Rotor walks and Markov chains

Jim Propp
(U. Mass. Lowell)

May 14, 2009

(joint work with Ander Holroyd)

(with thanks also to Hal Canary, Matt Cook, Dan Hoey, Michael Kleber, Lionel Levine, Yuval Peres, and Oded Schramm)

Slides for this talk are on-line at
`jamespropp.org/ndrsp.pdf`
and the article that the talk is based on is on the arXiv.

Goal: Replace a random process by a deterministic process that has the same average-case behavior but is more tightly concentrated around that average ("low discrepancy").

Tool: Simple local gadgets (rotor-routers).

What is discrepancy, and why are rotor-routers good at minimizing it?

I. Local discrepancy and global discrepancy

Let P be an n-by-n irreducible stochastic matrix with states a, b, c satisfying $p(b, a) = p(c, a) = 1$ s.t. the chain a.s. hits b or c starting from a. That is, we imagine a particle repeatedly walking from a source vertex (a) until it gets absorbed at a sink (b or c), restarting at a after each absorption.

For all v let $h(v)$ be the probability that a particle released from v will reach b before c, so that $h(b) = 1$ and $h(c) = 0$, and h is harmonic away from b and c (that is $\sum_w p(v, w)h(w) = h(v)$ for all $v \notin \{b, c\}$).

Consider a finite path

$$x_0, x_1, \ldots, x_{T-1}, x_T$$

in which the particle starts and ends at a, arriving K times at b and $N-K$ times at c.

If the path were a typical sample path of the random walk associated with the matrix P (a "P-random sample path"), we would expect $K \approx Np$ where $p = h(a)$.

Let's find a formula that expresses the discrepancy $K - Np$ in terms of how far from being P-distributed the path is.

Write

$$
\begin{aligned}
K - Np &= (K - Kp) - (Np - Kp) \\
&= K(1-p) + (N-K)(0-p) \\
&= K(h(b) - h(a)) \\
&\quad +(N-K)(h(c) - h(a)) \\
&= \sum (h(x_t) - h(x_{t-1}))
\end{aligned}
$$

where the sum is over all $1 \le t \le T$ with x_{t-1} not a sink.

We can gather together terms of the sum for which $x_{t-1} = v$ and $x_t = w$, obtaining the double sum

$$
\sum_v \sum_w N(v,w)(h(w) - h(v))
$$

where $N(v,w)$ is the number of times the particle moved from v to w up to time T, and where v is not a sink.

So the discrepancy $D := K - Np$ satisfies

$$D = \sum_v \sum_w N(v, w)(h(w) - h(v))$$

while the harmonicity of $h(\cdot)$ at v gives

$$0 = \sum_w (N(v)p(v, w))(h(w) - h(v))$$

(where $N(v) = \sum_w N(v, w)$); hence D equals

$$\sum_v \sum_w (N(v, w) - N(v)p(v, w))(h(w) - h(v)).$$

That is, the *global discrepancy* $K - Np$ can be written as the sum of the *local discrepancies*

$$(N(v, w) - N(v)p(v, w))(h(w) - h(v)).$$

If $x_0, x_1, \ldots$ is given by a random process, the local discrepancies are $\approx N^{1/2}$, so the global discrepancy is $\approx N^{1/2}$ too.

If $x_0, x_1, \ldots$ has smaller-than-random (*subrandom*) local discrepancies (say $\approx N^\alpha$ with $\alpha < 1/2$) then $K - Np$ will be subrandom too.

How subrandom can we get?

We can't do better than $\alpha = 0$; that is, choosing the sequence $x_0, x_1, \ldots, x_T$ so that the local discrepancies

$$N(v,w) - N(v)p(v,w)$$

are *bounded*.

II. Rotor-routing in general

Suppose we have a (not necessarily finite) Markov chain that is locally finite (for each v, $p(v, w) = 0$ for all but finitely many w) and "rational" (all $p(v, w)$ are rational numbers).

For each state v, in lieu of a d-sided die (where $d = d(v)$ is a common denominator of the $p(v, w)$'s), we use a periodic process of period d in which each w occurs with frequency $p(v, w)$.

