

Sharp thresholds for stopping times of Markov Chains (I)

Gabriel Istrate^a

^a *eAustria Research Institute*
Bd. V. Pârvan 4, cam 045B,
Timișoara, RO-300223, Romania

Abstract

Motivated by the problem of characterizing *cutoff phenomena in Markov chains*, and that of proving *sharp thresholds* for dynamic graph models, we investigate threshold properties for stopping times of Markov Chains. We propose an approach to this problem, extending results from the theory of thresholds of random graphs to stopping rules in Markov Chains. We give two such extensions:

1. We show that every monotonic stopping time has a threshold.
2. We give an extension (based on group representation theory) of Russo's lemma to stopping times.

Key words: Markov chains, cutoff properties, stopping times.

1 Introduction

Investigating *threshold properties of combinatorial optimization problems* [Fri05] has been a popular topic for recent research. The seminal result of Friedgut and Bourgain [Fri99] has allowed the classification of threshold properties in a large class of problems, such as *that of boolean constraint satisfaction problems (CSP)* [CD04,Ist05].

Threshold properties in CSP are usually studied using variants of probabilistic models from random graph theory [JLR00] (see also [Bol85]): the *binomial graph model* and the *uniform graph model*. An approach that, in the case of

Email address: gabrielistrate@acm.org (Gabriel Istrate).

random graphs, often provides stronger results is via *graph process models*: we start with a graph with no edges and repeatedly add edges according to a certain criterion. Motivated by the Statistical Physicists literature [NBW06], a variety of such processes has been recently investigated rigorously [Dur06]. A natural question is whether the Friedgut-Bourgain approach provides hints for the analysis of such processes as well.

A second motivation for investigating threshold properties of stopping times comes from the theory of Markov chains: Let M be a random walk on a set X with initial state $x_0 \in X$ and transition matrix Q . Assuming that M is irreducible and aperiodic, it is a classic theorem of the theory of stochastic processes that M has a unique stationary distribution π and, furthermore, π_t , the distribution obtained by running the Markov Chain for t steps converges (as $t \rightarrow \infty$) to π . A favorite theme of research in Theoretical Computer Science has been to estimate the *mixing time* of such chains, i.e. the number of steps t needed to make the distribution π_t be "close" to π (for a recent survey of the techniques used to prove such results see [Ran03]). A complementary way to describe the convergence of π_t to π is via the concept of a *cutoff*, studied and greatly popularized by Diaconis. A family of Markov chains displays a cutoff with respect to a distance d on probability distributions when $d(\pi_t, \pi)$ has (when $t \rightarrow \infty$) a "sudden jump" to 0, in an interval whose width is much smaller than the mixing time. Cutoffs are most often studied with respect to the total variation distance d_{TV} ; an example of a chain displaying a cutoff (with respect to this distance) is the random walk corresponding to top-to-random card shuffling. The proof [DS81] involves Fourier analysis on finite groups. There are other chains (e.g. the random walk on Z^n) for which the random walk does *not* display a cutoff. The characterization of those walks that have a cutoff is not known and was stated as an open problem in [Dia96] (see also [SC04]).

Cutoffs can be studied for metrics other than the total variation distance. Such distances include, for instance the L^p -norms and the so-called *separation distance* [Dia88]. Peres has conjectured [GNRS04] that the existence of a cutoff can be characterized using two parameters: the spectral gap and the mixing time. More precisely, a necessary and sufficient condition for the existence of a cutoff is that the product of the spectral gap and the mixing time goes to infinity. Chen and Saloff-Coste [CSC08] have proved Peres's conjecture for reversible Markov chains under L^p -distance, $p > 1$. However, for $p = 1$ the conjecture fails [CSC08]. For this distance (as well as the separation distance) a characterization of Markov chains with a cutoff is still an open problem. Similar concentration properties have been previously investigated by Aldous for the cover time [Ald94], who gave a complete characterization of random walks with a threshold for the cover time.

Motivated by these two problems in this note we propose study instead thresh-

old properties of (appropriately chosen) *stopping times*. We are, of course, inspired by the result of Friedgut-Bourgain characterizing monotone properties that have a sharp threshold [Fri99]. Our ultimate aim is to bring Fourier-theoretic methods similar to those used in the Friedgut-Bourgain result to the problem of characterizing sharp thresholds of stopping times.

The goal of this paper is to take some preliminary steps in this direction. In particular, we manage to recover a Fourier-theoretic interpretation of Russo's formula due to Kahn, Kalai and Linial [KKL88] for certain classes of stopping times.

2 Preliminaries

Definition 1 *Let G be a finite group and let X be a finite set. An action of a G on X is a mapping from $G \times X \rightarrow X$ which will be denoted by $(s, x) \rightarrow s \cdot x$. The orbits of the action are the equivalence classes of the equivalence relation $x \sim y \iff (\exists s \in G) s \cdot x = y$. G acts transitively on X if there is a single orbit of \sim . In this case the pair (G, X) is called a homogeneous space.*

When G acts transitively on X one can naturally represent X as follows: fix $x_0 \in X$. Define the isotropy subgroup of x_0 to be the group $N = \{s \in G \mid s \cdot x_0 = x_0\}$ (one can verify that the definition does not depend on the particular choice of x_0). We can, then, view the group G as acting on the coset space G/N , and there is an isomorphism between X and G/N that respects the action of G .

