ANALYSIS OF CONVERGENCE RATES OF SOME GIBBS
SAMPLERS ON CONTINUOUS STATE SPACES

AARON SMITH

1. ABSTRACT

We use a non-Markovian coupling and small modifications of techniques from the
theory of finite Markov chains to analyze some Markov chains on continuous state
spaces. The first is a Gibbs sampler on narrow contingency tables, the second a gen-
eralization of a sampler introduced by Randall and Winkler.

2. INTRODUCTION

The problem of sampling from a given distribution on high-dimensional continuous
spaces arises in the computational sciences and Bayesian statistics, and a frequently-
used solution is Markov chain Monte Carlo (MCMC); see [13] for many examples.
Because MCMC methods produce good samples only after a lengthy mixing period,
a long-standing mathematical question is to analyze the mixing times of the MCMC
algorithms which are in common use. Although there are many mixing conditions,
the most commonly used is called the mixing time, and is based on the total variation
distance:

For measures $\nu$, $\mu$ with common measurable $\sigma$-algebra $\mathcal{A}$, the total variation dis-
tance between $\mu$ and $\nu$ is

$$||\mu - \nu||_{TV} = \sup_{A \in \mathcal{A}} \mu(A) - \nu(A)$$

For an ergodic Markov chain $X_t$ with unique stationary distribution $\pi$, the mixing
time is

$$\tau(\epsilon) = \inf\{t | ||\mathcal{L}(X_t) - \pi||_{TV} < \epsilon\}$$

Although most scientific and statistical uses of MCMC methods occur in continu-
ous state space, much of the mathematical mixing analysis has been in the discrete
setting. The methods that have been developed for discrete chains often break down
when used to analyze continuous chains, though there have been efforts, such as [21],

Date: August 26, 2011.
to fix this. This paper extends the author’s previous work in [20] and work of Randall and Winkler [17], and attempts to provide some more examples of relatively sharp analyses of continuous chains similar to those used to develop the discrete theory.

The first case consists of narrow contingency tables. Beginning with the work of Diaconis and Efron [6], there has been interest in finding efficient ways to sample uniformly from the collection of integer-valued matrices with given row and column sums. A great deal of this effort has been based on Markov chain Monte Carlo methods. While some of the efforts have dealt directly with Markov chains on these integer-valued matrices, much recent success, including [10] [16], has involved using knowledge of Gibbs samplers on convex sets in $\mathbb{R}^n$ and clever ways to project from the continuous chain to the desired matrices [15].

Unfortunately, while the general bounds are polynomial in the number of entries in the desired matrix, they are often not of a small order; see [14]. In this note, we find some better bounds for very specific cases. Like the note [20], this is part of an attempt to make further use of non-Markovian coupling techniques [11] [3] [5] and also to expand the small set of carefully analyzed Gibbs samplers [17] [18] [7] [8]. In this case, the new techniques are two slight modifications of the path-coupling method introduced in [4]. In many path-coupling arguments, a burn-in argument is used to show that for most pairs of points in a metric space, there is a path along which the Markov transition kernel is contractive acting on any pair of points along the path. In this argument, we show that for all paths, the kernel is contractive acting on most pairs of points along the path. This type of modification seems likely to be useful only on continuous spaces.

We consider the following Gibbs sampler $X_t[i, j]$ on 2 by $n$ matrices satisfying the row sums $\sum_{i=1}^{n} X[i, j] = n$ for $1 \leq j \leq 2$ and column sums $\sum_{j=1}^{2} X[i, j] = 2$ for all $1 \leq i \leq n$. To make a step of the Gibbs sampler, choose two distinct integers $1 \leq i < j \leq n$ and update the four entries $X_{t+1}[i, 1], X_{t+1}[i, 2], X_{t+1}[j, 1]$ and $x_{t+1}[j, 2]$ to be uniform conditional on all other entries of $X_t$. We find the following reasonable bound on the mixing time of this sampler:

**Theorem 1** (Convergence Rate for Narrow Matrices). For $T > (31r + 81)n \log(n)$,

$$||\mathcal{L}(X_T) - U||_{TV} \leq 13n^{-r}$$

while for $T < n(\log(n) - r)$, and $n$ sufficiently large,

$$||\mathcal{L}(X_T) - U||_{TV} \geq 1 - 2e^{-r}$$

The next process is a Gibbs sampler on the simplex, with a very restricted set of allowed moves. Fix a group $G$ with symmetric generating set $S$. We consider the process $X_t[g]$ on the simplex $\Delta_G = \{X | \sum_{g \in G} X[g] = 1; X[g] \geq 0\}$. At each step,
choose $g \in G$, $s \in S$ and $\lambda \in [0, 1]$ uniformly, and set $X_{t+1}[g] = \lambda(X_t[g] + X_t[gs])$ and $X_{t+1}[gs] = (1 - \lambda)(X_t[g] + X_t[gs])$; for all other $h \in G$ set $X_{t+1}[h] = X_t[h]$. Also consider a simple random walk $S_t$ on $G$, where in each stage we choose $g \in G$ and $s \in S$ uniformly at random and set $S_{t+1} = gs$ if $S_t = g$, and $S_{t+1} = S_t$ otherwise. Then if $\hat{\gamma}$ is the spectral gap of the walk $S_t$,

**Theorem 2** (Convergence Rate for Gibbs Sampler with Geometry). For $T > \frac{4r+15}{\hat{\gamma}} \log(n)$,

$$||L(X_T) - U||_{TV} \leq 9n^{-r}$$

and conversely for $T < \frac{r}{\hat{\gamma}}$,

$$||L(X_T) - U||_{TV} \geq \frac{1}{2}e^{-r} - 3n^{-\frac{r}{3}}$$

This substantially generalizes [17] and [20], from samplers on the cycle or complete graph respectively to general Cayley graphs. In addition to being of mathematical interest, this process is an example of a gossip process with some geometry, studied by electrical engineers and sociologists interested in how information propagates through networks; see [19] for a survey.

The proof of the upper bound will use an auxiliary chain similar to that found in [17], a coupling argument improved from [20], and some elementary comparison theory of Markov chains. The lower bound is somewhat simpler than that in [17].

### 3. General Strategy and the Partition Process

Both of our bounds will be obtained using a similar strategy, ultimately built on the classical coupling lemma. We recall that a coupling of Markov chains with transition kernel $K$ is a process $(X_t, Y_t)$ so that marginally both $X_t$ and $Y_t$ are Markov chains with transition kernel $K$. Then we have the following lemma (see [12]):

**Lemma 3** (Fundamental Coupling Lemma). If $(X_t, Y_t)$ is a coupling of Markov chains, $Y_0$ is distributed according to the stationary distribution of $K$, and $\tau$ is the first time at which $X_t = Y_t$, then

$$||L(X_t) - L(Y_t)||_{TV} \leq P[t > \tau]$$

In each chain, then, we begin with $X_t$ started at a distribution of our choice, and $Y_t$ started at stationarity. For any fixed (large) $T$, we will then couple $X_t$ and $Y_t$ so that they will have coupled by time $T$ with high probability. Each coupling will have two phases: an initial phase in which $X_t$ and $Y_t$ get close with high probability, and a non-Markovian coupling phase in which we actually force them to collide. Unlike many coupling proofs, the time of interest $T$ must be specified before constructing the coupling.

