COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
60. SETS, GRAPHS AND NUMBERS, BUDAPEST (HUNGARY), 1991

Noise Perturbations of Random Mappings Tend to Produce Rapid Mixing

J. FRIEDMAN, C. GOTSMAN* and E. SHAMIR†

ABSTRACT

We show that noise operators composed with Markov chains induced by random mappings tend to cause rapid mixing of the chain.

1. Introduction

Denote $[N] = \{1, \ldots, N\}$. Consider a mapping $F : [N] \longrightarrow [N]$ and the $N \times N$ transition matrix $\mathbf{P}_0$, corresponding to F. The matrix $\mathbf{P}_0$ has zero entries, except for a single 1 on each row. Let $\{\mathbf{I}_\delta : \delta > 0\}$ be a family of stochastic matrices representing noise. By this we mean that $\mathbf{I}_0 = \mathbf{I}$ (identity matrix), and the diagonal elements of $\mathbf{I}_\delta$ are $1 - O(\delta)$. Now consider

$$\mathbf{P} = \mathbf{P}_0 \mathbf{I}_\delta$$

* Partially supported by an Eshkol fellowship, administered by The National Council for Research and Development - Israel Ministry of Science and Development.

† Partially supported by grant #438/89 of the Israel Academy of Sciences.

This is a (usually ergodic) Markov chain on $[N]$ for $\delta > 0$. One expects that it mixes rapidly. This means that the discrepancy between the rows of $\mathbf{P}^k$ and the stationary distribution of $\mathbf{P}$ decreases like q^k for some $q < 1$ (more rapidly if q is far from 1). The practical significance of this property is that stationarity is achieved after only a small fraction of the state space has been traversed (say, some power of $\log N$ states). If $\{\lambda_0, \lambda_1, \ldots, \lambda_{N-1}\}$ are the eigenvalues of $\mathbf{P}$ in decreasing order of their modulii, then $\lambda_0 = 1$ and $|\lambda_1|$ is an estimate for q [5]. Thus the goal is to show that λ_1 is well inside the unit disc as δ grows.

Example: $\mathbf{P}_0$ is a permutation matrix ($\mathbf{P}_0^t = \mathbf{P}_0{}^{-1}$) and $\mathbf{I}_\delta$ is symmetric such that $|\lambda_1(\mathbf{I}_\delta)| \leq 1 - b\delta$. It follows from the inequality concerning singular values of product matrices ([3] p. 246) that

$$|\lambda_1(\mathbf{P})| \leq |\lambda_1(\mathbf{P}_0)||\lambda_1(\mathbf{I}_\delta)| \leq 1 - b\delta \tag{1}$$

■

We shall concentrate on a specific form of symmetric noise matrix

$$\mathbf{I}_\delta = \bigtimes_{i=1}^{n} \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}.$$

where $\times$ denotes tensor product. Thus the size of $\mathbf{I}_\delta$ is $N = 2^n$. An illuminating interpretation of the Markov chain $\mathbf{P} = \mathbf{P}_0\mathbf{I}_\delta$ in this case is obtained as follows: Consider a network of n binary $(0-1)$ processors. A state $\mathbf{x} = (x_1, \ldots, x_n)$ of the chain is a boolean vector, where x_i is the value of processor i. The 2^n states of the chain are ordered and $\mathbf{P}_0$ represents n boolean functions $F = (f_1, \ldots, f_n)$, where $f_i : \{0,1\}^n \longrightarrow \{0,1\}$ describes how x_i evolves with time , i.e.

$$\mathbf{x}(t+1) = (f_1(\mathbf{x}(t)), \ldots, f_n(\mathbf{x}(t))). \tag{2}$$

However, in the chain $\mathbf{P}$, the transitions in time are stochastic, and the transition of each coordinate x_i as in (2) therefore occurs only with probability $1 - \delta$. A transition to the opposite boolean value occurs with probability δ. The complete state transition (2) occurs with probability $(1-\delta)^n$. Thus $\mathbf{I}_\delta$ introduces independent δ noise in each coordinate, and therefore its name "the hypercube noise".