Definition 2 *Given a homogeneous space (G, X) , an initial state $X_0 \in X$ and a probability distribution P on G , one can define a stochastic process on X specified as follows:*

- *Start at state X_0 .*
- *If X_t is the state of the process at time t then sample an element $s \in G$ according to P and define $X_{t+1} = s \cdot X_t$.*

The following simple result, Lemma 3 in Chapter 3F of [Dia88] motivates the choice of our framework:

Lemma 1 *The stochastic process in Definition 2 is a Markov chain M on X with transition matrix $Q_x(y) = P(xNy^{-1})$.*

Definition 2 is a generalization of *random walks on groups*. Such random walks can be defined by having group G acts on itself in the usual manner $(s, t) \rightarrow s \odot t$, where \odot is the group multiplication operation.

In what follows, Q^{*n} is the convolution of Q with itself n times, corresponding to taking exactly n jumps in the random walk. Under suitable conditions, as $n \rightarrow \infty$, Q^{*n} converges to the uniform distribution U , $U(s) = \frac{1}{|G|}$ on G .

The following are the two metrics we are going to be concerned with:

Definition 3 Let W be a probability on the finite group G .

(1) The n -step total variation distance is defined as

$$d_{TV}(n) = \frac{1}{2} \cdot \sum_{s \in G} \left| \frac{1}{|G|} - W^{*n}(s) \right|$$

(2) Define the n -step separation distance by

$$s(n) = |G| \cdot \max_s \left\{ \frac{1}{|G|} - W^{*n}(s) \right\}.$$

3 Trajectories on groups

Definition 4 Let $T(G) = G^*$ be the set of trajectories in G (when G is clear we will drop it from the notation, and write T instead of $T(G)$). The image $i(w)$ in X of trajectory $w \in T$ is the sequence $i(w) = X_0 X_1 \dots X_n \dots$ obtained starting at given node X_0 and repeatedly multiplying by elements of w .

If x is a prefix of a trajectory $w \in T$, let $\text{yield}(x) \in X$ be the state of the random walk after making the transitions specified by string x .

Definition 5 (1) Define partial order \lesssim on G^* by

$$x \lesssim y \iff x \text{ is a prefix of } y.$$

(2) For $r \in [0, 1]$, a coin toss with success probability r is the monotonically increasing function $f_r : [0, 1] \rightarrow \{0, 1\}$

$$f_r(x) = \begin{cases} 0, & \text{if } x < 1 - r, \\ 1, & \text{otherwise.} \end{cases}$$

We will denote by $\mathcal{CT} = \{f_r\}_{r \in [0, 1]}$ the set of such coin tosses, by $\mathbf{1}$ the "sure coin toss" $f_1(\cdot)$, and by $\mathbf{0}$ the "unlucky coin toss" $f_0(\cdot)$.

Definition 6 (1) A probabilistic stopping time for random walk M is function $F : T \rightarrow \mathcal{CT}$ such that

$$\text{if } (x \lesssim y) \text{ and } F(x) \equiv \mathbf{1} \text{ then } F(y) \equiv \mathbf{1}. \quad (1)$$

By slightly abusing the notation we will write, for $x \in T$, $F(x) = r$ instead of $F(x) = f_r$. Also define $\mu : T \rightarrow [0, 1]$ to be the probability that F does not accept at the last coin toss on path x .

A (deterministic) stopping time for the Markov chain is a probabilistic stopping time such that $F(x) \in \{\mathbf{0}, \mathbf{1}\}$ for every $x \in T$.

(2) A probabilistic stopping time is consistent iff

$$(\mu(x) > 0) \Rightarrow (\mu(x \cdot e) > 0) \text{ and } (\mu(x) < 1) \Rightarrow (\mu(x \cdot e) < 1)$$

(in other words we cannot turn the accepting/rejecting probabilities to 0/1 by multiplying with the neutral element of G). In particular, a deterministic stopping time is consistent iff $F(x) = F(x \cdot e)$ for all $x \in T$.

Example 1 (Thresholds of monotonically increasing properties as stopping times):

Let $A \in \{0, 1\}^n$ be a monotonically increasing property. Consider a Markov chain M with state space $\{0, 1\}^n$, that starts in state $\mathbf{0}$ and, at each step flips a random bit of the current state. M can be viewed in the context of the natural action A_n of $G = \mathbf{Z}_2^n$ onto the n -dimensional hypercube $X = \mathbf{Z}_2^n$, defined by $x \cdot y = x \oplus y$, the bitwise XOR of x and y . It is easy to see that $N = \{\mathbf{0}\}$. The moves in M correspond to multiplying by one of the generators \mathbf{e}_i .

