Pattern-avoiding permutons and a removal lemma
for permutations

Frederik Garbe† Jan Hladký‡ Gábor Kun§
Kristýna Pekárová¶

Abstract

The theory of limits of permutations leads to limit objects called permutons, which are certain Borel measures on the unit square. We prove that permutons avoiding a permutation \(A\) of order \(k\) have a particularly simple structure. Namely, almost every fiber of the support of the permuton consists only of atoms, at most \(k - 1\) many, and this bound is best-possible.

From this, we derive the following removal lemma. Suppose that \(A\) is a permutation. Then for every \(\varepsilon > 0\) there exists \(\delta > 0\) so that if \(\pi\) is a permutation on \(\{1, \ldots, n\}\) with density of \(A\) less than \(\delta\), then there exists a permutation \(\tilde{\pi}\) on \(\{1, \ldots, n\}\) which is \(A\)-avoiding and \(\sum_i |\pi(i) - \tilde{\pi}(i)| < \varepsilon n^2\).
1 Introduction

1.1 A permutation removal lemma

One of the most important consequences of Szemerédi’s regularity lemma is the graph removal lemma. It states that for any fixed graph $F$ on $\ell$ vertices we have that if a large $n$-vertex graph contains $o(n^\ell)$ copies of $F$ then it can be made $F$-free by deleting only $o(n^2)$ edges. The triangle version of the removal lemma was shown by Ruzsa and Szemerédi to imply (qualitatively) Roth’s theorem on 3-term arithmetic progressions.

Due to their wide applicability, removal properties were subsequently investigated for a range of further discrete structures. For hypergraphs of a fixed uniformity, this was done in a way analogous as for graphs, that is, first a hypergraph regularity lemma was established by Gowers [7] and Rödl and Skokan [18] from which then a corresponding hypergraph removal lemma was deduced by Nagle, Rödl and Schacht [16]. The line of research concerning graphs and hypergraphs was pushed substantially further, and the most notable highlights include so-called induced removal lemmas (in which deletions and additions of edges are necessary) for individual graphs due to Alon, Fischer, Krivelevich and Szegedy [1], their infinite families due to Alon and Shapira [2], as well as a counterpart of these results for hypergraphs due to Rödl and Schacht [17]. Another success story is that of removal properties of equations in groups with solutions restricted to a given set. The first result in this direction was obtained by Green [8] who also formulated the first regularity lemma for subsets in abelian groups. It was, however, later discovered by Král’, Serra and Vena [12] that one can obtain such a result using the ordinary (di)graph removal lemma without having to appeal to the complicated regularity lemma for abelian groups. This more direct combinatorial approach turned out to be also more flexible and allowed various extensions to non-abelian groups and finite fields [13,14].

One of the two main results of this paper is a removal lemma for permutations. Here, by a permutation we mean a bijection $\pi : [n] \to [n]$ for some $n \in \mathbb{N}$. We write $\mathcal{S}(n)$ for the set of all permutations on $[n]$. The fact that $[n]$ is equipped with the natural order allows us to define the notion of patterns. Namely, for $k \in [n]$ and for a $k$-element set $K \in \binom{[n]}{k}$ we say that $K$ induces a pattern $A \in \mathcal{S}(k)$ in $\pi$ if for all $i, j \in [k]$ we have that $\sigma(i) < \sigma(j)$ if and only if the image of the $i$th smallest element of $K$ under $\pi$ is smaller than the image of the $j$th smallest element of $K$ under $\pi$. In particular the density of $A$ in $\pi$, denoted by $t(A, \pi)$, is the proportion of $k$-tuples $K$ that induce $A$. We say that $\pi$ is $A$-avoiding if $t(A, \pi) = 0$. Pattern avoidance is one of the most vivid parts of the combinatorics of permutations, and indeed its treatise spans most of the standard books on permutations [3,10]. Most of the questions in the area of pattern avoidance concern a setting in which the pattern is fixed and $n$ is large.

We expect a permutation removal lemma to say that given a fixed pattern $A$ and a large permutation $\pi \in \mathcal{S}(n)$, if $t(A, \pi) = o(1)$ then

(♠) “all copies of $A$ can be removed from $\pi$ while changing $\pi$ only a little".
Let $\pi$ be defined by $\pi(2m - 1) = 2m$ and $\pi(2m) = 2m - 1$ for all $m \leq \frac{n}{2}$, and consider
the antiidentity permutation on two elements, $A = (21)$. Then we have $t(A, \pi) = \Theta(\frac{1}{n})$.
Observe that each element (except at most one) of $\pi$ is involved in a copy of $A$. Also,
the smallest set $R \subset [n]$ for which $\pi|_{[n] \setminus R}$ is $A$-avoiding has $\lfloor \frac{n}{2} \rfloor$ elements, not $o(n)$. So,
this example draws nontrivial limitations to removal features of permutations.

The first relevant removal result for permutations was proven by Cooper [4].

**Theorem 1.1.** Let $k \in \mathbb{N}$ and $A \in S(k)$. For every $\varepsilon > 0$ there exists $\delta > 0$ so that
the following holds. Suppose that $\pi \in S(n)$ is a permutation with $t(A, \pi) < \delta$. Then
there exists a set $P \subset \binom{[n]}{2}$ with $|P| < \varepsilon n^2$ such that for every $k$-tuple $K \in \binom{[n]}{k}$
which induces the pattern $A$ we have $|K \cap P| \geq 2$.

Cooper derives this result from a regularity lemma for permutations which he also
obtains. Note that the property in Theorem 1.1 is already quite far from that in
the original removal lemma in the sense that nothing is really removed, but rather
copies of the pattern are contained in a certain way. Actually, the very concept of
elementary removal operations is much less clear in permutations than it is, say in
graphs. This is because the usual elementary removal operation in graphs, namely
removing a single edge, obviously results again in a graph. On the other hand, the
category of permutations is not closed under many natural operations, such as changing
a value at a single position or deleting a single element from the domain. Yet quite
often we would like to obtain again a permutation after the “little change” in (♣).
The following result is a possible first step in this direction.