While the initial contraction phases are quite different for both chains, the final
coupling phase can be described in a unified way. The unifying device is the partition process $P_t$ on set partitions of $[n]$, introduced in [20] for a special case of the second sampler treated here (see that note for details). This partition process contains some information about the entire process $(Y_t)_{0 \leq t \leq T}$, and is the only source of information from the future that is used to construct the non-Markovian coupling. Critically, we don’t use any information about the random variables on $[0,1]$ used at each step.

Our process $P_t$ will consist of a set of nested partitions of $[n]$, $P_0 \leq P_1 \leq \ldots \leq P_T$, where I say partition $A$ is less than partition $B$ if every element of partition $B$ is a subset of an element of partition $A$. To construct $P_t$, begin by running $Y_t$ from time 0 to time $T$, and setting $P_T = \{\{1\}, \ldots, \{n\}\}$, $n$ singletons. While running the chain, we choose two privileged coordinates $i(t)$, $j(t)$ at each time $t$. To construct $P_{t-1}$ from $P_t$, we merge distinct sets $A$, $B$ in $P_t$ to a single set $A \cup B$ in $P_{t-1}$ if and only if one of $i(t)$, $j(t)$ is in $A$ and the other is in $B$. This defines the entire process.

We will be interested in the smallest time $\tau$ such that $P_{T-\tau} = [n]$, a single block. From classical arguments (see e.g. chapter 7 of [2]), it is easy to check that

\textbf{Lemma 4 (Connectedness). For the Gibbs sampler on narrow matrices,}

$$P[\tau > (\frac{1}{2} + \epsilon)n \log(n)] \leq 2n^{-\epsilon}$$

The analogous lemma for the other example will be proved in section 8.

For both of our walks, we will use two types of coupling, the ‘proportional’ coupling and the ‘subset’ coupling. In both cases, the choices of $i$, $j$ will be the same for both $X_t$ and $Y_t$ at all time steps. In the proportional coupling, we choose the uniform variable at time $t$ to minimize $||X_{t+1} - Y_{t+1}||_2$ given $X_t$, $Y_t$. In the simplex walk, this involves choosing the same variable $\lambda$ in both cases; for the other walk, the coupling will be almost as simple.

To discuss the subset coupling, we must define the weight of $X_t$ on a subset $S \subset [n]$, which we call $w(X_t, S)$. For the simplex walk, we define $w(X_t, S) = \sum_{s \in S} X_t[s]$. For narrow matrices, we define $w(X_t, S) = \sum_{s \in S} X_t[s, 1]$. A subset coupling of $X_t$ and $Y_t$ with respect to set $S$ is one that attempts to force $w(X_{t+1}, S) = w(Y_{t+1}, S)$ with maximal probability. We say that a subset coupling succeeds if that equality holds; otherwise, we say it fails.

In each case, the coupling of $X_t$ and $Y_t$ during the non-Markovian coupling phase will be as follows. At marked times $t_j$, we will perform a subset coupling of $X_{t_j}$, $Y_{t_j}$ with respect to $S(t_j, 1)$. At all other times, we will perform a proportional coupling. This leads to:
Lemma 5 (Final Coupling). Assume the non-Markovian coupling phase lasts from time \(T_1\) to \(T\), that \(P_{T_1} = \{[n]\}\), and that all subset couplings succeed. Then \(X_T = Y_T\).

Proof: At time \(T_1\), we have \(w(X_{T_1}, [n]) = w(Y_{T_1}, [n])\). I claim, inductively, that \(w(X_t, S) = w(Y_t, S)\) for all \(S \in P_t\) and all \(T_1 \leq t \leq T\). Note that it is true at time \(T_1\). By definition of the partition process, it cannot become untrue except at marked times, and at marked time \(t_j\) it can only become untrue at one of the marked sets \(S(t_j, 1)\) or \(S(t_j, 2)\). By assumption, all subset couplings have succeeded, so \(w(X_{t+1}, S(t_j, 1)) = w(Y_{t+1}, S(t_j, 1))\). By construction, \(w(X_{t+1}, S(t_j, 2)) = (X_{t+1}, S(t_j, 1) \cup S(t_j, 2)) - w(Y_{t+1}, S(t_j, 1))\) and similarly for \(Y_{t+1}\), so \(w(X_{t+1}, S(t_j, 2)) = w(Y_{t+1}, S(t_j, 2))\). Thus, the inductive claim has been proved. Finally, we note that if \(w(X_t, \{i\}) = w(Y_t, \{i\})\) for any singleton \(\{i\}\), then \(X_t[i] = Y_t[i]\) for all \(t \geq T_1\).

So, in both cases, showing that all subset couplings succeed with high probability is sufficient to show that coupling has succeeded.

4. Contraction and Narrow Matrices

We begin with some quick observations about the geometry of our space. It is the part of an \(n - 1\)-dimensional affine subspace of \(R^{2n}\) that lies in the upper orthant. Our updates are in fact moves along 1-dimensional pieces of this subspace, even though we are updating four entries. While the original motivation for this sampler comes from statistics, it is being treated here primarily as an example of a chain that is somewhere between the standard Gibbs sampler on the simplex and the standard Gibbs sampler on doubly-stochastic matrices or Kac’s famous walk on the orthogonal group.

In this section, we will prove the following contractivity estimate for the Gibbs sampler on narrow matrices:

Lemma 6 (Weak Convergence on Narrow Matrices). If \(X_t\) and \(Y_t\) are coupled under the proportional coupling until time \(T = (10.1 + 10.5)n \log(n)\), then

\[
P[||X_T - Y_T||_1 \geq \epsilon] \leq 3\epsilon^{-1}n^{-A}
\]

The above lemma is a contraction result, and it will be proved using a variant of the path-coupling argument introduced in [24]. In path-coupling arguments, the goal is to couple \(X_t\) and \(Y_t\) by constructing an interpolating chain, \(X_t = Z_t^{(0)}, Z_t^{(1)}, \ldots, Z_t^{(m)} = Y_t\), so that \(d(X_0, Y_0) \sim \sum_{j=1}^{m} d(Z_0^{(j-1)}, Z_0^{(j+1)})\) for some metric \(d\). We would then show that, in general, \(E[d(Z_t^{(j)}, Z_t^{(j)})] \leq \alpha d(Z_0^{(j)}, Z_0^{(j)})\) for some \(0 < \alpha < 1\). In most coupling arguments, we find such an \(\alpha\) that holds for all pairs \(Z_t^{(j)}, Z_t^{(j+1)}\) given a typical \(X_t\) and \(Y_t\); this immediately gives an estimate of \(E[d(X_t, Y_t)] \leq \alpha d(X_0, Y_0)\). In our argument, we show this only for
most pairs $Z^{(j)}_t, Z^{(j+1)}_t$. While this generally causes arguments for chains over finite chains to fail, it leads to only slightly worse bounds for this and (we conjecture) other chains on continuous spaces.

