Parameterized complexity:
A tool to handle multiobjective problems.
Application to lung cancer detection.

Directeur de thèse : Laetitia Jourdan
Courriel : laetitia.jourdan@univ-lille.fr

Co-encadrant de thèse : Julien Baste
Courriel : julien.baste@univ-lille.fr

Laboratoire : CRIStAL - 9189
Équipe : ORKAD

The goal of this PhD Thesis is to build a new approach to solve multi-objective optimization problems based on the parameterized complexity techniques. The parameterized complexity is an approach that appears in the 1990s [4]. This new approach appears to be really efficient to handle single-objective parameterized problems. In this PhD Thesis, we would like to extend the potential of parameterized complexity such that it can also handle multi-objective optimization problems in a way that is competitive with the current techniques already known in multi-objective optimization. The objective will be to develop this new approach and test it on medical data provided by CHU Lille on lung cancer with Pr. Sébastien Hulo.

1 Scientific context

In a multi-objective problem, a solution S dominates a solution S' if for each objectif of the problem, the solution S performs better than the solution S'. This notion of domination determine a partial order $\leq$ on the set of the solutions. When solving a multi-objective problem, it is rare that there exists one solution S that dominates every other solution. The goal is then to find the pareto set that consists of all the solutions that are minimal with regard to the partial order $\leq$.

The most used techniques in order to construct the pareto front consists in constructing *archives*. An archive is a set of solutions that are determined by the algorithm and stored in memory. The archive is progressively updated by adding new solutions and removing each solution of the archive that is dominated by another solution of the same archive. The way the archive is updated is crucial in order to obtain high-performance algorithms. The literature is really rich in techniques to perform this update (see [8, Chapter 4] sor some of them). For citing only the best known of them, one can cite for instance the Adaptive Grid Archiving [6] or the Hypervolume Archiving [7]. Recently, in [2], the author used a notion of diversity [1] coming from the parameterized complexity in order to update the archive. The positive results from [2] lead to a more general question: How much the ideas coming from parameterized complexity can be used to improve the field of multi-objective optimization?

A canonical problem studied in parameterized complexity is Vertex Cover that is known to be NP-hard (see [3] for instance). This problem can be solved in time $2^{O(n)} \cdot n^{O(1)}$ and cannot be solved in time $2^{o(n)} \cdot n^{O(1)}$, unless the exponential time hypothesis fails [3]. The main goal of parameterized complexity is to be able to catch the hardness of the problem within a parameter that can be a parameter describing the structure of the input (like the treewidth tw that describes how topologically close the input graph is from a tree, see [3] for a formal definition) or a parameter connected to the problem itself like the size k of the solution. In particular, Vertex Cover can be solved in time $2^{O(\mathsf{tw})} \cdot n^{O(1)}$ and $2^{O(k)} \cdot n^{O(1)}$ [3]. In these complexity results, we observe that the combinatorial explosion does not depend on n anymore but is caught by the parameter. More generally, if n is the size of the instance and x is a parameter, the main goal of the parameterized complexity is to find an FPT-algorithm, i.e., an algorithm with running time $f(x) \cdot n^{O(1)}$.

This field, emerging since the 90s, appears to be really promising. While Downley and Fellows, the first authors on parameterized complexity were more interested by the mathematical point of view, the number of implementation of parameterized based algorithms grows. This is led by a wish to spread the usefulness of the approach in practice. A typical example of this is the PACE challenge[1] that is focused on the implementation of parameterized algorithms.

In previous works, lung cancer detection from VOCs (volatile organic compounds) has been modelized as a multi-objective problem [9] and solved thanks to MOCAI [5] by using classical algorithms. Through this thesis, we hope to improve the performance of the resolution algorithms.

2 Goal of the thesis

The goal of this thesis would be to elaborate an extension of the parameterized complexity that can handle multi-objective problems. Several approaches should be considered:

(i) In an classical archiving approach, use the parameterized complexity in order to improve the way the archive is updated. This would be a continuation of the preliminary work done in [2].

(ii) Use parameterized complexity in order to determine whether a given part of the solution space contain a solution. In this situation, we would consider the parameter to be connected to the solution space we want to explore and as long as this space is small, the parameter will also be small and so, the used algorithm would be polynomial.

(iii) Elaborate a parameterized approach that uses (ii), in order to construct an approximation of the pareto front of good quality with affordable running time.

(iv) Application of (iii) to MOCAI for lung cancer detection with medical data from CHU Lille.

In a first hand, the preliminary work [2] shows that combining parameterized techniques and archiving approaches in a clever way should lead to improvement of the techniques solving multi-objective optimization problem. This means that it is expected that (i) is reasonable reachable.

In a second hand, (ii) and (iii) are much more ambitious as it implies a new approach to construct the pareto front. Being able to do (ii) in an affordable amount of time would be

[1] https://pacechallenge.org/

the crucial tool. This would need a wise combination of the multi-objective and parameterized complexity approaches. If (ii) can be done, then one could expect (iii) to be doable also, leading to the expected new techniques.