**Proposition 1.2.** Let $k \in \mathbb{N}$ and $A \in S(k)$. For every $\varepsilon > 0$ there exists $\delta > 0$ so that
the following holds. Suppose that $\pi \in S(n)$ is a permutation with $t(A, \pi) < \delta$. Then
there exists a permutation $\tilde{\pi} \in S(n)$ which is $A$-avoiding and for which we have
$$d_{\square}(\pi, \tilde{\pi}) < \varepsilon.$$

Here, $d_{\square}(\cdot, \cdot)$ is the so-called rectangular distance between two permutations, which
was introduced in [9] and which we recall in [2]. The rectangular distance is quite coarse,
which in turn makes Proposition 1.2 quite weak. Indeed, for example, two independently
uniformly generated permutations $\pi, \tilde{\pi} \in S(n)$ satisfy with high probability $d_{\square}(\pi, \tilde{\pi}) = o(1)$
even though their value at any particular entry differs by $\approx \frac{n}{2}$ in expectation. The
main result of this paper improves upon this by using a stronger metric.

**Theorem 1.3.** Let $k \in \mathbb{N}$ and $A \in S(k)$. For every $\varepsilon > 0$ there exists $\delta > 0$ so that
the following holds. Suppose that $\pi \in S(n)$ is a permutation with $t(A, \pi) < \delta$. Then
there exists a permutation $\tilde{\pi} \in S(n)$ which is $A$-avoiding and for which we have
$$\sum_{i=1}^{n} |\pi(i) - \tilde{\pi}(i)| < \varepsilon n^2. \quad (1)$$

If the rectangular metric is a counterpart to the cut metric in graph(on) theory,
then the metric involved in Theorem 1.3 is a counterpart to the edit distance. Since
the former is coarser than the latter, Proposition \[1.2\] immediately follows from Theorem \[1.3\].

1.2 The structure of pattern-avoiding permutons

The second result of this paper concerns limit properties of pattern avoiding permutations and as a fruitful application it will be used to prove our removal lemma. In the previous subsection, we defined the notion of $A$-avoidance of a pattern $A$ for a permutation $\pi$. A very broad question is how restrictive the property of avoiding a specific pattern is. The most famous result in this direction is due to Marcus and Tardos \[15\], formerly known as the Stanley–Wilf conjecture, and says, that for any fixed pattern $A$ there exists a constant $c_A$ so that the number of $A$-avoiding permutations of any order $n$ is at most $c_A n$. The monographs of Bóna \[3\] and Kitaev \[10\] contain many further results regarding the number and structure of pattern avoiding permutations.

Following the success of the theory of limits of dense graphs, Hoppen, Kohayakawa, Moreira, Ráth, and Sampaio \[9\] introduced a limit theory for finite permutations; the name and the following measure theoretic view on the limit object were introduced by Král’ and Pikhurko \[11\]. A permuton $\Gamma$ is a Borel probability measure on $[0, 1]^2$ with the uniform marginals property, that is, for any Borel set $Z \subset [0, 1]$ we have that $\Gamma(Z 	imes [0, 1]) = \Gamma([0, 1] \times Z)$ is equal to the Lebesgue measure of $Z$.

Recall that for $A \in S(k)$ and $\pi \in S(n)$ with $k \leq n$ we defined the density of $A$ in $\pi$ as the proportion of $k$-tuples of $[n]$ which induce the pattern $A$. To define pattern avoidance for permutons in a similar vein we will work with geometric representations of a pattern $A \in S(k)$. Let $S_A \subset ([0, 1]^2)^k$ be the collection of all sets of $k$ points $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ such that for all $1 \leq i < j \leq k$ we have $x_i < x_j$ and further that $y_i < y_j$ if and only if $A(i) < A(j)$. We call each element of $S_A$ a geometric representation of $A$. Vice versa each collection $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ of $k$ points such that for every $i < j$ we have $x_i < x_j$ and $y_i \neq y_j$ induces a unique permutation $\pi \in S(k)$ via $\pi(i) := |\{j \in [k] \mid y_j \leq y_i\}|$. Note that each element of $S_A$ induces $A$. We can now extend the notion of pattern density to permutons. The density of $A$ in a permuton $\Gamma$ is defined as $t(A, \Gamma) := k! \cdot \Gamma^{\otimes k}(S_A)$. A probabilistic interpretation of $t(A, \Gamma)$ is as follows. Suppose that we sample points $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ independently at random from the measure $\Gamma$. Then $t(A, \Gamma)$ is the probability that when reading these points from left to right, their vertical positions are consistent with $A$. In particular, we say that $\Gamma$ is $A$-avoiding if $t(A, \Gamma) = 0$.

Our theorem below says that the support of the measure of any avoiding permuton has a particularly simple one-dimensional structure. Let $\mathcal{M}$ be the space of all finite Borel measures on $[0, 1]$. For a given $\gamma \in \mathcal{M}$ and $\ell \in \mathbb{N}$ we say that $\gamma$ is an $\ell$-molecule if there exists a set $P \subset [0, 1]$ with $|P| \leq \ell$ so that $\gamma([0, 1] \setminus P) = 0$. We say that $\gamma$ is an $\ell$-molecule if it is an $\ell$-molecule but not an $(\ell - 1)$-molecule. Last, we recall the concept of disintegration of a measure. Suppose that $\{A_x \in \mathcal{M}\}_{x \in [0, 1]}$ is a collection

\[1\]A self-contained much shorter proof of Proposition \[1.2\] would be also possible.
of measures so that the map $x \mapsto \Lambda_x$ is measurable\(^2\). Then we can define a new Borel measure $\Lambda$ on $[0, 1]^2$ by

$$
\Lambda(B) := \int \Lambda_x \left( \{ y \in [0, 1] : (x, y) \in B \} \right) \, d\lambda(x) \quad \text{for each Borel } B \subset [0, 1]^2 .
$$

Then $\{\Lambda_x\}_{x \in [0, 1]}$ is called a disintegration of $\Lambda$ with respect to the Lebesgue measure $\lambda$. The Disintegration Theorem tells us that a disintegration always exists and is unique up to a nullset. We call the measures $\Lambda_x$ fibers.

We can now state our second main result.

