The algebra of the MinRank problem in post-quantum cryptography: Gröbner bases, complexity, and implementations

Abstract

This PhD project aims at foundational results on the complexity of the so-called MinRank problem which takes a key role in the design of post-quantum cryptosystems based either on error-correcting codes or on multivariate cryptography. The PhD candidate will leverage reductions to polynomial system solving and efficient linear algebra subroutines with an approach driven by the algebra of the MinRank problem.

Supervision team and scientific positioning. This PhD will be supervised by V. Neiger and L. Perret at Sorbonne University and É. Schost at Waterloo University, as a joint PhD at both universities through a "co-tutelle" agreement. V. Neiger is an expert on computer algebra and Gröbner bases. He recently studied the security of some rank metric code-based cryptosystems. L. Perret has a strong and renowned expertise in Post-Quantum Cryptography (PQC), in particular, on the use of computer algebra techniques to assess the security of cryptosystems. He already contributed to the development of MinRank-based attacks in multivariate cryptography. É. Schost is a top expert in computer algebra and polynomial system solving. He contributed to several results on determinantal systems which play a central role in this PhD project.

This PhD project is at the interface of computer science and computational mathematics. It requires a taste for computing, algorithm design, and complexity analysis, and at the same time for studying deeply the involved mathematical objects. We are seeking top candidates, preferably with an international experience who are willing to prepare a double degree PhD in France and in Canada.

1 Scientific context and objectives

1.1 The MinRank problem in post-quantum cryptography

Let $\mathbb{K}$ be a field and M be a $p \times q$ matrix with entries in the polynomial ring $\mathbb{K}[x_1, \ldots, x_n]$ (further, we assume that $p \leqslant q$). The MinRank problem consists in determining the minimum $r \in \mathbb{N}$ for which there exists $\boldsymbol{x} \in \mathbb{K}^n$ such that the evaluation of M at $\boldsymbol{x}$ has rank less than r. We call the set of points satisfying this property the realization set of the considered MinRank instance (in $\mathbb{K}^n$).

This problem is known to be $\mathcal{NP}$-hard and several post-quantum cryptosystems (i.e. which are expected to be immune against a quantum attacker) rely on the $\mathcal{NP}$-hardness of MinRank; see for example GeMSS and ROLLO which have been submitted to the NIST post-quantum cryptography standardization process. The post-quantum flavour of the MinRank problem is actually only one of its application domains. It has a long story and has been intensively studied from the mathematical viewpoint.

For instance, it is already known that the set of points realizing a MinRank instance has codimension $(p - r)(q - r)$ when the entries of the matrix M are chosen generically. The *degree* of such a set, which measures its complexity (for instance, when the set is finite, this degree coincides with its cardinality), is also known through an involved closed formula. We refer to [7] for further properties of determinantal ideals.

The connection of MinRank to polynomial system solving is rather immediate. The realization set of a MinRank instance can be seen as the simultaneous vanishing of all $r \times r$ minors of the matrix M. It can also be defined as the projection of the solution set to the kernel equations $K \cdot M = 0$ where K is a $(p - r + 1) \times p$ matrix with full rank, whose entries are unknown. Note that both modelings have been used in post-quantum cryptography to investigate the security of some cryptosystems [5, 11].

1.2 Polynomial system solving and Gröbner bases

As sketched above, solving a MinRank instance boils down to polynomial system solving. A classical and efficient framework for the latter problem is provided by the theory of Gröbner bases and the algorithms for computing them. Given a list of polynomial $f_1, \ldots, f_s$ in $\mathbb{K}[x_1, \ldots, x_n]$, we consider the ideal $\langle f_1, \ldots, f_s \rangle$ it generates, i.e. the set $\{q_1 f_1 + \cdots + q_s f_s \mid q_i \in \mathbb{K}[x_1, \ldots, x_n]\}$. Gröbner bases give appropriate bases of the *finite dimensional* vector spaces $E_d = \{q_1 f_1 + \cdots + q_s f_s \mid \deg(q_i f_i) \leqslant d\}$ for $d \in \mathbb{N}$, once an order $>$ on the monomials in $\mathbb{K}[x_1, \ldots, x_n]$, compatible with multiplication, is fixed. Thanks to the Noetherianity of the polynomial ring $\mathbb{K}[x_1, \ldots, x_n]$, it turns out that bases of these nested vector spaces for large enough $d \in \mathbb{N}$ allow us to define a *normal form* from $\mathbb{K}[x_1, \ldots, x_n]$ to the classes of the equivalence relation $f \simeq g \Leftrightarrow f - g \in \langle f_1, \ldots, f_s \rangle$. A consequence is that one can then compute "modulo the input equations" and then decide the existence of solutions with coordinates in an algebraic closure of $\mathbb{K}$, describe them through a triangular representation, etc.