One would hope to have an estimate like (1) for general (non-invertible) $\mathbf{P}_0$, with a fixed positive b. This is not true in general. Our thesis is

that a bound similar to (1) should remain true for a *random* mapping $F : \{0,1\}^n \longrightarrow \{0,1\}^n$, i.e. for *almost all* $\mathbf{P}_0$. We are unable to compute $|\lambda_1|$ exactly for arbitrary $\mathbf{P}$. However, we are able to compute $\frac{\partial|\lambda_1|}{\partial\delta}$ at $\delta = 0$ and show that it is approximately $-n$ with high probability. This gives an estimate of $1 - n\delta$ for $|\lambda_1|$ in the cases where δ is extremely small (e.g. exponentially small). For larger values of δ, an accurate estimate for $|\lambda_1|$ requires higher order derivatives. We have had no success in calculating these higher order derivatives. However, numerical results seem to indicate that the estimate $1 - n\delta$ for $|\lambda_1|$ is valid for $\delta = o(\frac{1}{n})$ (polynomially small noise), which would mean that rapid mixing is obtained.

The practical significance of this result is that deterministic graphs may be turned into stationary distributed (or nearly stationary distributed) graphs *very* quickly by *small* perturbations, yielding uninformative, unpredictable graphs. This could have applications in cryptography.

2. The Rapid Mixing Theorem

In this section, we estimate $\dfrac{\partial|\lambda_1|}{\partial\delta}$ at $\delta = 0$ for a random $F : \{0,1\}^n \longrightarrow \{0,1\}^n$. Recall that the transition graph corresponding to $\mathbf{P}_0$ is deterministic, characterized by a finite number of connected components. The successor of a state $\mathbf{x}$ will be denoted by $\mathrm{succ}(\mathbf{x})$. Each component C_i (of size $|C_i|$) has transient states $\mathrm{tr}(C_i)$ and a unique cycle $\mathrm{cyc}(C_i)$ of length l_i. To each cycle corresponds a primitive root of unity ω_i of order l_i. This cycle contributes all l_i distinct powers of ω_i to the spectrum of $\mathbf{P}_0$ (the transient states contribute zeros to the spectrum). The left eigenvector of $\mathbf{P}_0$ corresponding to ω_i^m supported only on C_i, is $\mathbf{x} = (\omega_i^{-m}, \omega_i^{-2m}, \ldots, \omega_i^{-ml_i} = 1)$ (coordinates not supported have been omitted).

For each specific eigenvalue $\omega \neq 0$ of $\mathbf{P}_0$ with multiplicity k, we define a set of vectors $\mathbf{z}_i$, $i = 1, \ldots, k$. Their definition and connection to $\mathbf{P}_0$ is outlined in the following two lemmas:

Lemma 1. *Let $\omega \neq 0$ be an eigenvalue of $\mathbf{P}_0$, and let C_i, $i = 1, \ldots, k$ be all the connected components of $\mathbf{P}_0$'s transition graph which contribute ω to the spectrum (all the components with cycle length l such that $\omega^l = 1$). Construct the vectors $\mathbf{z}_i$, each supported on C_i, as follows:*

- *Choose an arbitrary state* $\mathbf{x}$ *in the cycle* $\mathrm{cyc}(C_i)$. *Label* $\mathbf{x}$ *with* 1. *Label* $\mathrm{succ}(\mathbf{x})$ *with* ω, $\mathrm{succ}(\mathrm{succ}(\mathbf{x}))$ *with* ω^2 *and so on until the cycle is completed.*
- *Label each transient node in* C_i *at distance* r *from the closest cyclic node* $\mathbf{x}$ *(*r *transitions lead it into the cycle at* $\mathbf{x}$*) with* $\omega^{-r} \cdot \mathrm{label}(\mathbf{x})$.
- *Nodes not in* C_i *are labelled with* 0.
- *Define* $\mathbf{z}_i = (\mathrm{label}(\mathbf{x}_1), \ldots, \mathrm{label}(\mathbf{x}_{2^n}))$.

Then for any vector $\mathbf{u}$ *and any* $1 \le i \le k$, $\mathbf{u}\mathbf{P}_0\mathbf{z}_i^t = \omega\mathbf{u}\mathbf{z}_i^t$ *(*$\mathbf{z}_i^t$ *denotes the transpose of* $\mathbf{z}_i$*).*

Note that the labelling of the cycles is done in the opposite direction around the cycle, compared to the labelling of $\mathbf{P}_0$'s eigenvectors supported on these cycles. This fact will play an important role in the sequel.