**Theorem 1.4.** Suppose that $A$ is a pattern of order $k$ and $\Gamma$ is an $A$-avoiding permuton. Let us fix a disintegration $\{\Gamma_x\}_{x \in [0, 1]}$ of $\Gamma$. Then almost all fibers of $\Gamma$ are $\leq (k - 1)$-molecules.

This bound is optimal: for every pattern $A$ of order $k$ there exists an $A$-free permuton such that almost all of its fibers are $(k - 1)$-molecules.

Let us remark that Doležal, Máthé and the second author [5] proved a result in a similar spirit (but using completely different tools) for graphons: Every graphon avoiding a specific graph must be countably-partite.

Our construction for the second half of the theorem will be piecewise linear. One may expect that pattern-free permutons are similarly nice. However, this need not be the case as we show in Section 5.2.

### 1.3 Structure of the paper

In Section 2 we recall some facts about permutons and measure theory. Theorem 1.4 is proven in Section 3. Using Theorem 1.4 we then prove Theorem 1.3 in Section 4. In Section 5 we discuss optimality and potential strengthenings of both our main results.

### 2 Notation and preliminaries

#### 2.1 Permutons

We recall some concepts of the theory of permutation limits, as introduced in [9]. Earlier we already defined a permuton as a probability measure on the Borel sigma-algebra on $[0, 1]^2$ with the uniform marginal property. The relevant metric for the limit theory of permutations is the rectangular distance which is defined for two permutons $\Gamma_1$ and $\Gamma_2$ as

$$
d_{\square}(\Gamma_1, \Gamma_2) = \sup_{S, T \subseteq [0, 1]} |\Gamma_1(S \times T) - \Gamma_2(S \times T)|.
$$

**Remark 2.1.** Note that \(^2\) gives a metric for general measures (i.e., not necessarily permutons) as well.

\(^2\)For the concept of measurability, we need the structure of a sigma-algebra on the space $\mathcal{M}$. We shall get to this later.
Note that each finite permutation \( \pi \in S(n) \) has a **permuton representation** \( \Psi_\pi \) which is defined as follows. Take the union of rectangles \( S := \bigcup_{i=1}^n \left[ \frac{i-1}{n}, \frac{i}{n} \right] \times \left[ \frac{\pi(i)-1}{n}, \frac{\pi(i)}{n} \right] \). Now, for any Borel set \( B \subset [0,1]^2 \) we define \( \Psi_\pi(B) := n \cdot \lambda^2(B \cap S) \). Using permuton representations we can extend the rectangular distance to all finite permutations. That is, given a permuton \( \Gamma \) and finite permutations \( \pi_1 \in S(n_1) \) and \( \pi_2 \in S(n_2) \) we write \( d_{2\square}(\pi_1, \Gamma) := d_{2\square}(\Psi_{\pi_1}, \Gamma) \) and \( d_{2\square}(\pi_1, \pi_2) := d_{2\square}(\Psi_{\pi_1}, \Psi_{\pi_2}) \).

The following compactness result is the key result of the limit theory for permutations of which we will make use later.

**Theorem 2.2** ([9]). The space of permutons is compact with respect to the distance \( d_{2\square} \).

Furthermore, we will need that convergence in rectangular distance also induces convergence in pattern densities.

**Theorem 2.3** ([9]). Suppose that \( \pi_1, \pi_2, \ldots \) is a sequence of permutations of growing orders that converges to a permuton \( \Gamma \) in the rectangular distance. Then for any \( k \in \mathbb{N} \) and any \( A \in S(k) \) we have \( \lim_n t(A, \pi_n) = t(A, \Gamma) \).

### 2.2 Measure Theory

We use \( \lambda \) to denote the Lebesgue measure on \( \mathbb{R} \). For an arbitrary measure \( \gamma \) on a measure space \( X \), we write \( \gamma^{\otimes k} \) for its \( k \)-th power.

**2.2.1 The Lévy-Prokhorov metric**

We will make use of the Lévy-Prokhorov metric to compare probability measures on \([0,1]\).

**Definition 2.4** (Lévy-Prokhorov metric). For \( \varepsilon > 0 \) and a set \( A \subseteq [0,1] \) we define the \( \varepsilon \)-neighbourhood of \( A \) by

\[
A^{\varepsilon} = \{ p \in [0,1] \mid \exists q \in A : |p-q| < \varepsilon \}.
\]

Let \( \alpha \) and \( \beta \) be two Borel probability measures on \([0,1]\). The Lévy-Prokhorov distance between \( \alpha \) and \( \beta \) is defined by

\[
\text{dist}_{LP}(\alpha, \beta) = \inf\{ \varepsilon > 0 \mid \forall A \in B([0,1]) : \alpha(A) \leq \beta(A^{\varepsilon}) + \varepsilon, \beta(A) \leq \alpha(A^{\varepsilon}) + \varepsilon \}.
\]

Recall that a sequence of Borel probability measures \( \alpha_1, \alpha_2, \ldots \) on \([0,1]\) converges **weakly** to a measure \( \alpha \) if for every Borel set \( A \subset [0,1] \) we have \( \lim_n \alpha_n(A) = \alpha(A) \). It well-known that the Lévy-Prokhorov metric is a metrization of the weak convergence. It is also well-known that the Lévy-Prokhorov metric is separable, that is, there exists a countable set \( \{m_1, m_2, \ldots \} \) of Borel probability measures on \([0,1]\) so that for every Borel probability measure \( \gamma \) on \([0,1]\) and every \( \varepsilon > 0 \) there exists an index \( i \in \mathbb{N} \) so that \( \text{dist}_{LP}(m_i, \gamma) < \varepsilon \). Last, recall that for any metric space, the properties of being separable, having the property of Lindelöf and being second-countable are all
In particular, we will use that the Lévy-Prokhorov metric gives a second-countable space.

**Lemma 2.5.** Let \( \Gamma \) be a probability measure on \([0, 1]^2\). For all \( \delta > 0 \), there exists \( h \in \mathbb{N} \) and a partition \([0, 1] = J_1 \cup J_2 \cup \ldots \cup J_h\) of intervals such that for a disintegration \( \{\Gamma_x\}_{x \in [0, 1]} \) of \( \Gamma \) there exists \( X \subset [0, 1] \) with \( \lambda(X) < \delta \) such that \( \forall i \in [h], x, y \in J_i \setminus X: \) \( \text{dist}_{LP}(\Gamma_x, \Gamma_y) < \delta \).

