The square lattice shuffle

Johan Håstad
Royal Institute of Technology
Stockholm, Sweden

October 20, 2005

Abstract

We show that the operations of permuting columns and rows separately and independently mix a square matrix in constant time.

1 Introduction

The number of iterations needed to achieve a (close to) random permutation on a large set of elements by repeatedly performing a random permutation from a limited class is the central question of shuffling. This question has lead to interesting and beautiful mathematics [3]. For the most famous class of shuffles, the riffle shuffle, the time to get close to a random permutation has been determined up to a constant [2] and many other sets of limited permutations have been analyzed. One interesting case that has proved to be difficult to analyze is the Thorp shuffle, introduced in [5].

In the Thorp shuffle we have a deck of cards with an even number of cards which is split into two piles of equal size. Then we do a rather special riffle shuffle, we let go one card from one of the piles and then one from the other. Thus a step is given by releasing two cards, one from each pile, and the only randomness is the order in which we release the two cards. This is repeated until all cards are dropped.

If the number of cards is a power of two then, as also described in [4], this shuffle has an alternative, very convenient description. Put the cards at the corners of the hypercube. Consider all the edges along one fixed dimension and for each edge independently interchange the two cards at the endpoints. Now deterministically move each card to an address which, when written in binary, is one cyclic shift of its current address to the right and repeat this process. With this description of the Thorp shuffle it is clear that the interesting part to analyze is the random step exchanging elements along edges in the fixed dimension as this is the only probabilistic step.

Up to recently not much has been known on the number of iterations of the Thorp shuffle needed to get close to the uniform distribution on permutations but, for the case of $n = 2^d$, Morris [4] has obtained the bound $O(d^{44})$, proving that the mixing time is at most poly-logarithmic in the number of cards.

Motivated by the problem on the Thorp shuffle and as an appealing problem in general we propose to study the mixing of m^d elements distributed in a d-dimensional cube where we at each point in time make a random permutation of the m elements with all but one coordinate fixed. In this paper we analyze the case $d = 2$ and prove that a constant number of repetitions is sufficient for mixing, independently of the value of m.

2 Preliminaries

We study permutations on n elements organized in an $m \times m$ square created as follows.

Definition 2.1 *At each time step m permutations of $[m]$, $(\sigma_i)_{i=1}^m$ each on m elements, are picked independently and and uniformly at random. At even time steps, for $0 \leq i < m$, σ_i is applied to the elements in row i while at odd steps it is applied to column i. Repeating this process for t time steps with independent choices at each point in time creates a random permutation from the distribution Π_t.*

We want to prove that the distribution Π_t is not too different from the uniform distribution on permutations. Our measure of "different" is given by the statistical distance.

Definition 2.2 *The statistical distance between two probability distributions Π^1 and Π^2 is given by*

$$\Delta(\Pi^1, \Pi^2) = \sum_{\pi} |Pr_{\Pi^1}(\pi) - Pr_{\Pi^2}(\pi)|.$$

We let U_n be the uniform distribution on permutations on n elements. It is useful for us to study what happens when we compose a random permutation with a fixed permutation and we let $\Pi \circ \pi$ be the result of composing a random permutation from distribution Π with π.

Lemma 2.3 *We have*

$$\Delta(\Pi, U_n) \leq E_{\pi_1,\pi_2}[\Delta(\Pi \circ \pi_1, \Pi \circ \pi_2)],$$

where π_1 and π_2 are picked independently with the uniform distribution.