The F4 algorithm [8] basically provides efficient subroutines to compute bases of the nested vector spaces E_d by taking as generators the row vectors formed by the coefficients (sorted w.r.t. $>$) of the mf_i's where m ranges over a subset of well-chosen monomials such that $\deg(mf_i) \leqslant d$. The bulk of the computation is to perform Gaussian elimination on the matrix formed by stacking these row vectors. To achieve practical efficiency, it actually reuses the basis computed for E_d to find the one of E_{d+1}, enabling faster steps of Gaussian elimination.

It turns out that the matrices constructed this way are *generically* rank defective, leading to useless computations: a non-negligible number of the rows actually reduce to zero. Many such reductions to zero simply come from the *commutativity* of the polynomial ring: since $f_i f_j = f_j f_i$, there is a non trivial linear relation between the row vectors mf_i and $m' f_j$ when m (resp. m') ranges over the monomials in f_j (resp. f_i). The F5 algorithm [9] handles this general issue in Gröbner basis computations by introducing a module viewpoint (in the sense of commutative algebra) as well as a data structure called signature. This allows one to keep track of these linear dependencies, and in generic cases, to only build and work with full rank matrices, thus avoiding reductions to zero. This is currently the best known framework regarding the computation of Gröbner bases. Besides, an accurate complexity analysis of a variant of the F5 algorithm is given in [4].

1.3 Objectives of the PhD

The central issue which is at the genesis of this thesis is that polynomial systems coming from MinRank are *not* generic. Complexity studies of the F5 algorithm do not apply there. In [10], the authors present a preliminary study of the algebraic properties of the MinRank systems, and identify the largest size of the matrices considered during the Gröbner basis computation in order to deduce an upper complexity bound. Apart from this, only heuristic results are known, with no proved complexity upper bound. The goal of this PhD is to adapt the F5 algorithm to the MinRank systems, obtain the most accurate possible upper and lower bounds on the complexity, and apply the new derived algorithms to challenging instances coming from the future post-quantum standards.

2 Research program

2.1 The algebra of MinRank

The first key methodological ingredient to settle is to understand how the algebra of MinRank systems gives rise to rank-deficient matrices even when using the F5 algorithm. This requires a deep understanding of an algebraic object named *module of syzygies*, which is rather involved in the case of ideals generated by MinRank systems. In the case of the modeling based on the $r \times r$ minors of the matrix M, an easy observation shows that there are syzygies which actually do *not* come from the commutativity of polynomials. Indeed, considering the Laplace expansion of $(r + 1) \times (r + 1)$ minors, one sees that some algebraic combinations of the $r \times r$ minors lie in the ideal generated by the MinRank system under consideration. This will affect the behaviour of the F5 algorithm, leading rank-deficient matrices and reductions to zero.

Hence, a first step is to establish under which conditions these syzygies, combined with the aforementioned usual ones coming from commutativity, generate the full module of syzygies; and otherwise to find a description of the remaining syzygies that are not generated. One may investigate such theoretical issues in the context where $\mathbb{K}$ is a field of low characteristic, as this classically induces some difficulties. Last, but not least, in case the entries of the matrix M are not generic themselves, we will need to investigate how this affects the module of syzygies (we will concentrate on extra structures coming from post-quantum cryptography). To do so, we will start from classical results in textbooks of commutative algebra dealing with determinantal ideals [7].

2.2 Modification of the F5 algorithm

As sketched above, understanding the module of syzygies of the MinRank systems is the key to design F5-like algorithms adapted to these systems. The first step towards an efficient F5 variant for MinRank systems is to adapt accordingly the stored signatures of the classical F5 algorithm to avoid the reductions to zero which are induced by the syzygies (in particular, the ones which we already identified above). A next step, is to investigate extra algebraic properties of MinRank systems to circumvent difficulties which may arise from the over-determinacy of MinRank systems (when using the modeling through minors) – this is indeed an issue to consider since the F5 algorithm is incremental. Also, as above, when the base field $\mathbb{K}$ has small characteristic or the entries of M enjoy extra properties, dedicated variants will be designed.

2.3 Complexity of the MinRank problem

The next step will be to study as accurately as possible the complexity of these algorithms. This is an involved question; in particular an accurate complexity analysis should fit in the framework of amortized complexity theory to take into account rows which are already reduced when encoding the aforementioned nested vector spaces E_d. The machinery underlying this type of complexity analyses uses generating series "à la Flajolet", which are called Hilbert series.