**Proof.** We shall use Lusin’s Theorem in the following form. Suppose that \( M \) is a second-countable topological space. Suppose that \( f: [0, 1] \to M \) is measurable. Then for every \( \delta > 0 \) there exists a open set \( X \subset [0, 1] \) with \( \lambda(Y) < \delta \) so that \( f_{|[0,1]\setminus X} \) is continuous.

We are in this setting of Lusin’s Theorem as the space of all Borel probability measures on \([0, 1]\) equipped with the Lévy-Prokhorov metric is second-countable, and the map \( f : x \mapsto \Gamma_x \) is Borel. We use it with the same parameter as in the above statement. Since \([0, 1] \setminus X\) is compact and the function \( f \) is continuous on it, it is also uniformly continuous. That is, for the given \( \delta \) there exists a real \( \alpha > 0 \) so that for any two \( x, y \in [0, 1] \setminus X \) with \( |x - y| \leq \alpha \), we have \( \text{dist}_{LP}(\Gamma_x, \Gamma_y) < \delta \). The lemma therefore follows by choosing \( (J_i)_{i \in [k]} \) to be consecutive intervals of length \( \alpha \).

### 3 Proof of Theorem 1.4

#### 3.1 Proof of the upper bound

For the sake of contradiction suppose that the assertion does not hold. Then there exists a pattern \( A \) of order \( k \) and an \( A \)-avoiding permutation \( \Gamma \) together with a disintegration \( \{\Gamma_x\}_{x \in [0, 1]} \) and a set \( X \subset [0, 1] \) with \( \lambda(X) > 0 \) such that for every \( x \in X \) the fiber \( \Gamma_x \) is not \( \leq (k-1) \)-molecule.

Figure\(^3\) depicts the setting we introduce below. For \( t \in \mathbb{N} \) and \( m \in [t-1] \) we define \( J_{m,t} := [(m-1)/t, m/t] \) and \( J_{t,t} := [(t-1)/t, 1] \). For \( S \in \binom{[0]}{k} \) and \( i \in [k] \), write \( S_i \) for the \( i \)th smallest element of \( S \). For each \( S \in \binom{[0]}{k} \), set \( X_{S,t} := \{x \in X | \Gamma_x(J_{S,t}) > 0 \forall i \in [k]\} \).

Note that for \( x \in X \), since \( \Gamma_x \) is not \( \leq (k-1) \)-molecule, we have that there exist \( t \in \mathbb{N} \) and \( S \in \binom{[0]}{k} \) such that \( x \in X_{S,t} \). Hence \( X = \bigcup_{t \in \mathbb{N}, S \in \binom{[0]}{k}} X_{S,t} \). Recalling that \( \lambda(X) > 0 \), it follows from the countable subadditivity of measures that there exist \( t \in \mathbb{N} \) and \( S \in \binom{[0]}{k} \) such that \( \lambda(X_{S,t}) > 0 \). By Lebesgue’s density theorem, we can find a density point \( x_0 \) of \( X_{S,t} \) and choose \( \varepsilon > 0 \) small enough such that we have \( \lambda([x_0 - \varepsilon, x_0 + \varepsilon] \cap X_{S,t}) > 2\varepsilon(k-1)/k \). Then for \( i \in [k] \) define \( P_i := [x_0 - \varepsilon + 2\varepsilon \cdot (i-1)/k, x_0 - \varepsilon + 2\varepsilon \cdot i/k] \) and note that \( \lambda(X_{S,t} \cap P_i) > 0 \). Finally, define \( B_{i,j} := P_i \times J_{S,t} \), for \( i, j \in [k] \), and observe that

\[
\tau(A, \Gamma) \geq \Gamma^{\otimes k} \left( \prod_{i \in [k]} B_{i, A(i)} \right) \geq \prod_{i \in [k]} \left( \int_{x \in P_i \cap X_{S,t}} \Gamma_x(J_{S_{A(i)}, t}) \right) > 0,
\]

which is a contradiction.

\(^3\)We do not give definitions since the only way we use them is that we feed them into another theorem.
3.2 Optimality

Given a pattern $A \in S(k)$, we should construct an $A$-free permuton whose fibers with respect to a disintegration along the x-axis are almost all $(k - 1)$-molecules. The role of the two coordinates is exchangeable, and it turns out that it is notationally simpler to construct an $A^{-1}$-free a permuton whose horizontal fibers are $(k - 1)$-molecules.

Set $\pi = A^{-1}$. We consider the piecewise linear function $L : [0, 1] \to [0, 1]$ consisting of $k - 1$ pieces of slope $k - 1$ or $1 - k$ satisfying for $i = 1, \ldots, k - 1$ that $L(i/k - 1) = 0$ and $L(i/k) = 1$ if $\pi(i) > \pi(i + 1)$, while $L(i/k - 1) = 1$ and $L(i/k) = 0$ if $\pi(i) < \pi(i + 1)$. (At the endpoints this might not be well-defined: in this case we choose the value defined in the interval on the left.) An example is given in Figure 2.

Putting an appropriate multiple of the 1-dimensional Lebesgue measure on the graph of the function $L$ we obtain a permuton (that is, both uniform marginal properties are satisfied), which we call $\Gamma$. Firstly, almost all horizontal fibers of $\Gamma$ are $(k - 1)$-molecules. So, it remains to argue that $t(\pi, \Gamma) = 0$. To this end we prove the following.

Claim. Suppose that for some $\ell \in [k]$ we have $0 \leq x_1 < \cdots < x_\ell \leq 1$ such that for all $i \in [\ell]$, $x_i \not\in \{i/k : j \in \mathbb{Z}\}$, and such that the geometric configuration given by $x_1, \ldots, x_\ell$ (together with the unique $y$-coordinates on the respective fibers) generates $\pi_{|[\ell]}$. Then $x_\ell > \frac{\ell - 1}{k - 1}$.

