The Klee-Minty random edge chain moves with bounded speed

József Balogh[1,2],
Robin Pemantle [3,4]

Abstract: An infinite sequence of 0's and 1's evolves by flipping each 1 to a 0 exponentially at rate one. When a 1 flips, all bits to its right also flip. Starting from any configuration with finitely many 1's to the left of the origin, we show that the leftmost 1 moves right with bounded speed. Upper and lower bounds are given on the speed. A consequence is that a lower bound for the run time of the random-edge simplex algorithm on a Klee-Minty cube is improved so as to be quadratic, in agreement with the upper bound.

Keywords: Markov chain, ergodic, bit, flip, binary, simplex method

Subject classification: 60J27, 60J75

[1]Research supported in part by National Science Foundation grant # DMS 0302804 and partly supported by grant OTKA #049398

[2]The Ohio State University, Department of Mathematics, 231 W. 18th Avenue, Columbus, OH 43210, jobal@math.ohio-state.edu

[3]Research supported in part by National Science Foundation grant # DMS 0103635

[4]University of Pennsylvania Department of Mathematics, 209 S. 33rd Street, Philadelphia, PA 19104, pemantle@math.upenn.edu

1 Introduction

1.1 Motivation

The simplex algorithm is widely used to solve linear programs. It works well in practice, though often one cannot prove that it will. A parameter in the algorithm is the rule that selects one move among possible moves ("pivots") that decrease the objective function. Deterministic pivot rules are known to be possibly very far from optimal. For example, consider problems with the number m of constraints of the same order as the dimension, n. For virtually every deterministic pivot rule there is a problem for which the algorithm will take exponential time, although it is conjectured (the "strict monotone Hirsch conjecture" of [Z95]) that there exists a descending path whose length is $O(n)$. Variants of the original argument by Klee and Minty [KM72] are cited in [GHZ98]; for a summary of known results as of 1995, see [Z95].

Randomized pivot rules appear to do better. According to [GHZ98], several of the most popular randomized pivot rules appear to have polynomial – even quadratic – running time. Rigorous and general results on these, however, have been hard to come by. When one restricts to a narrow class of test problems, it becomes possible to obtain some rigorous results. Gärtner, Henk and Ziegler [GHZ98] consider three randomized pivot rules. Relevant to the present paper are their results on the *random edge* rule, in which the next move is chosen uniformly among edges leading to decrease the objective function. They analyze the performance of this rule on a class of linear programs, the feasible polyhedra for which are called *Klee-Minty cubes*, after [KM72]. Such cubes are good benchmarks because they cause some pivot rules to pass through a positive fraction, or even all, of the vertices. Recently Matousek and Szabó [MSZ04] used Klee-Minty cubes and a recursive construction to build an abstract cube (an acyclic uinique-sink orientation on the graph of the n-cube) for which for the random edge algorithm takes time $\exp(cn^{1/3})$.

Nevertheless, it is known [GHZ98] that the expected run time of the random edge algorithm on an actual Klee-Minty cube is quadratic, up to a possible log factor in the lower bound:

Theorem 1 (GHZ) *The expected number, E_n of steps taken by the random edge rule,* *started at a random vertex of a Klee-Minty cube, is bounded by*

$$\frac{n^2}{4(H_{n+1}-1)} \leq E_n \leq \binom{n+1}{2}.$$

Here, $H_n = \sum_{j=1}^{n} \frac{1}{j} \sim \log n$ *is the* n^{th} *harmonic number.*

Their lower bound rules out the possibility that $E_n \sim n\,\mathrm{polylog}(n)$ which was twice conjectured by previous researchers [PS82, page 29], [Kel81]. They guess that the upper bound is the correct order of magnitude, and state an improvement in the upper bound from $\frac{1}{2}n^2$ to $0.27\ldots n^2$, whose proof is omitted.

The method of analysis in [GHZ98] describes the progress of the algorithm as a random walk on an acyclic directed graph. In their model, vertices are bijectively mapped to sequences of 0's and 1's of length n, and each move consists of flipping a 1 (chosen uniformly at random) to a 0, and simultaneously flipping all bits to the right of the chosen bit. It was in this form that the problem came to our attention. Indeed, the remainder of the paper is framed in terms of a variant of this model, which we find to be an intrinsically interesting model. Our main result, Theorem 2, closes the gap left open in [GHZ98], proving that the upper bound is sharp to within a constant factor and obtaining upper and lower bounds differing by a factor of less than 3.

For several reasons, we have moved the model to continuous time and made n infinite: (1) from our view as probabilists this is the most natural way to frame such a model; (2) a heuristic we learned from David Aldous is that to understand a limit theorem it is often best to construct a limiting object or process; (3) we find natural generalizations in the limit that would not be apparent in the discrete setting. Nevertheless, in order to keep everything accessible to non-probabilists, we provide intuitive explanations of all probabilistic terminology as well as references to what might be considered standard facts from the probability literature. We emphasize that our results apply both to the setting of [GHZ98, Corollary 3] and to the continuous setting, and that the proofs have combinatorial interpretations as well. Although our model seems simple, we remark that we were unable to prove many things about the model, including whether certain limits exist.

1.2 Informal statement of the model

Place a 0 or a 1 at each point of the one-dimensional integer lattice, arbitrarily except that there must be some point to the left of which lie only 0's. There is a rate one Poisson process at every site, that is, an alarm clock which, independently of when it has gone off in the past and of all other sites, always has the same chance dt of going off in any time interval $[t, t + dt]$. Because of the infinite extent of the lattice, there will be infinitely many alarms going off in any time period. When a clock rings, if there is a 0 there nothing happens, but if there is 1 there, then it and all (infinitely many) of the sites to the right flip as well – 0's become 1's and 1's become 0's. It should be intuitively clear that flipping is happening at an unbounded rate as one moves to the right, but that each fixed site flips only finitely often in a bounded time period, due to the condition of there being all zeros to the left of some point. The description up to this point is of a continuous-time Markov process, which will be formalized below as a random variable M taking values in the space of functions from the time set $[0, \infty)$ to the space of legal configurations of zeros and ones; the value of M at time t is denoted M_t rather than $M(t)$.

The *shift times*, namely when the leftmost 1 flips to a 0 (and thus the location of the leftmost 1 moves to the right), will be particularly important. If we sample M_t only at shift times, and always describe the configuration relative to the position of the new leftmost 1, we get a discrete-time Markov chain $\{Y_n\}$ which will also be important. We now give formal constructions of these two processes. Readers who wish to avoid the formalities should at least understand the notation for the times $\{\xi_{i,j}\}$ at which alarms go off, the successive times σ_n at which the position of the leftmost 1 changes, and the size jump_n of the n^{th} shift.

1.3 Formal construction of the model

Let $\mathcal{S}$ be the subset of $\{0, 1\}^{\mathbb{Z}}$ consisting of the all-zero sequence and the sequences of 0's and 1's that have a leftmost 1 (equivalently, have finitely many 1's to the left of the origin). Let $\{N(j, t) : t \geq 0\}_{j \in \mathbb{Z}}$ be a collection of independent and identically distributed Poisson counting processes. The reader may consult [Dur05, Section 2.6] for the formal construction of such a collection on a generic probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Since these are at the heart of the construction, we pause to explain the relevant properties of such a collection. Each function $N(j, \cdot)$ is a step function, whose range is the nonnegative integers, and which increases by 1 at random times. Denoting the times of increase $\xi_{j,1} < \xi_{j,2}, \ldots$, we have $N(j,t) = 0$ for $0 \le t < \xi_{j,1}$, $N(j,t) = 1$ for $\xi_{j,1} \le t < \xi_{j,2}$ and so on[5]. The increments, $\xi_{j,i} - \xi_{j,i-1}$ are independent for all i and j and distributed as exponentials of mean 1. The exponential distribution, namely the distribution for which $\mathbb{P}(X > t) = e^{-t}$, is memoryless, ensuring that the probability of an alarm in any time interval $[t, t+dt]$ at any site j is always $\approx dt$, independent of what has happened in the past. Associated with such a process is a *filtration*. This is a collection $\{\mathcal{F}_t : t \ge 0\}$, where $\mathcal{F}_t$ is the σ-field of information known up to time t. Formally, $\mathcal{F}_t$ may be constructed as the collection of all measurable functions of the values $\{N(j,s) : s \le t\}$. The purpose of a filtration is to be able to make probability statements that are "conditional on what you know up to time t"; these are denoted $\mathbb{P}(\cdot \mid \mathcal{F}_t)$ and conditional expectations are denoted $\mathbb{E}(\cdot \mid \mathcal{F}_t)$; for any remaining details, the reader may consult [Dur05].

