Spectral gap and cut-off in Markov chains

M. Grinfeld and P. Knight
Department of Mathematics
University of Strathclyde
Livingstone Tower
26 Richmond Street
Glasgow G1 1XH, UK

A. Olde Daalhuis
Department of Mathematics and Statistics
King's Buildings
University of Edinburgh
Edinburgh EH9 3JZ, UK

February 17, 1999

Abstract

In this paper we consider an example of a family of Markov chains with a spectral gap and show that it exhibits $O(n)$ cut-off in the sense of Diaconis. We argue that this, and bandedness, is what lies at the heart of the $O(n)$ phase transition in the k-SAT problem.

1 Introduction

Consider a collection $\{A_n\}$ of $n \times n$, transition matrices corresponding to a family of Markov chains with n states, $n = 1, \ldots$. Assume that each of these has a stationary distribution $\boldsymbol{\sigma}_n$ (the eigenvector of A_n corresponding to the eigenvalue 1). Let $f(n)$ be an increasing function of n and, for an arbitrary vector $\boldsymbol{\tau}_n$ assume that the limit

$$g(c) = \lim_{n\to\infty} \|\boldsymbol{\tau}_n A^{cf(n)} - \boldsymbol{\sigma}_n\| \tag{1.1}$$

exists for all $c > 0$. If we have that

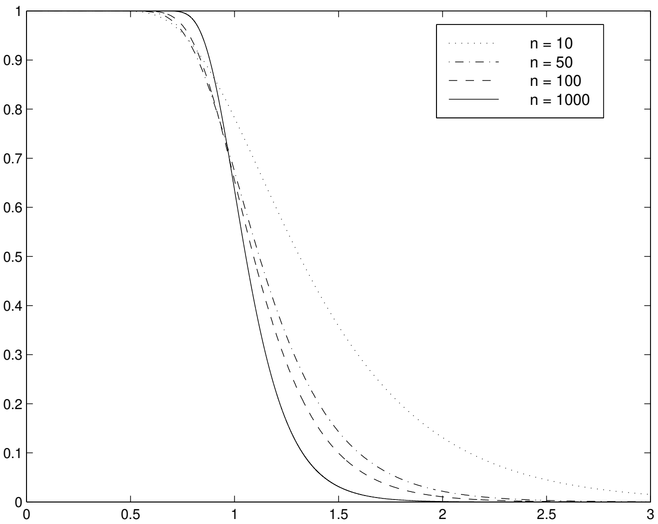
$$g(c) = \begin{cases} \alpha \neq 0 & \text{for } c < c_* \\ 0 & \text{for } c > c_*, \end{cases}$$

we say that the family of Markov chains has a cut-off of order $f(n)$ with critical value c_*.

For example, in riffle shuffling of cards [2, 3], in which case $\boldsymbol{\sigma}_n[k] = 1/n$, $k = 1, \ldots, n$, the cut-off is $O(\log n)$ with $c_* = 3/2$. For other examples we refer the reader to [2, 3, 9]; the last reference also contains an interesting discussion of the suitable norm in which (1.1) is to be considered. As a particularly simple example of a family of Markov chains exhibiting cut-off, consider the coupon collector's problem.

Example 1. As our transition matrix, we take a stochastic $n \times n$ matrix A, such that for $i = 1, \ldots, n$, $A[i,i] = i/n$, $A[i, i+1] = 1 - A[i,i]$. This is the coupon collector problem [1]. There is no spectral gap here and the cut-off is $O(n \log n)$ with $c_* = 1$. We illustrate the cut-off in Figure 1.

Figure 1: Cut-off in the coupon collector problem.



