Loria, INRIA, CNRS
Université de Lorraine

Mathilde Bouvel

Mathilde Bouvel
CNRS researcher
Loria, Campus Sciences
Nancy, France
Email: mathilde.bouvel@loria.fr

October 4, 2022

External report on the PhD thesis of Michal Opler,
entitled *Structural and algorithmic properties of permutation classes.*

To whom it may concern

The PhD thesis of Michal Opler is a major contribution to the study of the (parameterized) complexity of the Permutation Pattern Matching (denoted PPM) problem, when the pattern or the text are required to belong to a permutation class C. In the thesis, permutation class means a set of permutations that is downward-closed under taking patterns, or equivalently a set of permutations that is defined by the avoidance of a (possibly infinite) set of patterns. These permutation classes have been the topic of hundreds of research papers (see *e.g.* the annual conference *Permutation Patterns*), most of them in the field of enumerative and bijective combinatorics, but with some work in neighboring fields, the ones most relevant to the present thesis being algorithms, complexity and logic.

The work presented in this thesis is remarkable in two aspects. First, it is to my knowledge one of the first (series of) work(s) which takes a general approach to studying the PPM problem, as opposed to previous approaches which were usually specifically designed for each class C studied. Second, it allows to identify precisely the hardness of PPM on many classes, and in particular in almost all principal classes (those defined by the avoidance of a single pattern σ, *i.e.* when $C = Av(\sigma)$). This illustrates the applicability of the general results obtained. In my opinion, these applications justify the general theory successfully developed by Michal Opler.

Another aspect of the thesis that I have particularly liked is the interplay between permutations and graphs. While PPM is obviously a permutation problem (and arguably one of the most – if not *the* most – natural ones), its study as done by Michal Opler often uses concepts analog to or inspired by graph theory or graph algorithms. This should not be misunderstood: solving problems on permutations is not achieved by straightforwardly specializing or adapting similar results from graphs. However, building on the immense knowledge gained after decades of graph theory, Michal Opler has shown that some ideas of this field can be adapted fruitfully to the permutation setting.

The thesis contains 6 main chapters (numbered 2 to 7), preceded by an introduction, some preliminary notions gathered in Chapter 1, and followed by a brief conclusion. I offer below an overview of the results obtained in each chapter, obviously biased by my personal preferences. The document is over 200 pages long, and contains some but rather few typos compared to the length. It is generally well-written, although in some occasions too dense for my taste. There are numerous illustrations, which are essential to facilitate the reading. The organization is maybe not very friendly, starting with many definitions before we see their interest, but it serves the purpose of the thesis quite well in the end: it shows how to build a general structural theory on which the analysis of the complexity of PPM can be developed.

Introduction: A short introduction presents the results of the thesis and their organization. Striking in this introduction is the level of generality of the results obtained, especially compared to the previously known results on restricted variants of PPM.

Chapter 1: This preliminary chapter sets all the classical definitions needed later. The presentation is clear, although I find it sometimes heavier than necessary. I find Section 1.2 (a recap on all notions of generalized patterns in the literature) very handy. I also particularly like the way of building permutations illustrated by Figure 1.10: we start by allowing vertical of horizontal alignments between points, but then apply a rotation by a tiny angle in order to break these alignments. It is very natural and practical, but I don't remember seeing it elsewhere before.

Chapter 2: This long chapter first presents many notions of width for permutations, and establishes results on how they compare to each other. Essentially, the boundedness of one is equivalent to the boundedness of another, except for the twin-width (which is the smallest of all considered width). Then, Chapter 2 describes four properties of permutation classes, all based on the existence of grid subclasses satisfying certain conditions. These four properties are called (each being more restrictive that the previous one) Long Path property (LPP), Cycle property (CP), Deep Tree property (DTP) and Bicycle property (BP). It is proved that the LPP implies unbounded treewidth, and it is conjecture that this is actually an equivalence. This part is technical and notationally heavy, but the many figures are very helpful. The final part of the chapter investigates which of these four properties are satisfied by principal classes, resulting in an almost complete (only 5 excluded patterns are left aside) classification of the growth of treewidth in principal permutation classes. I should mention that bounded treewidth results in PPM being computationally easy to solve, so that this result is closely related to the study of PPM restricted to permutation classes. This final part flows very easily after the theory developed earlier in the chapter; to me, it fully justifies the efforts done in building this structural theory.

Chapter 3: Chapter 3 deviates a bit from the main focus of the thesis (which is the PPM problem) and deals with logic. The various width defined in Chapter 2 are useful here, making this chapter still quite well-connected with the rest of the thesis. The first part of this chapter recalls known results about the (in)expressibility of FO (first order logic) on permutations, while the second part investigates (in)expressibility in MSO (monadic second order logic). These results are new, and only partly inspired by the previous work on FO by other authors (including myself). In particular, they identify a property expressible in MSO but not in FO: the fact that a permutation is a merge of two permutations, each avoiding a fixed (but arbitrary) simple pattern α. Another contribution of this chapter is to provide FPT algorithms for model checking of formulas in MSO, for the parameter clique-width (recalling that it was already known for FO and twinwidth).