Proof. Observe that multiplication by $\mathbf{P}_0$ "pushes" u down the transients and around the cycle. Scalar product with $\mathbf{z}_i$ sums up the labels. This is equivalent to multiplying by ω and summing up. ■

Notation. *Let* $\mathbf{v}$ *be a* n*-vector and* $\mathbf{A}$ *a* $n \times n$ *matrix over a field* F. *Say that* $\mathbf{v} \in \mathbf{Im}(\mathbf{A})$ *iff there exists a* $\mathbf{u} \in F^n$ *such that* $\mathbf{v} = \mathbf{u}\mathbf{A}$.

Lemma 2. *If* $\mathbf{v} \in \mathbf{Im}(\mathbf{P}_0 - \omega\mathbf{I})$ *then* $\mathbf{v}\mathbf{z}_i^t = 0$ *for* $i = 1, \ldots, k$.

Proof. Since $\mathbf{v} \in \mathbf{Im}(\mathbf{P}_0 - \omega\mathbf{I})$, there exists a $\mathbf{u}$ such that $\mathbf{v} = \mathbf{u}(\mathbf{P}_0 - \omega\mathbf{I})$. Scalar product of the equation with $\mathbf{z}_i^t$ yields $\mathbf{v}\mathbf{z}_i^t = \mathbf{u}\mathbf{P}_0\mathbf{z}_i^t - \omega\mathbf{u}\mathbf{z}_i^t$, which vanishes by Lemma 1. ■

We are now ready to state the main result of this paper:

Theorem 1. *Let* $\mathbf{P}_0$ *be the transition matrix of a random mapping on* $N = 2^n$ *elements (i.e., each mapping is chosen with equal probability). Let* $\mathbf{P} = \mathbf{P}_0\mathbf{I}_\delta$, *and* λ_1 *be the complex random variable corresponding to the second eigenvalue of* $\mathbf{P}$. *Then*

$$\mathrm{Prob}\left[\frac{\partial|\lambda_1|}{\partial\delta}|_{\delta=0} \ge -n + O(2^{-\beta n})\right] \le \exp(-O(n^{-8}2^{2n(1/4-\beta)}))$$

for any $0 < \beta < 1/4$.

Proof. For small δ, we can regard $\mathbf{I}_\delta$ as a perturbation operating on $\mathbf{P}_0$, which perturbs the complex (roots of unity) spectrum of $\mathbf{P}_0$. Denote by ω

a typical eigenvalue of $\mathbf{P}_0$ with eigenvector $\mathbf{x}$ and multiplicity k. The same notations, with the subscript δ, are the perturbed values. Expanding as a power series (see [2], Chap. 2.2), we have:

$$\begin{aligned}\mathbf{I}_\delta &= \mathbf{I} + \delta\mathbf{\Delta} + \delta^2 \ldots \\ \omega_\delta &= \omega + \delta\omega' + \ldots \\ \mathbf{x}_\delta &= \mathbf{x} + \delta\mathbf{x}' + \ldots\end{aligned}$$

Here $\mathbf{\Delta} = \mathbf{A} - n\mathbf{I}$, where $\mathbf{A}$ is the adjacency matrix of the hypercube. We would like to estimate the possible values of $|\omega'|$. Since $\mathbf{x}_\delta \mathbf{P}_0 \mathbf{I}_\delta = \omega_\delta \mathbf{x}_\delta$, we derive the two first-order equations:

$$\mathbf{x}\mathbf{P}_0 = \omega\mathbf{x} \tag{3}$$

$$\mathbf{x}\mathbf{P}_0\mathbf{\Delta} + \mathbf{x}'\mathbf{P}_0 = \omega'\mathbf{x} + \omega\mathbf{x}' \tag{4}$$

Substituting (3) in (4), we have:

$$\mathbf{x}'(\mathbf{P}_0 - \omega\mathbf{I}) = \omega'\mathbf{x} - \omega\mathbf{x}\mathbf{\Delta}$$

or that $\omega'\mathbf{x} - \omega\mathbf{x}\mathbf{\Delta} \in \mathbf{Im}(\mathbf{P}_0 - \omega\mathbf{I})$, which has co-dimension k. Applying Lemma 2., we obtain

$$(\omega'\mathbf{x} - \omega\mathbf{x}\mathbf{\Delta})\mathbf{z}_i^t = 0 \qquad i = 1, \ldots, k \tag{5}$$

for all the $\mathbf{z}_i$ corresponding to ω, as defined in Lemma 1. Since $\mathbf{x}$ is an eigenvector of $\mathbf{P}_0$ corresponding to ω, it can be expressed as a linear combination of eigenvectors supported on the k distinct cycles whose lengths l_i satisfy $\omega^{l_i} = 1$:

$$\mathbf{x} = \sum_{j=1}^{k} \alpha_j \mathbf{x}_j. \tag{6}$$

Noting that $\mathbf{x}_j \mathbf{z}_i^t = l_j \delta_{ij}$, substituting (6) in (5), we obtain the k coupled equations:

$$l_i\omega'\alpha_i - \omega \sum_{j=1}^{k} \alpha_j \mathbf{x}_j \mathbf{\Delta} \mathbf{z}_i^t = 0 \qquad i = 1, \ldots, k.$$