In the examples we consider in this paper, once we have established that there is a cut-off, it is often relatively easy to pin-point precisely. The Markov chains we consider all have an absorbing state, and the cut-off is the time it takes to arrive in the final absorbing state, starting from the first state. Elementary theory of Markov chains [10] tells us that if

$$A_n = \begin{bmatrix} Q_n & r_n \\ 0 & 1 \end{bmatrix} \tag{1.2}$$

then the expected value of the time taken to arrive in the absorbing state, given that we started in state k, is the kth component of the vector

$$\tau^{(n)} = (I - Q_n)^{-1} e, \tag{1.3}$$

where e is a vector of ones. Thus the cut-off, if it occurs, can be found by looking at $\tau_1^{(n)}$ as $n \to \infty$. In the coupon collector's problem the computations are trivial and we find that

$$\tau_1^{(n)} = \sum_{i=1}^{n-1} \frac{n}{i}.$$

Proving the existence of cut-off in any particular case, however, is highly nontrivial, involving as it does tools of probability theory, combinatorics, asymptotics (as in th example here), group representation theory and so on. It is an even harder task to relate the order of the cut-off to the structure of the family of transition matrices. In [3] heuristic connections are made between the existence of the cut-off, the multiplicity of the second eigenvalue of the transition matrix and symmetry (it should be clear from the coupon collector's problem that neither is a necessary condition for cut-off). In the present paper we exhibit a family of upper-triangular transition matrices with a very simple spectrum that has an $O(n)$ cut-off. It differs from the coupon collector's problem in that it has a spectral gap.

More precisely, we consider transition matrices with real, positive spectrum. Let $1 = \lambda_1(n) \geq \lambda_2(n) \geq \cdots \geq \lambda_n(n) \geq 0$ be the eigenvalues of such an $n \times n$ member of the family. The family has a spectral gap if

$$\lim_{n \to \infty} (1 - \lambda_2(n)) \neq 0.$$

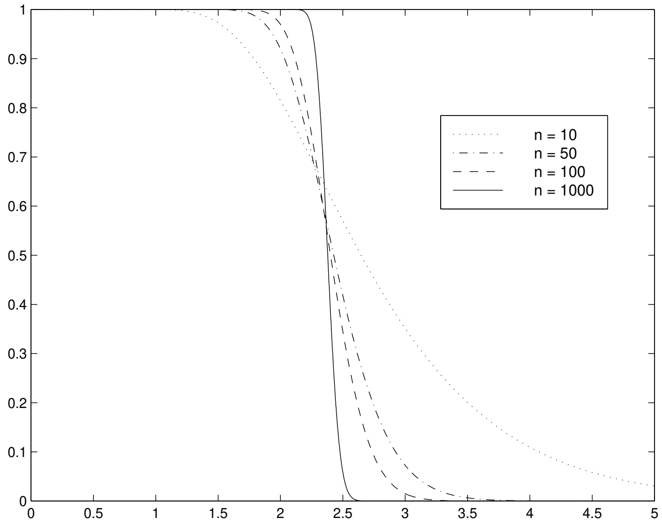
Clearly, in the coupon collector's problem there is no spectral gap.

Our interest in seeking families of Markov chains with $O(n)$ cut-off stems from the observation that an important problem in computer science, the k-SAT problem, can, on the one hand, be given a Markov chain formulation, and exhibits, on the other hand, what is called a "phase transition" [8], which is basically nothing else but a cut-off that turns out to be $O(n)$. These issues are considered in detail in section 3.

2 An example of a Markov chain with a spectral gap

Example 2. Consider stochastic $n \times n$ matrices A, such that for $i = 1, \ldots, n-1$, $A[i,i] = b*i/(n-1)$ $0 < b < 1$, $A[i,i+1] = 1 - A[i,i]$ and $A[n,n] = 1$. This is an upper triangular matrix with a spectral gap $1 - b$. This example shows a cut-off at $O(n)$, as illustrated in Figure 2 with $b = 7/8$.

Figure 2: Cut-off with $b = 7/8$.



Theorem 2.1 *For $0 < b < 1$, the family of transition matrices A in Example 2 exhibit $O(n)$ cut-off. Furthermore,*

$$c_*(b) = -\frac{\ln(1-b)}{b}. \tag{2.1}$$

Proof.

Defining $\boldsymbol{\pi}$ to be the n-vector $[1\,0\,\ldots\,0]$, we are clearly interested in the behaviour $(\boldsymbol{\pi}A^k)[n]$ as k increases. This expression can be computed explicitly.