One can define a stopping time T_A for M specified in the following way: Let X_0, \dots, X_t be the trajectory seen so far. Accept as soon as $Y_t = X_0 \vee X_1 \vee \dots \vee X_t$ (where the \vee is taken bitwise) is in A . It is easy to see that the stopping time T_A corresponds as well to the following process: start with an empty set of edges; at each step add a random edge (with repetitions); stop when the resulting graph belongs to A .

Definition 7 A coupling for random walk M is a stochastic process $C_{M,t} = (X_t, Y_t)$ on $G \times G$ such that

- (1) X_t has distribution π_t .
- (2) Y_t has uniform distribution U .
- (3) $X_t = Y_t$ implies $X_{t+1} = Y_{t+1}$.

Example 2 (Couplings as stopping times):

Couplings naturally define deterministic stopping times on $G^2 := G \times G$ by taking, for $t \in T(G \times G)$

$$C_M(t) = \begin{cases} 1, & \text{if } t \in \Delta_{G \times G}, \\ 0, & \text{otherwise.} \end{cases}$$

For Markov chains the existence of a total variation cutoff can be related to to the properties of stopping times associated to couplings using the following result from [Gri75] (see also [Lin92]):

Proposition 1 *If C_M is a coupling form random walk M then for all $n \in \mathbf{N}$*

$$\Pr[X_n \neq Y_n] \leq d_{TV}(n). \quad (2)$$

For every random walk M there exists an associated maximal coupling C_M for which equation (2) holds with equality.

Definition 8 *A strong uniform time F is a stopping time F such that, for every $k < \infty$,*

$$\text{Prob}_{|x|=k}[F(x, r_{|x|}) = 1, X_k = s] \text{ is constant in } s. \quad (3)$$

Relating Markov chain cutoffs to stopping times is also possible for the separation distance $s(n)$ because of the following result (see [AD87] or [Dia88], Theorem 4 on pp. 76):

Example 3 (Strong Uniform Times): *If F is a strong uniform time for the random walk generated by Q on G then for all k*

$$s(k) \leq \text{Prob}[F > k]. \quad (4)$$

Conversely, for every random walk there is a probabilistic strong uniform time such that (4) holds with equality.

These two observations suggest an approach to the problem of characterizing Markov chains cutoffs: **study threshold properties of associated stopping times** by extending of the Friedgut-Bourgain approach.

4 Stochastic processes on trajectories

Definition 9 *Let G be a group acting on set X . A G -process on X is a function $W : G^* \rightarrow [0, 1]$ that is a G -martingale, i.e. a function such that $W(\lambda) = 1$ and for all $x \in G^*$*

$$W(x) = \sum_{r \in G} W(xr).$$

Our main example of a G -process is specified as follows:

Example 4 (“random walk”): Let G be a group acting on set X and let Q be a probability measure on G . Q defines a G -process as follows: $W(\lambda) = 1$ and, for $x \in G^*$ and $r \in G$,

$$W(xr) = W(x) \cdot Q(r). \quad (5)$$

The previous example might fail to be ergodic because of *periodicity*. The usual alternative is the following:

Example 5 (“lazy random walk”): Let G be a group acting on set X and let Q be a probability measure on $G \setminus e$. Q defines a G -process as follows: $W(\lambda) = 1$ and, for $x \in G^*$ and $r \in G$,

$$\begin{cases} W(x \cdot e) = W(x) \cdot \frac{1}{|G|}, \\ W(x \cdot r) = W(x) \cdot \frac{(|G|-1)Q(r)}{|G|} \end{cases}$$

The previous examples motivate the following definition:

Definition 10 Define, for all x with $W(x) > 0$ and all $r \in G$, $W_x(r) = \frac{W(xr)}{W(x)}$. W is Markovian if for every r the quantity $W_x(r)$ (when defined) does not depend on x .

Couplings (when considered as processes on $G \times G$) are in general not Markovian processes. Only the *Markovian couplings* of [KR01] are Markovian in the sense of Definition 10:

Definition 11 A Markovian coupling for random walk M is a coupling $C_{M,t} = (X_t, Y_t)$ on $G \times G$ that is a Markov chain (when viewed as a stochastic process on $G \times G$).

Note however, that Markovian couplings can be substantially weaker than general ones [KR01].

Definition 12 A Markovian G -process is regular iff it is represented by a matrix P whose all diagonal elements $P_{i,i}$ are equal.

5 Threshold properties of stopping times

Definition 13 For stopping time F and $0 < \epsilon < 1$, define time t_ϵ via equation

$$\mu_{t_\epsilon}(F \neq \mathbf{1}) = \epsilon.$$

The interval $S_{\epsilon,T} = [t_{1-\epsilon}, t_\epsilon]$ is called the ϵ -scaling window of F . An assumption such as “the scaling window of F has property P ” will mean that for any constant $0 < \epsilon < 1$ $S_{\epsilon,T}$ has property P .