Having discussed the big picture, we will now make some basic remarks about the chain, beginning with an alternative description of the transition probabilities. Define $\delta_t[i, j] = 2 - X_t[i, 1] - X_t[j, 1]$, and $\epsilon_t[i, j] = 2 - Y_t[i, 1] - Y_t[j, 1]$. Then a step of the chain can be defined in the following way. Choose $i,j$ as before, and choose $\lambda \sim U[0, 1]$. If $\delta_t[i, j] \geq 0$, then we update according to:

$$
X_{t+1}[i, 1] = \lambda(2 - \delta_t[i, j]) \\
X_{t+1}[j, 1] = (1 - \lambda)(2 - \delta_t[i, j]) \\
X_{t+1}[i, 2] = 2(1 - \lambda) + \lambda\delta_t[i, j] \\
X_{t+1}[j, 2] = 2\lambda + (1 - \lambda)\delta_t[i, j]
$$

If $\delta_t[i, j] < 0$, we update according to:

$$
X_{t+1}[i, 1] = 2\lambda - (1 - \lambda)\delta_t[i, j] \\
X_{t+1}[j, 1] = (1 - \lambda)(2 - \lambda\delta_t[i, j]) \\
X_{t+1}[i, 2] = (1 - \lambda)(2 + \delta_t[i, j]) \\
X_{t+1}[j, 2] = \lambda(2 + \delta_t[i, j])
$$

Note that in both cases, a larger value of $\lambda$ means a larger value of $X_{t+1}[i, 1]$. We are now ready to describe the proportional coupling: as in the simplex case, we choose the same value of $\lambda$ for both chains in the above representation.

While studying contraction, we will primarily be interested in the $L^2$ norm $||X_t - Y_t||^2_2 = \sum_{j=1}^2 \sum_{i=1}^n (X_t[i, j] - Y_t[i, j])^2$, and occasionally in the $L^1$ norm $||X_t - Y_t||_1 = \sum_{j=1}^2 \sum_{i=1}^n |X_t[i, j] - Y_t[i, j]|$. We will need two initial contractivity lemmas. The first is:

**Lemma 7 (L^2 Contractivity).** If $X_t, Y_t$ are coupled under the proportional coupling for time $0 \leq t \leq T$, and $1_A$ is the indicator function $\delta_t[i, j] \epsilon_t[i, j] \geq 0$ for all $1 \leq i, j \leq n$ and all $0 \leq t \leq T$, then

$$
E[||X_T - Y_T||^2_2 1_A] \leq (1 - \frac{2}{3n})^T ||X_0 - Y_0||^2_2
$$
Proof: We begin the proof by calculating the change in the $L^2$ norm during a single move, and begin by conditioning on $i, j$. We find:

$$\Delta_1 = E[(X_{t+1}[i, 1] - Y_{t+1}[i, 1])^2 + (X_{t+1}[j, 1] - Y_{t+1}[j, 1])^2$$
$$+ (X_{t+1}[i, 2] - Y_{t+1}[i, 2])^2 + (X_{t+1}[j, 2] - Y_{t+1}[j, 2])^2]$$
$$= 2E[(\lambda(2 - \delta) - \lambda(2 - \epsilon))^2 + (\lambda\delta - \lambda\epsilon)^2]$$
$$= \frac{4}{3}(\delta - \epsilon)^2$$

It would be nice to calculate the sums of terms like $(\epsilon_t[i, j] - \delta_t[i, j])^2$ in terms of sums of terms like $(X_t[i, j] - Y_t[i, j])^2$. Fortunately, as in the simplex case, these are easy to relate. We first note that

$$0 = \sum_{i=1}^{n} (X_t[i, 1] - Y_t[i, 1])^2$$
$$= \sum_{i=1}^{n} (X_t[i, 1] - Y_t[i, 1])^2 + 2 \sum_{i < j} (X_t[i, 1] - Y_t[i, 1])(X_t[j, 1] - Y_t[j, 1])$$

Thus, we find that if $\delta_t[i, j] \epsilon_t[i, j] \geq 0$ for all $i, j$,

$$\sum_{i \neq j} (\delta_t[i, j] - \epsilon_t[i, j])^2 = \sum_{i \neq j} (X_t[i, 1] + X_t[j, 1] - Y_t[i, 1] - Y_t[j, 1])^2$$
$$= \sum_{i \neq j} [(X_t[i, 1] - Y_t[i, 1])^2 + (X_t[j, 1] - Y_t[j, 1])^2$$
$$+ 2(X_t[i, 1] - Y_t[i, 1])(X_t[j, 1] - Y_t[j, 1])^2]$$
$$= (n - 3) \sum_{j=1}^{n} \sum_{i=1}^{n} (X_t[i, j] - Y_t[i, j])^2$$
Let $E_{t}[i, j]$ be the event that coordinates $i, j$ are updated at time $t$. Then the final contraction is given by:

$$\Delta = \mathbb{E}[||X_{t+1} - Y_{t+1}||^2 | X_t, Y_t]$$

$$= \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{m=1}^{n} \sum_{i \neq j} E[(X_{t+1}[m, k] - Y_{t+1}[m, k])^2 | E_t[i, j]]$$

$$= \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{m=1}^{n} \sum_{i, j \neq m} (X_t[m, k] - Y_t[m, k])^2$$

$$+ 2 \sum_{j \neq m} E[(X_{t+1}[m, k] - Y_{t+1}[m, k])^2 | E_t[m, j])$$

$$= (1 - \frac{2}{n})||X_t - Y_t||^2 + \frac{1}{n(n-1)} \sum_{i \neq j} \frac{4}{3}(\delta_t[i, j] - \epsilon_t[i, j])^2$$

\[\square\]

**Lemma 8** ($L^1$ Contractivity). If $X_t, Y_t$ are coupled under the proportional coupling for time $0 \leq t \leq T$, then

$$||X_T - Y_T||_1 \leq ||X_0 - Y_0||_1$$

*Proof*: Follows immediately by induction on $t$ from the formula and the triangle inequality. \[\square\]

Having demonstrated contractivity for ‘nice’ pairs $X_t$ and $Y_t$, we must now look at ‘typical’ pairs $X_t$ and $Y_t$. The following burn-in lemma shows that, after a moderate number of steps, $X_t$ and $Y_t$ are unlikely to be too close to the boundary of our convex set.