Observe that the above claim indeed proves that $t(\pi, \Gamma) = 0$ since it asserts that we cannot find geometric configurations in the support of $\Gamma$ generating $\pi$, except the nullset of $k$-tuples where at least one point lies in $\{i/k : j \in \mathbb{Z}\}$.
Let us now prove the claim by induction on \(\ell\). The base case \(\ell = 1\) is obvious. Assume that the claim holds for \(\ell - 1\), and in particular for the \((\ell - 1)\)-tuple \(x_1, \ldots, x_{\ell - 1}\). If \(x_{\ell - 1} > \frac{\ell - 1}{k - 1}\) then we are done, since \(x_\ell > x_{\ell - 1}\). It remains to treat the case \(x_{\ell - 1} \in (\frac{\ell - 2}{k - 1}, \frac{\ell - 1}{k - 1})\). Then \(x_\ell \notin (\frac{\ell - 2}{k - 1}, \frac{\ell - 1}{k - 1})\) since \(x_\ell > x_{\ell - 1}\) and \(L\) is monotone increasing on this interval if \(\pi(\ell - 1) > \pi(\ell)\) and monotone decreasing otherwise. This completes the proof the theorem.

4 Proof of Theorem 1.3

We shall prove Theorem 1.3 in the following form.

**Theorem 4.1.** Let \(k \in \mathbb{N}\) and \(A \in S(k)\). Suppose that we have a sequence of permutations \((\pi_n \in S(n))_n\) converging to an \(A\)-free permuton \(\Gamma\) in the rectangular distance. Then there exists \(n_0 \in \mathbb{N}\) such that for every \(n > n_0\) there exists an \(A\)-free permutation \(\tilde{\pi}_n \in S(n)\) such that

\[
\sum_{i=1}^{n} |\pi_n(i) - \tilde{\pi}_n(i)| < \varepsilon n^2. \tag{3}
\]

**Deducing Theorem 1.3 from Theorem 4.1.** Suppose that Theorem 1.3 does not hold. That is, for a fixed pattern \(A\), there exists a number \(\varepsilon > 0\) and a sequence of permutations \(\pi_n\) with \(t(A, \pi_n) \to 0\) for which we cannot find similar but \(A\)-free permutations. By compactness of the space of graphons (Theorem 2.2), there exists a subsequence \((\pi_{n_t})\) which converges to a certain graphon \(\Gamma\) in the rectangular distance. By continuity of densities (Theorem 2.3), \(\Gamma\) is \(A\)-free. This setting contradicts Theorem 4.1 as was needed.
Figure 3: An example described in Section 4.1. In red, the support of a permuton \( \Gamma \) which is free of pattern \( A = (1234) \). Each fibre of the disintegration of \( \Gamma \) consists of two atoms, each of weight 0.5. Black dots show points \((x_i, y_i)\) representing a permutation \( \pi_{10} \) which represent an \( A \)-free permutation which might be an outcome of random resnapping on the fibers \( \Gamma_{x_i} \). The issue is that \(|y_i - \tilde{y}_i|\) is big for \( i = 1, 4, 6, 10 \). Right: Green diamonds correspond to resnapping to the closest atom on each fiber \( \Gamma_{x_i} \).

4.1 Motivating our proof of Theorem 4.1

Let us give an overview of our proof of Theorem 4.1. Our notation is consistent with Section 4.2 in which we give the proof in full detail.

Suppose that \( \pi_n \) is close to \( \Gamma \) in the rectangular distance. Fix a disintegration \( \{\Gamma_x\} \) of \( \Gamma \). We generate an \( n \)-tuple of points uniformly at random in \([0, 1]\), and read them from left to right as \( x_1 < x_2 < \cdots < x_n \). We set \( y_i := (\pi_n(i) - 0.5)/n \). Therefore, the points \((x_1, y_1), \ldots, (x_n, y_n)\) are a geometric representation of \( \pi_n \), and moreover their density in the unit square typically reflects \( \Gamma \).

The key idea is that any \( n \)-tuple of points \((\tilde{x}_1, \tilde{y}_1), \ldots, (\tilde{x}_n, \tilde{y}_n)\) whose \( x \)- and \( y \)-coordinates are pairwise distinct and where all the points lie in the support of the measure \( \Gamma \), is a geometric representation of an \( A \)-free permutation. We shall use this idea with \( \tilde{x}_1 = x_1, \ldots, \tilde{x}_n = x_n \). That is, we want to alter the \( y \)-coordinates of the points \((x_i, y_i)\), in such a way that, that the redefined point \( \tilde{y}_i \) lies in the support of the measure \( \Gamma_{x_i} \), which is roughly equivalent to the point \((x_i, \tilde{y}_i)\) lying in the support of \( \Gamma \).

One possible attempt would be to generate \( \tilde{y}_i \) at random according to \( \Gamma_{x_i} \). This would indeed result in an almost surely \( A \)-free geometric configuration \((x_1, \tilde{y}_1), \ldots, (x_n, \tilde{y}_n)\). The problem is that some of the points \( \tilde{y}_i \) might get moved far from \( y_i \). This in turn would lead to a violation of (3). See Figure 3 (left).

Here is where Theorem 1.4 comes into play. We know that almost every \( \Gamma_{x_i} \) is a collection of at most \( k - 1 \) atoms. Since \( \pi_n \) is close to \( \Gamma \) in the rectangular distance, we can also deduce that typically \( \Gamma_{x_i} \) must have positive mass around \( y_i \). Therefore there
exists an atom \( \tilde{y}_i \) of \( \Gamma_{x_i} \) which is close to \( y_i \). The points \((x_i, \tilde{y}_i)\) then induce our new permutation \( \tilde{\pi}_n \). See Figure 3 (right).