Definition 14 Stopping time F has a threshold if, for every $0 < \epsilon < 1/2$

$$\lim_{n \rightarrow \infty} \frac{t_\epsilon(n) - t_{1-\epsilon}(n)}{t_{1/2}(n)} < \infty.$$

Definition 15 Stopping time F has a sharp threshold if, for every $0 < \epsilon < 1/2$

$$\lim_{n \rightarrow \infty} \frac{t_\epsilon(n) - t_{1-\epsilon}(n)}{t_{1/2}(n)} = 0.$$

If, on the other hand the previous expression is bounded away from zero we say that F has a coarse threshold.

The result of Friedgut and Bourgain started from a property of monotonic sets that was a consequence of having a coarse threshold. The same property can be used for stopping times of Markov chains as well:

Observation 1 If stopping time has a coarse threshold then there exists a constant $\eta > 0$ that depends on ϵ but not on n and a family of points $t_C \in [t_{1-\epsilon}, t_\epsilon]$ such that

$$t_C \cdot \frac{d\mu_t(F)}{dt} \Big|_{t=t_C} < \eta,$$

Proof. Just as in the case of threshold of monotonic graph properties.

$$1 - 2\epsilon = \mu_{t_{1-\epsilon}} - \mu_{t_\epsilon} = (t_{1-\epsilon} - t_\epsilon) \cdot \frac{d\mu_t(F)}{dt} \Big|_{t=t_C},$$

for some $t_C \in [t_{1-\epsilon}, t_\epsilon]$.

Since the stopping time has a coarse threshold, there exists $D > 0$ such that $t_{1-\epsilon} - t_\epsilon > D \cdot t_{1/2} \geq D \cdot t_\epsilon$, in other words $t_\epsilon < \frac{1}{1+D} \cdot t_{1-\epsilon}$.

But then $t_{1-\epsilon} - t_\epsilon > \frac{D}{1+D} \cdot t_{1-\epsilon} \geq \frac{D}{1+D} t_C$, which implies that

$$t_C \cdot \frac{d\mu_t(F)}{dt} \Big|_{t=t_C} < \frac{1+D}{D} \cdot (1 - 2\epsilon),$$

therefore we can define $\eta = \frac{1+D}{D} \cdot (1 - 2\epsilon)$. □

In the case of random graphs Bollobás and Thomason have proved [BT86] that every monotonic graph property has a threshold. Such a property is known for a couple of other models, such as, for instance in the case of 1-dimensional geometric random graphs [McC04] (see also [McC98]) and, by a recent result of Graham and Grimmett [GG06], for monotone probability measures that satisfy the FKG inequality.

We now prove a theorem on the existence of thresholds for stopping times similar in spirit to the result of Bollobás and Thomason [BT86]:

Definition 16 *A stopping time is monotone if it has the following property: if x, y, z are trajectories and $T(y) = \mathbf{1}$ then for all x, z , $T(xy) = \mathbf{1}$. and $T(yz) = \mathbf{1}$.*

Theorem 1 *Let M be a Markov chain and let T be a monotone stopping rule. Then T has a threshold.*

Proof. Consider the following stopping time: run T up to time t_ϵ . If it has not accepted, run T for t_ϵ more time. From monotonicity we infer

$$t_{1-(1-\epsilon)^2} \leq 2t_\epsilon.$$

The rest of the proof closely mirrors the one in [BT86]. □

6 Russo's formula for stochastic dynamics on groups

We are interested in the following problem: given a monotone dynamics (F, W) , can one find an explicit expression for $\mu_t(F)$, the probability that simulating the Markov chain up to time t yields a sequence of moves that makes F accept? Rather than considering W in discrete time, we will assume that W (and the random walk M , when specializing our results to such walks) is a discrete-space continuous time stochastic process with jump probability parameterized by a Poisson process of rate 1 (it is well-known, see e.g. [Ald83] that, with respect to convergence and mixing time any ergodic Markov Chain in discrete time is equivalent to the corresponding continuous Markov Chain).

Definition 17 *Let (F, W) be a monotone dynamics on group G .*

- (1) $x \in T^*$ is F -minimal if $\mu_x > 0$ but $\mu_y < 1$ for all $y \lesssim x$. $\text{Min}(F)$ will represent the set of F -minimal trajectories.
- (2) For $t > 0$ T_t denotes the random variable on T^* specifying the set of transitions taken by M up to time t .

(3) For $t > 0$ and $x \in T^*$ denote by $\text{Path}_M(t, x)$ the event that $T_t = x$ and by $P_t(x) = \text{Prob}[\text{Path}_M(t, x)]$. In effect

$$P_t(x) = e^{-t} \frac{t^{|x|}}{|x|!} \cdot W(x).$$

(4) Define $U_F(x) = \prod_{y \prec x} (1 - \mu_y)$ to be the probability that F evaluates to 0 on all prefixes of x .

(5) Define $R_F(x)$ to be the probability that the first coin toss along x to yield a value of 1 is the one associated to x (and not one associated to one of its prefixes). Thus $R(x) = U_F(\hat{x}) - U_F(x)$, where \hat{x} is defined to be the prefix of x obtained by dropping the last element in T from x .

Claim 1 x is F -minimal if and only if $R_F(x) > 0$.