Proof: Let id denote the identity permutation. As

$$\Delta(\Pi \circ \pi_1, \Pi \circ \pi_2) = \Delta(\Pi \circ id, \Pi \circ \pi_2 \circ \pi_1^{-1}),$$

we can assume that $\pi_1 = id$. The probability, over a random π_2, that $\Pi \circ \pi_2$ takes any specific value is $(n!)^{-1}$ and hence

$$\begin{aligned}\Delta(\Pi, U_n) &= \sum_{\pi} |Pr_{\Pi}(\pi) - (n!)^{-1}| = \sum_{\pi} |Pr[\Pi \circ id = \pi] - E_{\pi_2}[Pr[\Pi \circ \pi_2 = \pi]]| \leq \\ & \sum_{\pi} E_{\pi_2}[|Pr[\Pi \circ id = \pi] - Pr[\Pi \circ \pi_2 = \pi]|] = E_{\pi_2}[\Delta(\Pi \circ id, \Pi \circ \pi_2)].\end{aligned}$$

■

We get the following consequence

Lemma 2.4 *We have*

$$\Delta(\Pi, U_n) \leq n \max_{\pi_1, \pi_2} \Delta(\Pi \circ \pi_1, \Pi \circ \pi_2),$$

where the maximum is taken over any pair of permutations that differ by a transposition.

Proof: This follows from Lemma 2.3, the triangle inequality, and the fact that we can move from any permutation π_1 to a permutation π_2 by n transpositions. ■

We prove that $\Delta(\Pi \circ \pi_1, \Pi \circ \pi_2)$ is small for any pair of permutations different by a transposition by a coupling argument. An introduction to coupling can be found in Chapter 4 of [3] but let us describe how it is used here. We have a probability distribution over pairs (τ_1, τ_2) of permutations such that the marginal probability distribution of the first component is that of Π and the same applies to the second component. Furthermore we have that the property that for two given permutations π_1 and π_2 different by a transposition, we have that $\tau_1 \circ \pi_1 = \tau_2 \circ \pi_2$ holds with probability at least $1 - q$ while, otherwise, $\tau_1 \circ \pi_1$ and $\tau_2 \circ \pi_2$ differ by a transposition. We call this property *coupling of permutations of distance two* with parameter q. The coupling is constructed in such a way that the number q is independent on the choice of the permutations π_1 and π_2. We have the following easy lemma.

Lemma 2.5 *Let Π^t be the composition of t independent copies of Π and suppose Π allows a coupling of permutations of distance two with parameter q. Then*

$$\Delta(\Pi^t \circ \pi_1, \Pi^t \circ \pi_2) \leq q^t.$$

for any pair of permutations π_1 and π_2 different by a transposition.

Proof: By straightforward induction on t, π_1 and π_2 couples under Π^t except with probability q^t. Let (τ_1^t, τ_2^t) give this coupling. Then

$$\begin{aligned}\Delta(\Pi^t \circ \pi_1, \Pi^t \circ \pi_2) &= \sum_{\sigma} |Pr[\tau_1^t \circ \pi_1 = \sigma] - Pr[\tau_2^t \circ \pi_2 = \sigma]| \leq \\ Pr[\tau_1^t \circ \pi_1 \neq \tau_2^t \circ \pi_2] &\leq q^t.\end{aligned}$$

■

3 The main result

Our main result is

Theorem 3.1 *Let Π_t be the distribution of Definition 2.1. Then*

$$\Delta(\Pi_t, U_n) \leq O(n^{1-\lfloor \frac{t}{3} \rfloor \frac{1}{4}}(\log n)^{\lfloor \frac{t}{3} \rfloor}).$$

Thus already for $t = 15$ the distance is small for large n and we have mixing in constant time independently of n.

We establish the theorem by, as discussed in the preliminaries, constructing a coupling and the key lemma is the following.

Lemma 3.2 *The distribution Π_3 allows a coupling of permutations of distance two with parameter $O((m/\log m)^{-1/2})$.*

As rows and columns are symmetric, Lemma 2.4, Lemma 2.5, and Lemma 3.2 jointly imply Theorem 3.1. We proceed to establish Lemma 3.2.

Proof: (Of Lemma 3.2) We let the coupling take place under three steps of the procedure, the first being a row-permutation. We have two permutations π_1 and π_2 that differ only by a transposition. First observe that if the positions in which the two permutations differ are located in the same row then we have an obvious coupling that couples the permutations in one step with probability one and hence we we can assume that positions in which the permutations differ are in different rows.