4.2 Proof of Theorem 4.1

Given \( \pi_n \xrightarrow{d} \Gamma \), where \( \Gamma \) is \( A \)-free, and \( \varepsilon > 0 \) we fix constants \( \zeta, \delta > 0 \) and \( h, n_0 \in \mathbb{N} \) according to the following hierarchy

\[
\frac{1}{n_0} \ll \zeta \ll \frac{1}{h} \ll \delta \ll \varepsilon.
\]

First, we apply Lemma 2.5 to \( \Gamma \) and \( \delta \) resulting in \( h \in \mathbb{N} \), a partition \([0, 1] = \bigcup_{i \in [h]} J_i \) into intervals, a set \( X \subset [0, 1] \) with \( \lambda(X) < \delta \), and a disintegration \( \{\Gamma_x\}_{x \in [0, 1]} \) of \( \Gamma \) such that for every \( i \in [h] \) and \( x, y \in J_i \setminus X \) we have

\[
\text{dist}_{LP}(\Gamma_x, \Gamma_y) < \delta.
\]

Furthermore, since \( \Gamma \) is \( A \)-free and by Theorem 1.4 we can fix a null set \( X_2 \subset [0, 1] \) such that \( \Gamma_x \) is \( \leq (k - 1) \)-molecule for every \( x \in [0, 1] \setminus X_2 \). We then set \( X := X_1 \cup X_2 \).

Now, fix a permutation \( \pi_n \) with \( n \geq n_0 \). We want to show that there exists an \( A \)-free permutation \( \pi_n \in \mathbb{S}(n) \) fulfilling (3). Note that by the constant hierarchy we can choose \( n_0 \) large enough such that we may assume that

\[
d_{\square}(\pi_n, \Gamma) < \zeta/2.
\]

Set \( T := \{i \in [n] \mid \lambda((i - 1)/n, i/n) \setminus X) > 0 \} \). Note that \(|T| > (1 - \delta)n\), as \( \lambda(X) < \delta \). We now generate a random \( n \)-tuple of points in the unit interval by choosing \( x_i \) uniformly at random from \(((i - 1)/n, i/n) \setminus X \) for every \( i \in T \). For \( i \in [n] \setminus T \) we choose \( x_i \) uniformly at random from \(((i - 1)/n, i/n) \setminus X_2 \). Furthermore we set

\[
y_i := (\pi_n(i) - 0.5)/n
\]

and denote the measure which has mass \( 1/n \) on each point \((x_i, y_i)\) by \( \Delta, \Delta = \frac{1}{n} \sum_{i \in [n]} \text{Dirac}(x_i, y_i) \). While \( \Delta \) is not a permutation, it can be thought of as a perturbation of the permutation representation \( \Psi_{\pi_n} \) in which each rectangle \( [\frac{i-1}{n}, \frac{i}{n}] \times [\frac{\pi_n(i)-1}{n}, \frac{\pi_n(i)}{n}] \) is transformed to a single point of the same measure and within that rectangle. In particular \( d_{\square}(\Delta, \pi_n) \leq \frac{2}{n} \). From the triangle inequality (c.f. Remark 2.1) and (5) we get

\[
d_{\square}(\Delta, \Gamma) \leq d_{\square}(\Delta, \pi_n) + d_{\square}(\pi_n, \Gamma) \leq \frac{2}{n} + \frac{\zeta}{2} \leq \zeta.
\]

Claim 1. With probability 1 with respect to the choice of the random points \((x_i)_{i \in [n]}\) the following holds.

(i) If \((\tilde{y}_i \in [0, 1])_n\) is collection of pairwise distinct points such that for each \( i \in [n], \)
\( \tilde{y}_i \) is an atom on \( \Gamma_{x_i} \) then the permutation induced by \((x_i, \tilde{y}_i)_{i \in [n]}\) is \( A \)-free.

(ii) If \( \Gamma_{x_i}(y) > 0 \) for some \( i \in [n] \) and \( y \in [0, 1] \), then \( \Gamma_{x_i}(y) = 0 \) for every \( i \neq j \).
Proof of Claim. For proving (i) suppose for the sake of contradiction that there exists $Q \subseteq [0,1]^n$ with $\lambda^{\otimes n}(Q) > 0$ such that for every $x \in Q$ there exists $\tilde{y}(x) \in [0,1]^n$ such that $\prod_{i=1}^n \Gamma_{x_i}(\tilde{y}(x)_i) > 0$ and such that $(x_i, \tilde{y}(x)_i)_{i \in [n]}$ contains the pattern $A$. Further, we can assume that the map $x \mapsto \tilde{y}(x)$ is measurable. Then the probability to sample the pattern $A$ from $\Gamma$ is at least

$$\int_Q \prod_{i=1}^n \Gamma_{x_i}(\tilde{y}(x)_i) d\lambda^{\otimes n}(x) > 0,$$

contradicting that $\Gamma$ is $A$-free.

For proving (ii) suppose for the sake of contradiction that there exists $y \in [0,1]$ such that for $Q_y := \{ x \in [0,1] \mid \Gamma_x(y) > 0 \}$ holds that $\lambda(Q_y) > 0$. By continuity of the Lebesgue measure there exists $\beta > 0$ and $Q'_y \subseteq Q_y$ with $\lambda(Q'_y) > 0$ such that for every $x \in Q'_y$ we have $\Gamma_x(y) > \beta$. Then for an arbitrary $\alpha > 0$ we have

$$\Gamma([0,1] \times [y-\alpha, y+\alpha]) \geq \beta \lambda(Q'_y),$$

which for $\alpha$ small enough contradicts the uniform marginal property of $\Gamma$.

We fix an outcome of our random selection of $x_1 < \cdots < x_n$ such that the above properties hold. Below, we abbreviate measures of intervals $[a,b]$ on fibers $\Gamma_{x_i}$ as $\Gamma_{x_i}(a,b) := \Gamma_{x_i}([a,b])$.

**Claim 2.** For all but at most $3\sqrt{n}$ many $i \in [n]$ it holds that $\Gamma_{x_i}(y_i-2\sqrt{\delta}, y_i+2\sqrt{\delta}) > \delta$.