Having constructed the Poisson processes $\{N(j,t) : t \ge 0\}_{j \in \mathbb{Z}}$, and given a starting configuration ω, we may construct a process $\{M_t\}_{t \ge 0}$ directly as a function of these, with no further randomness needed. Formally, for each t, M_t will be a random element of $\mathcal{S}$, i.e. a function from Ω to $\mathcal{S}$, which depends on $x \in \Omega$ only through the values of the already defined functions $\{N(j,s)\}$. In particular, the bit $M_t(i)$ in position i at time t will depend only on the times of alarms up to time t in positions in the interval $[i_0(\omega), i]$, where $i_0(\omega)$ is the position of the leftmost 1 of the starting configuration, ω. Since $N(j,t)$ is almost surely finite, the size of the collection

$$\mathcal{C}(i,t) := \{\xi_{j,k} : i_0 \le j \le i;\ k \le N(j,t)\}$$

is almost surely finite, so, except possibly on a subset of Ω of measure zero, the following inductive procedure produces a well-defined value for $M_t(i)$ for all $t \ge 0, i \in \mathbb{Z}$.

> For $i < i_0(\omega)$, set $M_t(i) := 0$ for all t. Assume now for induction that $M_t(j)$ has been defined for $i_0 \le j < i$, for all t, in terms of the quantities in $\mathcal{C}(j,t)$. To define $M_t(i)$, order the set of times in $\mathcal{C}(i,t)$ from least to greatest: call these times $0 := \mathsf{time}_0, \mathsf{time}_1, \ldots, \mathsf{time}_{Z(t)}$ and let $l(k)$ denote the location of the alarm at time time_k: the unique j such that $\mathsf{time}_k = \xi_{j,r}$ for some r. Over each interval $t \in [\mathsf{time}_k, \mathsf{time}_{k+1})$ each $M_t(j)$ is constant. For $t \in [0, \mathsf{time}_1)$ let $M_t(i) = \omega(i)$. Assuming that $M_t(i)$ has been defined up to time time_k^-, define $M_t(i)$ for $\mathsf{time}_k \le$

[5] By convention, these functions are taken to equal their limit from the right at discontinuity points.

$t < \mathsf{time}_{k+1}$ to be equal to $1 - M_{\mathsf{time}_{k-1}}(i)$ if $M_{\mathsf{time}_{k-1}}(j) = 1(k-1)$ and $j \le i$, and to be equal to $M_{\mathsf{time}_{k-1}}(i)$ otherwise. In other words, the bit flips if and only if the alarm goin off at time time_k is at a location where there is a 1 and is at or to the left of position i.

It should be clear that our formal construction corresponds to the informal process of defining the process on a finite interval, then adding positions to the right one at a time, observing that the new bits do not influence the old ones. One may formally verify that $\{M_t\}_{t\ge 0}$ satisfies the time-homogeneous Markov property, namely for $s < t$,

$$\mathbb{P}(M_t \in S \,|\, \mathcal{F}_s) = \mathbb{P}(M_t \in S \,|\, M_s) = \mathbb{P}_{M_s}(M_{t-s} \in S)\,.$$

This is either obvious or impenetrable depending on one's background, so we omit further discussion. The end result is that we have constructed a continuous time Markov process $\{M_t\}_{t\ge 0}$ whose jumps occur at a countable dense set of times. If one restricts attention to a finite interval $[i_0(\omega), i)]$, then there are finitely many jumps in any finite time interval, and renaming these times $1, 2, 3, \ldots$ recovers the version of Gärtner, Henk and Ziegler's Klee-Minty chain that includes supressed moves (when an alarm rings at a site with a zero).

We are interested in the speed at which the leftmost 1 drifts to the right. To motivate the upcoming construction of a discrete-time Markov chain, we remark that it is easy to see that the intervals between shift times are independent mean 1 exponentials. It would therefore suffice to determine the long-run average of the sizes of the *shifts*, that is, the moves to the right in the position of the leftmost 1. A natural way to do this is to show that the process reaches some kind of equilibrium, and that in equilibrium the mean jump size is finite. As will be seen, we cannot complete this but can go far enough on this path to prove bounded speed. Clearly, the process as defined above does not reach equilibrium: it is constantly drifting to the right. We therefore define a Markov chain $\{Y_n\}$ which is the "view from the leftmost 1".

Define the space $\Xi := \{1\} \times \{0,1\}^{\mathbb{Z}^+}$ to be the subspace of sequences of 0's and 1's indexed by the nonnegative integers consisting of those beginning with 1. We define two functionals zeros and ones on Ξ by letting $\mathsf{ones}(\mathbf{x}) \ge 1$ be the number of leading 1's:

$$\mathsf{ones}(\mathbf{x}) := \inf\{j \ge 1 : \mathbf{x}(j) = 0\}$$

and letting $\mathsf{zeros}(\mathbf{x}) \geq 0$ be the number of successive 0's after the first 1:

$$\mathsf{zeros}(\mathbf{x}) := -1 + \inf\{j \geq 1 : \mathbf{x}(j) = 1\} .$$

For $j = 1, \ldots, \infty$ let ones_j denote the set $\{\mathbf{x} : \mathsf{ones}(\mathbf{x}) = j\}$ which form a partition of Ξ, and for $j = 0, \ldots, \infty$ let $\{\mathsf{zeros}_j\}$ denote the analogous partition with respect to the values of zeros.

We now define a continuous-time Markov process $\{X_t\}$ on the space Ξ whose law starting from $\mathbf{x} \in \Xi$ is denoted $\mathbf{Q_x}$. Pick $\omega \in \mathcal{S}$ such that the leftmost 1 of ω is in some position i and $\omega(i + j) = \mathbf{x}(j)$ for all $j \geq 0$. With this ω as the starting position, construct the continuous-time process $\{M_t\}_{t \geq 0}$ as above. We construct the Markov process $\{X_t\}$ on Ξ as a function of $\{M_t\}$ as follows. First, define σ_0 to be 0, and i_0 to be the position of the leading 1 in ω. Now recursively we let σ_n be the first time after σ_{n-1} for which $N(i_{n-1}, \cdot)$ increases. We let jump_n denote $\mathsf{ones}(\sigma_n^-)$ and we let $i_n = i_{n-1} + \mathsf{jump}_n$. Informally, σ_n is the n^{th} shift time and i_n is the position of the leftmost 1 at (and just after) σ_n. For $\sigma_{n-1} \leq t < \sigma_n$ we let $X_t \in \Xi$ be the configuration defined by $X_t(j) = M_t(j - i_{n(t)})$ where $n(t) = \sup\{n : \sigma_n \leq t\}$. Informally, X_t is M_t shifted so that i_{n-1} is at the origin (and ignoring negative indices).

Finally, we sample the process $\{X_t\}$ at shift times to produce a discrete-time chain. Thus we let

$$Y_n := X_{\sigma_n}$$

which is now a discrete-time Markov chain. We let $\mathbb{P}_\mathbf{x}$ denote the law of this chain starting from $\mathbf{x}$ and $\mathbb{P}_\nu$ denote the law when the starting state is picked from the probability measure ν.

