Error Correcting Codes: Combinatorics, Algorithms and Applications (Spring 2009)

Lecture 5: Introduction to Asymptotic Analysis and Finite Fields

January 23, 2009

Lecturer: Atri Rudra *Scribe: Swapnoneel Roy*

1 Asymptotic notation

In this course, we will use asymptotic notations frequently. Typically, these notations will be used to bound the running time of an algorithm, which is defined in terms of a function whose domain is the set of natural numbers N = {0, 1, 2, ...}. Please refer to [1] for a more detail presentation of the subsequent definitions.

1.1 O-notation

We use O-notation as the asymptotic equivalent of "$\leq$". For a given function $g(n)$, we denote by $O(g(n))$ (pronounced "big-oh of g of n" or sometimes just "oh of g of n") the set of all functions $f(n)$, such that there exist positive constants c and n_0 such that $0 \leq f(n) \leq cg(n)$ for all $n \geq n_0$.

1.2 Ω-notation

Just as O-notation provides an asymptotic upper bound on a function, Ω-notation provides an asymptotic lower bound. In other words, Ω-notation is the asymptotic equivalent of "$\geq$". For a given function $g(n)$, we denote by $\Omega(g(n))$ (pronounced "big omega of g of n" or sometimes just "omega of g of n") the set of all functions $f(n)$, such that there exist positive constants c and n_0 such that $0 \leq cg(n) \leq f(n)$ for all $n \geq n_0$.

1.3 o-notation

The asymptotic upper bound provided by O-notation may or may not be "tight". The bound $2n^2 \leq O(n^2)$ [1] is asymptotically tight, but the bound $2n \leq O(n^2)$ is not. We use o notation to denote an upper bound that is not asymptotically tight. We formally define $o(g(n))$ ("little-oh of g of n") as the set of all functions $f(n)$ such that for every positive constant $c > 0$, there exists a constant $n_0 > 0$ such that $0 \leq f(n) < cg(n)$ for all $n \geq n_0$.
For example, $2n \leq o(n^2)$, but $2n^2 \not\leq o(n^2)$. The definitions of O-notation and o-notation are similar. The main difference is that in $f(n) \in O(g(n))$, the bound $0 \leq f(n) \leq cg(n)$

[1] Technically this should be stated as $2n^2 \in O(n^2)$ but we will henceforth overload the notation to say $2n^2 \leq O(n^2)$

holds for some constant $c > 0$, but in $f(n) = o(g(n))$, the bound $0 \leq f(n) < cg(n)$ holds for all constants $c > 0$. Intuitively, in the o-notation, the function $f(n)$ becomes insignificant relative to $g(n)$ as n approaches infinity; that is, $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. Some authors use this limit as a definition of the o-notation; this definition also restricts the functions to be asymptotically nonnegative.

1.4 ω-notation

By analogy, ω-notation is to Ω-notation as o-notation is to O-notation. We use ω-notation to denote a lower bound that is not asymptotically tight. One way to define it is by $f(n) = \omega(g(n))$ if and only if $g(n) = o(f(n))$. Formally, we define $\omega(g(n))$ ("little-omega of g of n") as the set of all functions $f(n)$ such that for every positive constant $c > 0$, there exists a constant $n_0 > 0$ such that $0 = cg(n) < f(n)$ for all $n \geq n_0$.
For example, $n^2/2 \geq \omega(n)$, but $n^2/2 \not\geq \omega(n^2)$. The relation $f(n) \in \omega(g(n))$ implies that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty$. That is, $f(n)$ becomes arbitrarily large relative to $g(n)$ as n approaches infinity.

2 Asymptotics of error-correcting codes

To correct t bits of error, we have seen that the distance of the code $\mathbf{C}$ should be at least $2t + 1$. To achieve good error correction capabilities of codes, we would like to maximize the distance of the code. However, to have a meaningful discussion about codes, we also need to keep track of the dimension, block length, and alphabet size. Definition 2.1 gives a compact way to represent the parameters of a code.

Definition 2.1 (q-ary code)**.** *An $(n, k, d)_q$ is a q-ary code where*

- *q is the alphabet size: $q = |\Sigma|$;*
- *n is the block length: $\mathbf{C} \subseteq \Sigma^n$;*
- *k is the dimension or message length: $|\mathbf{C}| = |\Sigma|^k$;*
- *d is the distance of the code $\mathbf{C}$: $d = \Delta(\mathbf{C})$.*

Hence $n - k$ bits are redundant in code $\mathbf{C}$. We need to minimize the redundancy, hence k needs to be maximized. The rate of a code $R(\mathbf{C}) = \frac{k}{n}$. Hence the goal translates to maximizing the rate $R(\mathbf{C})$.
Sometimes we will normalize the distance of a code to its block length, as we do with rate. To do so, we define the *relative distance* of a code $\mathbf{C}$.

Definition 2.2. *The relative distance denoted $\delta(\mathbf{C})$, is the ratio of distance of code $\mathbf{C}$ to its block length. $\delta(\mathbf{C}) = \frac{d(\mathbf{C})}{n}$.*

We want to study $R(\mathbf{C})$ vs $\delta(\mathbf{C})$ as $n \to \infty$. But we cannot do this for a fixed code, since its block length n is fixed. Hence we need a *family of codes* to do this.