**Lemma 9** (Burn-in). For any starting position $X_0$, $P[\inf_{i, j} X_t[i, j] \leq n^{-k}]$, $P[\sup_{i, j} X_t[i, j] \geq 2 - n^{-k}] \leq n^{-\frac{t}{n \log(n)} + 0.5} + o(n^{-k+2})$.

Our proof will be via comparison to a Gibbs sampler on the simplex, studied by the author in [20]. We let $\tilde{X}_t$ be a copy of our Gibbs sampler on $2$ by $n$ matrices, and we let $S_t$ be a Gibbs sampler on the simplex $\Delta_n = \{S|S[i] \geq 0, \sum_{i=1}^{n} S[i] = 1\}$. To make a move in this Gibbs sampler, choose distinct coordinates $1 \leq i, j \leq n$ and $0 \leq \lambda \leq 1$ uniformly at random, and update entry $S_t[i]$ to $\lambda(S_t[i] + S_t[j])$ and entry $S_t[j]$ to $(1 - \lambda)(S_t[i] + S_t[j])$, keeping all other entries fixed. This is identical to the other sampler in this note, with generating set $G = \{id\}$.

Since $\sum_i S_0[i] = 1$, for any given $X_0$ it is possible to choose a corresponding $S_0$ such that $X_0[i, 1] \geq S_0[i]$ for all $i$, without the column sum condition interfering. Next,
under our descriptions there is a natural proportional coupling of \( X_t \) and \( S_t \), given by always choosing \( i, j \) and \( \lambda \) to be the same. I claim that under this coupling, \( X_t[i, 1] \geq S_t[i] \) for all \( t > 0 \) and all \( 1 \leq i \leq n \). Assume inductively that this holds until time \( t \), and that coordinates \( i, j \) are updated at time \( t \).

\[
X_{t+1}[i, 1] \geq \lambda \min(X_t[i, 1] + X_t[j, 1], 2)
\]

\[
\geq \lambda \min(S_t[i] + S_t[j], 2)
\]

\[
= \lambda(S_t[i] + S_t[j])
\]

\[
= Y_{t+1}[i]
\]

Let \( S \) be drawn from the uniform distribution on the simplex. Then the above monotonicity tells us that

\[
P[\inf_i X_t[i, 1] \leq n^{-k}] \leq ||L(S_t) - U||_{TV} + P[\inf_i S[i] \leq n^{-k}]
\]

From [20], \( ||L(S_t) - U||_{TV} \leq n^{-\frac{1}{5}} \log(n)^{1.5} \), and it is easy to show that \( P[\inf_i S[i] \leq n^{-k}] = o(n^{-k+2}) \). This gives a good bound on \( P[\inf_i X_t[i, 1] \leq n^{-k}] \). Since all columns sum to 2, this gives the same bound on \( P[\sup_i X_t[i, 2] \geq 2 - n^{-k}] \). Since there is symmetry between the top and bottom rows, this completes the proof. \( \Box \)

The second step in the coupling is the creation of an interpolating sequence between \( X_{T_1} \) and \( Y_{T_1} \), for some \( T_1 = c_0 n \log(n) \) large enough that the burn-in lemma has taken effect. Since our sample space is convex, this will be simple. We define \( X_{T_1} = Z_{T_1}^1, Z_{T_1}^2, \ldots, Z_{T_1}^l = Y_{T_1} \) so that \( ||Z_{T_1}^m - Z_{T_1}^{m+1}||_1 \leq n^{-c_1} \) for some large constant \( c_1 \) to be determined later, and so that all of the \( Z_{T_1}^i \) are in order along the line between \( X_{T_1} \) and \( Y_{T_1} \). That is, \( Z_{T_1}^i \) is closer to \( X_{T_1} \) than \( Z_{T_1}^j \) if \( i < j \). For the following lemma and later use in this note, we define \( \delta_{t}^m[i, j] = 2 - Z_{T_1}^m[i, 1] - Z_{T_1}^m[j, 1] \), analogously to the definition of \( \delta_{t}^i[i, j] \) in the preliminary calculations. We also define \( D[Z_{t_1}^m, Z_{t_1}^{m+1}] = |\{(i, j)|\delta_{t_1}^{m+1}[i, j] - \delta_{t_1}^m[i, j] < 0\}|. \) Then:

**Lemma 10 (Interpolating Sequence).** The interpolating sequence described above satisfies:

1. \(|\{m|D[Z_{T_1}^m, Z_{T_1}^{m+1}] \geq 1\}| \leq n^2\)
2. \(\min(X_{T_1}[i, j], Y_{T_1}[i, j]) \leq Z_{T_1}^m[i, j] \leq \max(X_{T_1}[i, j], Y_{T_1}[i, j]) \) for all \( i, j, m \).

**Proof:** Part (2) follows from the fact that \( Z_{T_1}^m \) is on a line between \( X_{T_1} \) and \( Y_{T_1} \), and hence all coordinates are between those of \( X_{T_1} \) and \( Y_{T_1} \). For part (1), we observe that all linear functionals of the coordinates satisfy this property as well, and so in particular \( \delta_{T_1}^m[i, j] \) changes sign at most once as \( m \) changes, for any fixed pair \( i, j \). The inequality follows since the number of pairs \( i, j \) is less than \( \frac{n^2}{2} \). \( \Box \)
We are finally ready to show that that adjacent pairs in the interpolating sequence get closer, when the entire sequence is run under the proportional coupling. Let $A[m, t]$ be the event that $D[Z_m^m, Z_{m+1}^m] = 0$ for $s \leq t$. Then by lemmas 7 and 8, we have

$$E[|Z_t^m - Z_{t+1}^m|] \leq (1 - \frac{2}{3m})^t ||Z_0^m - Z_{0+1}^m||_2$$

which, combined with the elementary bound $||Z||_1 \leq n||V||_2$, tells us that

$$E[|Z_{2t}^m - Z_{2t+1}^m|] \leq n(1 - \frac{2}{3m})^t ||Z_0^m - Z_{0+1}^m||_2 + P[A[m, t]^c]||Z_0^m - Z_{0+1}^m||_1$$

So, it remains to bound $P[A[m, t]^c]$. To do so, let $G[t, m, k]$ be the event that $Z_t^m[i, j]$ and $Z_{t+1}^m[i, j]$ are all at least $n^{-k}$ away from 0 and 2, and that $D[Z_{t+1}^m, Z_{t+1}^m] = 0$. Then we find that

**Lemma 11** (Chamber Occupancy). $P[D[Z_{t+1}^m, Z_{t+1}^m] \geq 1 |G[t, m, k]] \leq n^{2+k}||Z_0^m - Z_{0+1}^m||_1$.

**Proof:** To prove this, just note that at the next move, $P[\delta_{t+1}^m[i, j] \delta_{t+1}^m[i, j] < 0]$ is bounded above by the ratio of the $L^1$ distance between $Z_{t+1}^m$ and $Z_{t+1}^m$ to the total range that can be travelled. But the former is bounded above by $||Z_0^m - Z_{0+1}^m||_1$, and the latter is bounded below by $n^{-k}$. This, combined with a union bound over all $\frac{n(n-1)}{2}$ pairs of distinct $i, j$ gives the bound. $\square$