When progressing in the different tasks, we plan to always test the progression using three specific problems. The first one is VERTEX COVER that is the canonical problem in parameterized complexity on which a lot is already known. The second one is the TRAVELLING SALESMAN PROBLEM that is a well-studied problem in multi-objective optimization du to the fact that when using an archiving approach, it quickly leads to archives of big size. The third one is the identification and quantification of VOCs for Cancer Prediction (iv). This problem has been modelized as a imbalanced classification problem and as a multi-objective optimization problem. The archives are big and computing effort to obtain results is important.

3 PhD schedule

Task 1: State of the art about both parameterized complexity and multi-objective optimization.

Task 2: Familiarisation with the toy problems that are VERTEX COVER and TRAVELLING SALESMAN PROBLEM with regard to both parameterized complexity and multi-objective optimization

Task 3: Determination of an algorithm for the archiving process that uses parameterized complexity. Answering (i).

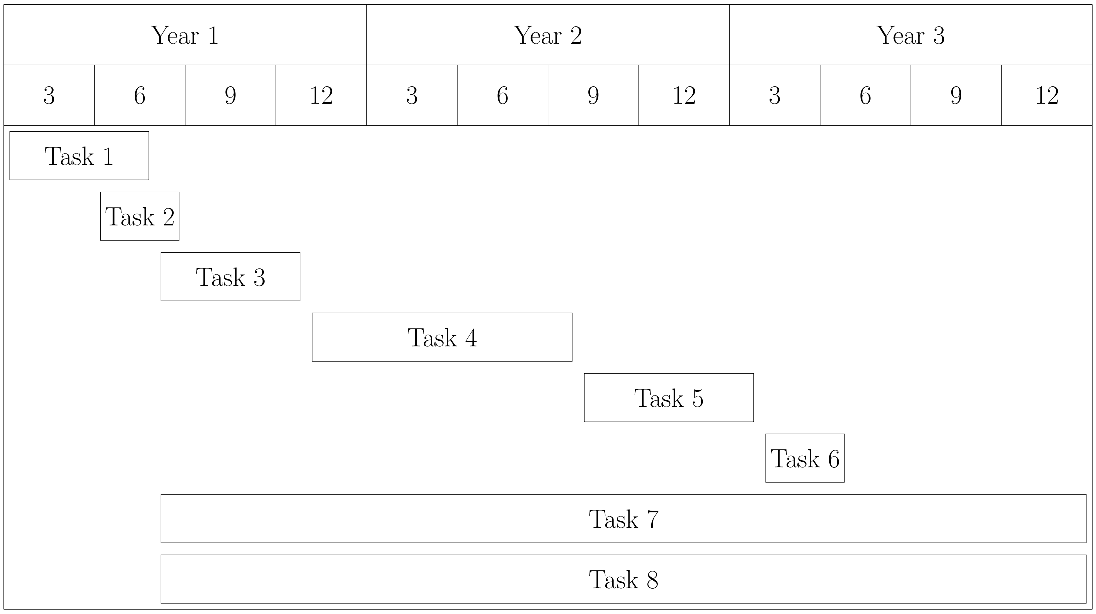
Task 4: Elaboration of parameterized complexity algorithms that determine whether a given part of the solution space contains a solution or not in an efficient way. Answering (ii)

Task 5: Elaboration of a new approach that uses parameterized complexity to construct a pareto front. Answering (iii)

Task 6: Implementation in MOCAI and experiments of Lung data

Task 7: Validation, Communication, Diffusion of the scientific results.

Task 8: Redaction of the PhD thesis. This will be a continuous work that will start after the state of the art until the end of the PhD.



References

[1] J. Baste, M. Fellows, L. Jaffke, T. Masarik, M. de Oliveira Oliveira, G. Philip, and F. Rosamond. Diversity of Solutions: An Exploration Through the Lens of Fixed-Parameter Tractability Theory. *Artificial Intelligence*, 303:103644, 2022.

[2] C. Boree, J. Baste, L. Jourdan, and L. Mousin. Impact of diversity on bounded archives for multi-objective local search. In preparation, 2023.

[3] M. Cygan, F. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh. *Parameterized Algorithms.* Springer, 2015.

[4] R. Downey and M. Fellows. *Parameterized Complexity.* Springer, 1999.

[5] J. Jacques, H. Martin-Huyghe, J. Lemtiri-Florek, J. Taillard, L. Jourdan, C. Dhaenens, D. Delerue, A. Hansske, and V. Leclercq. The detection of hospitalized patients at risk of testing positive to multi-drug resistant bacteria using moca-i, a rule-based "white-box" classification algorithm for medical data. *Int. J. Medical Informatics*, 142:104242, 2020.

[6] J. Knowles. *Local-search and hybrid evolutionary algorithms for Pareto optimization.* 2002.

[7] J. Knowles, D. Corne, and M. Fleischer. Bounded archiving using the lebesgue measure. In *The 2003 Congress on Evolutionary Computation (CEC)*, volume 4, pages 2490–2497, 2003.

[8] E. Talbi. *Metaheuristics: From Design to Implementation.* Wiley, 2009.

[9] S. Tari, L. Mousin, J. Jacques, M. Kessaci, and L. Jourdan. Impact of the discretization of vocs for cancer prediction using a multi-objective algorithm. In *Proceedings of the 14th International Conference Learning and Intelligent Optimization, (LION)*, volume 12096 of *Lecture Notes in Computer Science*, pages 151–157. Springer, 2020.