Recalling that $\mathbf{\Delta} = \mathbf{A} - n\mathbf{I}$, this reduces to the following matrix equation for the k-vector α:

$$\alpha\mathbf{M} = (\omega' + n\omega)\alpha \tag{7}$$

where $\mathbf{M}$ is a $k \times k$ matrix. This implies that $\omega' + n\omega$ are eigenvalues of $\mathbf{M}$, defined by $\mathbf{M}_{ab} = \frac{\omega}{l_a}\mathbf{x}_a\mathbf{A}\mathbf{z}_b^t$. Denoting by neigh($\mathbf{x}$) the neighborhood of the vertex $\mathbf{x}$ on the hypercube (all vertices at unit distance from $\mathbf{x}$), this is equivalent to:

$$\mathbf{M}_{ab} = \frac{1}{l_a} \sum_{\substack{\mathbf{x} \in \mathrm{cyc}(C_a) \\ \mathbf{y} \in \mathrm{neigh}(\mathbf{x})}} \begin{cases} 0 & \text{if } \mathbf{y} \notin C_b \\ \omega^r & \text{if } \mathbf{y} \in C_b, \text{ for some integer } r \end{cases} \tag{8}$$

The value ω^r is obtained as $\mathrm{label}_{\mathbf{z}}(y)\mathrm{label}_{\mathbf{x}}(x)$. Thus, there is a $k_i \times k_i$ matrix for each distinct eigenvalue ω_i of $\mathbf{P}_0$ with multiplicity k_i. The first (and largest) matrix is $\mathbf{M}^1$ of dimension k_1, the multiplicity of the eigenvalue $\omega_1 = 1$, i.e. the number of components in the transition graph. In effect, this describes the transitions *between components,* they being *metastates.* From (8), the rows of $\mathbf{M}^1$ all sum to exactly n, and consequently this is also the largest eigenvalue of $\mathbf{M}^1$. Substituting in (7), this corresponds to the case $\omega' = 0$, which is to be expected, as $\mathbf{P}$ also has a unique unit eigenvalue λ_0. We will now proceed to show that the remaining eigenvalues of $\mathbf{M}^1$ have very small modulii. This is because $\mathbf{M}^1$ is very close to a rank 1 matrix, i.e. its rows are almost identical, similar to the first eigenvector α. For this we need the following facts about almost all random mappings on sets of size N (see [1] Chap. 14 or [4] for a complete exposition).

1. For any $0 < c < 1$, the expected no. of trees (set of transients connected to a single cyclic state) of size $> cN$ is asymptotically (as $N \longrightarrow \infty$) $\approx c^{-\frac{1}{2}} - 1$.
2. The *total* number of *cyclic* states is asymptotically $O(N^{\frac{1}{2}})$.
3. The total number of cycles (components) is asymptotically distributed like Poisson($\frac{1}{2}\log N$).

These facts indicate that random mappings are rather peculiar: with high probability there are exceptionally large components. On the other hand, the number of cyclic states is only of order $N^{\frac{1}{2}}$.

Applying Fact 3 to our case ($N = 2^n$), we have that $k_1 = O(n)$. By definition, and because of the random nature of the hypercube structure relative to the random mapping,

$$\mathbf{M}^1_{ab} \approx \frac{1}{l_a} \sum_{j=1}^{nl_a} \chi_j \tag{9}$$

where χ_j are i.i.d. Bernoulli variables such that $Pr\{\chi_j = 1\} = \frac{|C_b|}{2^n}$. Calculating the first two moments of the entries of $\mathbf{M}^1$:

$$\mathbf{E}(\mathbf{M}^1_{ab}) = \frac{1}{l_a} n l_a \frac{|C_b|}{2^n} = \frac{n}{2^n}|C_b| \tag{10}$$

$$\mathbf{Var}(\mathbf{M}^1_{ab}) = \frac{1}{l_a^2} n l_a \frac{|C_b|}{2^n}(1 - \frac{|C_b|}{2^n}) \approx \frac{n}{l_a}\frac{|C_b|}{2^n} \tag{11}$$