Thus, using a union bound for lemma 11 over time from 0 to $n^2$, as well as lemma 9, we find that for any $k > 2$, $P[A[m, t]^c] \leq 2||Z_0^m - Z_{0+1}^m||_1n^{4+k} + n^{4-k}$ after a burn in period of at least $(7k + 6.5)n\log(n)$. Putting this together with the earlier calculation, we find that

$$E[|X_{2t}^m - Y_{2t}^m|] \leq n(1 - \frac{2}{3m})^t \sum_{m=1}^{l-1} ||Z_0^m - Z_{0+1}^m||_2$$

By the triangle inequality, $||X_{2t} - Y_{2t}||_1 \leq \sum_{m=1}^{l-1} ||Z_{2t}^m - Z_{2t+1}^m||_1$, and so

$$E[|X_{2t} - Y_{2t}||] \leq n(1 - \frac{2}{3m})^t \sum_{m=1}^{l-1} ||Z_0^m - Z_{0+1}^m||_2$$

$$+ 2 \sum_{m=1}^{l-1} ||Z_0^m - Z_{0+1}^m||_1(n^{4+k}||Z_0^m - Z_{0+1}^m||_1 + n^{4-k})$$
We note that this inequality holds for any choice of \( k, l \) possibly depending on \( t \), as long as the initial burn-in period is at least \( \Omega(nk \log(n)) \) as described earlier. In particular, if \( t = \frac{3}{2}(A + 1)n \log(n) \), then choosing \( k = A + 3 \) and \( l = n^{A + 7} \) gives

\[
E[||X_{2t} - Y_{2t}||_1] \leq 3n^A
\]

Finally, combining this with Markov’s inequality and adding the burn-in time proves lemma 6. \( \square \)

5. Coupling for Narrow Matrices

In this section, we show that subset couplings are likely to succeed, and finish the proof of theorem 1. The main lemma is:

Lemma 12 (Coupling for Nearby Points). If \( X_t, Y_t \) are two copies of the chain constructed as above, so that after a burn-in period of length \( 7(b + 7.5)n \log(n) \) during which \( X_t, Y_t \) evolve by proportional coupling we have \( ||X_0 - Y_0||_1 < n^{-a} \), then \( ||\mathcal{L}(X_t) - \mathcal{L}(Y_t)||_{TV} \leq 4n^{b + 3 - a} + 2n^{4 - b} + n^{-c} \) for \( t > (\frac{1}{2} + c)n \log(n) \).

Proof: Construct a partition process from time 0 to time \( T = (\frac{1}{2} + c)n \log(n) \). Our first step is to show that if \( X_t \) and \( Y_t \) are very close to each other and not too close to certain hyperplanes, that any subset couplings are likely to succeed. Let \( p(X, Y, i, j, S) \) be the maximal probability that a subset coupling of \( X, Y \) associated with subset \( S \) works given that coordinates \( i, j \) are updated.

Lemma 13 (Subset Coupling). For a pair of matrices \( (x, y) \) satisfying \( \sup_{m,k} |x[m, k] - y[m, k]| \leq n^{-a} \) and \( \inf_{m,k} (x[m, k], y[m, k], 2 - x[m, k], 2 - y[m, k]) \geq n^{-b} \), we have for all sufficiently large \( n \) that \( p(x, y, i, j, S) \geq 1 - 4n^{b + 2 - a} \) uniformly in \( S \) and possible \( i, j \).

Proof: The proof is nearly identical to the proof of lemma 4 in [20]. \( \square \)

For a subset \( S \) of \( [n] \), and vector \( v \in \mathbb{R}^n \), let \( ||v||_{1,S} = \sum_{s \in S} |v[s]| \). Then note that, after a successful subset coupling involving set \( S \) and \( R \) at time \( t \), we have \( ||X_{t+1} - Y_{t+1}||_{1,S} \leq ||X_t - Y_t||_{1,S \cup R} \). We are ready to prove the lemma. If all couplings until time \( t \) have been successful, then \( ||X_t - Y_t||_S \leq n^{-a} \) for any \( S \in P(t) \). In particular, by the subset coupling lemma, the probability of the next subset coupling succeeding is at least \( 1 - 4n^{b + 2 - a} - 2n^{2 - b} \). Using a union bound for the above probability and the connectedness lemma, the probability of all subset couplings succeeding and the partition process satisfying \( P(0) = [n] \) is at least \( 1 - 4n^{b + 3 - a} - 2n^{4 - b} - n^{-c} \). \( \square \)

Next, we put together lemmas 6 and 12. Using the constants in those lemmas, we set \( b = c + 4 \), \( a = 2c + 7 \) and \( A = 3c + 7 \), to find that
\[ \| \mathcal{L}(X_T) - \mathcal{L}(Y_T) \|_{TV} \leq 13n^{-c} \]

which is the upper bound in theorem 1. To prove the lower bound, we note that
\( X_t \) and \( Y_t \) couldn’t possibly have coupled until each coordinate has been touched at
least once. It then follows from the standard coupon collector bound.

6. CONTRACTION FOR GIBBS SAMPLERS ON THE SIMPLEX WITH GEOMETRY

In this section, we prove a contraction lemma for Gibbs samplers on the simplex
associated with a group \( G \) and symmetric generating set \( S \) of \( G \) (that is, \( S^{-1} = S \)),
where \(|G| = n, |S| = m\), and \( id \) is the identity element of \( G \). We recall briefly some
definitions. We write \( \Delta_G = \{ X \in \mathbb{R}^G | X[g] \geq 0, \sum_{g \in G} X[g] = 1 \} \).
If \( X_t \in \Delta_G \) is a
copy of the Markov chain, we take a step by choosing \( g \in G, s \in S \) and \( \lambda \in [0, 1] \)
uniformly and setting \( X_{t+1}[g] = \lambda (X_t[g] + X_t[gs]), X_{t+1}[gs] = (1 - \lambda) X_t[g] + X_t[gs] \),
and for all other entries \( X_{t+1}[h] = X_t[h] \). This walk is closely related to a slow simple
random walk on the group. In particular, we let \( Z_t \in G \) be the random walk that
follows at each time step a group element \( g \in G \) and generator \( s \in S \) uniformly at random, and setting \( Z_{t+1} = Z_t s \) if \( Z_t = g \), and \( Z_{t+1} = Z_t \) otherwise.
Let \( \hat{K} \) be the transition kernel associated with the random walk \( Z_t \). Since \( S \) is
symmetric, the random walk is reversible, so \( \hat{K} \) can be written in a basis of orthogonal
eigenvectors with real eigenvalues \( 1 = \hat{\lambda}_1 > \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_n \geq -1 \). Since it is \( \frac{1}{2} \)-lazy,
all eigenvalues are in fact nonnegative. Let \( \hat{\gamma} = 1 - \hat{\lambda}_2 \) be the spectral gap of \( \hat{K} \). In
this section we will show that

