mathcal{X}$ and $\mathcal{Y}$ are second countable topological spaces). They have much more general results and we will merely assume their conclusion

$$\overline{\sigma(X)} \cap \overline{\sigma(Y)} = \{\phi, \Omega\} \quad \text{up to } P \text{ null sets.}$$

It is natural to supplement the weak notion of convergence in (10) by convergence of $K^n((x, y), \cdot)$ to f in total variation distance for all (x, y). Here is what elementary manipulations from Aronszajn's bound give. We first treat the countable case.

THEOREM 6. *Consider the Gibbs sampling chain as described in the beginning of Section 4. Let* $\Omega = \mathcal{X} \times \mathcal{Y}$ *be countable. Assume* $\overline{\sigma(X)} \cap \overline{\sigma(Y)}$ *is trivial and let* c *be as in Theorem 5. Then for any* $\omega = (x, y) \in \Omega$, *and all* $n \geq 1$,

$$\|K^n_{(x,y)} - f\|_{\mathrm{TV}} \leq \frac{f(x,y)^{-\frac{1}{2}}}{2} c^{2n-1},$$

where $K^n_{(x,y)} = K^n \delta_{(x,y)}$.

Proof. Consider the self-adjoint operator $Q = P_2 P_1 P_2$ acting on $L^2_0(\sigma(X), f)$. By (9), $\|Q\|_0 = c^2$. Note that

$$(P_1 P_2)^n = P_1 Q^{n-1},$$

and c bounds the norm of P_1 from $L^2_0(\sigma(X), f)$ to $L^2_0(\sigma(Y), f)$. Hence, it follows that

$$\|(P_1 P_2)^n\|_{L^2_0(f) \to L^2_0(f)} \leq c^{2n-1}.$$

We observe initially that $(P_1 P_2)^n = (K^n)^*$, and then for $\omega = (x, y)$, using $g = f(\omega)^{-1} \delta_\omega - 1$ we get,

$$\begin{aligned}
&\|(P_1 P_2)^n g\|_2 \leq \|g\|_2 c^{2n-1} \\
&\quad \Rightarrow \left(\sum_{\omega' \in \Omega} ((K^n)^* g(\omega'))^2 f(\omega') \right)^{\frac{1}{2}} \leq \left(\sum_{\omega' \in \Omega} g(\omega')^2 f(\omega') \right)^{\frac{1}{2}} c^{2n-1} \\
&\quad \Rightarrow \left(\sum_{\omega' \in \Omega} \left(\frac{(K^n)^*(\omega', \omega)}{f(\omega)} - 1 \right)^2 f(\omega') \right)^{\frac{1}{2}} \leq f(\omega)^{-\frac{1}{2}} c^{2n-1} \\
&\quad \Rightarrow \left(\sum_{\omega' \in \Omega} \left(\frac{K^n(\omega, \omega')}{f(\omega')} - 1 \right)^2 f(\omega') \right)^{\frac{1}{2}} \leq f(\omega)^{-\frac{1}{2}} c^{2n-1} \\
&\quad \Rightarrow \sum_{\omega' \in \Omega} |K^n(\omega, \omega') - f(\omega)| \leq f(x,y)^{-\frac{1}{2}} c^{2n-1} \\
&\quad \Rightarrow \|K^n_{(x,y)} - f\|_{\mathrm{TV}} \leq \frac{f(x,y)^{-\frac{1}{2}}}{2} c^{2n-1}.
\end{aligned}$$

Hence, the result is proved. □

For general state space Ω, these bounds on total variation are not applicable because it may be that $P(\omega) = 0$ for all ω. However, in many examples the operators P_i, $i = 1, 2$, will have the property that

$$P_2 P_1 \delta_\omega = \psi_\omega$$

is in $L^2(f)$, for any fixed ω. Here $P_2 P_1 \delta_\omega$ is obtained as the limit of $P_2 P_1 \phi_n$ where ϕ_n is a sequence of (nonnegative) functions in $L^2(f)$ such that $\phi_n \to \delta_\omega$

(in total variation). In such cases, with $\omega = (x, y)$, one obtains

$$\left\|K^n_{(x,y)} - f\right\|_{\mathrm{TV}} \le \|\psi_{x,y}\|_2^{1/2} c^{2n-3}. \tag{11}$$

From the definition,

$$\begin{aligned}\psi_{x,y}(x', y') &= \int k(x', y'; x'', y'')\delta_{(x,y)}(x'', y'')\mu(dx'')\nu(dy'') \\ &= k(x', y'; x, y) \\ &= f(y|x')f(x|y),\end{aligned}$$

and

$$\|\psi_{x,y}\|_2^2 = f(x, y)^2 \int_{\mathcal{X}} \left|\frac{f(x'|y)}{m_X(x')}\right|^2 m_X(x')\, d\mu(x').$$