Whenever we arrive at a state v, we choose whichever w is next in succession.

E.g., if $p(a, b) = \frac{1}{3}$ and $p(a, c) = \frac{2}{3}$, our periodic process could be *cbccbccbc*....

The 1st time we leave a, we go to c;
the 2nd time we leave a, we go to b;
the 3rd time we leave a, we go to c;
etc.

Note that the local discrepancies associated with transitions from a to b and from a to c stay bounded, and indeed vanish whenever the number of visits to a is divisible by 3.

Equivalently, we can imagine that vertex a has a three-state **rotor** ρ associated with it that has one state pointing from a to b and two states pointing from a to c. The states are cyclically ordered. To move the particle forward one step, advance the rotor to its next state and move the particle to the state that the rotor points to.

E.g., if there is a particle at a and the rotor at a currently points from a to b, rotate the rotor so that it points from a to c and then route the particle from a to c.

More generally, we generate an infinite sequence $(x_0, \rho_0), (x_1, \rho_1), (x_2, \rho_2), \ldots$ where $x_0 \in V$ is some initial vertex, $\rho_0 : V \to V$ is some initial setting of the rotors, and (x_{t+1}, ρ_{t+1}) is obtained from (x_t, ρ_t) as follows:

1) **rotate** the rotor at x_t in ρ_t and leave the other rotors at all other vertices alone, obtaining ρ_{t+1}; and

2) **route** the particle at x_t to $x_{t+1} := \rho_{t+1}(x_t)$.

This is the **rotor-router process**, and the sequence $x_0, x_1, x_2, \ldots$ is a **rotor-router walk** in V.

Rotor-routing is deterministic, but more importantly, it minimizes the local discrepancy of the walk $x_0, x_1, x_2, \ldots$.

Theorem (Holroyd-Propp): Consider an irreducible recurrent Markov chain with states a, b, c satisfying $p(b, a) = p(c, a) = 1$, with $h(\cdot)$ defined as before. Suppose

$$C := 1 + \frac{1}{2} \sum_{\substack{u \in V \setminus \{b,c\}, \\ v \in V}} d(u) p(u, v) |h(u) - h(v)|$$

is finite. Then for any rotor walk and all t,

$$\left| h(a) - \frac{n_t(b)}{n_t(b) + n_t(c)} \right| \le \frac{C}{n_t(b) + n_t(c)},$$

where $n_t(b)$ (resp. $n_t(c)$) is the number of times the walk visits b (resp. c) before time t.

Instead of routing particles from b and c back to a, we can imagine that each time a particle gets absorbed at a sink, it stays there, and a new particle is released from the source.

Or: We can imagine that many particles start at the source, and we let them successively travel through the state-diagram of the Markov chain until absorption.

Abelian property: If there are multiple particles on the state-diagram of a Markov chain, so that at each instant you have a choice of which one to advance via rotor-routing, the choices you make don't matter. The end result is the same when all the dust settles (i.e., when all particles have been absorbed).

For instance, you can advance each particle one step, then advance each particle another step, and so on, until every particle has been absorbed ("tandem rotor-routing").

This gives us an easy way to see that the absorption frequencies in a rotor-router simulation agree with absorption probabilities for random simulation (though it doesn't explain why discrepancy is small):

If we put N particles at a and have them do tandem rotor-routing, a randomly-chosen particle does something close to a P-random walk.

E.g., if we start with $N = 999$ particles at a, with $p(a, v) = p(a, w) = \frac{1}{2}$, 500 of the particles will go one way and 499 will go the other way.

Part of what makes rotor-routing have such low discrepancy is that discrepancies like $|\frac{1}{2} - .499|$ fall like $1/N$.

But the other smart thing rotor-routing does is that the next time there are an odd number of particles at a, the rounding will go the other way.

The Holroyd-Propp article gives other examples of the concentration phenomenon for rotor-routing, where rotor-routing for N trials give approximations to probabilistic quantities (e.g. stationary measures and hitting times) that differ from the true values by $O(1/N)$, in contrast to ordinary random simulation, which gives errors on the order of $1/\sqrt{N}$.