1.4 Some problems associated with the model

Since the conditional distribution of $\sigma_n - \sigma_{n-1}$ given $\mathcal{F}_{\sigma_{n-1}}$ is exponential of mean 1, it follows that $\sigma_n/n \to 1$. The distance the leading 1 has moved to the right by time t is the sum $\sum_{n:\sigma_n \leq t} \mathsf{jump}_n$, and therefore the average speed $\mathsf{spd}(n)$ up to time σ_n is the random quantity

$$\mathsf{spd}(n) := \sigma_n^{-1} \sum_{j=1}^{n} \mathsf{jump}_j \sim n^{-1} \sum_{j=1}^{n} \mathsf{jump}_j . \tag{1}$$

It is not only *a priori* unclear, but also in fact we cannot prove that $\mathsf{spd}(n)$ has a limit as $n \to \infty$. Consequently we define

$$\begin{aligned}\mathsf{inf\text{-}spd} &:= \operatorname{liminf}_n \mathsf{spd}(n);\\ \mathsf{sup\text{-}spd} &:= \operatorname{limsup}_n \mathsf{spd}(n).\end{aligned}$$

Problem 1: Show that the limiting speed exists.

Note that computer simulations indicate that the limit exists and it might be around 1.76. The remainder of this section concerns ergodic properties of the continuous time model, and could be skipped by readers interested only in the discrete results and their proofs.

The space Ξ is compact in the product discrete topology. The Markov chain $\{Y_n\}$ is time-homogeneous, meaning that $\mathbb{P}(Y_{n+1} \in A \,|\, Y_n = \omega)$ is independent of n. Let μ_n denote the law of Y_n and $\nu_n := n^{-1} \sum_{k=1}^n \nu_k$; the set of probability measures on a compact space is compact, so although ν_n may not converge, there is at least one subsequential limit, call it ν, in the weak topology[6]. It is elementary to see that any such weak limit ν is a stationary distribution for the chain $\{Y_n\}$, meaning that under $\mathbb{P}_\nu$, each individual Y_n will have law ν.

Problem 2: Show that there is a unique stationary distribution for the chain $\{Y_n\}$.

This would imply a positive solution to Problem 1. To elaborate, a Markov chain with a unique stationary distribution is ergodic in the sense having a trivial invariant σ-field [Dur05, Section 6.1] (this implies the more commonly understood notion of ergodicity). From this it follows by Birkhoff's Ergodic Theorem [Dur05, Chapter 6 (2.1)] that the averages converge to the mean: with probability 1,

$$n^{-1} \sum_{j=1}^n \mathsf{jump}_j \to \int \mathsf{jump}_1 \, d\pi$$

where π is the unique stationary measure. Coupled with our indepedent proof that the speed is bounded, we could then conclude existence of the limiting speed, as well as the independence of this from the starting state. Although we do not have a solution to Problem 2, we believe something stronger may hold.

[6] The measure ν is a weak limit for $\{\nu_{n_k}\}$ in the product topology if $\nu_{n_k}(A) \to \nu(A)$ for any finitely determined set A of configurations, that is, any set of the form $\{\omega : \omega(j_k) = \epsilon_k, 1 \le k \le L\}$.

> **Problem 3a:** Assume there is a unique stationary measure, π, and let $T^j\pi$ denote the composition of the measure π with a translation by j bits, e.g., if A is the event that there is a 1 in position r, then $T^j\pi(A) = \pi(A^j)$, where A_j is the event that there is a 1 in position $r + j$. Prove that $T^j\pi \to M$, where M is independent fair coin-flipping.

Informally, this says we believe that once the process has been going a while, the bits far to the right of the leftmost 1 are nearly random.

> **Problem 3b:** Prove or disprove that the unique stationary measure π is equivalent (mutually absolutely continuous) to M.

Informally, this says that no definitive test can distiguish a single sample from π from independent fair coin-flips, though the two measures may give different likelihoods for the first few bits.

2 Statement of main result and lemmas

In this section we state the results that we do know how to prove, namely bounds on the lim inf and lim sup speeds.

Theorem 2 *Although the limiting speed of drift of the leftmost 1 is not known to exist, both its limsup and liminf are bounded on both sides by constants:*

$$1.646 < \mathsf{inf\text{-}spd} \le \mathsf{sup\text{-}spd} < 4.33\,.$$

Relating back to the performance of the random edge rule on Klee-Minty cubes, we have:

Corollary 3 *For sufficiently large n, starting from a uniform random vertex of the Klee-Minty cube,*

$$0.057n^2 \le E_n \le 0.152n^2\,.$$

Proof of Corollary 3 from Theorem 2: Gärtner, Henk and Ziegler consider another way of counting steps, where instead of choosing an edge at random among all those decreasing the objective function, they choose an edge at random from among all edges, but suppress the move if the edge increases the objective function. For a vector $\mathbf{x}$ of 0's and 1's of length n, let $L^*(\mathbf{x}^r) = L^*(\mathbf{x}^r, n)$ denote the expectation of the number $N^{(r)}$ of moves starting from $\mathbf{x}^r$, including the suppressed moves, before the minimum is reached, where $\mathbf{x}^r$ is the vector of length n, with the first $n - r$ digits 0 and the last r digits 1. They prove the following identity [GHZ98, Lemma 4].

$$E_n = \frac{1}{2n} \sum_{r=1}^{n} L^*(\mathbf{x}^r, n) \,. \tag{2}$$

Including suppressed moves in the count corresponds in our infinite, continuous-time model to counting the number of clock events (only among the first n vertices). Let $T^{(r)}$ be the time it takes starting from $\mathbf{x}^r$ to reach the minimum. To relate $T^{(r)}$ to $N^{(r)}$, note that the numbers of clock events between any two flips of the leftmost 1 are a sequence of independent geometric random variables of mean n: from the time the leftmost 1 enters a position, j, the locations of clock events are a sequence of uniform picks from $[1, n]$ ending when j is chosen. The strong law of large numbers implies that the average, A_k of the first k of these converges to n as $k \to \infty$; in fact it is easy to see that $A_k/(kn) \to 1$ as $k \to \infty$ uniformly in n. The quantity $N^{(r)}/T^{(r)}$ is such an average for a random k (the number of shift times before all n bits are zero) and since $k \to \infty$ in probability when $r \to \infty$, we have $N^{(r)}/(nT^{(r)}) \to 1$ as $r \to \infty$, except when $k = O(1)$, which happens with vanishing probability as $r \to \infty$. If the liminf and limsup speed are known to be in the interval (a, b), then with probability 1,

$$\frac{r}{b} < T^{(r)} < \frac{r}{a}$$

for sufficiently large r. Hence, for sufficiently large r, almost surely,

$$\frac{nr}{b} < N^{(r)} < \frac{nr}{a} \,.$$

Plugging into (2) and summing from $r = 1$ to n gives

$$\frac{n^2}{4b} < E_n < \frac{n(n+1)}{4a}$$

for sufficiently large n. Plugging in $b = 4.33$ and $a = 1.646$ proves the corollary. $\square$

The lower bound of Theorem 2 is proved in Section 4. The lower bound of 1.646 may in principle be improved so as to be arbitrarily near the actual speed. For the upper bound, we state some lemmas. Let

$$H_k := \sum_{j=1}^{k} \frac{1}{j}$$

denote the k^{th} harmonic number. Let $\{S_n\}$ be a random walk whose increments are equal k with probability $2/((k+1)(k+2))$ for each integer $k \geq 1$. Let $S := S_{G-1}$ where G is an independent geometric random variable with mean 2. Let Θ be a random variable satisfying

$$\mathbb{P}(\Theta \geq j) = 1 - F_\Theta(j-1) = \sum_{k=1}^{\infty} \frac{1}{k} \frac{1}{k+j} = \frac{H_j}{j} .$$

Assume $\{S_n\}$ and Θ are independent of each other and of $\{\mathcal{F}_t\}$; denote expectation with respect to $\mathbb{P}_{\mathbf{x}}$ by $\mathbb{E}_{\mathbf{x}}$ and let $\mathbb{E}$ denote expectation with respect to the laws of S and Θ. Analogously with $\mathsf{zeros}(\mathbf{x})$ we define the quantity

$$\mathsf{zeros}^*(\mathbf{x}) := \inf\{j \geq 0 : \mathbf{x}(\mathsf{ones}(\mathbf{x}) + j) = 1\}$$

to be the number of zeros after the first block of ones; thus $\mathsf{zeros}^*(\mathbf{x}) \geq 1$, $\mathsf{zeros}^*(\mathbf{x}) = \mathsf{zeros}(\mathbf{x})$ if and only if $\mathsf{zeros}(\mathbf{x}) \geq 1$ and $\mathsf{zeros}(\mathbf{x}) = 0$ if and only if $\mathsf{ones}(\mathbf{x}) \geq 2$.