Chapter 4: Here, the complexity of PPM (in its unrestricted variant) is studied. It has been known since the article of Bose, Buss and Lubiw in the nineties that PPM is NP-complete. The thesis provides a proof of this fact, as a preparation for later proofs. The given reduction is elegant and easy to follow (but I can't tell how new it is). In addition to NP-completeness, a lower bound for the complexity of PPM under ETH (exponential time hypothesis) is provided, which is exponential in the size of the pattern. We then move to the parameterized complexity of PPM.

The breakthrough result of Guillemot and Marx, stating that PPM is FPT for the size of the pattern, is presented. It appears like a little miracle in this landscape, where many variants of PPM are proven $W[1]$-hard (so, not FPT) for the size of the pattern. This is the case of the variant Left-PPM, introduced in the thesis, and proved to be $W[1]$-complete. As a consequence, it is established that the size of the pattern is essentially the only parameter which allows PPM to be FPT.

The NP-hardness of PPM is then extended to C-PPM (where both pattern and text are required to belong to C) in the case where C satisfies the LPP or the DTP. However, the complexity lower bounds for solving these variants (under ETH) are lower than in the general case. Finally, an important result of Berendsohn et al is stated, about the complexity of the counting version of PPM ("how many occurrences of the pattern in the text are there?"). It relates the bounded treewidth property to the solubility in polynomial time of $\sharp$PPM. The thesis proves that this complexity is best possible when C has the LPP.

All the reductions in this chapter are technical, but well-explained and illustrated to help the reader understand.

Chapter 5: Chapter 5 moves to the study of C-PPM, and reduces the gap between the cases known to be in P and those known to be NP-complete. After providing a very nice state of the are on the complexity of C-PPM, the main result is Theorem 5.6. It states that C-PPM is NP-complete when C satisfies some technical property (called the computable $\mathcal{D}$-rich path property, for some non-monotone-griddable class $\mathcal{D}$). The proof is a complicated reduction from 3-SAT, and it is **very** well presented, with first an overview and then the details organized with care. A consequence is the NP-completeness of C-PPM for all principal classes C defined by excluded patterns of size at least 4, except for five patterns where the problem remains open. (These are the same special cases as in Chapter 2, and for the same reason that the corresponding classes do not have the BP).

Chapter 6: The focus in Chapter 6 is on C-pattern PPM when C is a monotone grid class (of underlying gridding matrix M, *i.e.* $C = \mathsf{Grid}(M)$). It should be noted that $\mathsf{Grid}(M)$-PPM is proved to be in P by Theorem 5.5 in the previous chapter. So in Chapter 6, the pattern is still required to belong to $\mathsf{Grid}(M)$, but not the text.

The complexity of $\mathsf{Grid}(M)$-pattern PPM obeys a trichotomy, which follows one established for the growth of the treewidth, and which depends on the cycle structure of the matrix M (meaning of the underlying graph which connects non-empty entries of M which are closest in the same row or in the same columns). When M is acyclic, usual grid classes techniques are used to show that $\mathsf{Grid}(M)$ has bounded pathwidth (hence treewidth), yielding that $\mathsf{Grid}(M)$-pattern PPM is in P. When M is unicyclic in each connected component, it is shown that the treewidth grows like the squareroot of the size. The proof involves finding a drawing of a certain graph associated to the pattern in the projective plane, which I found nice and original. It follows that $\mathsf{Grid}(M)$-pattern PPM in NP-complete, with a lower bound for the complexity under ETH which is exponential in the squareroot of the size. Finally, when $\mathsf{Grid}(M)$ is polycyclic, the treewidth can grow linearly, and no improvement to the complexity of PPM is obtained compared to the general case.

This chapter is concluded by an extension of the previous results to classes $\mathsf{Grid}(M)$ where entries are not required to be monotone, but to have bounded treewidth. The cycle condition that drives the complexity of $\mathsf{Grid}(M)$-pattern PPM above is propagated to this setting, but need to be complemented by a technical property, called the bumper-ended path property.

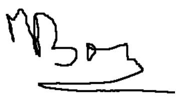
Chapter 7: This final chapter deals with the problem of recognizing whether a permutation belongs to a given class C, especially when C is a merge class. The topic studied here seemed to me mostly disconnected from the rest of the thesis. An exception is certainly the NP-hardness of recognizing whether a permutation is a merge of two permutations, each avoiding a fixed (but arbitrary) simple pattern α, which is the property expressible in MSO but not in FO encountered in Chapter 3.

While the recognizability problem is very natural, the motivation for focusing on merge classes is unclear to me. In addition, the results presented are mostly preliminary. (And this is also visible looking at the open problems corresponding to Chapter 7 listed in the conclusion.) While this shows that Michal Opler has perspectives for continuing his research work, I don't think it was a good choice to include the material of this chapter in the thesis (which is more than satisfactory even, without the results of Chapter 7).

Conclusion: It summarizes many open questions encountered along the thesis. They show the breadth of the results obtained in the thesis, and identify clearly which points are still missing to obtain an even more complete view of the complexity of PPM and its restricted variants. I also find that the questions on the expressivity of FO and MSO on permutations are very interesting.

In conclusion, the work presented by Michal Opler makes a very good PhD thesis, and I strongly recommend the admittance of the candidate to defend his thesis.

Yours sincerely,



Mathilde Bouvel