Also, if the rotor-and-particle system returns to a configuration it's already visited, then it will behave thereafter in a periodic fashion, and its behavior over the course of one period will give exact values of the probabilistic quantity in question.

History of rotor-routing:

- Eulerian walkers model
- load-balancing
- whirling tours

Recent applications:

- derandomized diffusion (Cooper, Doerr, Friedrich, Spencer, Tardos)
- diffusion-limited aggregation (Levine and Peres)
- rumor-spreading (Doerr, Friedrich, Künnemann, Sauerwald)

Markov chains with irrational transition probabilities can also be fit into the framework of discrepancy-minimization, but we can no longer use a finite-state router at each vertex.

E.g., if $p(1,2) = \alpha$ and $p(1,3) = 1 - \alpha$ with α irrational, then there is a unique protocol for routing the particle so that after the particle has left 1 for the nth time, the number of times it went to 2 is the integer closest to $n\alpha$ and the number of times it went to 3 is the integer closest to $n(1 - \alpha)$.

For out-degree > 2, things are a little more complicated; see the section of Holroyd-Propp on "stack-walk".

A nice example of rotor-routing is the "goldbug walk" on $\{-1, 0, 1, ...\}$ where states $b = -1$ and $c = 0$ are absorbing, all other states are transient, and $p(i, i-2) = p(i, i+1) = \frac{1}{2}$ for all $i \geq 1$.

This walk has leftward drift, so the probability of absorption in $\{b, c\}$ is 1.

To see what happens when all rotors initially point to the right, run

`http://jamespropp.org/rotor-router-model/`

with Graph/Mode set to 1-D Walk.

If we attend to where the successive particles end up, we see that the whole system, made of infinitely many $(1/2, 1/2)$ rotors, behaves like a single $(\alpha, 1-\alpha)$ rotor, with $\alpha = (-1 + \sqrt{5})/2$.

By the Abelian property, we could also put lots of particles at 0 at the start (N, say) and let them do rotor-walk in tandem; $N\alpha \pm O(1)$ of them will be absorbed at -1 and $N(1-\alpha) \pm O(1)$ will be absorbed at 0.

The same is true if the rotors initially point to the left, except that one particle will never get absorbed; it just wanders off to the right forever.

To take full advantage of the Abelian property in situations where some of the particles wander off to infinity, it's helpful to define simulation in "transfinite time".

E.g., in the goldbug system with all rotors initially pointing to the left, we let the first particle wander off to infinity, leaving leftward-pointing rotors in its wake, and "thereafter" continue to add other particles, all of which get absorbed at -1 and 0.

Specifically, we define rotor-simulation indexed by ordinals of the form $m\omega + t$ where m, t are non-negative integers.

This may be possible for some m and not others.

E.g., consider ordinary rotor-walk on $\mathbf{Z}$ where all rotors to the left of 0 point to the right and all other rotors point to the left. We can simulate from time 0 to time ω (the particle goes to $+\infty$) and from time ω to time 2ω (the particle goes to $-\infty$), but from time 2ω to time 3ω, each site gets visited infinitely often so it's impossible to say what state the rotors are in at time 3ω.

On the other hand, consider the Markov chain whose state space is $\{0, 1, 2, \dots\}$ where state $b = 0$ is absorbing and $p(i, i-1) = \frac{1}{3}$ and $p(i, i+1) = \frac{2}{3}$ for all $i > 0$ (biased random walk with rightward drift-rate $\frac{1}{3}$).

Regardless of the initial setting of the rotors, the particle runs off to infinity during its nth run if and only if it did *not* run off to infinity on the $n-1$st run, for all $n \geq 3$ (but not necessarily $n = 2$), in agreement with the fact that the escape probability is $\frac{1}{2}$.

In this case, the transfinite rotor-walk is defined for *all* times of the form $m\omega + t$.

What controls the well-definedness of transfinite rotor-walk is the following fact:

Recurrence/transience dichotomy:

An infinite rotor-walk on a connected graph either visits every vertex only finitely often or visits every vertex infinitely often.

(Proof: Every neighbor of a vertex that gets visited infinitely often must be visited infinitely often.)

If a vertex gets visited only finitely often, the rotor at that vertex has a limiting (indeed, eventual) setting; if a vertex gets visited infinitely often, the limiting setting does not exist.