Let $\{\lambda_i\}$, $i = 1, \ldots n$ be the eigenvalues of A arranged in decreasing order; thus $\lambda_1 = 1$ $\lambda_2 = b$ for all n etc. Note by the way that as n increases, $\lambda_2 - \lambda_3 \to 0$.

if we let ψ_i be the left (row) eigenvectors of A and ϕ_i be the respective right (column) eigenvectors, $i = 1, \ldots n$ $((\psi_i, \phi_j) = \delta_{ij})$, we easily obtain

$$\boldsymbol{\pi}A^k[n] = \sum_{i=1}^{n} \lambda_i^k \phi_i[1]\psi_i[n].$$

Computing the eigenvectors, this expression can be written in the following remarkable way:

$$\boldsymbol{\pi}A^{n+s-1}[n] = p(n,\, b) \sum_{j=0}^{s} B(j,\, n,\, b), \tag{2.2}$$

where

$$p(n,\, b) = \prod_{i=1}^{n-1} \left(1 - \frac{ib}{n-1}\right)$$

and

$$B(j,\, n,\, b) = \sum_{k=1}^{n-1} (-1)^{n-1-k} \frac{k^{n-2}}{(k-1)!(n-k-1)!} \left(\frac{bk}{n-1}\right)^j.$$

Since $p(n,\, b)$ is independent of s, to understand the asymptotic behaviour of $\boldsymbol{\pi}A^{n+s-1}[n]$ we need to understand the behaviour of $B(j,\, n,\, b)$ as a function of j (which we can consider to be real). In fact, we shall show that $B(j,\, n,\, b)$ has a unique maximum at $j = j_*$. Furthermore, we will show that

$$j_* \sim \left(-1 - \frac{\ln(1-b)}{b}\right) n + o(n),$$

so that the order of the cut-off is then precisely $O(n)$ and (see (2.2))

$$c_*(b) = \left(-\frac{\ln(1-b)}{b}\right).$$

To that end, we use Hankel's contour integral ((1.12) in Chapter 2 of [11])

$$k^{n+j-2} = \frac{\Gamma(n+j-1)}{2\pi i} \int_{-\infty}^{(0+)} e^{kz} z^{-n-j+1} dz,$$

and obtain after some manipulation

$$B(j, n, b) = \left(\frac{b}{n-1}\right)^j \frac{\Gamma(n+j-1)}{\Gamma(n-1)} I(j, n),$$

where

$$I(j,n)=\frac{1}{2\pi i}\int_{-\infty}^{(0+)} e^{z}\,(e^{z}-1)^{n-2}\,z^{-n-j+1}dz.$$

We write $j=\alpha n+\beta$, and obtain the integral representation

$$I(\alpha n+\beta,n)=\frac{1}{2\pi i}\int_{-\infty}^{(0+)}\frac{e^{z}z^{1-\beta}}{(e^{z}-1)^{2}}e^{-nf(z)}dz,$$

where

$$f(z)=(1+\alpha)\ln z-\ln(e^{z}-1).$$

The saddle points satisfy

$$f'(z)=\frac{1+\alpha}{z}-\frac{1}{1-e^{-z}}=0.$$

Hence, $f(z)$ has a unique saddle point $z_\alpha>0$. The previous line can be written as

$$\frac{z_\alpha}{1-e^{-z_\alpha}}=1+\alpha. \tag{2.3}$$

We let $n\to\infty$ and obtain from the method of steepest descents (see (7.05) in Chapter 4 of [11]):

$$I(\alpha n+\beta,n)\sim\frac{1}{\sqrt{2\pi n}}e^{-n(\alpha\ln z_\alpha-z_\alpha+\ln(1+\alpha))}e^{-z_\alpha}z_\alpha^{-\beta}(1+\alpha)^{3/2}(z_\alpha-\alpha)^{-1/2}.$$

Hence,

$$\begin{aligned}B(\alpha n+\beta,n,b)\sim\left(\frac{b}{n-1}\right)^{\alpha n+\beta}&\frac{\Gamma\left(n(1+\alpha)+\beta-1\right)}{\Gamma(n-1)}\frac{1}{\sqrt{2\pi n}}e^{-n(\alpha\ln z_\alpha-z_\alpha+\ln(1+\alpha))}\\&\times e^{-z_\alpha}z_\alpha^{-\beta}(1+\alpha)^{3/2}(z_\alpha-\alpha)^{-1/2}.\end{aligned}$$