Another ambitious goal is to obtain *lower bounds* for these techniques based on linear algebra, by determining the minimal sizes of objects which must be computed within this framework.

2.4 Applications in post-quantum cryptography

It is crucial to investigate the security of future post-quantum standards and the topic will remain vivid in the next few years. Since the hardness of MinRank is central for several post-quantum cryptosystems, one can anticipate that the expected fundamental results, which were sketched above, will greatly impact towards this goal. In a recent series of papers, including [1, 2, 3, 6, 12], authors proposed new approaches to attack post-quantum cryptosystems with MinRank as well as novel heuristic techniques for solving MinRank that led to new expected complexity bounds on MinRank. The outcome of this PhD project is then to provide a more rigorous framework to assess the new complexity results stated for MinRank, and to replace the unproven arguments underlying the existing heuristics by either proofs, or usual genericity assumptions, or classical conjectures.

References

[1] M. Bardet and P. Briaud. "An Algebraic Approach to the Rank Support Learning Problem". In: *Post-Quantum Cryptography - 12th International Workshop, PQCrypto 2021, Daejeon, South Korea, July 20-22, 2021, Proceedings.* Ed. by J. H. Cheon and J. Tillich. Vol. 12841. Lecture Notes in Computer Science. Springer, 2021, pp. 442–462.

[2] M. Bardet, P. Briaud, M. Bros, P. Gaborit, V. Neiger, O. Ruatta, and J. Tillich. "An Algebraic Attack on Rank Metric Code-Based Cryptosystems". In: *Advances in Cryptology - EUROCRYPT 2020, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part III.* Ed. by A. Canteaut and Y. Ishai. Vol. 12107. Lecture Notes in Computer Science. Springer, 2020, pp. 64–93.

[3] M. Bardet, M. Bros, D. Cabarcas, P. Gaborit, R. A. Perlner, D. Smith-Tone, J. Tillich, and J. A. Verbel. "Improvements of Algebraic Attacks for Solving the Rank Decoding and MinRank Problems". In: *Advances in Cryptology - ASIACRYPT 2020, Daejeon, South Korea, 2020, Proceedings.* Ed. by S. Moriai and H. Wang. Vol. 12491. Lecture Notes in Computer Science. Springer, 2020, pp. 507–536.

[4] M. Bardet, J.-C. Faugère, and B. Salvy. "On the complexity of the F_5 Gröbner basis algorithm". In: *J. Symbolic Comput.* 70 (2015), pp. 49–70.

[5] L. Bettale, J. Faugère, and L. Perret. "Cryptanalysis of HFE, Multi-HFE and Variants for Odd and Even Characteristic". In: *IACR Cryptol. ePrint Arch.* (2011), p. 399.

[6] W. Beullens. "Improved Cryptanalysis of UOV and Rainbow". In: *Advances in Cryptology - EUROCRYPT 2021, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part I.* Ed. by A. Canteaut and F. Standaert. Vol. 12696. Lecture Notes in Computer Science. Springer, 2021, pp. 348–373.

[7] W. Bruns and U. Vetter. *Determinantal rings.* Vol. 45. Monografias de Matematica [Mathematical Monographs]. Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 1988, pp. viii+236.

[8] J.-C. Faugére. "A new efficient algorithm for computing Gröbner bases (F_4)". In: vol. 139. 1-3. Effective methods in algebraic geometry (Saint-Malo, 1998). 1999, pp. 61–88.

[9] J.-C. Faugère. "A new efficient algorithm for computing Gröbner bases without reduction to zero (F_5)". In: *Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation.* ACM, New York, 2002, pp. 75–83.

[10] J.-C. Faugère, M. Safey El Din, and P.-J. Spaenlehauer. "On the complexity of the generalized MinRank problem". In: *J. Symbolic Comput.* 55 (2013), pp. 30–58.

[11] L. Goubin and N. T. Courtois. "Cryptanalysis of the TTM Cryptosystem". In: *Advances in Cryptology - ASIACRYPT 2000.* Ed. by T. Okamoto. Vol. 1976. Lecture Notes in Computer Science. Springer, 2000, pp. 44–57.

[12] C. Tao, A. Petzoldt, and J. Ding. "Efficient Key Recovery for All HFE Signature Variants". In: *Advances in Cryptology - CRYPTO 2021, August 16-20, 2021, Proceedings, Part I.* Ed. by T. Malkin and C. Peikert. Vol. 12825. Lecture Notes in Computer Science. Springer, 2021, pp. 70–93.