Lemma 4 *For any* $\mathbf{x} \in \mathsf{ones}_j$,

$$\mathbb{E}_{\mathbf{x}} \mathsf{jump}_1 = \sum_{k=1}^{j} \frac{1}{k} .$$

Equivalently, for any $\mathbf{x} \in \Xi$,

$$\mathbb{E}_{\mathbf{x}} \mathsf{jump}_1 = H_{\mathsf{ones}(\mathbf{x})} .$$

Our key lemma is the following statement. Recall that Y_1 is the view from the leftmost 1 right at the first shift time.

Lemma 5 *For any* $j \geq 1$, *any* $\mathbf{x} \in \mathsf{zeros}_j$, *and any integer* $L \geq 1$,

$$\mathbb{P}_{\mathbf{x}}(\mathsf{ones}(Y_1) \geq L) \leq \mathbb{P}(S + j \geq L) ,$$

When $\mathsf{zeros}(\mathbf{x}) = 0$ *then*

$$\mathbb{P}_{\mathbf{x}}(\mathsf{ones}(Y_1) \geq L) \leq \mathbb{P}(\Theta + \mathsf{zeros}^*(\mathbf{x}) \cdot B + S \geq L)$$

where B is a Bernoulli with mean $1/2$, and Θ, B and S are all independent.

Since $S + \Theta + \mathsf{zeros}^*(\mathbf{x})$ is an upper bound for both quantities $S + \mathsf{zeros}^*(\mathbf{x})$ and $S + \Theta + \mathsf{zeros}^*(\mathbf{x}) \cdot B$ appearing as stochastic upper bounds in Lemma 5, and since H_n increases in n, we may put this together with Lemma 4 to obtain

Corollary 6 *For any* $\mathbf{x}$*,*

$$\mathbb{E}_{\mathbf{x}} \mathsf{jump}_2 \leq \mathbb{E} H_{\Theta + \mathsf{zeros}^*(\mathbf{x}) + S} \, .$$

$\square$

Lemma 7 *The conditional distribution of* $\mathsf{zeros}^*(Y_n)$ *given* $\mathcal{F}_{\sigma_{n-1}}$ *is always bounded above stochastically by the law of* Θ*. In other words, for all* j*,*

$$\mathbb{P}(\mathsf{zeros}^*(Y_n) > j \,|\, \mathcal{F}_{\sigma_{n-1}}) \leq 1 - F_\Theta(j) \, .$$

Proof of upper bound in Theorem 2 from the lemmas: It suffices to show that for any $\mathbf{x} \in \Xi$,

$$\mathbb{E}_{\mathbf{x}} \mathsf{jump}_3 \leq 2.92 \, .$$

(To explain the appearance of the 3: it suffices to show this for any jump_k, but it is false for jump_1 and jump_2.) We simply iterate conditional expectations and compute. By the Markov property, and Corollary 6,

$$\begin{aligned} \mathbb{E}_{\mathbf{x}} \mathsf{jump}_3 &= \mathbb{E}_{\mathbf{x}} \mathbb{E}_{Y_1} \mathsf{jump}_2 \\ &\leq \mathbb{E}_{\mathbf{x}} \left(\mathbb{E}_y H_{\mathcal{S} + \Theta + \mathsf{zeros}^*(y)} \right) \Big|_{y = Y_1} \, . \end{aligned}$$

Since H_n is increasing in n we may use the stochastic upper bound in Lemma 7 for any $\mathbf{x}$ to see that this is at most $\mathbb{E} H_{S + \Theta^{(2)}}$ where $\Theta^{(2)}$ is the sum $\Theta_1 + \Theta_2$ of two independent copies of Θ. The upper bound in Theorem 2 is completed by computing an upper bound for this.

The function H is concave and $\Theta^{(2)} \geq 1$, so

$$H_{\Theta^{(2)} + j} - H_{\Theta^{(2)}} \leq H_{j+1} - 1$$

and we may therefore conclude that

$$\mathbb{E} H_{\Theta^{(2)} + S} \leq \mathbb{E} H_{\Theta^{(2)}} + \mathbb{E} H_{S+1} - 1 \, . \tag{3}$$

To compute the quantity $\mathbb{E}H_{\Theta^{(2)}}$, use the identity $\mathbb{E}H_Z = 1 + \sum_{l\geq 2}(1/l)\mathbb{P}(Z \geq l)$ with $Z = \Theta_1 + \Theta_2$. Breaking the event $\{Z \geq l\}$ into the disjoint union of events $\{Z \geq l, \Theta_2 = k\}$, we have

$$\begin{aligned}
\mathbb{E}H_{\Theta_1+\Theta_2} &= 1 + \sum_{l\geq 2, k\geq 1} \frac{1}{l}\mathbb{P}(Z \geq l, \Theta_2 = k) \\
&= 1 + \sum_{1\leq j,k} \frac{1}{j+k}\mathbb{P}(\Theta_1 \geq j)\mathbb{P}(\Theta_2 = k) + \sum_{k\geq l\geq 2} \frac{1}{l}\mathbb{P}(\Theta_2 = k) \\
&= 1 + \sum_{1\leq j,k} \frac{1}{j+k}\frac{H_j}{j}\left(\frac{H_k}{k} - \frac{H_{k+1}}{k+1}\right) + \sum_{l\geq 2}\frac{1}{l}\mathbb{P}(\Theta_2 \geq l) \\
&= 1 + \sum_{1\leq j,k} \frac{1}{j+k}\frac{H_j}{j}\frac{H_{k+1}-1}{k(k+1)} + \sum_{l\geq 2}\frac{H_l}{l^2}\,. \qquad (4)
\end{aligned}$$

In a previous draft of this paper, we evaluated these quantities numerically, both rigorously and nonrigorously. The rigorous bounds for the first sum on the right-hand side of (4) showed it to be between 1.9975 and 2.00093 and nonrigorous numerical estimates show a result even closer to 2. While it is not relevant to the problem of the run time of the random edge rule on Klee-Minty cubes, we were led to state an obvious question:

Problem 4 (now solved) Prove or disprove that

$$\sum_{1\leq j,k} \frac{1}{j+k}\frac{H_j}{j}\frac{H_{k+1}-1}{k(k+1)} = 1\,.$$

In a surprising demonstration of computer assisted identity-proving, Carsten Schneider [PS04] has since evaluated this sum as exactly

$$-4\zeta(2) - 2\zeta(3) + 4\zeta(2)\zeta(3) + 2\zeta(5) = 0.999222\ldots$$

where $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$ is the classical zeta function. Similarly, the second sum is shown there to equal $2\zeta(3) - 1 \approx 1.4041138\ldots$. Thus we have

$$\mathbb{E}H_{\Theta^{(2)}} = 2\zeta(5) - 4\zeta(2)(\zeta(3) - 1)) = 3.4033\ldots\ . \qquad (5)$$

For the other term on the RHS of (3), let $\phi(z) = \mathbb{E}z^S = \sum_{n=0}^{\infty} z^n \mathbb{P}(S = n)$ be the generating function for S, so $z\phi$ is the generating function for $S+1$. For any positive integer j there is an identity

$$\int_0^1 \frac{1-z^j}{1-z}\, dz = \int_0^1 (1 + \cdots + z^{j-1})\, dz = H_j\,.$$

Consequently, we may write

$$\mathbb{E}H_{S+1} - 1 = \int_0^1 \frac{z - \mathbb{E}z^{S+1}}{1-z}\, dz = \int_0^1 z\frac{1-\phi(z)}{1-z}\, dz\,. \tag{6}$$

To compute the generating function ϕ, first compute the generating function f for the increments of $\{S_n\}$:

$$f(z) = \sum_{k=1}^{\infty} \frac{2}{(k+1)(k+2)} z^k = \frac{2z - z^2 - 2(1-z)\log\frac{1}{1-z}}{z^2}\,.$$

Since ϕ is the sum of $G-1$ increments, with $\mathbb{P}(G-1=k) = 2^{-k}$, and since the sum of k independent increments has generating function f^k,

$$\phi(z) = \sum_{k\geq 0} 2^{-k} f(z)^k = \frac{1}{2 - f(z)}$$

and the integral in (6) becomes

$$\mathbb{E}H_{S+1} - 1 = \int_0^1 \frac{2z(\log\frac{1}{1-z} - z)}{3z^2 - 2z + 2(1-z)\log\frac{1}{1-z}}\, dz\,.$$

One may evaluate this numerically to approximately 0.918797. In the appendix, we obtain a rigorous bound which is only slightly worse: 0.91905. Adding this to the value of $\mathbb{E}H_{\Theta^{(2)}}$ gives $4.322\ldots$, proving the upper bound in Theorem 2. $\square$

3 Proofs of Lemmas

We employ below the usual notation M_{t-} for $\lim_{s\uparrow t} M_t$. The idea behind most of the arguments here, which is equally valid in the discrete and continuous models, is that events of interest may be described combinatorially in terms of the relative orders of corresponding clock events, which are invariant under permuting indices.

