Constructions of Codes from Number Fields

Venkatesan Guruswami
MIT Lab for Computer Science
200 Technology Square
Cambridge, MA 01239.

Abstract —

We define number-theoretic error-correcting codes based on algebraic number fields, thereby providing a generalization of Chinese Remainder Codes akin to the generalization of Reed-Solomon codes to Algebraic-geometric codes. Our construction is very similar to (and in fact less general than) the one given by Lenstra [1], but the parallel with the function field case is more apparent, since we only use the finite (non-archimedean) places for the encoding. We prove that over an alphabet size as small as 19, there even exist *asymptotically good* number field codes of the type we consider. This result is based on the existence of certain number fields that have an infinite class field tower in which some primes of small norm split completely.

I. Introduction

A family of algebraic codes that have received some study are number-theoretic redundant residue codes called the "Chinese Remainder codes" (henceforth called CRT codes). Here the messages are identified with integers with absolute value at most K (for some parameter K that governs the rate) and a message m is encoded by its residues modulo n primes $p_1 < p_2 < \cdots < p_n$. If $K = p_1 \cdot p_2 \cdots p_k$ and $n > k$, this gives a redundant encoding of m and the resulting code has distance $n-k+1$. CRT codes are the number-theoretic analogue of Reed-Solomon codes. Since Reed-Solomon codes are a specific example of the more general family of algebraic-geometric codes, it is natural to ask if there is a natural generalization of CRT codes akin to the generalization of Reed-Solomon codes to algebraic-geometric codes, and if so what properties do such codes have and what sort of parameters do they achieve. It is this question which is addressed in this work.

The algebraic codes (Reed-Solomon, AG, CRT codes) all have a common construction philosophy. They are defined based on an underlying integral domain R whose elements r come equipped with some notion of "size", denoted $\mathsf{size}(r)$. The messages of the code are the elements of the ring R whose size is at most a parameter B (this parameter governs the rate of the code). The encoding of a message $m \in R$ is given by

$$m \mapsto \mathsf{Enc}(m) = \langle m/I_1, m/I_2, \cdots, m/I_n \rangle$$

where I_j, $1 \le j \le n$ are n (distinct) prime ideals of R. (For instance, in the case of Reed-Solomon codes, we have $R = F[X]$ for some field F and $I_j = (X - \alpha_j)$ – the ideal generated by the polynomial $(X - \alpha_j)$ – for $1 \le j \le n$, where $\alpha_1, \dots, \alpha_n$ are *distinct* elements of F.)

There are two properties of a code that of primary concern in its design, namely (a) its rate, and (b) its minimum distance. The rate property of the code constructed by the above scheme follows from an estimate of the number of elements m of R that have $\mathsf{size}(m) \le B$. The distance of the code follows by using further properties of the $\mathsf{size}(\cdot)$ function such as:

1. For elements $a, b \in R$, $\mathsf{size}(a-b)$ is "small" whenever $\mathsf{size}(a)$ and $\mathsf{size}(b)$ are both "small".
2. If $f \neq 0$ belongs to "many" ideals among $I_1, I_2, \dots, I_n$, then $\mathsf{size}(f)$ cannot be "too small".

In this paper, we use the ring of integers of a number field (a finite extension of the field $\mathbb{Q}$ of rational numbers) to define a general family of codes called Number Field codes or NF-codes. These generalize CRT codes and are the parallel of AG-codes in the number-theoretic world. We define an appropriate notion of size and prove the properties above for this notion of size, and thus estimate the rate and minimum distance of our codes.

We nest briefly highlight how our codes are constructed and some of our results concerning asymptotically good NF-codes. NF-codes were already studied by Lenstra [1] and in the end we compare his results to ours.

II. Construction of NF-codes

Naturally, a description of our codes requires several concepts from algebraic number theory, which the reader can find in any standard text (eg. [2]).

Let $K/\mathbb{Q}$ be a number field of degree $[K : \mathbb{Q}] = M$. Then there are M distinct field homomorphisms (called embeddings) $\tau_i : K \to \mathbb{C}$ of the field K into $\mathbb{C}$ which leave $\mathbb{Q}$ fixed. Out of these let r of the embeddings be into the reals, say $\tau_1, \dots, \tau_r : K \to \mathbb{R}$, and let the remaining $2s = M - r$ embeddings be complex. We refer to this pair (r, s) as the *signature* of K. These $2s$ embeddings come in s pairs of complex conjugate non-real embeddings, say $\sigma_j, \bar{\sigma}_j : K \to \mathbb{C}$ for $1 \le j \le s$.