Consequently,

$$\frac{B(\alpha n+1,n,b)}{B(\alpha n+0,n,b)}\sim\frac{b\left(n(1+\alpha)-1\right)}{(n-1)z_\alpha}\sim\frac{(1+\alpha)b}{z_\alpha},$$

as $n\to\infty$. By definition of j^*, we are looking for α such that

$$\frac{B(\alpha n+1,n,b)}{B(\alpha n+0,n,b)}\sim 1.$$

Hence, we need α such that $z_\alpha=(1+\alpha)b$. We substitute this result into (2.3) and deduce that

$$\alpha=-\frac{\ln(1-b)}{b}-1.$$

□

2.1 Remarks

As for our previous example, once we have identified that a cut-off occurs it is easy to pin point by finding the mean arrival time in the absorbing state. Defining Q_n and $\tau^{(n)}$ as in (1.2) and (1.3) we find that

$$\tau_1^{(n)} = \sum_{i=1}^{n} \frac{n}{n - ib}$$

and it is simple to establish that

$$\tau_1^{(n)} = -\frac{n \log(1-b)}{b} + o(n).$$

Of course, in this case we have already established this result in the proof of the existence of a cut-off.

It has been observed [6] that the matrices in our examples can also arise in finite-difference discretizations of certain hyperbolic PDEs, and that the order of the cut-off can be inferred directly from the PDE. For example, the stochastic matrix in the coupon-collectors problem can be obtained by discretizing the problem

$$u_t = -(1-x)u_t + u.$$

Again, we are only able to identify the location of a cut-off with this technique if we already know that the cut-off exists. Whether this technique can be extended to more general problems involving cut-offs is an open question.

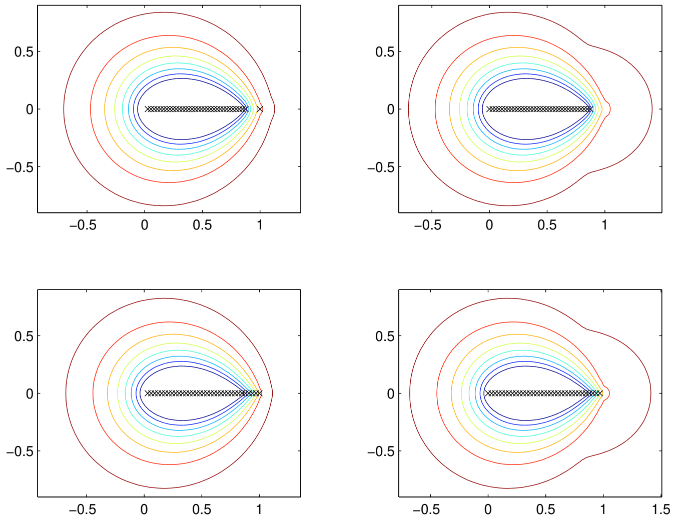
The matrices we consider in this paper are all highly nonnormal, and it is natural to ask whether pseudospectra [13] play a role in the cut-off phenomenon since it is well known that the ϵ-pseudospectrum of a matrix can explain transient growth in matrix powers, whereas the spectrum can only be relied on to predict asymptotic behaviour. In Figure 3 we plot the boundaries of the ϵ-pseudospectra ($\epsilon = 10^{-8}, 10^{-7}, \ldots, 10^{-1}$) for matrices associated with our two examples (the top two pictures are for example 2, the bottom two for example 1). The left two pictures show the pseudospectra for $n = 33$, and the right two pictures show the pseudospectra of the same matrices with the eigenspace associated with the eigenvalue 1 removed (this is to highlight the potential of transient growth in the modes associated with the non-dominant eigenvalues).

While these pictures are suggestive of transient growth, we have not been able to use the pseudospectra to pin-point the cut-off. We note that the pseudospectra for these examples are similar despite the fact that the cut-off appears in radically different places.

