The Hardness of Code Equivalence over $\mathbb{F}_q$ and Its Application to Code-Based Cryptography

Nicolas Sendrier$^1$ and Dimitris E. Simos$^{1,2}$

$^1$ INRIA Paris-Rocquencourt
Project-Team SECRET
78153 Le Chesnay Cedex, France
$^2$ SBA Research
1040 Vienna, Austria
{nicolas.sendrier,dimitrios.simos}@inria.fr,
dsimos@sba-research.org

Abstract. The code equivalence problem is to decide whether two linear codes over $\mathbb{F}_q$ are identical up to a linear isometry of the Hamming space. In this paper, we review the hardness of code equivalence over $\mathbb{F}_q$ due to some recent negative results and argue on the possible implications in code-based cryptography. In particular, we present an improved version of the three-pass identification scheme of Girault and discuss on a connection between code equivalence and the hidden subgroup problem.

Keywords: Code Equivalence, Isometry, Hardness, Zero-Knowledge Protocols, Quantum Fourier Sampling, Linear Codes.

1 Introduction

The purpose of this work is to examine the applications of the worst-case and average-case hardness of the Code Equivalence problem to the field of code-based cryptography. The latter problem is, given the generator matrices of two $q$-ary linear codes, how hard is it to decide whether or not these codes are identical up to an isometry of Hamming space? The support splitting algorithm (SSA) \cite{SSA} runs in polynomial time for all but a negligible proportion of all linear codes, and solves the latter problem by recovering the isometry when it is just a permutation of the code support.

The McEliece public-key cryptosystem \cite{McEliece} and Girault’s zero-knowledge protocol \cite{Girault}, both candidates for post-quantum cryptography, are related to the hardness of permutationally equivalent linear codes. For the McEliece cryptosystem, the SSA is able to detect some weak keys but a polynomial attack is infeasible due to the large number of possible private keys. However, the security of Girault’s zero-knowledge protocol is severely weakened and cannot longer be used with random codes but only with weakly self-dual codes (the hard instances of SSA).
Recently in [29], the worst-case and average-case hardness of code equivalence over $\mathbb{F}_q$ was studied and it was shown that in practice, SSA could be extended for $q \in \{3, 4\}$, and similarly solve all but an exponentially small proportion of the instances in polynomial time, when isometries are under consideration. However, for any fixed $q \geq 5$, the problem seems to be intractable for almost all instances.

In light of these new results, we repair Girault’s zero-knowledge protocol over $\mathbb{F}_q$, when $q \geq 5$, by showing that random codes are again a viable option. Moreover, the context of the framework built in [11] suggests that codes with large automorphism groups resist quantum Fourier sampling as long as permutation equivalence is considered. We examine whether it is possible to extend these results, when a more general notion of code equivalence over $\mathbb{F}_q$ is taken into account, in particular when the equivalence mapping is an isometry and not just a permutation of the code support.

The paper is structured as follows. In section 2 we define the different notions of equivalence of linear codes over $\mathbb{F}_q$ when isometries are considered, while in section 3 we formally define the Code Equivalence problem and present a thorough analysis of its hardness. In section 4 we review the protocol of Girault together with its weakness and repair its security using results based on the hardness of code equivalence, while in the last section we elaborate on the connection between code equivalence over $\mathbb{F}_q$ and the quantum Fourier sampling.

2 Equivalence of Linear Codes over $\mathbb{F}_q$

Code equivalence is a basic concept in coding theory with several applications in code-based cryptography; the McEliece public-key cryptosystem [23], Girault’s identification scheme [17] and the CFS signature scheme [10], to name a few. The notion of equivalence of linear codes used in code-based cryptography usually involves only permutations as the code alphabet is the binary field. However, this is by far the case in coding theory where for a more general notion of equivalence all isometries of the Hamming space have to be included. In this section, we review the concept of what it means for codes to be “essentially different” by considering the metric Hamming space together with its isometries, which are the maps preserving the metric structure. This in turn will lead to a rigorous definition of equivalence of linear codes and as we shall see later on may provide additional applications in cryptography. In fact, we will call codes isometric if they are equivalent as subspaces of the Hamming space.

Let $\mathbb{F}_q$ be a finite field of cardinality $q = p^r$, where the prime number $p$ is its characteristic, and $r$ is a positive integer. As usual, a linear $[n, k]$ code $C$ is a $k$-dimensional subspace of the finite vector space $\mathbb{F}_q^n$ and its elements are called codewords. We consider all vectors, as row vectors. Therefore, an element $\mathbf{v}$ of $\mathbb{F}_q^n$ is of the form $\mathbf{v} := (v_1, \ldots, v_n)$. It can also be regarded as the mapping $\mathbf{v}$ from the set $\mathcal{I}_n = \{1, \ldots, n\}$ to $\mathbb{F}_q$ defined by $v(i) := v_i$. The Hamming distance (metric) on $\mathbb{F}_q^n$ is the following mapping,

$$d : \mathbb{F}_q^n \times \mathbb{F}_q^n \rightarrow \mathbb{N} : (x, y) \mapsto d(x, y) := |\{i \in \{1, 2, \ldots, n\} \mid x_i \neq y_i\}|.$$