Let $\mathcal{O}_K$ be the ring of integers in K. We first define the notion of size of an element $x \in \mathcal{O}_K$. For a technical reason, we define the size in terms of a "shift" parameter $\mathbf{z} \in \mathbb{R}^r \times \mathbb{C}^s$. Given such a $\mathbf{z} = \langle z_1, z_2, \dots, z_{r+s} \rangle$ with $z_i \in \mathbb{R}$ for $1 \le i \le r$ and $z_j \in \mathbb{C}$ for $r < j \le r+s$, the $\mathbf{z}$-shifted size $\mathsf{size}_{\mathbf{z}}(x)$ of $x \in \mathcal{O}_K$ is defined as follows. For $i = 1, 2, \dots, r$, define $a_i^{(x)} = |\tau_i(x) - z_i|$, and for $1 \ge j \le s$, define $b_j^{(x)} = |\sigma_j(x) - z_{r+j}|^2$. Now define

$$\mathsf{size}_{\mathbf{z}}(x) = \sum_{i=1}^{r} a_i^{(x)} + \sum_{j=1}^{s} 2\sqrt{b_j^{(x)}} \tag{1}$$

A number field code (NF-code for short) $\mathcal{C} = \mathcal{C}_K$, based on a number field K, has parameters $(n, \mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n;\ B)$ where n is the blocklength of the code, $\mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n$ are distinct (non-zero) prime ideals of $\mathcal{O}_K$, and B is a positive real. The i^{th} position of the code $\mathcal{C}$ is defined over an alphabet of size $\|\mathfrak{p}_i\| \stackrel{\text{def}}{=} |\mathcal{O}_K/\mathfrak{p}_i|$ for $1 \le i \le n$. Assume w.l.o.g that $\|\mathfrak{p}_1\| \le \|\mathfrak{p}_2\| \le \cdots \le \|\mathfrak{p}_n\|$. We now formally define our construction of NF-codes.

Definition 1 *The code $\mathcal{C} = \mathcal{C}_K$ based on a number field K with parameters $(n, \mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n;\ B;\ \mathbf{z})$ is defined as follows. The message set of $\mathcal{C}$ is $\{m \in \mathcal{O}_K : \mathsf{size}_{\mathbf{z}}(m) \le B\}$. The encoding function is $\mathsf{Enc}_{\mathcal{C}}(m) = \langle m/\mathfrak{p}_1, m/\mathfrak{p}_2, \ldots, m/\mathfrak{p}_n\rangle$.*

The following lemma lower bounds the distance of a code as defined above.

Lemma 2 *For a number field code $\mathcal{C} = \mathcal{C}_K$ based on a field K with $[K : \mathbb{Q}] = M$ with parameters $(n, \mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n;\ B;\ \mathbf{z})$, if t $(1 \le t \le n)$ is such that $\|\mathfrak{p}_1\| \times \|\mathfrak{p}_2\| \cdots \times \|\mathfrak{p}_t\| > (2B/M)^M$, then the distance $d(\mathcal{C})$ of $\mathcal{C}$ is* at least $(n - t + 1)$. *In particular, $d(\mathcal{C}) > n - \frac{M \log(2B/M)}{\log \|\mathfrak{p}_1\|}$.*

To estimate the rate of the code we need a lower bound (or good estimate) of the number of elements x of $\mathcal{O}_K$ with $\mathsf{size}_{\mathbf{z}}(x) \le B$. The key quantity in such a lower bound is the *root discriminant* of a number field K, denoted rd_K. The root discriminant of K is defined as $|D_K|^{1/M}$ where D_K is the discriminant of K and $M = [K : \mathbb{Q}]$ is the degree of K. In terms of the root discriminant, such a lower bound on the number of elements of small size is given by *Minkowski theory.*