As we do independent random permutations of the rows, the relative starting positions between different rows are of no consequence. We may hence assume that the two positions in which π_1 and π_2 differ are in different columns. As all rows and columns are equivalent we may hence assume that one position is $(0, 0)$ and the other is $(1, 1)$.

Before we continue we need to make some preliminary observations and to establish some notation.

The main information we are interested in after one row-permutation and one column-permutation is the row in which each element ends up. We define a *pattern* to be an $m \times m$ square filled with numbers from 0 to $m - 1$. An r in position (i, j) means that the element in the ith row and jth column ends up in row r.

Each number appears exactly m times in a pattern and any pattern with this property is possible. To see the latter consider the following bipartite multigraph on $2m$ vertices. It has m vertices to the left labeled $\{0^-, 1^-, \ldots (m-1)^-\}$ and m vertices to the right labeled $\{0^+, 1^+, \ldots (m-1)^+\}$. There is an edge for each element in the matrix and if the element is in row i and takes the value r the edge is between i^- to r^+. As several elements in the same row might take the same value this creates a multi-graph and to be able to speak of edges in an unambiguous way we label each edge by the starting column of its element. As each row has m positions and each r appears m times we have an m-regular bipartite graph.

To achieve a particular pattern, use the fact that each regular bipartite graph can be written as a union of matchings. Write the graph as the union of m matchings, $\overline{M} = M_0,\, M_1, \ldots M_{m-1}$. We call $\overline{M}$ a *matching cover* and it defines a set of row permutations and column-permutations as follows. Each (labeled) edge in $\overline{M}$ corresponds to an element in a unique way, the left hand endpoint giving the starting row and the label giving the starting column. If this element has value r and belongs to M_j then it is moved to column j by the row permutation and then to the row r by the column permutation.

A matching M_i can, in the natural way, be thought of as a permutation. We use the same notation for this permutation as there is no need to distinguish the two interpretations of the object.

We conclude that one matching cover corresponds uniquely to one set of row and column permutations while the same pattern is obtained by many different matching covers. The probability space on patterns we consider is the one given by picking a random matching cover uniformly from all possible matching covers and then using the corresponding pattern.

As a change in the column-permutation on column i results in the corresponding change in the permutation corresponding to M_i and vice versa, we conclude that the the set of permutations (disregarding labels) occurring as M_i $1 \leq i \leq m$ are uniformly random and independent of each other.

We proceed to construct a coupling of patterns by appending a row-permutation. The patterns that give the same value to $(0,0)$ and $(1,1)$ move, by definition, our two special elements to the identical rows. The subsequent row permutation will give a perfect coupling and hence we need to study patterns giving different values to $(0,0)$ and $(1,1)$.

Construct randomly a partial pattern p containing m^2-2 values to all squares except $(0,0)$ and $(1,1)$. To be precise p is constructed by picking a full random pattern (as induced by picking a random matching cover) and then erasing the contents of these two positions. We want to prove that for most p the two ways of completing the pattern are approximately equally likely. To be more specific, let p_1 and p_2 be the two ways to complete the pattern, then we have the following lemma.

Lemma 3.3 $\sum_p |Pr[p_1] - Pr[p_2]| \leq O\left(\left(\frac{\log m}{m}\right)^{1/2}\right)$.

Let us first see that Lemma 3.3 is sufficient to establish Lemma 3.2 by defining a suitable coupling.

First with probability $(Pr[p_1] + Pr[p_2])$ decide that the patterns of both chains are either p_1 or p_2. Suppose for concreteness that $Pr[p_1] \geq Pr[p_2]$. Then with probability $Pr[p_2](Pr[p_1] + Pr[p_2])^{-1}$ choose the pattern to be p_1 in the first chain and to be p_2 in the second chain and with equal probability choose p_2 in the first chain and p_1 in the second chain. Finally with the remaining probability, i.e., $(Pr[p_1] - Pr[p_2])(Pr[p_1] + Pr[p_2])^{-1}$, choose p_1 in both chains.