Whenever $\|\psi_{x,y}\|_2^2 < \infty$, the bound (11) is a useful bound. We further note that $\|\psi_{x,y}\|_2^2 < \infty$ for most of the examples considered in [20]. It is important to note that all of these bounds just use the second eigenvalue β_1. For a Markov chain M with stationary distribution π, a typical bound based on the second largest eigenvalue β of the operator corresponding to $\hat{K}$ is given by

$$\|M^n_\omega - \pi\|_{\mathrm{TV}} \le \frac{\pi(\omega)^{-\frac{1}{2}}}{2}\beta^n. \tag{12}$$

Here is an example showing that just using the second eigenvalue can give results which are off. Consider the Poisson–Exponential example introduced at the beginning of this section. In [20], [21] it is shown that for any starting values (x, y), the following bound holds.

$$\|K^n_{x,y} - f\|_{\mathrm{TV}} \le 2^{-(c+1)} \quad \text{for } n = \log_2(x + 1) + c, c > 0.$$

This is a sharp form of the informal statement "order $\log_2 x$ steps suffice for convergence." The reference also shows this is sharp in various senses.

It is further shown that the convergence of the bivariate chain is controlled sharply by the marginal x-chain. The self-adjoint operator corresponding to the marginal x-chain Q has second largest eigenvalue $\frac{1}{2}$. The stationary distribution is given by $m_X(x) = (\frac{1}{2})^{x+1}$. Hence, the bound (12) becomes,

$$\|Q^n_x - m_X\|_{\mathrm{TV}} \le \frac{((\frac{1}{2})^{x+1})^{-\frac{1}{2}}}{2}\left(\frac{1}{2}\right)^n = 2^{\frac{x-1}{2}-n}.$$

The last inequality implies that the marginal x-chain (and by the results in [20], the bivariate chain) is close to the stationary distribution after order $\frac{x-1}{2}$ steps. A careful use of Theorem 6 would also lead to the same conclusion. This bound is "off" compared to the correct answer $\log_2(x)$.

4.4. Applications of Burkholder's theorem. With notation as above, the most naive application of Burkholder's theorem to the Gibbs sampler gives the following result. Suppose that $\overline{\sigma(X)} \cap \overline{\sigma(Y)} = \{\phi, \Omega\}$. Then, for all U such that $\int_{\mathcal{X}\times\mathcal{Y}} |U(x,y)| \log(1 + |U(x,y)| f(x,y)\, d\mu(x)\, d\nu(y) < \infty$,

$$K^n U(x,y) \to \int_{\mathcal{X}\times\mathcal{Y}} U(x,y) f(x,y)\, d\mu(x)\, d\nu(y). \tag{13}$$

This clearly refines (10) of Section 4.3. Further, using our results from Section 4.1, and Burkholder's theorem, Berti–Pratelli–Rigo [7] give the following result.

THEOREM 7. *Let $(X_i, Y_i)_{0\le i<\infty}$ be successive states of the Gibbs sampler with $(X_0, Y_0) \sim f$. For $U \in L^1(f)$, let*

$$m_n(U) = \frac{1}{n} \sum_{i=0}^{n-1} U(X_i, Y_i).$$

Then

$$m_n(U) \to \int_{\mathcal{X}\times\mathcal{Y}} U(x,y) f(x,y)\, d\mu(x)\, d\nu(y) \quad \text{for all } U \in L^1(f),$$

if and only if $\overline{\sigma(X)} \cap \overline{\sigma(Y)}$ is trivial.

Acknowledgments. We thank Patricia Berti, Luca Pratelli, Pietro Rigo, David Freedman, Ben Weiss, and especially Don Burkholder for their help with this paper.

REFERENCES

[1] Y. Amit, *Convergence properties of the Gibbs sampler for perturbations of Gaussians*, Ann. Statist. **24** (1996), 122–140. MR 1389883

[2] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404. MR 0051437

[3] K. Athreya, H. Doss and J. Sethuraman, *On the convergence of the Markov chain simulation method*, Ann. Statist. **24** (1996), 89–100. MR 1389881

[4] H. Bauschke, *Projection algorithms: Results and open problems*, Inherently parallel algorithms in feasibility and optimization and their applications (Haifa 2000) (D. Butnariu, Y. Censor and S. Reich, eds.), North-Holland, Amserdam, 2001, pp. 11–22. MR 1853213

[5] H. Bauschke, P. Combettes and S. Reich, *The asymptotic behaviour of the composition of two resolvents*, Nonlinear Anal. **60** (2005), 283–301. MR 2101879

[6] H. Bauschke, P. Combettes and D. Luke, *A strongly convergent reflection method for finding the projection onto the intersection of two closed convex sets in a Hilbert space*, J. Approx. Theory **127** (2006), 178–192. MR 2058156

[7] P. Berti, L. Pratelli and P. Rigo, *Trivial intersections of σ-fields and Gibbs Sampling*, Ann. Probab. **36** (2008), 2215–2234. MR 2478681

[8] P. Berti, L. Pratelli and P. Rigo, *Atomic intersection of σ-fields and some of its consequences*, Probab. Theory Related Fields **148** (2010), 269–283. MR 2653229

