Rannsóknasjóður

2021

Émile Nadeau

Extending the Combex framework

The Icelandic Research Fund 2021
Doctoral Student Grant — New proposal
Detailed project description

rannís

GUIDELINES

Please note that *all proposals and appendices* should be in English.

The text in angle brackets (<text>) in this document is provided for guidance purposes only, and should be overwritten, by applicant, with the requested information. Please insert appropriate material (i.e. text, pictures and/or tables) and add the principal investigator's name to the header and the project's name to the footer.

The detailed project description should provide the following information, and be divided into the following sections (the order and titles should not be changed):

- *A.* *Specific aims of the project, research questions/hypotheses, feasibility, originality and impact*
- *B.* *Present state of knowledge in the field*
- *C.* *Research plan (time and work plan, present status of project, methodology and milestones) and deliverables*
- *D.* *Management and co-operation (domestic/foreign)*
- *E.* *Proposed publication of results and data storage (including open access policy)*
- *F.* *Contribution of doctoral and master's degree students to the project*

The project description should not exceed **5 pages** plus front page and guidelines (1.5 line spacing, 12 point Times/Times New Roman, or similar, 2.5 cm side margins). The bibliography is submitted in a separate file with no length limits. Convert the file to pdf before submitting. Please note:

→ ***Incomplete proposals will be rejected***
→ ***Proposals which exceed length limitations will be rejected without review***
→ ***Corrections or amendments after the previously announced deadline will be rejected***
→ ***The clarity and the overall quality of the presentation are taken into consideration when proposals are reviewed***
→ ***Applications that are not submitted using the most recent application form available will be rejected***

For further information please refer to the ***Icelandic Research Fund Handbook 2021.***

Please do not delete, overwrite, or amend this Guideline page in your submission

A. Specific aims of the project, research questions/hypotheses, feasibility, originality and impact

Combinatorial exploration is an experimental approach that rigorously derives structural results about mathematical objects. When a human has discovered the structure of an object there are several tools, often automatic, which allow various properties of the object to be computed. However, the steps from the problem statement and the original object to the structure is often ad-hoc. A framework is being developed at Reykjavik University to automate and formalize these steps. A beta version of the framework, called Combex (Combinatorial Exploration) exists. It has mainly been applied in the field of permutation patterns and is called in that context the PermScope algorithm. This algorithm has been able to discover new theorems and rediscover several results spanning dozens of papers in the research literature.

We propose to integrate techniques from machine learning to enhance the beta, as well as improving the algorithm to find more general descriptions of mathematical objects. The outcome of this proposal will be publications in journals and presentations at international conferences. The implementation of these theoretical algorithms will be made available free and open source and will be accompanied by a comprehensive manual.

The Combex framework

Combex performs combinatorial exploration in two main stages. First is the expansion stage, where *strategies* are applied to the objects of interest, resulting in a universe of rules relating the objects. The second stage is to check if there exists a *specification* for the original object of interest within this set of rules. The term *specification* has a technical meaning, but we can think of it as a system of equations describing the structure of an object. If a specification exists in the universe Combex is guaranteed to find it. If no specifications are found, Combex re-enters the expansion stage, enlarging the universe, until searching again, *etc*.

We will show below how Combex has been applied to the field of permutation patterns, but ask the reader to remember that the framework can be applied in other fields. As is to be expected with a beta version there is room for improvement. In particular, even though the framework is guaranteed to find a specification in the universe if one exists, there are cases where the universe contains enough information to describe the original object, without any specification existing for it.

There is no existing theory for this phenomenon but we say that the universe contains a *combinatorial system* for the object. Even if some combinatorial systems have been found by hand [8], there is currently no way to find them automatically. The proposer will develop heuristics and algorithms to find combinatorial systems.

