APPLICABLE ANALYSIS AND DISCRETE MATHEMATICS
available online at http://pefmath.etf.rs

APPL. ANAL. DISCRETE MATH. **x** (xxxx), xxx–xxx. doi:10.2298/AADMxxxxxxxx

THRESHOLD PROPERTIES FOR STOPPING TIMES (I)

Gabriel Istrate

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.

1. INTRODUCTION

Investigating *threshold properties of combinatorial optimization problems* [**17**] has been a popular topic for recent research. The seminal result of Friedgut and Bourgain [**16**] has allowed the classification of thresholds in a large class of problems, e.g. *boolean constraint satisfaction problems (CSP)* [**9, 20**].

2010 Mathematics Subject Classification. 68R05.
Keywords and Phrases. Markov chains, cutoff properties, stopping times.

Threshold properties in CSP are usually studied using variants of probabilistic models from random graph theory [**21**]: the *binomial graph model* and *the uniform graph model*. An approach that 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 [**26**], a variety of such processes has been recently investigated rigorously [**15**]. 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 an unique stationary distribution π and, furthermore, π_t, the distribution obtained by running the Markov Chain for t steps converges (as $t \to \infty$) 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 [**29**]). 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 $t \to \infty$) 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 [**14**] 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 was stated as an open problem in [**13**] (see also [**31**]).

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* [**12**]. Peres has conjectured [**19**] 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. On the positive side, Chen and Saloff-Coste [**10**] have proved Peres's conjecture for reversible Markov chains under L^p-distance, $p > 1$. However, for $p = 1$ the conjecture fails [**10**]: Aldous [**19**] provided a counterexample, and proposed a modified version of the original conjecture. Proving this conjectured version, as well as characterizing Markov chains with a cutoff in L^1 and separation distance is still open. Similar concentration properties for the cover time have been previously investigated by Aldous [**4**], 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 threshold 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 [**16**]. 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 recover a Fourier-theoretic interpretation of Russo's formula due to Kahn, Kalai and Linial [**22**] for certain classes of stopping times.

2. PRELIMINARIES

To make the paper self-contained we will give the basic notions we need in the sequel. Further background is provided by [**12**], [**11**].

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 $G \times X \to X$ which will be denoted by $(s, x) \to s \cdot x$. 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 $X_{t+1} = s \cdot X_t$.*

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 $(s, t) \to s \odot t$, 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 $n \to \infty$, 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*

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

2. *Define* the n-step separation distance *by*

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

3. TRAJECTORIES ON GROUPS

Given an alphabet W (our target examples will be G^* and W^*), we will view $T(A) = A^*$ as the set of *trajectories in A*, endowed with partial order $\lesssim$ on $T(A)$ by $x \lesssim y \iff x$ is a prefix of y.

Let $\alpha : G \times X \to X$ be a group action. We can extend action α from group G to G^* of all *words over G* in the natural way. Furthermore, we will consider the effect of α over X^*, rather than simply X. Formally, we define $\overline{\alpha} : G^* \times X \to X^*$ by

$$\overline{\alpha}(g, x) = \alpha(g, last(x)), \forall g \in G. \tag{1}$$

$$\overline{\alpha}(g_1 g, x) = \alpha(g, x)\alpha(g_1, last(\alpha(g, x)), \forall w \in G^*, z \in G, x \in X. \tag{2}$$

The extension of α naturally maps trajectories over G onto corresponding "trajectories" in X:

Definition 4. *The* image $i(w)$ in X *of trajectory* $w \in T$ *is the sequence* $i(w) = X_0 X_1 \ldots X_n \ldots$ *obtained starting at given node* X_0 *and repeatedly multiplying by elements of* w.