Once we have chosen the patterns we choose matching covers corresponding to the given patterns. We couple these choices such that if the same pattern is chosen in both chains we choose the same matching covers in both chains. If different patterns are chosen, the matching covers are paired in some arbitrary way.

If the opposite patterns were chosen, the two squares, after the row and column-permutations in both chains we have the same elements in each row and in the subsequent row-permutation we can couple the permutations with probability one. If we have the same pattern for both chains then the permutations remain different by a transposition.

We conclude that the probability of not coupling is

$$\sum_p |Pr[p_1] - Pr[p_2]|$$

and we have established Lemma 3.2 using Lemma 3.3. ■

We proceed to establish Lemma 3.3.

Proof: (Lemma 3.3) We can without loss of generality assume that p_1 takes $(0,0)$ to 0 and $(1,1)$ to 1. This has created an edge from 0^- to 0^+ with label 0 and an edge from 1^- to 1^+ with label 1. We call these edges the *green* edges. In p_2 we instead have an edge from 0^- to 1^+ with label 0 and an edge from 1^- to 0^+ with label 1. We call these edges the *red* edges.

Now take any matching cover $M_0, M_1, \ldots M_{m-1}$ corresponding to p_1 and suppose that the first green edge (the one with label 0) appears in M_i and the other edge appears in M_j. If $i = j$ then way say that we have a *special matching cover*. For such covers we can replace the green edges by the red edges and we have a one-to-one correspondence of the special matching covers of p_1 and p_2 and we write it as $p_1 = C(p_2)$.

Let us consider $i \neq j$ and look at the graph induced by M_i and M_j. Direct the edges of M_i going left to right and the ones of M_j right to left. We get a graph of in-degree and out-degree one and hence it is a union of disjoint cycles. Since the graph is bipartite all cycles are of even length. We say that graph is *good* if the two green edges are on the same cycle and *bad* otherwise. We have a similar definition for matching covers corresponding to the pattern p_2 with "red" replacing "green".

We now proceed to define a one-to-one correspondence of the good matching covers. Start with a good matching cover corresponding to p_1.

We have the two green (directed) edges $(a^-, b^+) \in M_i$ and $(c^+, d^-) \in M_j$. Note that a, b, c and d takes values in $\{0, 1\}$ and in fact a is chosen to be the 0, however keeping the letters makes the argument easier to follow.

The two (undirected) red edges are (a^-, c^+) and (b^+, d^-). The directed cycle in the union of M_i and M_j can be written as

$$a^- - b^+ - P_1 - c^+ - d^- - P_2 - a^-,$$

for paths P_1 and P_2. Consider now the cycle

$$a^- - c^+ - P_1^R - b^+ - d^- - P_2 - a^-,$$

where P_1^R is the reverse of P_1. This is a directed cycle with one of the red edges going from left to right and the other from right to left. We split the graph including the red edges into two matchings in the following way. Edges not in the changed cycle remain in their original matching. The changed cycle is split into two partial matchings and the red edge containing a^- joins with its partial matching into M_i. This gives a matching cover corresponding to the pattern p_2. As we have a cycle containing both red edges this is a good matching cover and it is easy to see that we have a one-to-one mapping between good matching covers corresponding to p_1 and p_2, respectively.

Now let us analyze a bad matching cover corresponding to p_1. The two green edges are on the different cycles

$$a^- - b^+ - P_1 - a^-$$

and

$$c^+ - d^- - P_2 - c^+.$$

Taking the green edges out and putting in the red edges we get the one cycle

$$a^- - c^+ - P_2^R - d^- - b^+ - P_1 - a^-.$$

Note that we get the same direction of the two red edges. We create a matching cover corresponding to the pattern p_2 in the same way as above, i.e. edges not in the affected cycle stay where they are and the affected cycle is split into two pieces. In this split both red edges go to the same partial matching and we include them in M_i, the matching that originally contained the green edge (a^-, b^+). We have created a special matching cover corresponding to the pattern p_2 and, not surprisingly, the mapping is many-to-one.