3 The k-SAT problem and Markov chains

The k-SAT problem, a good introduction to which is in [8], is as follows: Assume that we are given a set of n literals (Boolean variables) $\{s_1, \ldots, s_n\}$. Choosing each time k of these, we create m clauses (conjunctions). The question is to find an assignment of these variables such that the disjunction of all the clauses is true.

Figure 3: Pseudospectra of matrices with cut-off.



Example Find values of P, R, Q such that

$$(P \vee \neg Q) \wedge (Q \vee R) \wedge (\neg P \vee \neg R) \tag{3.1}$$

is true. Here $k = 2$, $n = 3$, $m = 3$. In this case the possible assignments are: P and Q being true while R is false and vice versa.

There is a very intriguing observation. Define $\mathcal{P}(n, k, m)$ to be the probability that a k-SAT problem in n literals having m clauses has a solution. Then the limit

$$f(c, k) = \lim_{n\to\infty} \mathcal{P}(n,\, k,\, nc)$$

exists for all $c > 0$. Furthermore, the function $f(c,\, k)$ is piecewise constant, taking the value 1 for $c \leq c_*(k)$ and 0 for $c > c_*(k)$. It has been proved by Goerdt [5] that this situation obtains for $k = 2$, and that $c_*(2) = 1$. There are strong indications that $c_*(3) = 4.25$, but it was only recently proved that for $k \geq 3$ a "phase transition" of this sort occurs at all [4].

A different way of looking at it is as follows. Define $m(n, k) = \{m \,|\, \mathcal{P}(n, k, m) = 1/2\}$. Clearly,

$$c_*(k) = \lim_{n\to\infty} \frac{m(n,\, k)}{n}.$$

In other words, in this problem $m(n, k)$ scales with n. Obviously, the problem is hardest to solve when the probability that the problem has precisely one solution is at a maximum.

3.1 Mapping onto a vertex painting problem

An assignment of the literals can be thought as an integer in the set $\{0, \ldots, 2^n - 1\}$ or alternatively as a vertex in the unit cube $I_n \subset \mathbb{Z}^n$. Obviously, $|I_n| = 2^n$.

The clauses C_i, $i = 1, \ldots, m$ can be regarded as functions $C_i : I_n \mapsto \{0, 1\}$. For each $i = 1, \ldots, m$ define

$$S_i = \{x \in I_n \mid C_i(x) = 0\}.$$

In other words, S_i is the set of the vertices on which C_i is false. For example, if $n = 4$, $C_1 = s_1 \vee s_2 \vee \neg s_3$, S_1 can be seen as either the set $\{2, 3\}$ or the set of vertices $(0010,\ 0011)$.

It easy to see that for k-SAT problem $|S_i| = 2^{n-k}$ for all i, and that there are $2^k C_n^k$ such distinct sets. Furthermore, it is clear that each set S_i is an $(n-k)$-face of the cube I_n. Hence solving a k-SAT problem in n literals having m clauses is the same as taking an n-cube with white vertices, m times choosing at random an $(n-k)$-face and applying, say, black paint to the set of its vertices. The number of solutions of a k-SAT problem is then precisely the cardinality of the set of unpainted vertices.

3.2 The associated Markov chain

Having established the equivalence of the k-SAT problem and the vertex painting problem, we will consider the latter more closely. More precisely, let a (l, n)-**problem** be the problem of understanding the statistics of painting the vertices of an n-cube by applying paint at each time to all the vertices contained in an l-face of the cube. Clearly, the k-SAT problem is an $(n-3, n)$-problem.

The (l, n)-problem (apart from the $l = 0$ case which is trivial) seems to be quite hard for the following reason: it is not easy to work out the combinatorics of the intersections of the sets S_i.

It is clear that the k-SAT problem can be formulated as a Markov chain problem on the power set of $(n-k)$-faces of an n-cube. However, the resulting Markov chain with $2^{2^k C_n^k}$ states is obviously unmanageable. Hence one may try to lump the state space [10, Chap. 6]. Note that in the case of Ehrenfest urns [9], the 2^n states are lumpable to n. In [7] it is shown that instead of considering an $2^{2^k C_n^k} \times 2^{2^k C_n^k}$ transition matrix we can consider one with $O(2^k C_n^k)$ states, i.e. one that grows as n^k with n. To be precise, the following theorem is proved.