Proof. x is not minimal if and only if $\mu_x = 0$ or $\mu_{\hat{x}} = 1$. In the first case $U_F(x) = U_F(\hat{x}) \cdot 1 = U_F(\hat{x})$. In the second $U_F(\hat{x}) = 0$, therefore $R_F(x) = 0$.

The converse is just as easy: $U_F(x) = U_F(\hat{x}) \cdot (1 - \mu_x)$. If $R_F(x) = 0$, that is $U_F(x) = U_F(\hat{x})$, and $U_F(\hat{x}) > 0$ then $1 - \mu_x = 1$, that is $\mu_x = 0$. Otherwise we have $U_F(\hat{x}) = 0$, so $\mu_y = 1$ for some prefix y of x . By condition 1 it follows, therefore, that $\mu_{\hat{x}} = 1$ as well. A Markovian coupling of matrix P is *regular* if all diagonal elements $P_{i,i}$ are equal. \square

Theorem 2 Assume F is a monotone dynamics for random walk M with $F(\lambda) = 0$ (that is F is nontrivial). Then we have:

$$t \frac{d\mu_t(F)}{dt} = \sum_{x \in T^*} |x| \cdot P_t(x) \cdot R_F(x). \quad (6)$$

Observation 2 A problem with the previous formula is that it contains an infinite sum. However, if all x with $|x| > n$ are discarded from the sum on the right hand side of formula 6 the value of the sum changes by no more than $t \cdot \text{Prob}[Po(1, t) \geq n]$.

In particular, for any $\epsilon > 0$ there exists $C_\epsilon > 0$ such that the contribution in the sum of all x with $|x| > C_\epsilon t$ is at most ϵ .

Proof. The value of the sum for all x of length i is at most $i \cdot \text{Pr}[Po(1, t) = i] = t \text{Pr}[Po(1, t) = i - 1]$.

The second statement follows from tail inequalities for the Poisson distribution (see for example [AES92]). \square

Proof.

To stop, the random walk has to make a sequence of transitions corresponding to some $x \in \text{Min}(F)$. This requirements has two components:

- The number of jumps of the Markov chain is at least $|x|$.
- The first $|x|$ transitions conform to x .
- F only accepts x , and none of its prefixes.

The first event is the event that a Poisson process of rate 1, observed at time t makes at least $|x|$ jumps. The second event has probability $W(x)$. The third event has probability $R_F(x)$. Because of Claim 1 we can, in fact, sum up over all values of x , not only those that are F -minimal.

Thus

$$\mu_t(F) = \sum_{x \in T^*} W(x) \cdot R_F(x) \cdot [e^{-t} \cdot \sum_{i \geq |x|} \frac{t^i}{i!}].$$

Differentiating we get

$$\begin{aligned} \frac{d\mu_t(F)}{dt} &= \sum_{x \in T^*} W(x) \cdot R_F(x) \cdot \frac{d}{dt} [e^{-t} \cdot \sum_{i \geq |x|} \frac{t^i}{i!}] \\ &= \sum_{x \in T^*} W(x) \cdot R_F(x) \cdot [-e^{-t} \cdot \sum_{i \geq |x|} \frac{t^i}{i!} + e^{-t} \cdot \sum_{i \geq |x|} \frac{t^{i-1}}{(i-1)!}] \\ &= \sum_{x \in T^*} W(x) \cdot R_F(x) \cdot [e^{-t} \cdot \frac{t^{|x|-1}}{(|x|-1)!}]. \end{aligned}$$

Hence

$$t \cdot \frac{d\mu_t(F)}{dt} = \sum_{x \in T^*} |x| \cdot W(x) \cdot R_F(x) \cdot [e^{-t} \cdot \frac{t^{|x|}}{(|x|)!}] = \sum_{x \in T^*} |x| \cdot P_t(x) \cdot R_F(x).$$

□

We want to compare formula 6 with the original form of Russo's formula for monotone properties:

$$p \frac{d}{dp} \mu_p(A) = \sum_{i=1}^n \mu_p(A_i) = \sum_{i=1}^n \mathcal{I}_i(1_A), \quad (7)$$

where

$$A_i = \{x \in A \mid \omega_i(x) \in \overline{A}\}.$$

and

$$\mathcal{I}_i(f) = \sum_x \text{Prob}[f(x) \neq f(\omega_i x)],$$

the *influence of the i 'th coordinate on function f* (see [KKL88]).

Let's try to get our formula (6) to look more closely like (7): define, for $r \in G$

$$\omega_r : G^* \rightarrow G^*,$$

$$\omega_r(x) = \begin{cases} y & \text{if } (x = yr) \text{ for some } x \in G^*r, \\ x & \text{otherwise. } \mathcal{I}_{r,t}^n(T) \end{cases}$$

and, for $F : G^* \rightarrow \mathbf{C}$,

$$\mathcal{I}_{r,t}^{(n)}(f) = \sum_{|x|=n} \mu_t(x) \cdot [U_F(\omega_r x) - U_F(x)], \quad (8)$$

Then we can rewrite (6) as

$$t \frac{d\mu_t(F)}{dt} = \sum_{r \in G} \left(\sum_{n \geq 0} n \cdot \mathcal{I}_{r,t}^{(n)}(F) \right) = \sum_{n \geq 0} n \cdot \mathcal{I}_t^{(n)}(F), \quad (9)$$

where

$$\mathcal{I}_t^{(n)}(F) = \sum_{r \in G} \mathcal{I}_{r,t}^n(F).$$

The extra factor n in (9) is only a consequence of commutativity of the monotone property setup: to fully mimic (9) in the commutative case one would have to choose an ordering of the bits of $x \in A_i$ such that the i 'th bit of x is the last chosen one.