Similarly we can define a mapping from bad matching covers corresponding to p_2 to special matching covers corresponding to p_1.

Let us sum up the facts so far. We have a pair of patterns p_1 and p_2 different in only a pair of elements. Matching covers corresponding to either of these two patterns can be partitioned as follows.

1. We have equally many special matchings covers and a one-to-one mapping, C which is a pairing of these matching covers.

2. We have equally many good matchings covers and a one-to-one mapping of these matchings covers.

3. We have a mapping of bad matchings covers corresponding to p_1 to special matchings covers corresponding to p_2.

4. We have a mapping of bad matchings covers corresponding to p_2 to special matchings covers corresponding to p_1.

For a special matching cover $\overline{M}$ corresponding to p_2 let $G_{\overline{M}}$ be the number of bad matching covers corresponding to p_1 that are mapped to $\overline{M}$. Similarly let

$R_{\overline{M}}$ be the number of bad matching covers corresponding to p_2 that are mapped onto $C(\overline{M})$. If follows from the above facts that

$$\sum_p |Pr[p_1] - Pr[p_2]| \leq \sum_{\overline{M}} Pr[\overline{M}] |R_{\overline{M}} - G_{\overline{M}}|, \tag{1}$$

where the sum is over special matching covers.

Let us consider $G_{\overline{M}}$ for a special matching cover $\overline{M}$ and suppose that both red edges appear in M_i. We claim that, for each $j \neq i$, there is one unique bad matching cover coming from operating on matchings M_i and M_j that is mapped to $\overline{M}$ iff, when looking at the graph containing M_i and M_j, the two red edges are on the same cycle, and otherwise $\overline{M}$ does not arise from operating on M_i and M_j. To see this, first note that the procedure described above always creates a special matching cover where the two red edges are on the same cycle in the graph given by M_i and M_j. Uniqueness and existence of the preimage follows if we can prove that, once the value of j is given, the procedure is reversible. We establish this by explicitly describing how to find a preimage.

In the graph given by M_i and M_j take out the red edges and put in the green edges. This splits the cycle into two different cycles each containing one green edge. The two cycles are split into partial matchings. Combine the matching pairwise to make each set contain one green edge. Finally including the part with the green edge adjacent to 0^- in M_i and the other part in M_j we obtain a matching cover. By inspection this is a preimage of the given pattern. It is the only possible preimage as the green edge adjacent to 0^- must be included in M_i and there must be one green edge in each of M_i and M_j.

We now turn to computing, for a random special matching cover, the number of j's such that the two red edges belong to the same cycle when M_j is combined with M_i.

Lemma 3.4 *Let M_i be a fixed matching and suppose another matching M_j is chosen randomly. Fix two edges e_1 and e_2 of M_i. The probability that e_1 and e_2 belongs to the same cycle in the graph given by the union of M_i and M_j is 1/2.*

This is just a restatement of the fact that the probability that two elements belong to the same cycle in a random permutation is 1/2, but for completeness let us give a proof.

Proof: Let us reveal M_j one edge at the time. Let us start by traversing e_1 going left to right. Every time we traverse an edge of M_i left to right we ask what edge of M_j is adjacent to this node. If this edge does not attach to the left hand side of e_1 or e_2 we continue the process with another edge from M_i. If we hit e_1 before we hit e_2 the two edges are in different cycles while if we hit e_2 before e_1 the two edges are in the same cycle. By symmetry the two cases are equally likely and the lemma follows. ■

As the event of the two red edges being on the same cycle is independent for different j, when $\overline{M}$ is chosen randomly from from all special matchings the number of preimages of $\overline{M}$ is Binomial$(m-1, 1/2)$ distributed. The same argument applies to $R_{\overline{M}}$. The two numbers are not independent and hence it is convenient to use the upper bound

$$\sum_{\overline{M}} Pr[\overline{M}]|R_{\overline{M}} - (m-1)/2| + \sum_{\overline{M}} Pr[\overline{M}]|G_{\overline{M}} - (m-1)/2| \tag{2}$$

for (1).