Theorem 3.2 *The $2^{2^k C_n^k}$ states of the k-SAT problem can be lumped into $N = 2^k C_n^k - C_n^k + 1$ sets $R_1, \ldots R_N$, where R_i, $i = 1, \ldots, N-1$ contains all possible collections of precisely i $(n-k)$-faces of an n-cube that do not cover the entire cube, and R_N contains all the collections of k faces that cover the whole cube.*

In the case of $k = 2$, $n = 3$ this allows us to lump from 4096 states to just 10. It should be clear that it is not possible to lump any more. Clearly, $B_{ii} = i/(2^k C_n^k)$, $i = 1, \ldots, N-1$, $B_{NN} = 1$, so that we have a gap of $1/2^k$, as promised. It is non-trivial to compute the remaining terms. For the case of $k = 2$, $n = 3$ the resulting matrix 10×10 can be worked out by a straightforward algorithm and is shown below.

$$\begin{bmatrix} 1/12 & 11/12 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/6 & 5/6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 81/110 & 0 & 0 & 0 & 0 & 0 & 3/220 \\ 0 & 0 & 0 & 1/3 & 95/162 & 0 & 0 & 0 & 0 & 13/162 \\ 0 & 0 & 0 & 0 & 5/12 & 143/342 & 0 & 0 & 0 & 113/684 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 161/572 & 0 & 0 & 125/572 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7/12 & 4/23 & 0 & 67/276 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & 1/12 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that the entries in the last column form the solution of a problem that is interesting in its own right. We formulate it in terms of the k-SAT problem. What is the probability that a k-SAT problem with m clauses in N Boolean variables has a solution, given that a) all clauses are different and b) every $m-1$-clause subproblem has a solution?

There is circumstantial evidence that the entries in the last column also exhibit a cut-off for large N. By making a number of assumptions on the entries, one can show, using (1.2) and (1.3), that this cut-off is $O(N)$. Further research is necessary to add mathematical rigour to this observation.

References

[1] D. Aldous. *Probability Approximations via the Poisson Clumping Heuristic*. Springer-Verlag, New York 1989.

[2] P. Diaconis. Applications of non-commutative Fourier analysis to probability problems. LNM **1362**, Springer-Verlag, Berlin 1989, pages 53–100.

[3] P. Diaconis. The cutoff phenomena in finite Markov chains. PNAS **93**:1659–1664 (1996).

[4] E. Friedgut. Necessary and sufficient conditions for sharp thresholds of graph proprties, and the k-sat problem. Preprint, available from `http://www.ma.huji.ac.il/~ehudf/` (1998).

[5] A. Goerdt. A threshold for unsatisfiability. J. Comp. System Sci. **53**:469–486 (1996).

[6] D. F. Griffiths. Private communication. 1998.

[7] M. Grinfeld and P. A. Knight. Weak lumpability in the k-SAT problem. Preprint, 1999.

[8] B. Hayes. Can't get no satisfaction. American Scientist **85**:108-112 (1997).

[9] G. F. Johnsson and L. N. Trefethen. A numerical analyst looks at the "cutoff phenomenon" in card shuffling and other Markov chains. Technical Report NA-97/12, Numerical Analysis Group, Oxford University Computing Laboratory, 1997.

[10] H. G. Kemeny and J. L. Snell. *Finite Markov Chains*. van Nostrand, Princeton, NJ 1960.

[11] F. W. J. Olver. *Asymptotics and Special Functions*. Academic Press, New York 1974. Reprinted by AK Peters, Wellesley 1997.

[12] J. S. Rosenthal. Convergence rates for Markov chains. SIAM Review **37**:387–405 (1995).

[13] L. N. Trefethen. Pseudospectra of matrices. In *Numerical Analysis 1991, Proceedings of the 14th Dundee Conference*, D. F. Griffiths and G. A. Watson editors, volume 260 of *Pitman Research Notes in Mathematics*. Longman Scientific and Technical, Essex, UK, 1992, pages 234–266.