ESTIMATING THE SPECTRAL GAP OF A REVERSIBLE MARKOV CHAIN FROM A SHORT TRAJECTORY

DAVID A. LEVIN AND YUVAL PERES

ABSTRACT. The spectral gap $\gamma$ of an ergodic and reversible Markov chain is an important parameter measuring the asymptotic rate of convergence. In applications, the transition matrix $P$ may be unknown, yet one sample of the chain up to a fixed time $t$ may be observed. Hsu, Kontorovich, and Szepesvari [1] considered the problem of estimating $\gamma$ from this data. Let $\pi$ be the stationary distribution of $P$, and $\pi_* = \min_x \pi(x)$. They showed that, if $t = O(\frac{1}{\gamma^3 \pi_*})$, then $\gamma$ can be estimated to within multiplicative constants with high probability. They also proved that $\tilde{\Omega}(\frac{n}{\gamma})$ steps are required for precise estimation of $\gamma$. We show that $\tilde{O}(\frac{1}{\gamma \pi_*})$ steps of the chain suffice to estimate $\gamma$ up to multiplicative constants with high probability. When $\pi$ is uniform, this matches (up to logarithmic corrections) the lower bound in [1].

1. INTRODUCTION

Consider an ergodic and reversible Markov chain $\{X_t\}$ on a finite state space of size $n$, with transition matrix $P$ and stationary distribution $\pi$. We will assume that $P$ is positive definite, to avoid complications arising from eigenvalues close to $-1$. The spectral gap of the chain is $\gamma = 1 - \lambda_2$, where $\lambda_2$ is the second largest eigenvalue of $P$. The spectral gap is an important parameter of intrinsic interest, as it governs the asymptotic rate of convergence to stationarity.

Suppose one does not know $P$, but is able to observe the chain $\{X_t\}_{t=1}^T$. Can one estimate $\gamma$ with precision from this data? This question was studied by Hsu, Kontorovich, and Szepesvari in [1]. Their estimator is the spectral gap of the (suitably symmetrized) empirical transition matrix. They show that $t = \tilde{O}(\frac{1}{\gamma^3 \pi_*})$ observations of the chain are enough to estimate $\gamma$ to within a constant factor. See Theorem 2 for a precise statement. In the case where $\pi$ is uniform, the authors of [1] also show that $\tilde{\Omega}(\frac{1}{\pi \gamma})$ steps are needed to estimate $\gamma$. Here we show that $t = \tilde{O}(\frac{1}{\gamma \pi_*})$ is a sufficient number of observations to estimate $\gamma$ to within a constant factor. In particular, we prove:

Theorem 1. Fix $\delta > 0$. There is an estimator $\hat{\gamma}$ of $\gamma$ based on $\{X_t\}_{t=0}^T$, and a polynomial function $\mathcal{L}$ of the logarithms of $\gamma$, $\pi_*^{-1}$, $\delta^{-1}$, and $n$, such that, if $t > \frac{1}{\pi \gamma \mathcal{L}}$, then we have $|\hat{\gamma} - \gamma| < \varepsilon$ with probability at least $1 - \delta$.

The definition of $\mathcal{L}$ is in (4).

The proof of Theorem 1 applies the estimator of Hsu, Kontorovich, and Szepesvari to estimate the gap $\gamma_A$ of the "skipped chain" $\{X_{A_i}\}_{i=1}^{T/A}$. By successively doubling $A$, with
high probability one can identify the first value \( A \) such that \( \gamma_A \) is uniformly bounded below. Once this \( A \) is identified, the estimate of \( \gamma_A \) can be transformed to an estimate of \( \gamma \).

While \( \gamma \) is a parameters of intrinsic interest, it is also related to another important parameter, the mixing time. The mixing time \( t_{\text{mix}}(\epsilon) \) is the first time such that (from every starting state) the distribution of the chain is within \( \epsilon \) from \( \pi \) in total-variation. Always \( \gamma^{-1} \leq t_{\text{mix}}(1/4) + 1 \), however, if \( \pi_* = \min_x \pi(x) \), then \( t_{\text{mix}}(\epsilon) \leq |\log(\epsilon \pi_*)| \cdot \gamma^{-1} \). See [2] for background on the spectral gap and the mixing time.

2. Proof of Theorem 1

We will repeatedly apply the following estimate of Hsu, Kontorovich, and Szepesvari:

**Theorem 2** ([1]). There is an estimator \( \hat{\gamma} \) of \( \gamma \), based on \( t \)-steps of the Markov chain, such that for some absolute constant \( C \), with probability at least \( 1 - \delta \),

\[
|\hat{\gamma} - \gamma| \leq C \left( \frac{\log \left( \frac{\delta}{\epsilon} \right) \cdot \log \left( \frac{1}{2^t} \right) \cdot \log(1/\gamma)}{\pi_* \gamma^2} + \frac{\log(1/\gamma)}{\gamma^t} \right) := M(t; \delta, \pi, \gamma). \tag{1}
\]

Thus, Theorem 2 says \( t = \tilde{O}(\frac{1}{\gamma^2}) \) steps suffice for \( \hat{\gamma} \gamma \) to be near 1.