The probability that a matching cover is special is exactly $1/m$. This follows since with probability $1/m$ our two special entries go to the same column. We conclude that we have a stochastic variable X which is distributed according do Binomial$(m-1, 1/2)$ and we are summing $|X - (m-1)/2|$ over a fraction $1/m$ of the space. We have the following lemma.

Lemma 3.5 *Let X be distributed according to Binomial$(m-1, 1/2)$. Let Ω be any event subset of total probability $1/m$, then*

$$\sum_{\omega \in \Omega} |X(w) - (m-1)/2| Pr(\omega) \leq O((m/\log m)^{-1/2}).$$

Proof: Clearly the sum is maximized if Ω is chosen as the event $|X - (m-1)/2| \geq a$ where a is chosen to make Ω have total probability $1/m$. By a standard result ([1], Theorem A.1.1) we have $a = O(\sqrt{m \log m})$. By the same result

$$Pr[|X - (m-1)/2| \geq 2\sqrt{m \log m}] \leq m^{-2}$$

and we conclude

$$\begin{aligned} \sum_{\omega \in \Omega} |X(w) - (m-1)/2| Pr(\omega) &\leq Pr(\Omega) 2\sqrt{m \log m} + m^{-2} \max |X - (m-1)/2| \\ &\leq O((m/\log m)^{-1/2}). \end{aligned}$$

■

By the above discussion, Lemma 3.5 gives an upper bound for (2) and this concludes the proof of Lemma 3.3. ■

3.1 Improving the estimate for the rate of mixing

Looking more closely that at the proof we can get a better bound for the mixing time. Let us sketch the argument.

We have essentially analyzed Π_2 and Lemma 3.2 was proved by a coupling that after two steps had moved all elements in both chains to identical rows. Assuming this was successful the third step produced coupling with probability one.

If the two key elements are not in the same row after the second step the third step does not contribute anything. Thus we should immediately start the argument from the scratch.

In other words if we let the success condition, not be “equal permutations” but rather “permutations with the difference in only one row” we couple except with probability $O((m/\log m)^{-1/2})$ in *two* time steps. Coupling in the old sense is achieved one step later. We get the following theorem.

Theorem 3.6 *Let Π_t be the distribution of Definition 2.1. Then*

$$\Delta(\Pi_t, U_n) \leq O(n^{1-\lfloor\frac{t-1}{2}\rfloor\frac{1}{2}}(\log n)^{\lfloor\frac{t-1}{2}\rfloor}).$$

4 Discussion

The goal of this paper was to prove that our process mixes in constant time and we have not made a strong effort to determine the exact constant.

Clearly, two steps are not sufficient as the elements that start in the same row end up in different columns with probability one. Thus at least three steps are needed. On the other hand, since any pattern is possible we know that Π_3 has full support. It is quite possible that Π_3 is in fact close to the uniform distribution on permutations. To prove this one would have to prove that most patterns are almost equally likely. In other words that the induced probability distribution on patterns is close to uniform on possible patterns. We have not been able to determine whether this is the case and we leave it as an open problem.

Acknowledgements. I am indebted to Svante Janson, Alistair Sinclair, Balint Virag and Peter Winkler for discussions on the topic of this paper. I am also grateful to an anonymous referee for some comments that helped to improve the quality of the paper and catching at least one essential typo.

References

[1] N. Alon and J. Spencer. *The probabilistic method.* Wiley Interscience, second edition edition, 2000.

[2] D. Bayer and P. Diaconis. Tracing the dovetail shuffle to its lair. *Annals of applied probability*, 2:294–313, 1992.

[3] P. Diaconis. *Group representations in probability and statistics.* Lecture Notes Monograph Series Vol. 11, Institute of Mathematical Statistics, Hayward, California, 1988.

[4] B. Morris. On the mixing time for the Thorp shuffle. Manuscript.

[5] E. Thorp. Nonrandom shuffling with applications to the game of Faro. *Journal of the American Statistical Association*, 68:842–847, 1973.