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

Definition 5. *1. A* probabilistic stopping time *for random walk* M *is function* $F : T(G) \to [0, 1]$ *such that*

$$\textit{if } (x \lesssim y) \textit{ and } F(x) \equiv \mathbf{1} \textit{ then } F(y) \equiv \mathbf{1}. \tag{3}$$

Also define $\mu_F : T \to [0, 1]$ *as follows:* $\mu_F(x)$ *is the probability that* F not accept right 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*

$$(\mu_F(x) > 0) \implies (\mu_F(x{\cdot}e) > 0) \textit{ and } (\mu_F(x) < 1) \implies (\mu_F(x{\cdot}e) < 1)$$

(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 $A \subseteq \{0,1\}^n$ be a monotonically increasing property. Consider a Markov chain M with state space $\{0,1\}^n$, that starts in state **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 $x \cdot y = x \bigoplus y$, 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 $X_0, \ldots, X_t$ be the trajectory seen so far. Accept as soon as $Y_t = X_0 \vee X_1 \vee \ldots \vee X_t$ (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 6. *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 $w \in T(G \times G)$, $w = (X_0, Y_0), \ldots, (X_t, Y_t)$,

$$C_M(w) = \begin{cases} 1, & \textit{if } X_t = Y_t, \\ 0, & \textit{otherwise.} \end{cases}$$

For Markov chains the existence of a total variation cutoff can be related to to the properties of stopping times associated to couplings.

Definition 7. *A* strong uniform time *F is a stopping time F such that, for every $k < \infty$,*

$$\Pr_{|x|=k}[F(x, r_{|x|}) = 1, X_k = s] \textit{ is constant in } s. \tag{4}$$

Markov chain cutoffs for the separation distance $s(n)$ relate to properties of strong uniform times (see [**1**] or [**12**], Theorem 4 on pp. 76).

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

4. STOCHASTIC PROCESSES ON TRAJECTORIES

Definition 8. *Let G be a group acting on set X. A G-*processs* on X is a function $W : G^* \to [0,1]$ such that $W(\lambda) = 1$ and $\forall x \in G^*$, $W(x) = \sum_{r \in G} W(xr)$.*

Our main example of a G-process is the following:

Example 3. ("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 random walk is lazy to prevent periodicity. It motivates the following

Definition 9. *Define, for all x with $W(x) > 0$ and $r \in G$, $W_x(r) = \frac{W(xr)}{W(x)}$. W is* Markovian *if for every r $W_x(r)$ (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 [**23**] are Markovian in the sense of Definition 9. Note however, that Markovian couplings can be substantially weaker than general ones [**23**].

Definition 10. *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

We are interested in the following problem: given a monotone dynamics (F, W), can one characterize the threshold properties of $\mu_t(F)$, the probability that simulating the Markov chain for t steps makes F accept ?

Rather than considering W in discrete time, we will assume from now on 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. [**3**] that, with respect to convergence and mixing time any ergodic Markov Chain in discrete time is equivalent to the corresponding continuous Markov Chain).

Definition 11. *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 $y \lesssim x$. $Min(F)$ 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 $Path_M(t, x)$ the event that $T_t = x$ and by $P_t(x) = \Pr[Path_M(t, x)]$. In effect*

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

4. *Define $U_F(x) = \prod_{y \lesssim x}(1 - \mu_y)$ 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 $R_F(x) = U_F(\hat{x}) - U_F(x)$, where $\hat{x}$ is defined to be the prefix of x obtained by dropping the last element in T from x.*

A technical issue we have to deal with is to make sure that the monotone dynamics is specified by a “nice enough” function. In particular we require that the natural series for $\mu_t(F)$ is differentiable, and the derivative can be obtain by summing up the derivatives of all terms in the series.

$$\mu_t(F) = \sum_{x \in Min(F)} P_t(x) = \sum_{x \in Min(F)} e^{-t} \frac{t^{|x|}}{|x|!} \cdot W(x).$$

All these requirements are true in general (we will justify this statement later, in the course of the proof of Theorem 2).