Proof of Lemma 4: By definition, each $\mathbf{x} \in \mathsf{ones}_j$ begins with j 1's in positions $0, \ldots, j-1$ followed by a zero. The evolution of $\{M_t\}$ decreases the binary representation $\sum_k 2^{-k}\mathbf{x}(k)$, whence $M_{\mathsf{jump}_1^-} \in \mathsf{ones}_k$ for some $k \leq j$, that is, there is always a zero in some position in $[0, j]$. Furthermore, once there is a zero in position k for some $k < j$, then there is always a zero at or to the left of position k. It follows that jump_1 is equal to the least $k < j$ for which $\xi_{j,1} < \xi_{0,1}$, that is, for which the clock in position k goes off before the clock at position 0. The minimum is taken to be j if there is no such k.

It follows that $\mathbb{P}_{\mathbf{x}}(\mathsf{jump}_1 = j) = 1/j$ and that for $0 < k < j$,

$$\mathbb{P}_{\mathbf{x}}(\mathsf{jump}_1 = k) = \frac{1}{k(k+1)} \,. \tag{7}$$

To see this, note that $\mathsf{jump}_1 = k$ if and only if $\xi_{0,1}$ is the minimum of the variables $\{\xi_{k,1} : 0 \leq k < j\}$. Similarly, for $0 < k < j$, $\mathsf{jump}_1 = k$ if and only if $\xi_{k,1}$ is the minimum of the variables $\{\xi_{r,1} : 0 \leq r \leq k\}$ and $\xi_{0,1}$ is the next least of the values. Computing expectations via (7) proves the lemma. $\square$

Proof of Lemma 7: Again let us denote $q = \mathsf{ones}(\mathbf{x})$. We recall that $\mathbb{P}(\mathsf{jump}_1 = k) = 1/q$ for $k = q$ and $1/(k(k+1))$ for $1 \leq k \leq q-1$. We claim that for any $k \leq q$,

$$\mathbb{P}(\mathsf{zeros}^*(Y_1) \geq j \,|\, \mathsf{jump}_1 = k) \leq \frac{k+1}{k+j} \,. \tag{8}$$

If we can show this, then we will have

$$\mathbb{P}(\mathsf{zeros}^*(Y_1) \geq j) \leq \frac{1}{q}\frac{q+1}{q+j} + \sum_{k=1}^{q-1} \frac{1}{k(k+1)}\frac{k+1}{k+j} \,.$$

Changing q to $q+1$ increases this by

$$\frac{j-1}{(q+1)(q+j)(q+j+1)}$$

which is nonnegative. Setting $q = \infty$ then yields the upper bound in the lemma, and it remains to show (8).

Observe first that it suffices to show this for $k = q$. This is because when $k < q$, the event $\{\mathsf{jump}_1 = k\}$ necessitates $\xi_{k,1} = \min\{\xi_{r,1} : 0 \leq r \leq k\}$. Thus to evaluate

$\mathbb{P}(\mathsf{zeros}^*(Y_1) \geq j \,|\, \mathsf{jump}_1 = k)$ we may wait until time $\xi_{k,1}$, at which point if no bit to the left of k has flipped yet, the new conditional probability $\mathbb{P}(\mathsf{zeros}^*(Y_1) \geq j \,|\, \mathcal{F}_{\xi_{k,1}}, \mathsf{jump}_1 = k)$ is always at most $(k+1)/(k+j)$ by applying the claim for $q = k$.

Assuming now that $k = q$, we note that the event $\{\mathsf{jump}_1 = k\}$ that we are conditioning on is just the event

$$A := \{\xi_{0,1} = \min_{0 \leq i \leq k-1} \xi_{i,1}\}$$

that the clock at 0 goes off before any clock in positions $1, \ldots, k-1$. Conditioning on A then makes the law of $\xi_{0,1}$ an exponential of mean $1/k$ without affecting the joint distribution of $\{\xi_{r,s} : r > k\}$. Now let m_1 be the position at time $t_0 := 0$ of the first 1 to the right of k, and let t_1 be the time this 1 flips. Inductively, define m_{r+1} to be the position of the first 1 to the right of k after time t_r and let t_{r+1} be the first time after t_r that this 1 flips.

If the positions $m_r, \ldots, m_r + j - 1$ are not filled with ones at time t_{r-1} (define $t_0 = 0$) then it is not possible to have $\mathsf{zeros}^*(Y_1) \geq j$ and A and $t_{r-1} < \xi_{0,1} < t_r$. That is, one cannot get from fewer than j ones in the first block of ones to the right of k to at least j ones at the time of the flip at 0 without having the leftmost one in this block flip. On the other hand, if these j positions are filled with ones at time t_{r-1}, then

$$\mathbb{P}(\mathsf{zeros}^*(Y_1) \geq j, \xi_{0,1} < t_r \,|\, A, \mathcal{F}_{t_{r-1}}, \xi_{0,1} > t_{r-1}) \leq \frac{k}{k+j}$$

since the event $\{\mathsf{zeros}^*(Y_1) \geq j, \xi_{0,1} < t_r\}$ requires that the clock at 0 go off before the clocks in positions $m_r, \ldots, m_r + j - 1$ (recall that conditioning on A has elevated the rate of the clock at 0 to rate k). Similarly, $\mathbb{P}(\xi_{0,1} < t_r \,|\, A, \mathcal{F}_{t_{r-1}}, \xi_{0,1} > t_{r-1}) = k/(k+1)$. Therefore,

$$\frac{\mathbb{P}(\mathsf{zeros}^*(Y_1) \geq j, \xi_{0,1} < t_r \,|\, A, \mathcal{F}_{t_{r-1}}, \xi_{0,1} > t_{r-1})}{\mathbb{P}(\xi_{0,1} < t_r \,|\, A, \mathcal{F}_{t_{r-1}}, \xi_{0,1} > t_{r-1})} \leq \frac{k+1}{k+j}\,. \tag{9}$$

The LHS of (9) is $\mathbb{P}(\mathsf{zeros}^*(Y_1) \geq j \,|\, A, \mathcal{F}_{t_{r-1}}, t_r > \xi_{0,1} > t_{r-1})$, and considering (9) for every $r \geq 1$, we obtain that the RHS of (9) is an upper bound for the probability of $\mathsf{zeros}^*(Y_1) \geq j$ conditioned only on $\mathsf{jump}_1 = k$. $\square$

Proof of Lemma 5: Let $\mathbf{x} \in \mathsf{zeros}_j$ and first assume $j \geq 1$. We prove the statement for $n = 1$, the proof for greater n being identical, conditioned on $\mathcal{F}_{\sigma_{n-1}}$. It is simple to check whether $\mathsf{ones}(Y_1) = j$. The bits in positions $1, \ldots, j$ will remain 0's until the leading 1 flips at time σ_1, so the only thing to check is whether $\sigma_1 = \xi_{0,1}$ is less than or greater than $\xi_{j+1,1}$. With probability $1/2$, $\xi_{0,1} < \xi_{j+1,1}$ and in exactly this case $\mathsf{ones}(Y_1) = j$.