7 Fourier analysis of stopping times

Let G be a group and let Q be a probability measure on G . One can organize the set of functions on G as a vector space by defining

$$\langle f, g \rangle_Q = \sum_{s \in G} Q(s) \cdot f(s)g(s).$$

If Q was the uniform measure on G , an orthonormal basis for the space of functions on G with the above scalar product would be given ([Dia88] pp. 13) by the set of matrix entries of the unitary irreducible representations of G . On the other hand, for $G = Z_2^n$ and Q the n -time tensor product of the measure μ_p on Z_2 (that is the probability space corresponding to random graph model $G(n, p)$) an orthonormal basis is given (see [Tal94]) by the set of Fourier-Walsh functions $\{r_S\}_{S \subseteq [n]}$

$$r_S(x) = \prod_{i \in S} r_i(x),$$

where

$$r_i(x) = \begin{cases} \sqrt{\frac{1-p}{p}} & \text{if } x_i = 1, \\ -\sqrt{\frac{p}{1-p}} & \text{if } x_i = 0. \end{cases}$$

This result highlights some of similarities, as well as differences between the case of the random walk on \mathbf{Z}_2 and on a general group. In the case of monotonic graph properties the support of measure μ_p is the entire group \mathbf{Z}_2 . In the general case, of course, the probability measure Q will *not* have its support on all the elements of group G , but only on a smaller set Σ (that generates G , otherwise there existed some nodes that could never be reached by the random walk).

We can generalize these observation as follows:

Definition 18 *The diameter of group G with respect to the generator set Σ is the smallest number d such that each element in G can be written as the product of d elements in Σ .*

The following "folklore" result provides a solution to the technical problem we described:

Claim 2 *Let G_1, \dots, G_n be groups, Q_i measures on G_i and $\{b_j^{(i)}\}_{i=1, |G_i|}$ be an orthonormal basis for (G_i, Q_i) . Then an orthonormal basis for the group $G^n = G_1 \times \dots \times G_n$ with scalar product induced by $Q_1 \times \dots \times Q_n$ is given by $B^n = \{b_w\}$, where the elements in B^n are indexed by words $w = w_1 \dots w_n \in [|G|]^n$ and defined as*

$$b_w(x) = \prod_{i=1}^n b_{w_i}^{(i)}(x_i).$$

Thus, the suggested approach is the following: whenever we are dealing with a probability measure Q on G that does not have full support, consider instead

the random walk Q^{*d} that “takes d steps at a time” . This is reasonable (i.e. by doing so we are not missing information about the original walk Q) because of the following “folklore” result:

Proposition 3 *Let F be a strong uniform time that is optimal with respect to separation distance. Then the width of the scaling window of F is $\Omega(d)$.*

Proof.

Because of optimality $s(T_\epsilon) = \epsilon$, $s(T_{1-\epsilon}) = 1 - \epsilon$. By the submodularity of separation distance, similar to the reasoning in Example 3, $T_\epsilon - T_{1-\epsilon} = O(T_{1-\epsilon}) = O(T_{1/2})$. But it is easy to see that $d \leq T_{1/2}$, since at $r = T_{1/2}$ we have to have $K^r(x, y) > 0$ for all x, y . \square

Since we are trying to reconnect (in this more general framework) Fourier analysis with formula (6) let us review this connection in the case of monotone graph properties. In this case the group used for Fourier analysis is \mathbf{Z}_2^n . The orthonormal basis of Fourier-Walsh functions can be rewritten as the set of functions $\{u_S\}_{S \subseteq \{1, 2, \dots, n\}}$ given by

$$u_S(x_1, x_2, \dots, x_n) = \left(\prod_{i \in S} e_1(x_i) \right) \times \left(\prod_{i \notin S} e_0(x_i) \right), \quad (10)$$

where $\text{Char}(\mathbf{Z}_2) = \{e_0, e_1\}$ is the group of characters of \mathbf{Z}_2 .

A Fourier-theoretic interpretation of Russo’s formula was first observed in [KKL88] as

$$p \cdot \frac{d\mu_p(A)}{dp} = \sum_{i=1}^n \alpha_{\{i\}}, \quad (11)$$

where $1_A = \sum_{S \subseteq [n]} \alpha_S u_S$ is the Fourier decomposition of the characteristic function of A in the orthonormal basis (10). To obtain a similar formula we restrict ourselves to a particular class of actions and their associated stopping times.