Definition 12. *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 13. *Stopping time F has a* threshold *if* $\forall 0 < \epsilon < 1/2$

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

Definition 14. *Stopping time F has a* sharp threshold *if* $\forall 0 < \epsilon < 1/2$

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

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*

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

Bollobás and Thomason have proved [**8**] that every monotonic property of a random graph has a threshold. Such a property is known for a couple of other models, such as, for instance, 1-dimensional geometric random graphs [**25**] (see also [**24**]) and, by a recent result [**18**], for monotone probability measures that satisfy the FKG inequality. An extension of the result of Bollobás and Thomason [**8**] to stopping times is easy to give:

Definition 15. *A stopping time is* monotone *if for all trajectories x, y, z with $T(y) = \mathbf{1}$ we have $T(xy) = \mathbf{1}$ 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 rule: run T up to time t_ϵ. If it has not accepted, run T for t_ϵ more time.

The probability that the algorithm does not accept in the first phase is at most $1 - \epsilon$. The probability that the chain does not accept in the second phase, given that it has not accepted in the first phase is at most $1-\epsilon$. Hence the chain accepts with probability at least $1 - (1 - \epsilon)^2$.

From monotonicity we infer

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

The rest of the proof closely mirrors the one in [**8**].

6. RUSSO'S FORMULA FOR STOCHASTIC DYNAMICS ON GROUPS

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 3 it follows, therefore, that $\mu_{\hat{x}} = 1$ as well.

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

$$t\frac{d\mu_t(F)}{dt} = \sum_{x\in T^*} |x| \cdot P_t(x) \cdot R_F(x). \tag{5}$$

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 5 the value of the sum changes by no more than $t \cdot Prob[Po(1,t) \geq n]$.*

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 $i \cdot Pr[Po(1,t) = i] = tPr[Po(1,t) = i-1]$. The second statement follows from tail inequalities for the Poisson distribution (see for example [**2**]).

Proof.

To stop, the random walk has to make a sequence of transitions corresponding to some $x \in Min(F)$. 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 term-by-term the right hand side we get

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

where we have used the equality

$$\frac{d}{dt}[e^{-t}\cdot\sum_{i\geq|x|}\frac{t^i}{i!}] = e^{-t}\cdot\frac{t^{|x|-1}}{(|x|-1)!}$$

which follows from computing the derivative of a product, absolute convergence and term rearrangement.

The series of derivatives of the right hand side terms is composed of non-negative terms and is absolutely convergent. Hence, by standard results in calculus it is equal to $\frac{d}{dt}\mu_t(F)$. Thus

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

We want to compare formula 5 with the original form of Russo's formula for monotone properties:

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

where

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

and

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

the *influence of the i'th coordinate on function* f (see [**22**]).

Let's try to get (5) to look more closely like (6): define, for $r\in G$ $\omega_r : G^*\to G^*$,

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

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

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

Then we can rewrite (5) as

$$t\frac{d\mu_t(F)}{dt} = \sum_{r\in G}(\sum_{n\geq 0} n\cdot \mathcal{I}_{r,t}^{(n)}(F)) = \sum_{n\geq 0} n\cdot \mathcal{I}_t^{(n)}(F), \tag{8}$$

where

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

The extra factor n in (8) is only a consequence of commutativity of the monotone property setup: to fully mimic (8) 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 $< f, g >_Q = \sum_{s\in G} Q(s)\cdot f(s)g(s)$. 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 ([**12**] 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 [**32**]) by the set of Fourier-Walsh functions $\{r_S\}_{S\subseteq [n]}$, $r_S(x) = \prod_{i\in S} r_i(x)$, where $r_i(x) = \sqrt{\frac{1-p}{p}}$ if $x_i = 1$ and $r_i(x) = -\sqrt{\frac{p}{1-p}}$ if $x_i = 0$.

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 16. *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 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 and $\{b_j^{(i)}\}_{i=1,|G|}$ 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, whenever we are dealing with a probability measure Q on G without 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 1. *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 **??**, $T_\epsilon - T_{1-\epsilon} = O(T_{1-\epsilon}) = O(T_{1/2})$. 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.

Since we are trying to reconnect (in this more general framework) Fourier analysis with formula (5) 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

$$u_S(x_1, x_2, \dots, x_n) = (\prod_{i \in S} e_1(x_i)) \times (\prod_{i \notin S} e_0(x_i)), \tag{9}$$

where $Char(\mathbf{Z}_2) = \{e_0, e_1\}$ is the group of characters of $\mathbf{Z_2}$.