Now condition on this inequality going the other way: $\xi_{0,1} > \xi_{j+1,1}$. Let $t_1 := \xi_{j,1}$. Let $j+1+k_1$ be the position of the first 0 of $\mathbf{x}$ to the right of position $j+1$. Then at time t_1, the position of the first one to the right of $j+1$ is $j+Z_1$, where Z_1 is the least $l \in [1, k_1-1]$ for which $\xi_{j+1+l,1} < \xi_{j+1,1}$. If no such l exists, then $Z_1 = k_1$. We compute $\mathbb{P}_{\mathbf{x}}(Z_1 = l)$ as follows.

The variables $\{\xi_{r,1} : r \in \{0\} \cup [j+1, j+k_1]\}$ are independent exponentials. For $1 \le l < k_1$, the event that $\mathsf{ones}(Y_1) \neq j$ and $Z_1 = l$ is the intersection of the event A that $\xi_{j+1,1}$ is less than $\xi_{0,1}$ and $\xi_{r,1}$ for all $j+2 \le r \le j+l$ with the event B that $\xi_{j+1+l,1} < \xi_{j+1,1}$. In other words, among $l+2$ independent exponentials, the index of the least and second least must be $j+2$ and $j+1$ respectively. The unconditional probability of this is $1/((l+1)(l+2))$. Having conditioned on the larger event $\{\xi_{j+1,1} < \xi_{0,1}\}$, the conditional probability is therefore equal to $2/((l+1)(l+2))$. This holds for $l < k_1$, where $Z_1 = k_1$ with the complementary probability. To sum up, Z_1 is distributed as $S_1 \wedge k_1$ where S_1 has the distribution of the random walk increments described in the lemma.

The last step is to invoke the Markov property. Condition on $\mathcal{F}_{t_1}$. The chain from here evolves under the law $\mathbb{P}_{X(t_1)}$. Iterating the previous argument, there are two cases. The first case, which happens with probability 1/2 is that the clock at the origin goes off before the next alarm at location $j+1+Z_1$. In this case, $\mathsf{ones}(Y_1) = j + Z_1$. In the alternative case, we let t_2 be the time at which the clock at location $j+1+Z_1$ next goes off. We let Z_2 be the number of consecutive 1's at time t_2^- starting from position $j+1+Z_1$. Then conditional on $\mathcal{F}_{t_1}$, Z_2 is distributed as $S_1 \wedge k_2$ where k_2 is the number of consecutive 1's at time t_1 starting at position $j+1+Z_1$.

Iterating in this way, we have the following inductive definitions. Let $t_0 = 0$. Let τ be the least r for which the clock at the origin goes off after time t_r but before the first alarm at location $j+1+\sum_{i=1}^r Z_i$. For each $r \le \tau$, we may define k_r to be the number of consecutive 1's at time t_{r-1} starting at location $j+1+\sum_{i=1}^{r-1} Z_i$. We may then define t_r to be the first time after t_{r-1} that the alarm at location $j+1+\sum_{i=1}^{r-1} Z_i$ goes off, and we may define Z_r so that $j+1+\sum_{i=1}^r Z_i$ is the location of the first zero to the right of $j+1+\sum_{i=1}^{r-1} Z_i$ at time t_r^-.

The upshot of all of this is that

$$\mathsf{ones}(Y_1) = j + \sum_{i=1}^{\tau} Z_i$$

and that the joint distributions of τ and $\{Z_i : 1 \le i \le \tau\}$ are easily described. Conditioned on $\tau \ge r$ and on $\mathcal{F}_{t_r}$, the probability of $\tau = r+1$ is always $1/2$; as well, Z_{r+1} given $\tau \ge r+1$ and $\mathcal{F}_{t_r}$ is always distributed as a truncation of S_1. We conclude that $\mathsf{ones}(Y_1)$ is stochastically dominated by the sum of τ independent copies of S_1, hence as S_{G-1} (we say that X is *stochastically dominated* by Y, denoted $X \preceq Y$, if for every t the inequality $\mathbb{P}(X > t) \le \mathbb{P}(Y > t)$ holds).

Finally, we consider the case where $\mathsf{zeros}^*(\mathbf{x}) = l > \mathsf{zeros}(\mathbf{x}) = 0$. Let $q = \mathsf{ones}(\mathbf{x})$, so that $\mathbf{x}$ begins with q ones, followed by l zeros, followed by a one in position $q+l$. A preliminary observation is that if we begin with a one at the origin, the position $W(t)$ of the leading one at a later time t is an increasing function of t; hence, if T_μ is an independent exponential with mean μ, the distribution of $W(T_\mu)$ is stochastically increasing in μ.

Begin by writing

$$\mathbb{P}_{\mathbf{x}}(\mathsf{ones}(Y_1) \ge j) = \sum_{k=1}^{q} \mathbb{P}(\mathsf{ones}(Y_1) \ge j, \mathsf{jump}_1 = k)\,.$$

Let l^* denote the number of zeros consecutively starting from position k at time $\xi_{k,1}$ if $\mathsf{jump}_1 = k < q$, and $l^* = l$ if $k = q$. In other words, $l^* = \mathsf{zeros}^*(\mathbf{x}')$ where $\mathbf{x}'$ is the word at the last time t that ones changes before the leading bit flips ($t = \xi_{\mathsf{jump}_1,1}$ if $\mathsf{jump}_1 < q$ and $t = 0$ otherwise). We may then describe $\mathsf{ones}(Y_1)$ as $l^* + W$, where W is the number of consecutive positions starting from position $\mathsf{jump}_1 + l^*$ that turn to zeros between time t and $\xi_{0,1}$. Now we break into two cases.

Condition first on $\{\mathsf{jump}_1 = q\}$. The time $\xi_{0,1}$ is now an exponential of mean $1/q$, and before this time, the bits from position $q+l$ onward evolve independently. We may describe $\mathsf{ones}(Y_1)$ as $l + W(\xi_{0,1})$, where W is the number of consecutive positions starting at $q+l$ which have become zeros in the time from 0 to $\xi_{0,1}$. The first part of this lemma established that when ξ has rate 1, then $W(\xi) \preceq S$. Our preliminary observation now shows, conditional on $\{\mathsf{jump}_1 = q\}$, that $W(\xi_{0,1}) \preceq S$).

Next, let us condition on $\mathsf{jump}_1 = k < q$, observing that then $l^* \le q - k$. In order to have $l^* \ge r$, it is necessary that $\xi_{k,1} = \min\{\xi_{s,1} : k \le s \le k+r-1\}$. Having conditioned on $\mathsf{jump}_1 = k$, the distribution of $\xi_{k,1}$ becomes an exponential of rate $k+1$, so that the conditional probability of this clock going off before $r-1$ other conditionally independent clocks of rate 1 is just $(k+1)/(k+r)$. Since the event $\{\mathsf{jump}_1 = k < q\}$ has probability

$1/(k(k+1))$ if $k < q$ and zero otherwise, we may remove the conditioning and sum to get

$$\mathbb{P}(l^* \geq r, \mathsf{jump}_1 < q) \leq \sum_{k=1}^{q-1} \frac{1}{k(k+1)} \frac{k+1}{k+r} \leq \mathbb{P}(\Theta \geq r)\,.$$

We also still have in this case $W \preceq S$.

Putting together the cases $\mathsf{jump}_1 = q$ and $\mathsf{jump}_1 < q$, we see that $l^* = l$ with probability $1/q$ and otherwise $l^* \preceq \Theta$. The crude bound $1/q \leq 1/2$ gives $l^* \preceq \Theta + \mathsf{zeros}^*(\mathbf{x}) \cdot B$. Since $l^* \in \sigma(\mathcal{F}_t)$ and the bound $W \preceq S$ holds conditionally on $\mathcal{F}_t$, we arrive at $\mathsf{ones}(Y_1) \preceq \Theta + \mathsf{zeros}^*(\mathbf{x}) \cdot B + S$. $\square$

4 Lower bound

The argument for the lower bound in Theorem 2 is a generalization of the argument for the easiest nontrivial lower bound, which goes as follows.