Divide the components of the random mapping into two disjoint sets: $\mathcal{C}_1 = \{C_i : |C_i| > 2^{\frac{2n}{3}}\}$, $\mathcal{C}_2 = \{C_i : |C_i| \leq 2^{\frac{2n}{3}}\}$. Because of Fact 1 about random mappings, $\mathcal{C}_1$ is not empty with extremely high probability, and because of Facts 2 and 3, the cycle lengths of components in $\mathcal{C}_1$ are $O(2^{\frac{n}{2}})$. After renaming the states, write the matrix $\mathbf{M}^1$ in block form:

$$\mathbf{M}^1 = \begin{pmatrix} \mathbf{M}^{11} & \mathbf{M}^{12} \\ \mathbf{M}^{21} & \mathbf{M}^{22} \end{pmatrix}$$

where $\mathbf{M}^{ij}$ describes the transitions from $\mathcal{C}_i$ to $\mathcal{C}_j$. Applying (10) and remembering that the rows sum to n, the expected values of the $\mathbf{M}^1_{ab}$ are (in block form):

$$\mathbf{E}(\mathbf{M}^1) = \begin{pmatrix} O(n) & O(n2^{-\frac{n}{3}}) \\ O(n) & O(n2^{-\frac{n}{3}}) \end{pmatrix}$$

and from (11), the variances are:

$$\mathbf{Var}(\mathbf{M}^1) = \begin{pmatrix} o(n2^{-\frac{n}{2}}) & o(n2^{-\frac{n}{3}}) \\ o(n2^{-\frac{n}{2}}) & o(n2^{-\frac{n}{3}}) \end{pmatrix}.$$

We now show that the matrices $\mathbf{M}^{12}$ and $\mathbf{M}^{22}$ almost vanish by bounding their Euclidean norm $||\mathbf{M}|| = (\sum_{a,b} \mathbf{M}^2_{ab})^{\frac{1}{2}}$. This also bounds the spectral radius of $\mathbf{M}^{i2}$:

$$||\mathbf{M}^{i2}|| \leq \sum_{a,b} |\mathbf{M}_{ab}| = O(n^3 2^{-\frac{n}{3}}) \qquad i = 1, 2. \tag{12}$$

This shows that $\mathbf{M}^1$ is effectively lower block-triangular. It's spectrum is the union of the spectra of $\mathbf{M}^{11}$ and $\mathbf{M}^{22}$. We have shown that the spectrum of $\mathbf{M}^{22}$ is exponentially small, so we turn our attention to that of $\mathbf{M}^{11}$.

Denote $\sigma^2 = n2^{-\frac{n}{2}}$. The modulus of the second eigenvalue of $\mathbf{M}^{11}$ can be bounded by the following (weak but useful) inequality ([5] p. 63): [1]

$$|\lambda_1| = |n + \omega'| \leq \frac{1}{2} \max_{a_1,a_2} \sum_b |\mathbf{M}^{11}_{a_1,b} - \mathbf{M}^{11}_{a_2,b}|. \tag{13}$$

It is easily verified that:

$$\mathbf{E}(|\mathbf{M}^{11}_{a_1,b} - \mathbf{M}^{11}_{a_2,b}|) = o(\sigma)$$

$$\mathbf{Var}(|\mathbf{M}^{11}_{a_1,b} - \mathbf{M}^{11}_{a_2,b}|) = o(\sigma^2).$$

Since $\max_{i,j} |a_{ij}| \leq \sum_{ij} |a_{ij}|$, the first two moments of the right side of (13) are of order:

$$\mu_1 = o(n^3\sigma) \quad ; \quad \sigma_1^2 = o(n^3\sigma^2)$$

which obviously can be bounded by

$$\mu_1 = \sigma_1 = o(n^4 2^{-\frac{n}{4}}).$$

Chernoff's bound on the deviation of a sum of i.i.d. random variables implies:

$$\text{Prob } [n + |\omega'| \geq |n + \omega'| > \mu_1 + \alpha\sigma_1] \leq \exp(-\frac{1}{2}\alpha^2)$$

for any $\alpha > 0$. Taking $\alpha = O(2^{(1/4-\beta)n})$ where $0 < \beta < 1/4$ yields

$$\text{Prob } [|\omega'| > -n + O(2^{-\beta n})] \leq \exp(-O(n^{-8} 2^{2n(1/4-\beta)}))$$

as required.