Russo's formula has a well known Fourier-theoretic interpretation [**22**]:

$$p \cdot \frac{d\mu_p(A)}{dp} = \sum_{i=1}^n \alpha_{\{i\}}, \tag{10}$$

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 (9). To obtain a similar formula we restrict ourselves to a particular class of actions and stopping times.

Definition 17. *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 18. *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 19. *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 $< \cdot, \cdot >_U$ has an orthonormal basis consisting of the set of characters of irreducible representations of G (see e.g. [**12**]).

We now consider the vector space of central functions on the n-power $G^{(n)} = G \times G \times \ldots G$. An easy application of Claim 2 yields an orthonormal basis for this space, similar to the basis (9):

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

where

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

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 $r_{reg}(x) = |G|$, if $(x = e)$, 0, otherwise, 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 $\mathbf{i} = (i_1, \ldots, i_n)$ define representation

$Z_{\mathbf{i}} = Z_{i_1} \oplus Z_{i_2} \ldots \oplus Z_{i_n}$ of $G^{(n)}$, and let $\rho_{\mathbf{i}}$ be its character. Finally, define the *Fourier coefficients* $\alpha_{\mathbf{i}} =< F, \rho_{\mathbf{i}} >_Q$.

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

$$\mathcal{I}^{(n)}(F) =< F, r_e^{n-1} \otimes \Big[\frac{1}{W(e)|G|} r_{reg} - r_e\Big] >_Q .$$

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

Then

$$\mathcal{I}^{(n)}(F) = \sum_{i\neq e} deg(Z_i) \cdot \alpha_{(e,e,\ldots,i)} \tag{11}$$

Proof.

1. We have

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

2. A simple consequence of the well-known formula (see [**12**])

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

When Q is the uniform measure characters $\rho_i^{(n)}$ are orthogonal, and formula (11) is, of course, quite similar to formula (10), 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 20. *Let G be a group. The* lamplighter group of G, *$G^{\diamondsuit}$ is defined as follows. Elements of $G^{\diamondsuit}$ are pairs (f,x), $x \in G$, $f : G \to \mathbf{Z}_2$. Multiplication in $G^{\diamondsuit}$ is defined by $(f,x)\cdot(h,y) = (\psi, xy)$, where $\psi(i) = f(i) + h(x^{-1}i)$.*

Example 4. *Random walks on lamplighter groups [**28**] provide natural examples of stopping times with the properties in Theorem 3.*

*Indeed, the total variation mixing time of a random walk on a group G is the hitting time of all lamps in the lamplighter group [**28**]. So we can investigate the stopping time T defined by hitting all lamps in G. This hitting time is deterministic by definition and consistent since the set of already hit lamps does not change when multiplying with e. For a very similar reason the stopping time is regular: diagonal elements correspond to multiplying with e, hence in matrix P the diagonal elements have the same value.*

8. APPLICATION TO GELFAND PAIRS

The result in Theorem 3 further simplifies in one case of particular interest, that of *Gelfand pairs*:

Definition 21. *Consider a group G acting transitively on a set X, with isotropy subgroup N. (G,N) is a* Gelfand pair *if the convolution of N-bi-invariant functions $f : X \to \mathbf{C}$ is commutative.*

If (G,X) is a Gelfand pair then the Fourier theory of biinvariant functions is simpler: there exist functions (the so-called *spherical functions*) $\Phi_0 \equiv 1, \Phi_1, \ldots, \Phi_{N+1}$ a basis for $L(X)$ satisfying the orthonormality relations

$$\sum_{x\in X} \Phi_i(x)\overline{\Phi_j(x)} = \frac{|X|}{d_i}\cdot \delta_{i,j}, \tag{12}$$

where d_i is the degree of the subspace in the representation $L(X) = \oplus_{i=0}^{N} L_i$ corresponding to Φ_i and $N+1$ is the number of orbits of G in X.