We call the estimator \( \hat{\gamma} \) the HKS estimator. Note that if

\[
t_1 = t_1(\epsilon; \delta, \gamma) := \frac{1}{\pi_* \gamma} \frac{12C^2 \log(n/\delta) \log(12C^2/(\epsilon^2 \pi_*^2 \gamma \delta))}{\epsilon^2},
\]

then

\[
M(t_1; \delta, \pi, \gamma) \leq \frac{\epsilon}{2} \left( \frac{\log \left( \frac{12C^2}{\epsilon^2 \pi_*^2 \gamma \delta} \right) + \log(\log(n/\delta)) + \log \log(\frac{12C^2}{\epsilon^2 \pi_*^2 \gamma \delta})}{3 \log \left( \frac{12C^2}{\epsilon^2 \pi_*^2 \gamma \delta} \right)} + \frac{\epsilon}{2} \right) \leq \epsilon.
\]

(Each term in the numerator under the radical is at most a third of the denominator. We have used that \( \pi_* \leq 1/n \) in comparing the second term in the numerator to the denominator.)

For \( a > 0 \), the gap of the chain with transition matrix \( P^a \) is denoted by \( \gamma_a \), and the HKS estimator of \( \gamma_a \), based on \( t/a \) steps of \( P^a \), is denoted by \( \hat{\gamma}_a \). Note that

\[
\gamma_a = 1 - (1 - \gamma)^a.
\]

Let \( \delta_{\gamma} = \frac{\delta}{\lfloor \log(1/\gamma) \rfloor + 1} \).

**Proposition 3.** Fix \( \delta > 0 \) and \( \epsilon < 0.01 \). If \( t > t_1(\epsilon/2; \delta_{\gamma}, \gamma) \), then there is an integer-valued random variable \( A \), based on \( t \) steps of the Markov chain, and an event \( G(\epsilon) \) having probability at least \( 1 - \delta \), such that on \( G(\epsilon) \),

\[
0.30 < \gamma_A < 0.54 \quad \text{if} \quad \gamma < 1/2,
\]

\[
A = 1 \quad \text{if} \quad \gamma \geq 1/2.
\]

Moreover, on \( G \), the HKS estimator \( \hat{\gamma}_A \) applied to the chain \( \{X_{sA} \}_{s=0}^{t/A} \) satisfies

\[
|\hat{\gamma}_A - \gamma_A| < \epsilon.
\]

Define

\[
G(a; \epsilon) = \{ |\gamma_a - \hat{\gamma}_a| < \epsilon \}.
\]

**Lemma 4.** Fix \( t > t_1(\epsilon/2; \delta, \gamma) \). If \( \alpha \gamma \leq 1 \), then \( \mathbb{P}(G(a; \epsilon)) > 1 - \delta. \)
Proof. Recall the bound \(M(t; \delta, \pi, \gamma)\) on the right-hand side of (1). If \(\gamma_a \geq \frac{1}{2} \gamma a\), then
\[
M(t; \alpha; \delta, \gamma, \pi, \gamma_a) \leq 2M(t; \delta, \pi, \gamma) \leq 2 \frac{\epsilon}{a} = \epsilon,
\]
and the lemma follows from applying Theorem 2 to the \(P^a\)-chain. We now show that \(\gamma_a \geq \frac{1}{2} \gamma a\). Expanding \((1 - \gamma)^a\), there exists \(\xi \in [0, a^{-1}]\) such that
\[
\gamma_a = 1 - (1 - \gamma)^a = \gamma a - \frac{a(a - 1)(1 - \xi)^{a-2} \gamma^2}{2} \geq \frac{\gamma a}{2}.
\]
(We have used the hypothesis \(a \gamma \leq 1\) in the inequality.) \(\square\)

Proof of Proposition 3. Let \(\delta_{\gamma} = \frac{\delta}{\log (1/\gamma)}\). Fix \(t > t_1(\epsilon/2; \delta_{\gamma}, \gamma)\).

Set \(K_\gamma := \left\lfloor \log_2 \left( \frac{1}{\delta_{\gamma}} \right) \right\rfloor\), and let \(\{X_{i+1}^t\}_{i=1}^t\) be \(t\) steps of the Markov chain. Consider the following algorithm:

Begin by setting \(k = 0\). Let \(a = 2^k\). Using the “skipped” chain \(\{X_{i+1}^t\}_{i=1}^{t/\alpha}\) observed for \(t/\alpha\) steps, form the HKS estimator \(\hat{\gamma}_a\) of the spectral gap \(\gamma_a\) of the skipped chain. If \(\hat{\gamma}_A > 0.31\) then set \(A = a = 2^k\), and stop. Otherwise, increment \(k\) and repeat.