The PermScope algorithm

A *permutation* is a reordering of the integers $1, 2, \ldots, n$ for some n. Although these are simple objects, they have deep connections to other fields. Famous examples include Schubert varieties [25], posets [12], water waves [23], genome rearrangements [17], planar maps [15] meanders [18], and sorting operators [22]. A permutation π *contains* the permutation p if certain technical conditions are satisfied. We only give an example here: The permutation $\pi = 526413$ contains $p = 132$, in the subsequences 264, 263 and 243; see Figure 1 where we have additionally circled the occurrences of p.

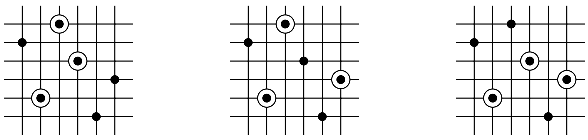

Figure 1: The permutation 526413 and three occurrences of the 132

The same permutation *avoids* 123, as it has no increasing subsequence of length three. In this context we refer to p as a *pattern*. The main object of study in this field is the (usually infinite) set of permutations that avoid a given pattern, denoted $\mathrm{Av}(p)$, and called a *permutation class*.

Domain-specific knowledge from the field of permutation patterns was added to Combex to create the PermScope algorithm. When PermScope succeeds it produces a specification in the form of a tree, shown in Figure 2, for the permutations avoiding the patterns 123 and 1432, first enumerated by West [36]. So far, the PermScope algorithm has proved to be a excellent

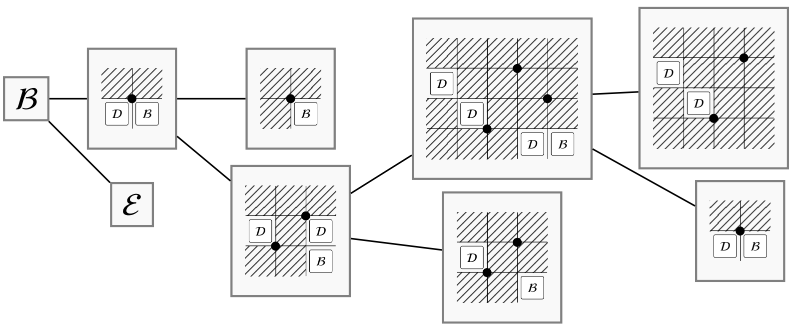

Figure 2: The structure of $\mathrm{Av}(123, 1432)$ in terms of simpler substructures

tool, providing structural and enumerative results that together subsume over a dozen of existing research papers.

The proposer will undertake the creation of a generalization of the PermScope algorithm that describes the structure of permutation classes in a finer way by accounting for permutation statistics. The study of permutation statistics is an active area of research in combinatorics [5; 13; 14; 16]. and similar research is being done more widely on statistics of other combinatorial objects [10; 21; 30]. One important example of a permutation statistic is the inversion number. An *inversion* is a pair of entries in a permutation such that the leftmost entry of the pair has value greater than the rightmost entry on the pair. The *inversion number* of a permutation is the number of pairs of entries it has that form an inversion; as such, the inversion number is a function from the set of all permutations to $\mathbb{N}$. For example, the inversions of the permutation 526413 (see Figure 1) are $52, 54, 51, 53, 21, 64, 61, 63, 41$ and 43. Hence, the inversion count of 526413 is 10. More generally, a *permutation statistic* is any function from the set of all permutations to $\mathbb{N}$.

Machine learning and heuristics

The Combex beta is a naive breadth-first search algorithm. The general objectives which the algorithm is trying to accomplish have clear analogies with other tree-search paradigms. For instance *iterative deepening* [24; 31], i.e., pursuing a search deeper at promising nodes before returning to a generally breadth-first approach. Since there is no existing expert knowledge in the field to determine how to evaluate what constitutes a "promising" node this approach would need to be coupled with machine learning of such evaluation functions based on the beta's experience over a wide variety of training examples, both successful and unsuccessful.

Feasibility, originality, and impact

Given a specification for a mathematical object there are methods, often automatic, for understanding the object structurally and enumeratively. To find a specification for an object, mathematicians have often used ad-hoc techniques and the theory has not been developed as much. Combex and the theory of combinatorial exploration will bridge this gap and automate the process of going from an object to a specification. Rather than rediscovering similar methods and repeating the same ideas, Combex will push mathematical developments in new directions.