Lemma 14 (Contraction Estimate for Gibbs Sampler on Cayley Graphs). Let \( X_t, Y_t \) be two copies of the Gibbs sampler on the simplex associated with \( G \) and \( S \), with
joint distribution given by a proportional coupling at each step. Then

\[ E[\| X_t - Y_t \|_2] \leq 4ne^{-\frac{15}{16}\hat{\gamma} t} \]

Proof: We will construct an auxiliary Markov chain on \( G \) associated with \( X_t \),
and compare it to the standard random walk \( Z_t \). Let \( X_t, Y_t \) be two copies of the
walk, and couple them at each step with the proportional coupling. For \( h \in G \), let
\( S^h_t = \sum_{i=1}^n (X_t[i] - Y_t[i])(X_t[hi] - Y_t[hi]) \). We will analyze the evolution of the vector
\( S_t = (S^id_t, \ldots) \).
There are three cases to analyze: \( h \in S, h \notin S \) and \( h \neq id \), and \( h = id \). Let \( F_t \) be
the $\sigma$-algebra generated by $X_s$ and $Y_s$, $0 \leq s \leq t$. For case 1, we have

$$E[S_{t+1}^h|F_t] = (1 - \frac{4}{n} + \frac{2}{mn})S_t^h + \frac{1}{2mn} \sum_{i=1}^n \sum_{s \in S, s \neq h, h^{-1}} [ (X_t[i] + X_t[s] - Y_t[i] - Y_t[s]) \cdot (X_t[h] - Y_t[h]) ]$$

$$+ (X_t[s] + X_t[i] - Y_t[i] - Y_t[i]) \cdot (X_t[hs] - Y_t[hs])$$

$$+ (X_t[h^{-1}i] - Y_t[h^{-1}i]) \cdot (X_t[i] + X_t[s] - Y_t[i] - Y_t[s])$$

$$+ (X_t[h^{-1}si] - Y_t[h^{-1}si]) \cdot (X_t[i] + X_t[s] - Y_t[i] - Y_t[s])$$

$$+ \frac{2}{mn} \sum_{i=1}^n \left[ \frac{1}{6} (X_t[i] + X_t[hi] - Y_t[i] - Y_t[hi])^2 \right.$$

$$+ (X_t[i] - Y_t[i]) \cdot (X_t[hi] - Y_t[hi]) + (X_t[i] - Y_t[i]) \cdot (X_t[h^2i] - Y_t[h^2i])$$

$$= (1 - \frac{2}{n} + \frac{2}{3mn})S_t^h + \frac{2}{3mn} S_t^{id} + \frac{2}{mn} S_t^{id}$$

$$+ \frac{1}{2mn} \sum_{s \in S, s \neq h, h^{-1}} (S_t^{sh^{-1}} + S_t^{hs} + S_t^{sh} + S_t^{sh^{-1}})$$

and we note that the sum of the coefficients is $1 - \frac{2}{3mn}$. For case 2, we have

$$E[S_{t+1}^h|F_t] = (1 - \frac{4}{n})S_t^h + \frac{2m}{mn} S_t^h$$

$$+ \frac{1}{2mn} \sum_{s \in S} (S_t^{hs^{-1}} + S_t^{hs} + S_t^{sh} + S_t^{sh^{-1}})$$

$$= (1 - \frac{2}{n})S_t^h + \frac{1}{2mn} \sum_{s \in S} (S_t^{hs^{-1}} + S_t^{hs} + S_t^{sh} + S_t^{sh^{-1}})$$

where the sum of the coefficients is 1. Finally, in case 3, we have

$$E[S_{t+1}^{id}|F_t] = (1 - \frac{2}{n})S_t^{id} + \frac{2}{3mn} \sum_{s \in S} \sum_{i=1}^n (X_t[i] + X_t[s] - Y_t[i] - Y_t[s])^2$$

$$= (1 - \frac{2}{3n})S_t^{id} + \frac{4}{3mn} \sum_{s \in S} S_t^s$$

and here the sum of the coefficients is $1 + \frac{2}{3n}$. If we rewrite $U_t^{id} = \frac{1}{2} S_t^{id}$, and otherwise $U_t^g = S_t^g$, then we find the following transformations. For case 1, we have

$$E[U_{t+1}^h|F_t] = (1 - \frac{2}{n} + \frac{2}{3mn})U_t^h + \frac{4}{3mn} U_t^{id} + \frac{2}{mn} U_t^{id}$$

$$+ \frac{1}{2mn} \sum_{s \in S, s \neq h, h^{-1}} (U_t^{hs^{-1}} + U_t^{hs} + U_t^{sh} + U_t^{sh^{-1}})$$
For case 2, we have
\[ E[U_{t+1}^h|F_t] = (1 - \frac{2}{n}) U_t^h + \frac{1}{2mn} \sum_{s \in S} (U_t^{hs-1} + U_t^{hs} + U_t^{sh} + U_t^{sh-1}) \]

Finally, in case 3, we have
\[ E[U_{t+1}^{id}|F_t] = (1 - \frac{2}{3n}) U_t^{id} + \frac{2}{3mn} \sum_{s \in S} U_t^s \]

where the sum of the coefficients is now 1 in all three cases. In particular, the above three equations now define a Markov chain on \( G \). This random walk sends the identity to itself with probability \( 1 - \frac{2}{n} \), and to a uniformly chosen element of \( S \) with the remaining probability; the two more complicated formulae above describe transitions for \( h \in S \) and \( h \notin S \). Call the transition operator \( K \).

Before analyzing the chain, we note that \( \sum_{g \in G} (X_t[g] - Y_t[g]) = 0 \), and so
\[
0 = \left( \sum_{i=1}^{n} (X_t[i] - Y_t[i]) \right)^2 \\
= \sum_{i=1}^{n} (X_t[i] - Y_t[i])^2 + \sum_{i \neq j} (X_t[i] - Y_t[i])(X_t[j] - Y_t[j]) \\
= S_t^{id} + \sum_{h \neq id} S_t^h
\]

From this calculation, if \( \langle v, (2, 1, 1, \ldots, 1) \rangle = 0 \), then \( \langle K v, (2, 1, 1, \ldots, 1) \rangle = 0 \) as well (where \( K \) is the operator described above). By direct computation, \( \pi = \frac{1}{n+1}(2, 1, 1, \ldots, 1) \) is a reversible measure for \( K \). It is also clear that the distribution \( \hat{\pi} = \frac{1}{n}(1, 1, \ldots, 1) \) is the reversible measure for \( \hat{K} \).

