A Simple Proof of a Theorem of Schmerl and Trotter for Permutations

Robert Brignall
Department of Mathematics and Statistics
The Open University
Milton Keynes, England

Vincent Vatter
Department of Mathematics
University of Florida
Gainesville, Florida USA

When specialized to the context of permutations, Schmerl and Trotter's Theorem states that every simple permutation which is not a parallel alternation contains a simple permutation with one fewer entry. We give an elementary proof of this result.

An interval in the permutation $\pi$ (thought of in one-line notation) is a contiguous set of entries whose values also form a contiguous set. Every permutation of length $n$ has trivial intervals of lengths 0, 1, and $n$, and permutations with only trivial intervals are called simple. We take a graphical view of permutations, in which we identify a permutation $\pi$ with its plot, the set of points $(i, \pi(i))$ in the plane. Three examples of simple permutations are plotted in Figure 1. Note that 1, 12, and 21 are all simple and that there are no simple permutations of length three. A bit more examination shows that 2413 and 3142 are the only simple permutations of length four.

The second and third simple permutations in Figure 1 are called parallel alternations. A parallel alternation is, formally, a permutation whose entries can be divided into two halves of equal length, either both increasing or both decreasing, such that the entries of the halves interleave perfectly. In particular, 12, 21, 2413, and 3142 are parallel alternations. While parallel alternations need not be simple (1324 is a nonsimple parallel alternation), from any parallel alternation we may obtain a simple permutation by removing at most two entries.

Specialized to permutations\(^3\), the main result of Schmerl and Trotter [6] is as follows.

The Schmerl-Trotter Theorem for Permutations [6]. Every simple permutation which is not a parallel alternation contains an entry whose removal leaves a simple permutation.

Schmerl and Trotter’s theorem has found wide application both theoretically and practically. For example, it has been used in [2, 3] to show that certain classes of permutations are defined by a finite set of restrictions, and in Albert’s PermLab package [1] to efficiently generate the simple permutations in a permutation class.

---

\(^{1}\) Both authors were partially supported by the EPSRC Grant EP/J006130/1.

\(^{2}\) Vatter was partially sponsored by the National Security Agency under Grant Number H98230-12-1-0207 and the National Science Foundation under Grant Number DMS-1301692. The United States Government is authorized to reproduce and distribute reprints not-withstanding any copyright notation herein.

\(^{3}\) The notions of intervals and simplicity extend naturally to all relational structures (though with different names, such as modules and primitivity). Schmerl and Trotter proved their result for simple, irreflexive, binary relational structures.
In this note we give a short, self-contained, proof of the Schmerl-Trotter Theorem. We use only a few definitions in this proof. First, in order to simplify the discussion, we say that an entry of the simple permutation $\sigma$ is inessential if its removal leaves a permutation which is still simple (entries are essential otherwise). The other concept we need is separation: given a permutation $\sigma$ and entries $x_1$, $x_2$, and $x_3$, we say that $x_1$ separates $x_2$ and $x_3$ if $x_1$ lies between $x_2$ and $x_3$ either horizontally or vertically, but not both, as in Figure 2. We extend this to sets of entries by saying that $x$ separates the entries $X$ if it lies outside the rectangular hull of $X$ and separates any two of its entries.

**Proof.** We prove the theorem by induction on the length of the simple permutation. It is vacuously true for permutations of length four, as both such simple permutations are parallel alternations, so suppose that $\sigma$ is a simple permutation of length at least five which is not a parallel alternation and that the theorem holds for all shorter permutations. As we are done otherwise, we assume throughout the proof that every entry of $\sigma$ is essential.

We begin by assuming, for the sake of contradiction, that removing any entry of $\sigma$ creates a minimal proper interval containing precisely two entries. Let $x_1$ be an arbitrary entry of $\sigma$ and suppose that $\sigma - x_1$ contains the interval $\{x_2, x_3\}$. By symmetry, we may assume that these entries are in the relative order of $231$, with $x_1$ on the left, as shown on the left of Figure 2.

Consider the doubleton interval in $\sigma - x_3$. Because $\sigma - x_3$ still contains $x_2$, this doubleton interval must contain at least one of $x_1$ or $x_2$, together with a new entry $x_4$. Figure 3 shows the three possibilities: the doubleton interval can consist of $\{x_1, x_4\}$ with $x_4$ either to the left or right of $x_1$, or it can consist of $\{x_2, x_4\}$, but only if $x_4$ lies below and to the left of $x_2$. In this and all later figures, the gray areas indicate regions which cannot contain entries (because of the interval conditions).