The beta version has been very successful and found specifications for classes that had entire papers devoted to them, including but not limited to [1; 3; 11; 26; 29]. Despite these successes, numerous interesting cases still remain, in particular, Av(1324), the biggest open problem in the field, against which every new proof method is tested. If we are successful in finding a structure and generating function for unenumerated classes, in particular Av(1324), it would be considered a major breakthrough. Even tighter bounds on the enumeration of this set are worthy of publication.

B. Present state of knowledge in the field

There exist specialized algorithms for enumerating permutation classes, such as enumeration schemes of Zeilberger [37] with a generalization that accounts for statistics of Baxter [7]; regular insertion encoding of Albert, Linton, and Ruškuc [4]; substitution decomposition of classes with finitely many simple permutations of Bassino et al. [6]; and guessing a disjoint union of grid classes of Bean, Gudmundsson, and Ulfarsson [2; 9]. These algorithms are deterministic and either succeed or fail on their inputs. The Combex framework is different in that it generates a universe of true statements about the object under investigation, and from this universe one can often extract enough information to reveal the structure of the object. What we are proposing in this application is strengthening the framework in such a way that it can reveal the structure of the object with a smaller universe, therefore making it applicable to more problems.

C. Research plan (time and work plan, methodology, milestones, present status of project, etc.) and deliverables. Refer to more detailed description of milestones and deliverables in part 3 in the IRF electronic application system. Explain if consents and/or permits are needed

Time and work plan. The work packets are detailed under milestones below, as well as when they will be attempted.

Methodology. Some methods used will be traditional, coming from combinatorics, algorithms and computer science. We will also use newer tools, such as generalized grid classes and mesh patterns. We will also introduce new strategies, and apply methods from computer science, such as probabilistic search and machine learning.

Milestones.

WP1 Develop algorithm to find combinatorial systems automatically. (Jan-Dec ’21)

WP2 Implement iterative deepening search in the framework. (Jan-Jun ’22)

WP3 Extend the PermScope algorithm to find specifications and combinatorial systems that describe permutation statistics in classes. (Jul-Dec '22)

Present status of the project. As we noted in the subsection on feasibility, originality, and impact of our work, our beta has been successful with several cases. As can be expected from a beta, there are many obvious avenues for improvements, such as searching in a more clever manner, looking for combinatorial systems and finding specifications that describe statistics.

Deliverables. Results obtained will be turned into papers and talks at conferences, as is customary with mathematical research. We hope that these results will lead to more work by others in the area. The developed code base will be made publicly available for others to use and extend.

D. Management and co-operation (domestic/foreign)

The proposer has a master's degree in Mathematics and Informatics from Université du Québec à Montréal. He has demonstrated great research potential by completing original research during his undergraduate and master programs. The student already has two peer-reviewed publications and a third publication devoted to permutation patterns under review. He is also a good programmer interested in algorithms

The proposer will be advised by Henning Ulfarsson, an assistant professor at Reykjavik University. Ulfarsson leads a research group in the field and has a large network of collaborators. He has solved two long-standing open problems in the field: the description of local complete intersection Schubert varieties, with Woo [35], and the description of West-3-stack-sortable permutations [33]. He co-created the BiSC [34] and Struct algorithms [2; 9]. He has advised many students on projects in the theory of permutations, e.g. [8; 19; 20; 27; 28; 32].

The PhD student will also collaborate with Jay Pantone who is an assistant professor in the Department of Mathematical and Statistical Sciences at Marquette University. Pantone was involved in the developments of the beta version of Combex and is an expert in analytic combinatorics and permutations patterns.

E. Proposed publication of results and data storage (including adherence to open access policy)

New results will be published in top-tier refereed open access international journals in discrete mathematics and related fields, as well as disseminated at seminars and conferences.