Since the number of shift times in $[0, T]$ is asymptotically T, so $\mathsf{inf\text{-}spd}$ is the same as the liminf average of the sizes of the first n shifts. The trivial lower bound is 1 since each shift has size at least 1. We can improve this by showing a greater jump happens a positive fraction of the time. States ω with $\mathsf{ones}(\omega) \geq 2$ are helpful because from such a state there is a chance of $1/2$ that $\mathsf{jump}_1 \geq 2$. On the other hand, if $\mathsf{ones}(\omega) = 1$ then the prefix of length 3, which we will denote $\omega|_3$, is either 100 or 101. The more favorable prefix is 100: although the first shift will have size one, the new state will have prefix 11, which generates a shift of size at least 2 half the time. The least favorable prefix is 101, but here it can at least be said that if the 1 at the end flips before the 1 at the beginning, which happens half the time, we arrive at the prefix 100. Summarizing, we partition into the three prefix classes $11*, 100*, 101*$, and record the sequence of these. The long run transition frequencies are not all known, but can be summarized in the following diagram, where transitions back to the same state are permitted.

$$\begin{tikzcd}
 & 101* \arrow[d, "p \geq \frac{1}{2}"] \\
11* & 100* \arrow[l, "p = 1"]
\end{tikzcd}$$

It is easy to show that the worst case consistent with this is $11* \to 101*$ with probability 1 and $101* \to 101*$ with probability $1/2$. One finds that one prefix in four is $11*$, generating a shift of size at least 2 one eighth of the time, so $\mathsf{inf\text{-}spd} \geq 9/8$.

To improve this argument, we do two things: (i) replace the prefix partition by a larger one and (ii) take advantage of the fact that not every change in prefix counts in the denominator of

$$\frac{\text{size of shift}}{\text{number of shifts}} .$$

In fact the $1/8$ above is immediately improved to $1/6$ by noticing that there is no shift when the prefix changes from $101*$ to $100*$ via a flip of the 1 at the end.

To carry out (i), consider a tree whose vertices are positive integers, identified with their binary expansions. The root is 1, and the children of x are $2x$ and $2x+1$. Let $\mathcal{T}$ be any finite binary rooted subtree, meaning that any vertex in the subtree has either zero or two children in the subtree. The leaves of $\mathcal{T}$ are the prefix partition and we will pay attention to precisely those bit-flips that change which element of the partition we are in. It will remain to choose $\mathcal{T}$ judiciously. Associated with each node x is a set $U(x)$ of other nodes to which transitions are possible from x. The cardinality of $U(x)$ is the number of 1's in the binary expansion of x and $y \in U(x)$ if the bit string for y is obtained from the bit string for x by flipping one of the 1's in x, simultaneously flipping all bits to the right of this 1, and if a shift occurs, also shortening the new bit string by the size of the shift. By convention, if x is all 1's and a shift occurs, the new string is 1 rather than the empty string. Note that all transitions from x are to numbers less than x; in fact the reason we have encoded as integers as well as bit strings is so that we may proceed by induction on x.

To carry out (ii), we define for each transition from a state x to a state y both a reward and a counter as to whether we have observed a shift. The reward, denoted $r(x, y)$ is defined to be one less than the number of leading 1's of x if a shift occurs, and zero otherwise. The time counter increments by $t(x, y)$, defined to be 1 if a shift occurs and zero otherwise.

The idea is now to assign recursively a least expected reward per time from each node of $\mathcal{T}$. We do this by assuming the worst possible values for information beyond the prefix we are keeping track of. We improve on the naive argument in one more way: if the pessimistic mean reward per time we have calculated at an internal node x is worse than it is for the worst prefix, we can go ahead and look at some more information, arriving at some leaf of

$\mathcal{T}$. We will have to assume arrival at the worst one, but that is still an improvement. The set of nodes, B, at which we do this will also be chosen judiciously.

We are now ready to state and prove the validity of the computing apparatus. Assign to each x a mean reward and shift counter $(r(x), t(x))$ defined recursively as follows. Fix any subset B of internal nodes of $\mathcal{T}$ that contains the root (on first reading, one may imagine B to contain only the root). The base step of the recursion is to let $(r(x), t(x)) := (0, 0)$ if $x \in B$. The recursive step is to define

$$(r(x), t(x)) := \frac{1}{|U(x)|} \sum_{y \in U(x)} ((r(x, y), t(x, y)) + (r(y), t(y))) \ .$$

Lemma 8 *An almost sure lower bound for the lim inf speed from any starting configuration is given by the minimum over leaves x of T of $1 + r(x)/t(x)$.*

The lower bound in Theorem 2 will follow from Lemma 8 together with an implementation of the recursion. At the URL

`http://www.math.upenn.edu/~pemantle/papers/C-link`

is some code written in C that implements the recursion for a complete binary tree of depth 15, with the set B chosen to give a good bound without much trouble. A look at the data shows the minimum value of $r(x)/t(x)$ on each level to be obtained when the binary expansion of x alternates. In particular, the global minimum is at $x = 349525$ and has value $0.646\ldots$, which proves the lower bound, assuming the lemma.

Proof of Lemma 8: Any finite rooted binary subtree induces a prefix rule, that is, a map η from infinite sequences beginning with a 1 to leaves of $\mathcal{T}$, defined by $\eta(\mathbf{x}) = w$ for the unique leaf of $\mathcal{T}$ that is a prefix of η.

Given a trajectory of the Markov process $\{X_t\}$, define a sequence of elements of $\mathcal{T}$ as follows. Let $x_0 := \eta(X_0)$ be the prefix of the initial state of the trajectory. Let $\tau_0 := 0$. As the definition proceeds, verify inductively that for $\tau_k \leq t < \tau_{k+1}$, the string x_k will be an initial segment of X_t. The recursion is as follows. Let τ_{k+1} be the first time after τ_k that a 1 flips in the initial segment x_k of X_t. Let x'_{k+1} be the string gotten from x_k by flipping this bit and all bits to its right, removing any leading zeros and setting the string equal to 1 if it is empty. If $x'_{k+1} \notin B$ then let x_{k+1} be x'_{k+1}. $x'_{k+1} \in B$, then let $x_{k+1} = \eta(X_{\tau_{k+1}})$; we say the $\{x_k\}$ chain is "bumped" back into the prefix set at time $k+1$.

Given $j \geq 0$, let $\rho(j) = \inf\{k > j : x'_k \in B\}$ be the next time that the $\{x_k\}$ chain is bumped back to the prefix set. Given $x \in \mathcal{T} \setminus B$ and $j \geq 0$ for which $x_j = x$,

$$R(x, j) = \sum_{i=j}^{\rho(j)-1} R(x_i, x_{i+1})$$

denote the total reward from time τ_j until it is bumped back to the prefix set. Similarly, let

$$T(x, j) = \sum_{i=j}^{\rho(j)-1} T(x_i, x_{i+1})$$

denote the total shift count until being bumped back.

Claim: The expected accumualations of reward and shift count between τ_j and $\tau_{\rho(j)}$ are given by r and t. That is, if $x_j = x$, then

$$\mathbb{E}(R(x, j) \,|\, \mathcal{F}_{\tau_j}) = r(x)$$

and

$$\mathbb{E}(T(x, j) \,|\, \mathcal{F}_{\tau_j}) = t(x)\,.$$

Proof of claim: This is just induction on x. For $x \in B$ both sides of both equations are zero. Assuming this for nodes in $U(x)$, observe that $R(x, j) = R(x, x_{j+1}) + R(x_{j+1}, j+1)$ so by the Markov property, $\mathbb{E}(R(x, j) \,|\, \mathcal{F}_{\tau_j})$ is the average of $r(x, y) + \mathbb{E}R(y, j+1)$ over $y \in U(x)$, which by the induction hypothesis is the average over $y \in U(x)$ of $r(x, y) + r(y)$, which is equal to $r(x)$ by the recursive definition. The same holds for T.

Returning to the proof of the lemma, the sequence of times defined by $s_0 = 0$ and $s_{i+1} = \rho(s_i)$ breaks the path of the $\{x_k\}$ chain into the disjoint union of segments $[s_i, s_{i+1} - 1]$. For each $i > 0$, the state s_i is a leaf of $\mathcal{T}$. Among those leaves x occurring infinitely often, the strong law of large numbers implies that the average reward collected over intervals $[s_i, s_{i+1} - 1]$ for which $s_i = x$ is equal to $r(x)$. Similarly, the average shift count increment over such intervals is $t(x)$. Therefore,

$$\liminf_{i \to \infty} \frac{\sum_{j=0}^{s_i} r(x_j, x_{j+1})}{\sum_{j=0}^{s_i} t(x_j, x_{j+1})} \geq \min \frac{r(x)}{t(x)}$$

where the minimum is over leaves x of $\mathcal{T}$.