Definition 19 *Consider a group G acting transitively on a set X , with isotropy subgroup N . A function $f : G \rightarrow \mathbf{C}$ is called N -bi-invariant if*

$$f(n_1 s n_2) = f(s),$$

for all $s \in G$, $n_1, n_2 \in N$.

Definition 20 A central function f defined on a group W is a function that is constant on the conjugacy classes of W , i.e. for all $x, y \in W$

$$f(y^{-1}xy) = f(x).$$

The scenarios we will be most interested is that of Markov chains induced by a N -bi-invariant probability distribution on a group G . In particular, for random walks on groups we have the following definition:

Definition 21 A random walk M on group G of matrix Q is bi-invariant if the transition probability is a central function.

The reason for considering such a restriction is the set of central function on a group G with the uniform measure U and scalar product $\langle \cdot, \cdot \rangle_U$ has an orthonormal basis consisting of the set of characters of irreducible representations of G (see e.g. [Dia88]).

We now consider the vector space of central functions on the n -power $G^{(n)} = G \times G \times \dots \times G$. An easy application of Claim 2 yields an orthonormal basis for this space, similar to the basis (10):

$$F = \sum_{S:[n] \rightarrow \text{Char}(G)} \alpha_S u_S,$$

where

$$U_S(x_1, \dots, x(n)) = \prod_{i=1}^n u_{S(i)}(x_i).$$

Consider the regular R_{reg} representation of G , decomposable as a direct sum of the irreducible representations Z_i of G , with the corresponding multiplicity equal to the dimension of the representation Z_i . Let

$$r_{reg}(x) = \begin{cases} |G|, & \text{if } (x = e), \\ 0, & \text{otherwise,} \end{cases}$$

be its character and let r_i be the character of Z_i . In particular $r_e(x) = 1, x \in G$ is the character of the trivial representation of G . Also, for $\mathbf{i} = (i_1, \dots, i_n)$ define representation $Z_{\mathbf{i}} = Z_{i_1} \oplus Z_{i_2} \dots \oplus Z_{i_n}$ of $G^{(n)}$, and let $\rho_{\mathbf{i}}$ be its character. Finally, define the *Fourier coefficients*

$$\alpha_{\mathbf{i}} = \langle F, \rho_{\mathbf{i}} \rangle_Q,$$

Theorem 3 *Suppose F is a deterministic, consistent, regular, central stopping time. Then the following relation is valid for every $n \geq 1$:*

$$t \frac{d\mu_t(F)}{dt} = \sum_{n \geq 0} n \cdot \mathcal{I}_t^{(n)}(F),$$

with

$$\mathcal{I}^{(n)}(F) = \langle F, r_e^{n-1} \otimes \left[\frac{1}{|W(e)|G|} r_{reg} - r_e \right] \rangle_Q.$$

Suppose that, additionally, we also have $W(e) = \frac{1}{|G|}$.

Then

$$\mathcal{I}^{(n)}(F) = \sum_{i \neq e} \deg(Z_i) \cdot \alpha_{(e, e, \dots, i)} \quad (12)$$

Proof.

(1) We have

$$\begin{aligned} \sum_{r \in G} \mathcal{I}_r^{(n)}(F) &= \sum_{|y|=n-1} \sum_r \mu_t(yr) \cdot [F(y) - F(yr)] = \\ &= \sum_{|y|=n-1} \mu_t(y) \cdot \frac{t}{n} \cdot \left(\sum_r W_y(r) \cdot [F(y \cdot e) - F(yr)] \right) = \\ &= \sum_y \mu_t(ye) \cdot \frac{1 - W_y(e)}{W_y(e)} \cdot F(ye) - \sum_y \sum_{r \neq e} \mu_t(yr) \cdot F(yr) = \\ &= \frac{1}{W(e) \cdot |G|} \left[\sum_{|y|=n-1} \mu_t(ye) \cdot F(ye) \cdot |G| \right] - \left[\sum_{|y|=n} \mu(y) \cdot F(y) \cdot 1 \right]. \end{aligned}$$

(2) A simple consequence of the well-known formula

$$r_{reg} = \sum_{i \in G} \deg(W_i) \cdot r_i$$

(see [Dia88]).

□

When Q is the uniform measure characters $\rho_i^{(n)}$ are orthogonal, and formula (12) is, of course, quite similar to formula (11), that connects influences and Fourier coefficients in the boolean case. This is not necessarily true otherwise, since there is no guarantee that vectors $\rho_i^{(n)}$ are orthonormal.

Definition 22 *Let G be a group. The lamplighter group of G , G^\diamond is defined as follows:*

- (1) Elements of G^\diamond are pairs (f, x) , where $x \in G$ and $f : G \rightarrow \mathbf{Z}_2$.
(2) Multiplication in G^\diamond is defined by

$$(f, x) \cdot (h, y) = (\psi, xy), \text{ where } \psi(i) = f(i) + h(x^{-1}i).$$

Example 6 Random walks on lamplighter groups [PR04] provide natural examples of stopping times with the properties in the theorem. Indeed, the total variation mixing time of a random walk on a group is the hitting time of all lamps in the lamplighter group [PR04]. So the stopping time associated to hitting all lamps is deterministic.