We are now ready to compare the chains. Recall from [9] that the Dirichlet form associated to a Markov chain with transition kernel \( Q \) and stationary distribution \( \nu \) is given by
\[
\mathcal{E}(\phi) = \frac{1}{2} \sum_{h,g \in G} \nu(g)Q(g,h)(\phi(x) - \phi(y))^2
\]

Let \( \mathcal{E} \) and \( \hat{\mathcal{E}} \) be the Dirichlet forms associated with \( K \) and \( \hat{K} \) respectively. Then by comparing terms, it is clear that \( \mathcal{E}(\phi) \geq \frac{1}{2}\hat{\mathcal{E}}(\phi) \) for any \( \phi \) and \( \frac{\pi}{\hat{\pi}}, \frac{\pi}{\hat{\pi}} \leq 2 \). By e.g. lemma 13.12 of [12], this implies \( \gamma \geq \frac{1}{8}\hat{\gamma} \).

By comments above, if \( \langle v, \pi \rangle = 0 \), then \( \langle K^m v, \pi \rangle = 0 \) as well, and in particular \( K \) applied to the subspace orthogonal to \( \pi \) has \( L^2 \to L^2 \) operator norm at most \( 1 - \gamma \). Thus, we have for any \( v \) in that subspace
\[
\|K^m v\|_2 \leq e^{-[\gamma m]}\|v\|_2
\]
going back to our original situation, we are interested in the vector \((S^g_t)\). At time 0, \(S^{id}_0 \leq 2\), and by Cauchy-Schwarz \(|S^{id}_0| \leq 4\). Thus, \(||(U^g_0)||_2 \leq 4n\), and of course \(|S^{id}_t| \leq ||(S^g_t)||_2 \leq ||(U^g_t)||_2\). So we find that
\[
E[|S^{id}_t|] \leq 4ne^{-\frac{1}{8}}
\]
which is the contraction we were interested in. □

7. Coupling for Gibbs Samplers on the Simplex with Geometry

Having shown contraction, we must now show convergence in total variation distance. First, the analogue to the connectedness lemma:

Lemma 15 (Connectedness for Gibbs Sampler on Cayley Graphs). Let \(\tau, \hat{\gamma}\) be as above. Then for \(t > \frac{(C+3)\log(n)}{\gamma}\), we have
\[
P[\tau > t] \leq 2n^{-C}
\]

Proof: We consider a graph-valued process \(G_t\), where \(G_0\) is a graph with no edges, and vertex set equal to the group \(G\). To construct \(G_{t+1}\) from \(G_t\), choose elements \(g \in G\) and \(s \in S\) uniformly at random, and add the edge \((g, gs)\) if it isn’t already in \(G_t\). We note that \(\tau > t\) if and only if \(G_t\) is not connected, so we would like to estimate the time at which it becomes connected.

First, fix two elements \(x\) and \(y\) in the group. We’d like to see if \(x\), \(y\) are in the same component of \(G_t\). To do so, let \(X_t, Y_t\) be two copies of the Gibbs sampler described in the last section, with \(X_0 = \delta_x\), \(Y_0 = \delta_y\). Couple \(X_t, Y_t\) and \(G_t\) by choosing the same edge and uniform variable \(\lambda\) at each step. Then assume \(x\), \(y\) are in different components \(C_x\), \(C_y\) at time \(t\). We would have
\[
\sum_g |X_t[g] - Y_t[g]|^2 \geq \sum_{g \in C_x} \frac{1}{|C_x|^2} + \sum_{g \in C_y} \frac{1}{|C_y|^2} \\
\geq \frac{4}{|G|}
\]
By Markov’s inequality, then,
\[
P[C_x \neq C_y] \leq \frac{n}{4}E[\sum_g |X_t - Y_t|^2]
\]
and so, by standard union bound for fixed \(x\) over all \(y\), if \(A_t\) is the event that \(G_t\) is disconnected,
\[
P[A_t] \leq \frac{n^2}{4} \sup_{\mu, \nu} E[\sum_g |X_t - Y_t|^2 |X_0 = \mu, Y_0 = \nu] \\
\leq 2n^3e^{-[t\gamma]}
\]
Next, we discuss the probability of success for subset couplings. For a pair of points \((x, y)\) in the simplex, a pair of update entries \((i, j)\), and a subset \(S \subset [n]\) of interest such that \(i \in S\) and \(j\) not in \(S\), we define \(p(x, y, i, j, S)\) to be the probability that the associated subset coupling succeeds. Then the following lemma from [20] gives a lower bound on this probability:

**Lemma 16 (Subset Coupling).** For a pair of vectors \((x, y)\) satisfying \(\sup_i |x_i - y_i| \leq n^{-e}\) and \(\inf_i x_i, \inf_i y_i \geq n^{-b}\), for \(e > b\), we have for all sufficiently large \(n\) that \(p(x, y, i, j, S) \geq 1 - 2n^{b+1-e}\) uniformly in \(S\) and possible \(i, j\).

In general, it is possible to choose \(x, y, i, j, S\) so that the probability of success is 0 under any coupling, and the lemma is quite restrictive. Having bounded the probability of failure when \(X_t, Y_t\) are close, we must show that they remain close with high probability. Define \(||X||_{p,S} = (\sum_{s \in S} |X[s]|^p)^{\frac{1}{p}}\) for \(p \geq 1\) and \(S \subset [n]\). Then:

**Lemma 17 (Smallness).** Let \(X_t, Y_t\) be coupled as described above, and assume that \(P_0\) is a single set, that all maximal couplings up to time \(t\) have succeeded, and that \(||X_0 - Y_0||_1 < \epsilon\). Then \(||X_t - Y_t||_{S,1} < \epsilon\) for every \(S\) in \(P_t\).

**Proof:** There are two types of coupling to take care of. For a proportional coupling between \(i\) and \(j\), we note that the error \(\Delta\) satisfies:

\[
\Delta = |X_{t+1}[i] - Y_{t+1}[i]| + |X_{t+1}[j] - Y_{t+1}[j]|
= \mu_t |X_t[i] + X_t[j] - Y_t[i] - Y_t[j]| + (1 - \mu_t) |X_t[i] + X_t[j] - Y_t[i] - Y_t[j]|
\leq |X_t[i] - Y_t[i]| + |X_t[j] - Y_t[j]|
\]

Since \(i\) and \(j\) always connect elements of the same set in \(P_t\), this shows that proportional couplings never increase \(||X_t - Y_t||_{S,1}\).