[9] A. Bufetov, *Markov averaging and ergodic theorems for several operators*, Topology, ergodic theory, real algebraic geometry, AMS Transl. Ser. 2, vol. 202, Amer. Math. Soc., Providence, RI, 2001, pp. 39–50. MR 1819180

[10] D. L. Burkholder, *Sufficiency in the undominated case*, Ann. Math. Stat. **32** (1961), 1191–1200. MR 0131287

[11] D. L. Burkholder, *Successive conditional expectations of an integrable function*, Ann. Math. Stat. **33** (1962), 887–893. MR 0143246

[12] D. L. Burkholder, *On the order structure of the set of sufficient subfields*, Ann. Math. Stat. **33** (1962), 596–599. MR 0137227

[13] D. L. Burkholder and Y. S. Chow, *Iterates of conditional expectation operators*, Proc. Amer. Math. Soc. **12** (1961), 490–495. MR 0142144

[14] D. Butnariu, *The expected projection method: Its behaviour and applications to linear operator equations and convex optimatization*, J. Appl. Anal. **1** (2005), 93–108. MR 1390423

[15] G. Casella and E. George, *Explaining the Gibbs sampler*, Amer. Statist. **46** (1992), 167–174. MR 1183069

[16] G. Cohen, *Iterates of a product of conditional expectation operators*, J. Funct. Anal. **242** (2006), 658–668. MR 2274825

[17] B. Delyon and F. Delyon, *Generalization of von Neumann's spectral sets and integral representation of operators*, Bull. Soc. Math. Fr. **127** (1999), 25–41. MR 1700467

[18] F. Deutsch, *The method of alternating orthogonal projections*, Approximation Theory, Spline, Functions and Applications, Kluwer Academic, Dordrecht, 1992, pp. 105–121. MR 1165964

[19] F. Deutsch, *Best approximation in inner product spaces*, Springer-Verlag, New York, 2001. MR 1823556

[20] P. Diaconis, K. Khare and L. Saloff-Coste, *Gibbs sampling, exponential families and orthogonal polynomials*, Statist. Sci. **23** (2008), 151–178. MR 2446500

[21] P. Diaconis, K. Khare and L. Saloff-Coste, *Gibbs sampling, exponential families and coupling*, Sankhya A **72** (2010), 136–169. MR 2658168

[22] P. Diaconis and L. Saloff-Coste, *Comparison techniques for reversible Markov chains*, Ann. Appl. Probab. **3** (1993), 696–730. MR 1233621

[23] I. Halperin, *The product of projection operators*, Acta Sci. Math. (Szeged) **23** (1962), 96–99. MR 0141978

[24] G. Jones and J. Hobert, *Honest exploration of intractable probability distributions via Markov chain Monte Carlo*, Statist. Sci. **16** (2001), 312–334. MR 1888447

[25] M. Kassabov, *Subspace arrangements and property T*, Groups Geom. Dyn. **5** (2011), 445–477.

[26] H. J. Landau, *An overview of time and frequency limiting*, Fourier Techniques and Applications, Plenum, New York, 1985, pp. 201–220. MR 0819297

[27] J. Liu, *Monte Carlo strategies in scientific computing*, Springer, New York, 2001. MR 1842342

[28] S. P. Meyn and R. Tweedie, Markov chains and stochastic stability, 2nd ed., Cambridge Univ. Press, Cambridge, UK, 2009. MR 2509253

[29] A. Netyanun and D. Solomon, *Iterated products of projections on Hilbert spaces*, American Mathematical Monthly **113** (2006), 644–648.

[30] J. Neveu, *Mathematical foundations of the calculus of probability*, Holden-Day, San Francisco, 1965. MR 0198505

[31] D. Ornstein, *On the pointwise behaviour of iterates of a self adjoint operator*, J. Math. Mech. **18** (1968), 473–477. MR 0236354

[32] J. Rosenthal, *Minorization conditions and convergence rates for Markov chain Monte Carlo*, J. Amer. Statist. Assoc. **90** (1995), 558–566. MR 1340509

[33] G. C. Rota, *An "alternierende Verfahren" for general positive operators*, Bull. Amer. Math. Soc. **68** (1962), 95–102. MR 0133847

[34] H. A. Schwarz, *Ueber einen Grenzũbergang durch alternirendes Verfahren*, Vierteljahrsschrift der Naturforschenden Gessellschaft in Zurich **15** (1870), 272–286.

[35] R. Zaharopol, *On products of conditional expectation operators*, Canad. Math. Bull. **33** (1990), 257–260. MR 1077092

Persi Diaconis, Stanford University, Department of Statistics, Sequoia Hall, 390 Serra Mall, Stanford, CA 94305, USA

E-mail address: diaconis@math.stanford.edu

Kshitij Khare, University of Florida, Department of Statistics, 102 Griffin-Floyd Hall, Gainesville, FL 32606, USA

E-mail address: kdkhare@stat.ufl.edu

Laurent Saloff-Coste, Cornell University, Department of Mathematics, 310 Malott Hall, Ithaca, NY 14853, USA

E-mail address: lsc@math.cornell.edu