Lemma 3 *Let K be a number field of degree $[K : \mathbb{Q}] = M$ and signature (r, s). Let $\mathcal{C} = \mathcal{C}_K$ be a number field code defined with parameters $(n, \mathfrak{p}_1, \mathfrak{p}_2, \ldots, \mathfrak{p}_n;\ B;\ \mathbf{z})$ with $\|\mathfrak{p}_1\| \le \cdots \le \|\mathfrak{p}_n\|$. Then there exists a choice of the "shift" $\mathbf{z}$ for which the rate $R(\mathcal{C})$ of $\mathcal{C}$ is at least*

$$R(\mathcal{C}) \ge \frac{\log\left(2^r \pi^s B^M\right) - \log M! - M \log \sqrt{\mathrm{rd}_K}}{n \log \|\mathfrak{p}_n\|}\ . \tag{2}$$

III. Asymptotically good NF-codes

From Lemmas 2 and 3, we note that in order to construct NF-codes of good rate and distance over a reasonably small alphabet, we need a number field K such that (a) it has small root discriminant rd_K, and (b) its ring of integers $\mathcal{O}_K$ has several prime ideals of small norm. In particular, we are interested in whether it is possible to achieve non-vanishing rate and relative distance, i.e., whether we can get asymptotically good NF-codes.

The answer to this question is yes, though the constructions end up using deep results from algebraic number theory concerning the existence of certain infinite *Hilbert class field towers.* We refer the reader to [2] for background on class field theory. Infinite class field towers provide a tower of function fields with bounded root discriminant. For us, we need to add further "splitting" conditions in the tower as we require several primes of small norm in the fields. The formal notion we use is that of a *T-decomposing p-class field tower.* For any number field k and a set of primes T in $\mathcal{O}_k$, the T-decomposing p-class field tower of k is obtained by repeatedly taking maximal unramified abelian p-extensions in which every prime in T splits completely. Using a result of [3] on existence of such infinite class field towers, we obtain the following specific construction:

Lemma 4 *Let $d' = 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 23 \cdot 29 \cdot 37 \cdot 41$, and let $K' = \mathbb{Q}(\sqrt{-d'})$. Then: (i) $\mathrm{rd}_{K'} = \sqrt{4d'} \simeq 246515.72$, (ii) $\mathcal{O}_{K'}$ has a set T' of four prime ideals, two of which lie above 17 (and have norm 17) and two of which lie above 19 (and have norm 19), and (iii) K' has an infinite T'-decomposing 2-class field tower.*

Using the above construction and the prime ideals that lie above 17 and 19 in the tower of number fields for encoding, we get a family $\{\mathcal{C}_n\}_{N \ge 1}$ of NF-codes over GF(19) whose rate $R(\mathcal{C}_n)$ and relative distance $\delta(\mathcal{C}_n)$ satisfy:

$$R(\mathcal{C}_n) > \frac{\log 17}{\log 19}\Big(1 - \delta(\mathcal{C}_n) - \frac{\log 14.5}{\log 17}\Big)\ .$$

Thus we can get asymptotically good codes over GF(19) for any value of relative minimum distance that is at most $(1 - \frac{\log 14.5}{\log 17}) \simeq 0.056$.

Theorem 5 *There exist asymptotically good number field codes over an alphabet of size 19.*

Comparison with [1]: Lenstra [1] had long back already considered the construction of codes from algebraic number fields. The construction in [1] is actually more general, and also allows the *infinite* places to be used for encoding, while our definition uses only the *finite* places for encoding. However, we were able to get asymptotically good codes even with this restriction. Moreover, our focus is on getting good codes over a small alphabet, while the focus in [1] was on the performance in the limit of large (but constant) alphabet size q. See the full version of this paper for a more detailed comparison.

References

[1] H. W. Lenstra. Codes from algebraic number fields. In: M. Hazewinkel, J.K. Lenstra, L.G.L.T. Meertens (eds), *Mathematics and computer science II, Fundamental contributions in the Netherlands since 1945*, CWI Monograph 4, pp. 95-104, North-Holland, Amsterdam, 1986.

[2] J. Neukirch. *Algebraic Number Theory.* Grundlehren der mathematischen Wissenschaften, Volume 322, Springer-Verlag, 1999.

[3] M. A. Tsfasman and S. G. Vlădut. *Asymptotic properties of global fields and generalized Brauer-Siegel Theorem.* Preprint IML-98-35 (Marseille), 1998.