This shows that ω' is exponentially close to $-n$ in all except an increasingly small percentage of the cases.

We now show that the complex matrices $\mathbf{M}^j$ have almost vanishing norms for $j > 1$. This means that their spectrum also almost vanishes, yielding perturbations of almost $-n\delta$ on ω_j, the j'th eigenvalue of $\mathbf{P}_0$. If ω_j is a s'th root of unity, we first note that the dimension k_j of $\mathbf{M}^j$ is the number of components C_i such that $l_i \ (mod\ s) = 0$. This is approximately $\frac{k_1}{s}$.

---

[1] The expression on the right side of the inequality is the *coefficient of ergodicity*. In some cases, this inequality is strong enough to be of use in bounding the convergence rate of ergodic stochastic matrices. An example may be found in [6].

The matrix $\mathbf{M}^j$ describes the transitions between these cycles. Similarly to (9),

$$\mathbf{M}^j_{ab} = \frac{1}{l_a} \sum_{m=1}^{nl_a} \chi_m$$

where χ_m are i.i.d. random variables such that $Pr\{\chi_m = \omega^r\} = \frac{1}{s}\frac{|C_b|}{2^n}$, $r = 0, \ldots, s-1$. Calculating the first two moments:

$$\mu = \mathbf{E}(\mathbf{M}^j_{ab}) = 0$$

$$\sigma^2 = \mathbf{Var}(\mathbf{M}^j_{ab}) = \frac{1}{l_a^2} nl_a \frac{|C_b|}{2^n} = o(2^{-\frac{n}{2}}) \tag{14}$$

we can see that the norm (and therefore also spectral radius) of $\mathbf{M}^j$ is exponentially small, similarly to (12):

$$||\mathbf{M}^j|| \leq k_j^2 \sigma = o(n^2 2^{-\frac{n}{4}}). \quad \blacksquare$$

We have shown that $\mathbf{I}_\delta$ almost always perturbs the second unit eigenvalue of $\mathbf{P}_o$ with a derivative of $-n$. This gives an estimate of $1 - n\delta$ for $|\lambda_1|$ in the cases where δ is extremely small (eg. exponentially small). For larger values of δ, an accurate estimate for $|\lambda_1|$ may involve higher order derivatives. We have had no success in calculating these higher order derivatives. However, numerical results seem to indicate that the estimate $1-n\delta$ for $|\lambda_1|$ is valid for $\delta = o(\frac{1}{n})$ (polynomially small noise), which would mean that rapid mixing is obtained.

3. Conclusion

Two properties of random mappings on $N = 2^n$ elements are essentially responsible for the apparent rapid mixing property: the small (logarithmic) number of connected components in the transition graph and the lack of correlation between these transitions and the topology of the n-dimensional hypercube. Many forms of noise operators are possible, the one used here ($\mathbf{I}_\delta$) naturally derived from an underlying boolean processor network model, where each processor may err independently. It seems that any other noise operator uncorrelated with the transition graph should result in similar rapid mixing behavior. This phenomenon may find applications in cryptography.

Acknowledgement. We would like to thank Daniel Lehmann and Yuri Kifer for helpful discussions during this work.

References

[1] B. Bollobás, *Random Graphs,* Academic Press Inc., London, 1985.

[2] T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators,* Springer-Verlag, New York, 1982.

[3] A. W. Marshall and O. Olkin, *Inequalities: Theory of Majorization and its Applications,* Academic Press, New York, 1979.

[4] L. R. Mutafchiev, Limit theorems concerning random mapping patterns, *Combinatorica* **8**(1988), 345–356.

[5] E. Seneta, *Non-negative Matrices and Markov Chains (Second Edition),* Springer-Verlag, New-York, 1981.

[6] P. E. Wright, Statistical complexity of the power method for Markov chains, *Journal of Complexity* **5**(1988), 119–143.

Joel Friedman

Department of Computer Science,
Princeton University,
Princeton, NJ 08544, USA

Craig Gotsman

Department of Computer Science,
Hebrew University,
Jerusalem 91904, Israel

Eli Shamir

Department of Computer Science,
Hebrew University,
Jerusalem 91904, Israel