Rotor-walks can be *more recurrent* than random walks: Landau and Levine show that for rotor-walk on an infinite binary tree with source at the root, if we set the rotors so that the first time a particle leaves a vertex v it goes toward the root, the particle will visit the root infinitely often.

Rotor-walks can be more more transient than random walks in the short run: rotor-walks on $\mathbf{Z}^d$ can go off to infinity (this a.s. doesn't happen for random walk on $\mathbf{Z}^d$).

However, a rotor-walk cannot be more transient than its random counterpart in terms of asymptotic escape-frequency.

Theorem (Schramm): For transfinite rotor-walk, let I_n be the number of times the walk goes to infinity before the nth return to a. Then $\limsup_{n\to\infty} I_n/n$ is at most the probability that random walk started from a never returns to a.

In particular, if the walk is recurrent, $I_n/n \to 0$.

III. Rotor-routing on $\mathbf{Z}^2$

For rotor-walk in $\mathbf{Z}^2$, it's natural to have rotors that cycle through the four directions as N,E,S,W,N,E,S,W,...

Cooper and Spencer proved that there is a small finite constant C (less than 10) such that if one starts N particles at the origin and lets them execute T steps of tandem rotor-walk, the discrepancy between

# of particles at site v at time T

and

N times $p^{(T)}(0, v)$

is at most C, for *all* v, N, T and all initial configurations of the rotors!

In fact, Cooper and Spencer showed that an analogous bounded discrepancy property holds for any initial distribution of the N particles on the even sublattice $\{(i,j) \in \mathbf{Z}^2 : i+j \text{ is even}\}$ (not just the distribution where all the particles start at 0).

The bound is *independent* of the initial distribution of the particles.

Boundedness also holds if the rotors cycle in the pattern N,S,E,W,N,S,E,W,... (thought the constant C is different).

Boundedness also holds for $\mathbf{Z}^d$ for all d (though the constants are worse, and might grow quickly with d).

Back to absorption probabilities:

Let $a = (0,0)$, $b = (1,1)$, $c = (0,0)$.

(Technically we need two copies of $(0,0)$, one a target and one a source, so $p(b,a) = p(c,a) = 1$ and $p(a,b) = p(a,c) = 0$.)

It's known that the probability that a particle emitted from a arrives at b before it arrives at c is $p = \pi/8$.

To see how closely rotor-walk concentrates around this value, see
`http://jamespropp.org/rotor-router-model/`
with Graph/Mode set to 2-D Walk.

(What sort of stable structures does the rotor-configuration exhibit?)

Theorem (Holroyd-Propp): With rotors that cycle clockwise and initial conditions

```
      . . . .
  N N N N N N N E
  W N N N N N E E
. W W N N N E E E .
. W W W N E E E E .
. W W W W S E E E .
. W W W S S S E E .
  W W S S S S S E
  W S S S S S S S
      . . . .
```

we have $D = K - Np = O(\log N)$, where N is the number of particles emitted from a, K is the number of particle absorbed at b, and $p = \pi/8$.

Proof sketch: Recall from the start of the talk that the discrepancy D equals

$$\sum_v \sum_w (N(v,w) - N(v)p(v,w))(h(w) - h(v))$$

(with v not a sink). The rotor-router protocol guarantees that the differences $N(v,w) - N(v)p(v,w)$ are uniformly bounded, so, using standard facts about the potential kernel for two-dimensional random walk, we can show that the inner sum is on the order of $1/|v|^2$.

If we had to sum over all $v \in \mathbf{Z}^2$, this would diverge like the harmonic series (since the number of points at distance $n \pm \frac{1}{2}$ from the origin is on the order of constant time n).

However, the initial conditions we picked guarantee that when N particles have gone through the system and been absorbed, the sites that have been visited lie in the $2N$-by-$2N$ square centered at $(\frac{1}{2}, \frac{1}{2})$ (the combinatorial details are omitted here), and for sites that have not been visited, the contribution to the discrepancy D is 0.

Hence, the global discrepancy is bounded by the harmonic series truncated after $O(N)$ terms, which is $O(\log N)$.