Next consider the doubleton interval in $\sigma - x_4$. In the leftmost case of Figure 3, there is only one possibility, as the doubleton could only be $\{x_3, x_5\}$ for a new entry $x_5$. Then we see that the doubleton...
interval in $\sigma - x_5$ must be $\{x_4, x_5\}$ for a new entry $x_5$. Continuing this process until we have run out of entries of $\sigma$ (as indicated on the left of Figure 4) yields the desired contradiction. The case where $x_4$ lies to the right of $x_1$ (the middle case in Figures 3 and 4) yields a similar contradiction. This leaves only the rightmost case of Figure 3. In this case we instead consider the doubleton interval in $\sigma - x_2$. This doubleton must consist of a new entry $x_5$ together with $x_3$, as shown on the right of Figure 4. However, in that case, $\sigma - x_4$ cannot contain a doubleton interval, which completes the contradiction to our assumption that $\sigma - x$ contains a doubleton interval for every entry $x$.

![Figure 4: The three possible configurations after finding $x_5$.](image)

The next case we consider is when there is an entry $x$ for which $\sigma - x$ contains a minimal proper interval, say $\iota$, which is a simple parallel alternation of length at least four. First suppose that $\iota$ has precisely four entries. By symmetry, we may further assume that $x$ lies to the left of $\iota$, which is itself a copy of 3142, leaving us with three cases. These are depicted in Figure 5. In each of the three pictures in that figure, the circled entry is a potential inessential entry of $\pi$, which can only be essential if the entry labeled $y$ exists. In the first case, where $x$ separates the ‘1’ and the ‘2’ of $\iota$, the ‘4’ of $\iota$ would be inessential unless its removal were to result in an interval involving the ‘3’. That possibility could only happen if there were an entry $y$ lying directly above and to the left of $\iota$ in $\sigma - x$. However, in that case $y$ could not separate any set of entries not already separated by either $x$ or $\iota$, and thus $y$ would be inessential. In the second case, the ‘3’ of $\iota$ would be inessential unless there were an entry $y$ immediately below and to the left of $\iota$. In that situation, there must be other entries of $\sigma$ (because $\sigma$ is not a parallel alternation), and so there must be an entry other than $y$ separating $x$ from $\iota$, and in particular from the ‘3’ of $\iota$, thereby showing that $y$ is inessential. The third and final case is similar to the first.

![Figure 5: Up to symmetry, the three cases where $\iota$ has length four.](image)

Next suppose that $\sigma - x$ contains a minimal proper interval $\iota = \iota(1) \cdots \iota(2\ell)$ which is a parallel alternation of length at least six. By symmetry we may assume that $\iota$ is in the same relative order as $(\ell + 1)(\ell + 2)\cdots(2\ell)\ell$ and that $x$ lies either to the left of or above $\iota$. First suppose that $x$ lies above $\iota$. If $x$ lies horizontally between $\iota(1)$ and $\iota(2)$, then we have the situation shown on the left of Figure 6, and $\iota(2)$ is inessential. Otherwise, $x$ lies horizontally between two entries from the “bottom half” of $\iota$, say $\iota(2i)$ and $\iota(2i + 2)$, and in this case the entry $\iota(2i + 1)$ is inessential (the second picture in Figure 6 shows an example of this).

We may now suppose that $x$ lies to the left of $\iota$. As in the previous case, if $x$ lies vertically between two entries of the same “half” of $\iota$, say $\iota(i)$ and $\iota(i + 2)$, then the entry $\iota(i + 1)$ is inessential (the third picture in Figure 6 shows an example of this). The only remaining case is when $x$ lies vertically between the first and last entries of $\iota$. In this case the only way $\iota(1)$ can be essential is if its removal creates an interval involving $\iota(2)$ and an entry $y$ lying immediately below and to the left of $\iota$ (as
shown on the right of Figure 6). This case is similar to the analogous case where \( \iota \) is of length four; \( \iota \) must have other entries because it is not a parallel alternation and in particular, it must contain an entry separating \( x \) from \( \iota \) (because there is no other position in which \( \iota \cup \{x, y\} \) could be separated). This implies that \( y \) is inessential, thus completing the analysis of the case where the interval of \( \sigma - x \) contains an interval which is a parallel alternation.

\[ x \quad \ldots \quad x \quad \ldots \quad x \quad y \quad \ldots \quad x \quad \ldots \quad x \]

Figure 6: Up to symmetry, the cases where \( \iota \) is a parallel alternation of length at least six.