7.1 Conclusions and Acknowledgments

Theorem 3 is especially relevant for analyzing threshold properties of Markovian couplings and deterministic strong uniform times: just as in the case of monotone graph properties, for such couplings the existence of a coarse threshold shows up in the low-weight coefficients in the Fourier decomposition of the associated stopping time. The next step in the development of a toolkit for analyzing threshold properties of stopping times is to assess under what conditions does the crucial isoperimetric inequality due to Bonami and Beckner [Bon68, Bon70, Bec75] extend to powers of arbitrary groups. We aim to do this in subsequent work.

I thank V.S. Anil Kumar for useful discussions. This work has been supported by a Marie Curie International Reintegration Grant within the 6th European Community Framework Programme.

References

- [AD87] D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, 8:69–97, 1987.
- [AES92] N. Alon, P. Erdős, and J. Spencer. *The probabilistic method*. John Wiley and Sons, second edition, 1992.
- [Ald83] D. Aldous. *Random walks on finite groups and rapidly mixing Markov chains*, pages 243–297. Seminar on probability XVII, Lecture Notes in Mathematics 986. Springer Verlag, 1983.
- [Ald94] D. Aldous. Threshold limits for cover times. *Journal of Theoretical Probability*, 4(1):197–211, 1994.
- [Bec75] W. Beckner. Inequalities in fourier analysis. *Annals of Mathematics*, pages 159–182, 1975.

- [Bol85] B. Bollobás. *Random Graphs*. Academic Press, 1985.
- [Bon68] A. Bonami. Ensembles $\lambda(p)$ dans le dual de d^∞ . *Ann. Inst. Fourier*, 18(2):193–204, 1968.
- [Bon70] A. Bonami. Etudes des coefficients Fourier des fonctions de $l^p(g)$. *Ann. Institute Fourier*, 20(2):335–402, 1970.
- [BT86] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1986.
- [CD04] N. Creignou and H. Daudé. Coarse and sharp thresholds for random generalized satisfiability problems. In M. Drmota, P. Flajolet, D. Gardy, and B. Gittenberger, editors, *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities*, pages 507–517. Birkhauser, September 2004.
- [CSC08] G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electronic Journal of Probability*, 13:Paper 3, 26–78, 2008.
- [Dia88] P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, 1988.
- [Dia96] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.
- [DS81] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.
- [Dur06] R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2006.
- [Fri99] E. Friedgut. Necessary and sufficient conditions for sharp thresholds of graph properties, and the k-SAT problem. With an appendix by J. Bourgain. *Journal of the A.M.S.*, 12:1017–1054, 1999.
- [Fri05] E. Friedgut. Hunting for sharp thresholds. *Random Structures and Algorithms*, 26(1–2):34–51, 2005.
- [GG06] B. Graham and G. Grimmett. Influence and sharp-threshold theorems for monotonic measures. *Annals of Probability*, 34:1726–1745, 2006.
- [GNRS04] S. Goel, A. Nachmias, P. Ralph, and D. Shah. (scribes) ARCC workshop: Sharp thresholds for mixing times (list of open problems). December 20–23 2004.
- [Gri75] D. Griffeath. A maximal coupling for Markov chains. *Z. Wahrsch. Verw. Gebiete*, 31:95–100, 1975.
- [Ist05] G. Istrate. Threshold properties of random boolean constraint satisfaction problems. *Discrete Applied Mathematics*, 153:141–152, 2005.

- [JLR00] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. John Wiley & Sons, 2000.
- [KKL88] J. Kahn, E. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science (FOCS'88)*, pages 68–80, 1988.
- [KR01] V. A. Kumar and H. Ramesh. Coupling vs. conductance for the Jerrum-Sinclair chain. *Random Structures and Algorithms*, 18(1):1–17, 2001.
- [Lin92] T. Lindvall. *Lectures on the coupling method*. Wiley series in Probability and Mathematical statistics. John Wiley, 1992.
- [McC98] G. McColm. A splitting inequality. *The Ramanujan Journal*, 2:511–519, 1998.
- [McC04] G. McColm. Threshold functions for random graphs on a line segment. *Combinatorics, Probability, Computing*, 13:373–387, 2004.
- [NBW06] M. Newman, A.L. Barabási, and D. Watts, editors. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.
- [PR04] Y. Peres and D. Revelle. Mixing times for random walks on finite lamplighter groups. *Electron. Journal Probability*, 9:825–845, 2004.
- [Ran03] D. Randall. Mixing. In *Proceedings of the 43rd Annual I.E.E.E. Symposium on Foundations of Computer Science*, 2003.
- [SC04] L. Saloff-Coste. *Probability on Discrete Structures (H. Kesten, editor)*, volume 110 of *Encyclopedia of Mathematical Sciences*, chapter Random Walks on Finite Groups. Springer, 2004.
- [Tal94] M. Talagrand. On Russo’s approximate zero-one law. *The Annals of Probability*, 22(3):1576–1587, 1994.