Otherwise, assume that at time \(t\) we had a successful maximal coupling between subsets \(S, T\) along edge \(i, j\), with \(i\) in \(S\) and \(j\) in \(T\). Then we note that

\[
X_{t+1}[i] - Y_{t+1}[i] = \sum_{s \in S/i} (Y_t[s] - X_t[s])
= Y_t[i] - X_t[i] + \sum_{s \in T} (X_t[s] - Y_t[s])
\]

and so

\[
|X_{t+1}[i] - Y_{t+1}[i]| \leq |X_t[i] - Y_t[i]| + ||X_t - Y_t||_{T,1}
\]

which immediately implies that

\[
||X_{t+1} - Y_{t+1}||_{S,1} \leq ||X_t - Y_t||_{S \cup T,1}
\]
inductively, this shows that $|X_{t+1} - Y_{t+1}|_{S,1} \leq |X_0 - Y_0|_1$. □

Related to this, the following lemma from chapter 13 of [1] shows that $X_t$, $Y_t$ rarely have entries close to 0:

**Lemma 18 (Largeness).** $P[\inf_{1 \leq i \leq n} \inf_{0 \leq t \leq T} Y_t[i] \leq n^{-b}] = O(Tn^{2-b})$

This lets us complete the calculation. Assume that the initial contractive phase is of length $T_1 = \frac{C_1}{n}n \log(n)$, and that the second coupling phase is of length $T_2 = \frac{C_2}{n}n \log(n)$.

By lemma 14, $E[\sum_{g \in G}|X_{T_1}[g] - Y_{T_1}[g]|^2] \leq 4n^{1-C_1}$, and by Markov’s inequality, $P[\sum_{g \in G}|X_{T_1}[g] - Y_{T_1}[g]| \geq n^{-a}] \leq 4n^{1+a-C_1}$.

By lemma 18, $P[\inf_{1 \leq i \leq n} \inf_{0 \leq t \leq T} Y_t[g] \leq n^{-b}] \leq \frac{1}{2}n^{3-b}$. By lemma 15, the probability that $P_{T_1}$ consists of a single block is at least $1 - 2n^{3-C_2}$. Finally, by lemmas 16 and 17 the probability that any of the maximal couplings fail is less than $2n^{b+2-a}$. Thus, the sum of the failure probabilities is at most $4n^{1+a-C_1} + \frac{1}{2}n^{3-b} + 2n^{3-C_2} + 2n^{b+2-a}$. To come close to minimizing this, we choose, for $C_1 + C_2 = 4x + 9 + \frac{\log(\hat{\gamma})}{\log(n)}$, $C_1 = 3x + 6 + \frac{\log(\hat{\gamma})}{\log(n)}$, $C_2 = x + 3$, $a = 2x + 5 + \frac{\log(\hat{\gamma})}{\log(n)}$, and $b = x + 3 + \frac{\log(\hat{\gamma})}{\log(n)}$ and find that the probability of failure is at most $9n^{-x}$.

Thus, we have shown that for the simplex walk, for $t > (4x + 9)\frac{\log(n)}{\gamma} + \frac{\log(\hat{\gamma})}{\gamma}$,

$$||\mathcal{L}(X_t) - \mathcal{L}(Y_t)||_{TV} \leq 9n^{-x}$$

It is easy to show that $\hat{\gamma} \geq \frac{1}{2n^2}$ for simple random walk on any Cayley graph (e.g. by naive bounds with theorem 13.14 of [12]), and using that in the inequality above proves theorem 2.

8. LOWER BOUNDS FOR GIBBS SAMPLERS ON THE SIMPLEX WITH GEOMETRY

In this section, we prove lower bounds on the mixing time of the Gibbs sampler on the simplex. The results are similar to those of [17], though the method is different and elementary. We begin by calculating

$$E[X_{t+1}[g]|X_t] = (1 - \frac{2}{n})X_t[g] + \frac{1}{n|S|} \sum_{s \in S} \frac{1}{2} (X_t[g] + X_t[gs])$$

$$= (1 - \frac{1}{n})X_t[g] + \frac{1}{n|S|}X_t[gs]$$

In particular, let $K$ be the transition matrix on $G$ given by $K[g,g] = 1 - \frac{1}{n}$, and $K[g,gs] = \frac{1}{n|S|}$ for $s \in S$. This is the standard ‘edge’-based random walk on $G$ with generating set $S$ described above, and we have shown that $E[X_t] = K^tX_0$. Note that this is not the same $K$ as in the previous two sections.

Next, let $v$ be a top eigenvector of $K$, normalized so that $||v||_2 = 1$ and $||v - Kv||_2 = \gamma$, and...
the spectral gap of $K$. Let $\Pi$ be the collection of vectors with nonnegative entries summing to 1, and let $w \in \Pi$ maximize the inner product $\langle v, w \rangle$ among such vectors. Let $X_t$ be a copy of the Markov chain begun from $X_0 = w$, then $E[\langle X_t, v \rangle] = \lambda^t(w, v)$. On the other hand, if $A_{t,d}$ is the event that $\langle X_t, v \rangle > d$,

$$
E[\langle X_t, v \rangle] = E[\langle X_t, v \rangle 1_{A_{t,d}}] + E[\langle X_t, v \rangle 1_{A_{t,d}^c}]
\leq \langle X_0, v \rangle P[A_{t,d}] + d
$$

and so

$$
P[A_{t,d}] \geq \frac{E[\langle X_t, v \rangle] - d}{\langle X_0, v \rangle}
$$

putting the two inequalities together,

$$
P[A_{t,d}] \geq \lambda^t - \frac{d}{\langle X_0, v \rangle}
$$

Next, we prove that $\langle X_0, v \rangle \geq \frac{2}{\sqrt{n}}$. Let $P \subset [n]$ be the collection of indices so that $v[p] \geq 0$ for $p \in P$. Without loss of generality, we may assume $\sum_{p \in P} v_p^2 > \frac{1}{2}$. Now set $\lambda^{-1} = \sum_{p \in P} v_p \leq \sqrt{n}$. Then consider the distribution $\nu_p = \lambda v_p$ for $p \in P$, 0 otherwise. We have

$$
\langle \nu, v \rangle = \lambda \sum_{p \in P} v_p^2
\geq \frac{2}{\sqrt{n}}
$$

and so

$$
P[A_{t,d}] \geq \lambda^t - \frac{1}{2}d\sqrt{n}
$$

Now, let $Y \in \Delta_n$ be chosen according to the uniform distribution. Then $E[(Y, v)] = 0$, and

$$
E[(Y, v)^2] = E[(\sum_{i=1}^n Y[i]v_i)^2]
= \sum_{i=1}^n v_i^2 E[Y[i]^2] + 2 \sum_{1 \leq i < j \leq n} v_i v_j E[Y[i]Y[j]]
\leq \frac{2}{n^2} + 0
$$
So, by Chebyshev’s inequality, \( P[\langle Y, v \rangle > d] \leq \frac{2}{dn^2} \). Putting this together with the inequality above, letting \( d = n^{-\frac{5}{6}} \) and defining \( P[A_{\infty, d}] = \lim_{t \to \infty} P[A_{t, d}] \),

\[
P[A_{t, d}] - P[A_{\infty, d}] \geq \lambda t - 3n^{-\frac{1}{3}}
\]

And the lower bound follows immediately.

**REFERENCES**


**Department of Mathematics, Stanford University, Stanford, CA 94305**

E-mail address: asmith3@math.stanford.edu