Define the event \(G = G(\epsilon) \overset{\text{def}}{=} \bigcap_{k=0}^{K_\gamma} G(2^k; \epsilon)\). If \(k \leq K_\gamma\), then \(\gamma 2^k \leq 2^{-\log_2 (1/\gamma)} \leq 1\) and Lemma 4 implies that
\[
P(G^c) \leq \sum_{k=0}^{K_\gamma} P(G(2^k; \epsilon)^c) \leq (K_\gamma + 1) \frac{\delta}{K_\gamma + 1} = \delta.
\]

On \(G\), if \(\gamma \geq 1/2\), then \(\hat{\gamma} - \gamma < 0.01\), and consequently \(\hat{\gamma} \geq 0.49 > 0.31\). In this case, \(A = 1\) on \(G\).

On the event \(G\), if the algorithm has not terminated by step \(k - 1\), then:

- \(\gamma_{2^k} \leq 0.30\), then the algorithm does not terminate at step \(k\);
- \(\gamma_{2^k} > 0.32\), then the algorithm terminates at step \(k\).

Also,
\[
\gamma_{2^k} \geq 1 - (1 - \gamma)^{1/2} \geq 1 - e^{-1/2} \geq 0.39,
\]
so the algorithm always terminates before \(k = K_\gamma\) on \(G\).

Finally, on \(G\), if \(A > 1\), then \(\gamma_{2^k} \leq 0.32\), whence
\[
\gamma_A = 1 - (1 - \gamma_{2^k}) \leq 1 - (0.68)^2 < 0.54.
\]

If \(\gamma < 1/2\) and \(A = 1\), then \(\gamma_A = \gamma \leq 1/2\). \(\square\)

Proof of Theorem 1. For \(C_0 = 23232 \cdot C\), where \(C\) is the constant in (1), let
\[
t_0(\epsilon; \delta, \gamma, \pi_+) = t_0(\epsilon) := \left( \frac{1}{\epsilon \pi_+} \right) \frac{C_0}{\epsilon^2 \mathcal{L}^*},
\]
where
\[
\mathcal{L}^* = \log \left( \frac{C_0 \left( \left\lfloor \log_2 (1/\gamma) \right\rfloor + 1 \right)}{\epsilon^2 \pi_+^2 \gamma \delta} \right) \log \left( \frac{n \left( \left\lfloor \log_2 (1/\gamma) \right\rfloor + 1 \right)}{\delta} \right).
\]

Fix \(t > t_0(\epsilon) = t_1(\epsilon/44; \delta_{\gamma}, \gamma)\). Let \(A\) and \(G\) be as defined in Proposition 3. Assume we are on the event \(G = G(\epsilon/22)\) for the rest of this proof.

Suppose first that \(\gamma < 1/2\). We have \(0.30 < \gamma_A < 0.54\), and
\[
|\hat{\gamma}_A - \gamma_A| < \frac{\epsilon}{22} < 0.01,
\]
so both \(\gamma_A\) and \(\hat{\gamma}_A\) are in \([0.29, 0.55]\), say.
Let \( h(x) = 1 - (1 - x)^{1/A} \), so \( \gamma = h(\gamma A) \). Note that on \([0.29, 0.55]\),

\[
\frac{d}{dx} \log h(x) = \frac{1}{1 - (1 - x)^{1/A}} \cdot \frac{1}{A} (1 - x)^{1/A - 1} \leq \frac{1}{A (1 - (1 - 0.29)^{1/A})} \leq \frac{1}{0.45}.
\]

Since \( (1 - x)^{1/A} \leq 1 - x/A + x^2/(2 A^2) \),

\[
\frac{d}{dx} \log h(x) \leq \frac{1}{0.29 - 0.29^2/(2)(0.45)} < 11.
\]

Thus, \(|\frac{d}{dx} \log h(x)\)| is bounded (by 11) on \([0.29, 0.55]\). Write \( \tilde{\gamma} = h(\tilde{\gamma} A) \). We have

\[
|\log(h(\tilde{\gamma} A)/\gamma)| = |\log h(\gamma A) - \log h(\tilde{\gamma} A)| \leq 11 |\gamma A - \tilde{\gamma} A| \leq 11 \frac{\varepsilon}{22} \leq \frac{\varepsilon}{2}.
\]

Thus,

\[
\frac{h(\tilde{\gamma} A)}{\gamma} \leq e^{\varepsilon/2} \leq 1 + \varepsilon.
\]

Similarly, \( \frac{\gamma}{h(\gamma A)} \leq e^{\varepsilon/2} \), so

\[
\frac{h(\tilde{\gamma} A)}{\gamma} \geq e^{-\varepsilon/2} \geq 1 - \varepsilon.
\]

Suppose that \( \gamma \geq 1/2 \). Then \( A = 1 \) on the event \( G \), and

\[
|\tilde{\gamma} - \gamma| < \frac{\varepsilon}{22},
\]

so

\[
\left| \frac{\tilde{\gamma}}{\gamma} - 1 \right| < \frac{\varepsilon}{22 \gamma} \leq \varepsilon.
\]

\[\square\]

**Remark 1.** If \( t < t_0(\varepsilon) \), then our estimation procedure is not guaranteed to produce a sensible estimate.

**Acknowledgements**

We thank Jian Ding for helpful conversations on this topic.

**References**

[1] Daniel J. Hsu, Aryeh Kontorovich, and Csaba Szepesvari. Mixing time estimation in reversible markov chains from a single sample path. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., 2015.

[2] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.