Proof of Claim. We call a point $(x_i, y_i)$ bad if $\Gamma_{x_i}(y_i-2\sqrt{\delta}, y_i+2\sqrt{\delta}) \leq \delta$ and let $Y$ be the set of such bad points. We denote by $J(x)$ the interval $J_j$ (where $j \in [h]$) such that $x \in J_j$. Consider a system of rectangles $C = \{ J(x_i) \times [y_i - \sqrt{\delta}, y_i + \sqrt{\delta}] \mid (x_i, y_i) \text{ bad} \}$ covering $Y$ in $[0,1]^2$. Choose a minimal system $\mathcal{M} \subseteq \mathcal{C}$ which covers $\bigcup \mathcal{C}$. For $j \in [h]$, let $\mathcal{M}_j \subset \mathcal{M}$ be the collection of sets of the form $J_j \times U$. By the minimality, we have

$$|\mathcal{M}_j| \leq 2/(2\sqrt{\delta}) = 1/\sqrt{\delta}. \quad (8)$$

Therefore $\bigcup \mathcal{C}$ can be written as the disjoint union of at most $h/\sqrt{\delta}$ many rectangles. By using (7) we get

$$|\Gamma(\bigcup \mathcal{C}) - \Delta(\bigcup \mathcal{C})| < \zeta h/\sqrt{\delta}. \quad (9)$$

For every rectangle $J(x_i) \times [y_i - \sqrt{\delta}, y_i + \sqrt{\delta}] = J_j \times B = M \in \mathcal{M}_j$ we have that

$$\Gamma(M \setminus (X \times [0,1])) = \int_{J_j \setminus X} \Gamma_x(B) d\lambda(x) \overset{\text{(9)}}{=} \lambda(J_j) \cdot (\Gamma_{x_i}(B^{\delta}) + \delta).$$

In particular, since $(x_i, y_i)$ is bad we get

$$\Gamma(M \setminus (X \times [0,1])) \leq 2\delta \lambda(J_j). \quad (10)$$
Thus the triangle inequality implies that
\[
\Delta(\cup C) \leq \Gamma(\cup C \setminus (X \times [0, 1])) + \Gamma(X \times [0, 1]) + |\Gamma(\cup C) - \Delta(\cup C)|
\]
\[
\leq \sum_{j \in [n]} \sum_{M \in M_j} \Gamma(M \setminus (X \times [0, 1])) + \delta + \zeta h/\sqrt{\delta}
\]
\[
\leq \sum_{j \in [n]} \frac{1}{\sqrt{n}} \cdot 2\delta \lambda(J_j) + \sqrt{\delta}
\]
\[
\leq 3\sqrt{\delta},
\]
and therefore $|Y| \leq \Delta(\cup C)n \leq 3\sqrt{\delta}n$.

Now set $S := \{i \in T \mid \Gamma_{x_i}(y_i-2\sqrt{\delta}, y_i+2\sqrt{\delta}) > \delta\}$. Observe that by Claim 2 and the fact that $|T| \geq (1-\delta)n$ we have $|S| \geq n-3\sqrt{\delta}n-\delta n \geq (1-4\sqrt{\delta})n$. For every $i \in S$, since $\Gamma_{x_i}$ is $\leq (k-1)$-molecule there exists $\tilde{y}_i \in [y_i-2\sqrt{\delta}, y_i+2\sqrt{\delta}]$ such that $\Gamma_{x_i}(\tilde{y}_i) > 0$. For every $i \in [n] \setminus S$ choose an arbitrary atom $\tilde{y}_i$ of $\Gamma_{x_i}$ which is possible, as $x_i \notin X_2$ for every $i \in [n]$. By Claim 3(ii) we have that the points $(x_i, \tilde{y}_i)_{i \in [n]}$ induce a permutation which we denote by $\tilde{\pi}_n$. Furthermore, by Claim 3(ii) $\tilde{\pi}_n$ is $A$-free. It remains to establish the key property (3). We split the summands $\sum_{i=1}^n |\pi_n(i) - \tilde{\pi}_n(i)|$ according to whether $i \in S$ or $i \notin S$. In the latter case, it is enough to use the trivial bound $|\pi_n(i) - \tilde{\pi}_n(i)| \leq n$. So, we need to have a good bound for $|\pi_n(i) - \tilde{\pi}_n(i)|$ when $i \in S$. Recall (6) which tells us that the spacing in the y-coordinates of the geometric representation of $\pi_n$ was exactly $\frac{1}{n}$. Thus, if we change each point $(y_j)_{j \in S}$ by at most $2\sqrt{\delta}$, the point $y_i$ changes its relative position with at most $4\sqrt{\delta}n$ other points $(y_j)_{j \in S}$. Taking into account also elements which are not in $S$, we get $|\pi_n(i) - \tilde{\pi}_n(i)| \leq 4\sqrt{\delta}n + |[n] \setminus S|$. Putting all this together,

$$\sum_{i=1}^n |\pi_n(i) - \tilde{\pi}_n(i)| \leq \sum_{i \in S} (4\sqrt{\delta}n + |[n] \setminus S|) + \sum_{i \in [n] \setminus S} n \leq (4\sqrt{\delta} + 4\sqrt{\delta})n^2 + 4\sqrt{\delta}n^2 < \varepsilon n^2.$$ 

5 Discussion and further questions

5.1 Removal lemma for Latin squares

One might investigate removal properties of similar combinatorial structures, of which the closest one are Latin squares. A limit theory of Latin squares was worked out by Hancock, Sharifzadeh, the first and the second author in [6]. Actually, this project was originally motivated by a question of László Lovász about removal properties of Latin squares.

5.2 Nondifferentiable pattern-avoiding permuton

Theorem 1.4 together with Lusin’s theorem tells us that we can think of the support of a pattern-avoiding permuton as a union of graphs of partial functions that are continuous

\footnote{that is ‘bigger than/smaller than’}
on a set with complement of arbitrary small positive measure. The following example shows that continuity cannot be strengthened to differentiability. More precisely, we construct a function \( f : [0, 1] \to [0, 1] \) whose restriction to any subset of \([0, 1]\) of positive measure is not differentiable and the permuton supported on the graph of this function is avoiding the pattern \((3142)\).

Consider the quaternary expansion (with digits 0, 1, 2 and 3) of the numbers in \([0, 1]\). This is well-defined for all, but countably many numbers, the rational numbers whose denominator is a power of two. We will ignore this countable set. Let \( f \) be the function that swaps 1 and 2 for every coordinate. This function is injective (up to a nullset) and measure preserving; in particular we can indeed construct a measure whose support is the graph of \( f \) and which satisfies both uniform marginals properties.