Having eliminating special cases, we may now suppose that there is an entry \( x_1 \) such that \( \sigma - x_1 \) contains a minimal proper interval, say \( \iota_1 \), which is not a parallel alternation (and thus consists of at least five entries). We construct a sequence \( x_1, x_2, \ldots \) of essential entries and a sequence \( \iota_1, \iota_2, \ldots \) of intervals such that \( \iota_k \) is a minimal proper interval of \( \sigma - x_k \) for every \( k \). Thus \( x_k \) separates \( \iota_k \) in \( \sigma \), and the minimality of \( \iota_k \) and simplicity of \( \sigma \) imply that \( \iota_k \) is itself simple. We are done by the above if any \( \iota_k \) is a parallel alternation, so \( \iota_k \) must contain an inessential entry by induction, which we take to be \( x_{k+1} \). Next we consider two possibilities. First, if \( x_k \) separates \( \iota_k - x_{k+1} \), as in the left of Figure 7, then \( x_{k+1} \) is inessential for \( \sigma \) and we are done. Otherwise, as shown on the right of Figure 7, \( x_{k+1} \) must separate \( x_{k+1} \) from the rest of \( \iota_k \), i.e., from \( \iota_k - x_{k+1} \). In this case \( \iota_k - x_{k+1} \) is a proper interval of \( \sigma - x_{k+1} \), so we take it to be \( \iota_{k+1} \).

\[ x_k \quad \iota_k \quad x_{k+1} \]

\[ x_k \quad \iota_k \quad x_{k+1} \]

Figure 7: Examples of the two cases that can happen after choosing \( x_{k+1} \).

This process must terminate because \( |\iota_{k+1}| = |\iota_k| - 1 \). Moreover, because each \( \iota_k \) is simple and there are no simple permutations of length three, when this process does terminate it must be because we have either found an inessential entry of \( \sigma \) or because some \( \iota_k \) is a parallel alternation of length at least four, and in either case the theorem is proved.

Now that we have proved that “almost all” simple permutations have an inessential entry, how many inessential entries should we expect them to have? The answer is almost all of them. This fact seems to have been first observed by Pierrot and Rossin [5], but we include a short sketch below.

It has been shown (see, for example, Corteel, Louchard, and Pemantle [4]) that the number of nontrivial intervals in a random permutation of length \( n \) is asymptotically Poisson distributed with mean 2 (in fact this is true for nontrivial intervals of size 2; the probability of a random permutation having a larger nontrivial interval tends to 0 as \( n \) tends to infinity). Therefore the number of simple permutations of length \( n \) is asymptotic to \( n!/e^2 \). Now we double-count pairs \( (\sigma, x) \) where \( \sigma \) is a
simple permutation of length \(n + 1\) and \(x\) is an inessential entry of \(\sigma\). On one hand, the number of such pairs is asymptotic to

\[
\frac{(n + 1)!}{e^2} \cdot \mathbb{E}[\text{number of inessential entries}].
\]

On the other hand, the number of such pairs is equal to the number of pairs \((\sigma, \tau)\) where \(\sigma\) and \(\tau\) are simple and \(\tau = \sigma - x\). Consider inserting a new entry \(x\) into a simple permutation \(\tau\) of length \(n\). There naively \((n + 1)^2\) different places to insert \(x\). However, \(2n\) of these places will create intervals of size two with entries of \(\tau\), while the 4 places on the corners will create an interval of size \(n\). Each of the remaining \(n^2 - 3\) places to insert \(x\) gives a different permutation, and if inserted into one of those positions, \(x\) cannot lie in an interval of size strictly between two and \(n\) because otherwise \(\tau\) would contain a proper interval. Therefore the number of such pairs is asymptotic to

\[
\frac{n!}{e^2} (n^2 - 3),
\]

showing that the expected number of inessential entries is asymptotic to \(n\), as desired.

Acknowledgements. We are grateful to Jay Pantone for his comments and corrections.

REFERENCES