The method of proof works for any finite target set: Let p be the probability that a random walk in $\mathbf{Z}^2$ that walks from source vertex $(0,0)$ until it hits the finite target set B stops at a particular vertex b in B. If one performs N successive runs of a rotor-router walk in $\mathbf{Z}^2$ from $(0,0)$ to B, the number of runs that stop at b is $Np \pm O(\log N)$.

What's wrong with this theorem:

It's not general enough.

E.g., the concentration phenomenon seems to be just as strong if we use the initial configuration

```
          .  .  .  .
    E  E  E  E  E  E  E  S
    N  E  E  E  E  E  S  S
 .  N  N  E  E  E  S  S  S  .
 .  N  N  N  E  S  S  S  S  .
 .  N  N  N  N  W  S  S  S  .
 .  N  N  N  W  W  W  S  S  .
    N  N  W  W  W  W  W  S
    N  W  W  W  W  W  W  W
          .  .  .  .
```

even though the proof given above doesn't apply.

What's wrong with this theorem:

It's not sharp enough.

The observed discrepancy D_N after N trials seems to be a lot less than $\log N$.

In 10,000 trials, $|D_N| < 0.5$ for $5,070$ of the trials. That is, more than half the time, the number of absorptions at b during the first N trials is equal to the integer closest to Np.

We have $|D_N| < 2.05$ for all $N \leq 10^4$.

Is $|D_N|$ bounded? Unknown!

Yuval Peres points out that if the summands in our truncated harmonic series are uncorrelated, we would expect the global discrepancy to behave like the random sum $\pm 1 \pm \frac{1}{2} \pm \frac{1}{3} \pm ...$, which is a.s. bounded.

IV. Open problems

For the initial conditions in the Holroyd-Propp theorem on hitting probabilities in $\mathbf{Z}^2$, how does the discrepancy D grow as a function of N?

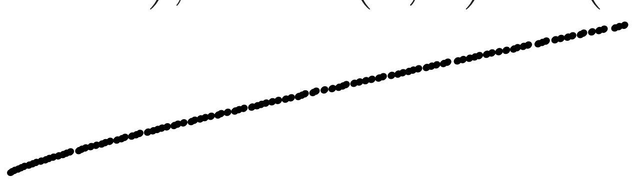
We know it's $O(\log N)$, but the data suggest a better bound is possible.

How many escapes to infinity can happen if we use other initial conditions?

We know the number of escapes must be $o(N)$ (by Schramm's result).

We know it need not be $O(1)$ (e.g., consider the initial rotor-setting with all rotors aligned).

For rotor walk on $\mathbf{Z}^2$ with no sinks, and all rotors initially aligned, the data for small N wouldn't lead you to guess the $o(N)$ result. Here's a plot of number of visits to the origin (horizontal axis) versus number of escapes to infinity (vertical axis), from $(0, 0)$ to $(1884, 458)$:



What about $\mathbf{Z}^3$?

About five years ago I did a transfinite rotor-router simulation of walk on $\mathbf{Z}^3$ with $N = 10^6$, with all rotors initially aligned; it gave the escape probability to four (but not five) significant figures. So in this case discrepancy is almost certainly not bounded, but it might be smaller than $O(\sqrt{N})$.

Joel Spencer and I ask what happens if the rotors are set up in $\{(x, y, z) : x, y, z \in \mathbf{Z}\}$ so that the first time a particle visits (x, y, z) it moves to the neighbor for which $|z|$ is smallest, unless z is already 0, in which case the particle moves to the neighbor for which $|y|$ is smallest, unless y is already 0, in which case the particle moves to the neighbor for which $|x|$ is smallest.

Is this rotor-walk recurrent?

V. Pure fun (popcorn not included)

Lionel Levine: rotor walk on $\mathbf{Z}^2$ with all arrows initially aligned

Ander Holroyd: ditto, but with biased random initial conditions: $p(\text{North}) = p(\text{East}) = \frac{1}{2}$, $p(\text{South}) = p(\text{West}) = 0$.

Tobias Friedrich: directed DLA (Diffusion-Limited Aggregation)

- fully random
- rotor-router, with random initial conditions
- rotor-router, with simple conditions