First we argue that the graph of \( f \) is free of \((3142)\). Suppose for a contradiction that there are four numbers \( x_1 < x_2 < x_3 < x_4 \) such that \( f(x_2) < f(x_4) < f(x_1) < f(x_3) \). Consider the first digit where \( x_1 \) and \( x_2 \) differ. Clearly this digit should be 1 for \( x_1 \) and 2 for \( x_2 \). The number \( x_4 \) cannot differ in an earlier digit, since \( f(x_2) < f(x_4) < f(x_1) \). Hence \( x_4 \) being less than \( x_4 \) can also not differ in an earlier digit. We will get a contradiction when checking the critical digit: \( f(x_3) > f(x_1) \), \( x_3 > x_2 \) implies that this digit of \( x_3 \) should be 3, but then \( x_4 \) also has digit 3 here contradicting \( f(x_1) > f(x_4) \).

It remains to show that the restriction of \( f \) to any set of positive measure cannot be differentiable: if there was such a set then on a subset of positive measure for every pair the inequality \( |\frac{f(x) - f(y)}{x - y} - f'(x)| < \frac{1}{2} \) would hold. However, given a natural number \( n, i \in \{-3, -2, -1, 1, 2, 3\} \) and \( x, x + i4^{-n} \) such that the numbers agree until the first \( n - 1 \) digits note that \( f(x) - f(x + i4^{-n}) \in \{-2, -1, -\frac{1}{2}, \frac{1}{2}, 1, 2\} \), and this value depends on \( i \) and the \( n \)th digit of \( x \) (but not on \( n \)), and for every \( x \) and \( n \) we get three different values (depending on \( i \)). By the Lebesgue density theorem any set of positive measure contains more than the three quarter of a small interval, so we can find four numbers in the set agreeing in all, but one digit. Hence the restriction of \( f \) to any set of positive measure cannot be differentiable.

### 5.3 Lower bounds on the Stanley–Wilf constant

For a permutation \( A \), let \( \mathcal{S}(A; n) \subset \mathcal{S}(n) \) be the set of \( A \)-avoiding permutations of order \( n \). The Stanley–Wilf constant is defined as \( c_A := \lim_{n \to \infty} \sqrt[\nu]{|\mathcal{S}(A; n)|} \). The Stanley–Wilf constant is known exactly for several permutations and permutation classes (see §4, Chapter 4): trivially for \( A \in \mathcal{S}(2) \) we have \( c_A = 1 \), for all \( A \in \mathcal{S}(3) \) we have \( c_A = 4 \), for each \( k \in \mathbb{N} \) we have \( c_{\text{identity}_k} = (k - 1)^2 \), and then the Stanley–Wilf constant is known for several sporadic examples, including \( c_{(1342)} = 8 \), \( c_{(12453)} = 9 + 4\sqrt{2} \).

Suppose that \( \Gamma \) is an \( A \)-free permuton. We can use \( \Gamma \) to generate many different \( A \)-free permutations of order \( n \). A general treatise of the topic is a subject of ongoing research but let us illustrate this with the \((1234)\)-avoiding graphon \( \Gamma \) from Figure 3. Let us fix points \( x_1 < \ldots < x_n \) on \([0, 1]\), say equispaced. Each fiber \( \Gamma_{x_i} \) has exactly 2 atoms. So, we choose \( y_i \) to be one of these 2 atoms. In total we have \( 2^n \) options of generating a geometric configuration. It can be shown that most of these choices represent different
permutations, that is, we obtain $c_{(1234)} \geq 2$.

6 Acknowledgments

This work was initiated during the workshop Interfaces of the Theory of Combinatorial Limits at the Erdős center of the Rényi Institute. We thank the organisers for creating a very productive atmosphere. We thank Martin Doležal for discussions about real analysis, and Lara Pudwell for information about the relevant literature.

References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. Combinatorica, 20(4):451–476, 2000.

[2] N. Alon and A. Shapira. A characterization of the (natural) graph properties testable with one-sided error. SIAM J. Comput., 37(6):1703–1727, 2008.

[3] M. Bóna. Combinatorics of permutations. Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, second edition, 2012. With a foreword by Richard Stanley.

[4] J. N. Cooper. A permutation regularity lemma. Electron. J. Combin., 13(1):Research Paper 22, 20, 2006.

[5] M. Doležal, J. Hladký, and A. Máthé. Cliques in dense inhomogeneous random graphs. Random Structures Algorithms, 51(2):275–314, 2017.

[6] F. Garbe, R. Hancock, J. Hladký, and M. Sharifzadeh. Limits of latin squares. arXiv:2010.07854.

[7] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. Ann. of Math. (2), 166(3):897–946, 2007.

[8] B. Green. A Szemerédi-type regularity lemma in abelian groups, with applications. Geom. Funct. Anal., 15(2):340–376, 2005.

[9] C. Hoppen, Y. Kohayakawa, C. G. T. de A. Moreira, B. Ráth, and R. M. Sampaio. Limits of permutation sequences. J. Comb. Theory, Ser. B, 103(1):93–113, 2013.

[10] S. Kitaev. Patterns in permutations and words. Monographs in Theoretical Computer Science. An EATCS Series. Springer, Heidelberg, 2011. With a foreword by Jeffrey B. Remmel.

[11] D. Král’ and O. Pikhurko. Quasirandom permutations are characterized by 4-point densities. Geom. Funct. Anal, 23(2):570–579, 2013.
[12] D. Král’, O. Serra, and L. Vena. A combinatorial proof of the removal lemma for groups. *J. Combin. Theory Ser. A*, 116(4):971–978, 2009.

[13] D. Král’, O. Serra, and L. Vena. A removal lemma for systems of linear equations over finite fields. *Israel J. Math.*, 187:193–207, 2012.

[14] D. Král’, O. Serra, and L. Vena. On the removal lemma for linear systems over abelian groups. *European J. Combin.*, 34(2):248–259, 2013.

[15] A. Marcus and G. Tardos. Excluded permutation matrices and the Stanley–Wilf conjecture. *J. Comb. Theory, Series A*, 107(1):153–160, 2004.

[16] B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular $k$-uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.

[17] V. Rödl and M. Schacht. Generalizations of the removal lemma. *Combinatorica*, 29(4):467–501, 2009.

[18] V. Rödl and J. Skokan. Regularity lemma for $k$-uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004.