8.1 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 hypercontractive inequalities are powerful enough to prove results on threshold properties on such as the one due due extend to powers of arbitrary groups. Rossignol [**30**] showed that one can bypass the hypercontractive inequality due to Bonami and Beckner [**6, 7, 5**] and independently rederive some of the results on thresholds using a version of the log-Sobolev inequality. A similar tool has been used in [**27**] to analyze properties of certain functions defined on Schreier graphs. We aim to further investigate such an approach for our purposes in subsequent work.

This work has been supported by CNCS IDEI Grant PN-II-ID-PCE-2011-3-0981 ”Structure and computational difficulty in combinatorial optimization: an interdisciplinary approach”.

REFERENCES

1. D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, 8:69–97, 1987.

2. N. Alon, P. Erdős, and J. Spencer. *The probabilistic method*. John Wiley and Sons, second edition, 1992.

3. 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.

4. D. Aldous. Threshold limits for cover times. *Journal of Theoretical Probability*, 4(1):197–211, 1994.

5. W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, pages 159–182, 1975.

6. A. Bonami. Ensembles $\lambda(p)$ dans le dual de d^{∞}. *Ann. Inst. Fourier*, 18(2):193–204, 1968.

7. A. Bonami. Etudes des coefficients Fourier des fonctions de $l^p(g)$. *Ann. Institute Fourier*, 20(2):335–402, 1970.

8. B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1986.

9. N. Creignou and H. Daudé. Coarse and sharp thresholds for random generalized satisfiability problems. In M. Drmota et al., editor, *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities*, pages 507–517. Birkhauser, 2004.

10. G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electronic Journal of Probability*, 13:26–78, 2008.

11. T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli. *Harmonic Analysis on Finite Groups: Representation Theory, Gelfand Pairs and Markov Chains*. Cambridge University Press, 2008.

12. P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, 1988.

13. P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

14. P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

15. R. Durrett. *Random Graph Dynamics*. Cambridge Univ. Press, 2006.

16. 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.

17. E. Friedgut. Hunting for sharp thresholds. *Random Structures and Algorithms*, 26(1–2):34–51, 2005.

18. B. Graham and G. Grimmett. Influence and sharp-threshold theorems for monotonic measures. *Annals of Probability*, 34:1726–1745, 2006.

19. S. Goel, A. Nachmias, P. Ralph, and D. Shah. (scribes) ARCC workshop: Sharp thresholds for mixing times. December 20–23 2004.

20. G. Istrate. Threshold properties of random boolean constraint satisfaction problems. *Discrete Applied Mathematics*, 153:141–152, 2005.

21. S. Janson, T. Luczak, and A. Ruczinski. *Random Graphs*. Wiley, 2000.

22. 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.

23. V. A. Kumar and H. Ramesh. Coupling vs. conductance for the Jerrum-Sinclair chain. *Random Structures and Algorithms*, 18(1):1–17, 2001.

24. G. McColm. A splitting inequality. *Ramanujan Journal*, 2:511–519, 1998.

25. G. McColm. Threshold functions for random graphs on a line segment. *Combinatorics, Probability, Computing*, 13:373–387, 2004.

26. M. Newman, A.L. Barabási, and D. Watts, editors. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

27. R. O'Donnell and K. Wimmer. Kkl, Kruskal-Katona and monotone nets. In *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*, pages 725–734, 2009.

28. Y. Peres and D. Revelle. Mixing times for random walks on finite lamplighter groups. *Electron. Journal Probability*, 9:825–845, 2004.

29. D. Randall. Mixing. In *Proceedings of the 43rd Annual I.E.E.E. Symposium on Foundations of Computer Science*, 2003.

30. Raphaël Rossignol. Threshold for monotone symmetric properties through a logarithmic Sobolev inequality. *The Annals of Probability*, 34(5):1707–1725, November 2006.

31. 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.

32. M. Talagrand. On Russo's approximate zero-one law. *The Annals of Probability*, 22(3):1576–1587, 1994.