Finally, we note that the number of times the leftmost 1 moves up to time τ_{s_i} is exactly the total shift count $\sum_{j=0}^{s_i} t(x_j, x_{j+1})$, while the total distance it has moved is at least $\sum_{j=0}^{s_i} t(x_j, x_{j+1}) + r(x_j, x_{j+1})$. Thus the liminf speed is at least $1 + \min r(x)/t(x)$ and the lemma is proved. $\square$

We remark that the only place this last inequality is not sharp is when the number of 1's flipping together exceeds the number recorded, because the present knowledge of the prefix was a string of all 1's and there were more 1's after this that also flipped. Thus by making the tree $\mathcal{T}$ big enough, even without increasing B, we can get arbitrarily close to the true value.

5 Further observations

The following argument almost solves Problem (3a), and perhaps may be strengthened to a proof. Lemma 4 of [GHZ98] is proved by means of a duality result. The result is that the probability, starting from a uniform random state, of finding a 1 in position r after t steps (counting suppressed transitions), is equal to half the probability that $\mathbf{x}^r$ has not reached the minimum yet after t steps (again counting suppressed transitions). The argument that proves this may be generalized by introducing a simultaneous coupling of the process from all starting states. The probability, from a uniform starting state, of finding a 1 in every position in a set A after t steps, is then the expectation of the function that is zero if the column vector of all 1's is not in the span of the columns of the matrix whose rows are the states reached at time t starting at $\mathbf{x}^r$, as r varies over A, and is 2^{-u} if the column vector of all 1's is in the span and the matrix has rank u. The kernel of the matrix is the set of starting configurations that reach the minimum by time t (the simultaneous coupling is linear). Hence, as long as A and t are such that the probability of reaching the minimum from any $\mathbf{x}^r$ by time t goes to zero, the rank of the matrix will be $|A|$ and the probability of finding all 1's in positions in A at time t will go to $2^{-|A|}$. In particular, if a window of fixed size moves rightward faster than the limsup speed, then what one sees in this window approaches uniformity. This is not good enough to imply uniformity of a window a fixed distance to the right of the leftmost 1.

Acknowledgments: We would like to thank Attila Pór for sharing this problem, and to Jiři Matoušek giving references about the Klee-Minty cube. Thanks also to Tibor Szabó for pointing out an error in our understanding of [GHZ98, Lemma 4]. We also are indebted to the referees for their many comments and corrections.

6 Appendix: rigorous upper bounds for integrals

by Kate Davidson

To bound

$$\int_0^1 \frac{2z(\log \frac{1}{1-z} - z)}{3z^2 - 2z + 2(1-z)\log \frac{1}{1-z}}\, dz$$

from above, we will use a trapezoidal approximation. For this to give a valid upper bound we need to show

Lemma 9 *The function*

$$g(z) := \frac{2z(\log \frac{1}{1-z} - z)}{3z^2 - 2z + 2(1-z)\log \frac{1}{1-z}}\, dz$$

is convex on $(0,1)$.

Proof: Differentiate g twice with respect to z, plug in $z = 1 - \exp(-x)$, and multiply top and bottom by e^{3x} to obtain the quantity $h(x) = N(x)/D(x)$ where

$$\begin{aligned} N(x) \;\; := \;\; & 2e^{5x} + (20x - 42)e^{4x} + (108 + 48x - 72x^2 + 16x^3)e^{3x} \\ & \qquad + (-92 - 168x + 72x^2)e^{2x} + (18 + 112x)e^x + 6\,. \end{aligned}$$

and $D(x)$ is a power of e^x times a power of $e^{2x} + (2x-4)e^{-x} + 3$.

The Taylor series for $D(x)$ is easily seen to have all nonnegative coeffiecients (just check that the zeroth and first coefficients vanish and note that the general coefficient is $(2^n + 2n - 4 + 3\delta_{n,0})/n!$. In particluar, $D(x) > 0$ on $(0, \infty)$. To see that $N(x)$ is positive, note that the coefficient of $x^n/n!$ in the Taylor expansion is simply

$$108{\cdot}3^n - \frac{88}{9}\,3^n n^2 + \frac{680}{27}\,3^n n + \frac{16}{27}\,3^n n^3 + 5\,4^n n - 92\,2^n - 42\,4^n - 102\,2^n n + 2\,5^n + 18 + 18\,2^n n^2 + 112\,n$$

When $n \geq 6$, we have $(18n^2 - 102n)2^n > 0$. When $n \geq 14$, we have $2 \cdot 5^n - 42 \cdot 4^n > 0$. When $n \geq 17$, we have $((16/27)n^3 - (88/9)n^2)3^n > 0$. Since $108 \cdot 3^n - 92 \cdot 2^n > 0$ for all n, we see that the coefficients are positive once $n \geq 17$. Computing the first 16 terms, one sees they are positive as well, finishing the proof of the lemma. $\square$

Having proved this, we conclude that the integral on $[0, 0.999]$ may be bounded above by a polygonal path, with vertices at every multiple of 0.001 in the ordinate. This gives an upper bound of 0.90523 for the integral on $[0, 0.999]$. For the upper end of the integral we use the fact that the denominator of g is bounded below by 1 on roughly the interval $[0.85, 1]$ and provably containing $[0.999, 1]$. Thus

$$\begin{aligned}
& \int_{0.999}^{1} g(z)\,dz \\
\leq\ & \int_0^1 2z \left(\log \frac{1}{1-z} - z\right) dz \\
=\ & \left[2\,(z-1)\ln\left(-(z-1)^{-1}\right) + z - 3/2 + \ln\left(-(z-1)^{-1}\right)(z-1)^2 + 1/2\,z^2 - 2/3\,z^3\right]\Big|_{z=0.999}^{z=1} \\
\leq\ & 0.01382
\end{aligned}$$

Thus we have shown that the whole integral is at most 0.91905. Finally, putting this together with the upper bound for the double sum yields an upper bound of $2.91998 < 2.92$.

The numerical computations showing that

$$0.9975 < \sum_{1 \leq j,k} \frac{1}{j+k} \frac{H_j}{j} \frac{H_{k+1} - 1}{k(k+1)} < 1.00093$$

are available from the authors, though we have omitted them because the exact value is now known.

References

[AC78] Avis, D. and Chvatal, V. (1978). Notes on Bland's pivoting rule. In: Polyhedral Combinatorics, *Math. Programming Study* **8**, 24–34.

[Dur05] Durrett, R. (2005). *Probability, theory and examples.* Third Edition. Brooks/Cole: Belmont, CA.

[GHZ98] Gärtner, B., Henk, M. and Ziegler, G. (1998). Randomized simplex algorithms on Klee-Minty cubes. *Combinatorica* **18**, 349–372.

[GS01] Grimmett, G. R. and Stirzaker, D. R. (2001). Probability and Random Processes. Third ed. Oxford University Press.

[Kel81] Kelly, D. (1981). Some results on random linear programs. *Methods of Operations Research* **40**, 351–355.

[KM72] Klee, V. and Minty, G. (1972). How good is the simplex algorithm? In: *Inequalities III* (O. Sisha, Ed.). Academic Press: New York.

[MSZ04] Matoušek J. and Szabó T. (2004). Randomized edge can be exponential on abstract cubes. to appear FOCS 2004.

[PS82] Papadimitriou, C. and Steiglitz, K. (1982). *Combinatorial Optimization: Algorithms and Complexity*. Prentiss-Hall, Inc.: Englewood Cliffs, NJ.

[PS04] Pemantle, R. and Schneider, C. (2004). When is $0.999\ldots$ equal to 1? *Preprint.*

[Z95] Ziegler, G. M. (1995). Lecture on Polytopes. Graduate Texts in Mathematics **152**. Springer-Verlag, New York 1995.