Definition 2.3 (Family of q-ary codes)**.** *A family of q-ary codes* $\mathbf{C}$ *is just an infinite collection of codes* $\{\mathbf{C_i}\}_{i=1}^{\infty}$ *where each* $\mathbf{C_i}$ *is an* $(n_i, k_i, d_i)_q$ *code (typically* $n_{i+1} > n_i$*). We further define* $\mathbf{R}(\mathbf{C}) = \lim_{i\to\infty} \frac{k_i}{n_i}$ *and* $\delta(\mathbf{C}) = \lim_{i\to\infty} \frac{d_i}{n_i}$*.*

From now on, when we say a code, we would really mean a family of codes. As an example, the question if there is a code with positive R and δ is really the question that is there an infinite family $[n_i, k_i, d_i]$ q-ary codes of increasing block length n_i where $\frac{k_i}{n_i} \geq R$ and $\frac{d_i}{n_i} \geq \delta$ for some $R, \delta > 0$? The rate versus distance in the context of a family of codes is better studied as the *tradeoff* between R and δ.From now on, this is the form we will study the error correction versus redundancy question.
The case of $\mathbf{R}(\mathbf{C}) \geq \Omega(1)$, and $\delta(\mathbf{C}) \geq \Omega(1)$ is an interesting special case of the general R vs δ question. We call a code having this property an *asymptotically good code*.

Definition 2.4 (Asymptotically good code)**.** *A family* $\mathbf{C}$ *is called asymptotically good if* $\mathbf{R}(\mathbf{C})$, $\delta(\mathbf{C}) \geq \Omega(1)$.

3 Fields

Informally, a field is a set $\mathbb{F}$ on which two binary operations, called *addition* and *multiplication* are defined and which contains two distinguished elements 0 and 1 with $0 \neq 1$. The results of the operations addition, subtraction, multiplication, and division performed on the elements of $\mathbb{F}$ gives results contained in $\mathbb{F}$. The elements 0 and 1 are called the additive and multiplicative identities respectively of the field $\mathbb{F}$.

Formally a field is defined as follows:

Definition 3.1 (Field)**.** *A field* $\mathbb{F}$ *is the 5-tuple* $(S, +, \times, 0, 1)$ *with the following properties:*

- **Associativity***: For all* $a, b, c \in S$, $(a + b) + c = a + (b + c)$ *and* $(a\times b)\times c = a\times(b\times c)$*.*
- **Commutativity***: For all* $a, b \in S$, $a+b = b+a$ *and* $a\times b = b\times a$*.*
- **Identities***: For all* $a \in S$, $a+0 = a$*, and* $a\times 1 = a$*.*
- **Inverses***: For all* $a \in S$, $\exists\ -a \in S$ *such that* $a+(-a) = 0$ *and for all* $b \in S \setminus \{0\}$, $\exists\ b^{-1} \in S \setminus \{0\}$ *such that* $b\times b^{-1} = 1$*.*

The set of real numbers $\mathbb{R}$ is a field with the operations being normal addition and multiplication. 0 and 1 are the additive and multiplicative identities respectively. The additive inverse of an element is its negative, while the multiplicative inverse is its reciprocal.
However, set of integers $\mathbb{Z}$ is not a field since the division of two integers could result in a real number not in $\mathbb{Z}$.

3.1 Finite Fields

Finite fields are fields that contain only finitely many elements. As an example consider the field $\mathbb{Z}_p$ which is a set of integers modulo the prime integer p i.e. the elements are $\{0, 1, \cdots, p-1\}$. Addition here is defined as $+ \bmod p$ and multiplication as $\times \bmod p$.

The finite field $\mathbb{Z}_p$ has zero element 0, and identity 1. Computing with elements of $\mathbb{Z}_p$ is ordinary arithmetic of integers with reduction modulo p. It can be verified that $\mathbb{Z}_p$ is indeed a field.

Definition 3.2. *A finite field of size q is denoted by $\mathbb{F}_q$.*

A simple but important example is the finite field $\mathbb{F}_2 = \{ \{0, 1\}, \oplus, \wedge, 0, 1 \}$, where $\oplus$, and $\wedge$ are the *exclusive-or* and the *and* operations respectively. In fact, $\oplus$, and $\wedge$ are same as $+ \bmod 2$ and $\times \bmod 2$ respectively for $\{0, 1\}$.
The next two theorems state two important properties of the finite field $\mathbb{F}_q$. The first theorem states the *existance* of $\mathbb{F}_q$ while the second one states its *uniqueness*. For the proofs and more details on these, the reader is asked to refer [2].

Theorem 3.3. *For all finite fields $\mathbb{F}_q$, $q = p^s$ for some prime p and integer $s \geq 1$.*

The prime p is called the *characteristic* of $\mathbb{F}_q$.

Theorem 3.4. *There is only one finite field $\mathbb{F}_q$ for any prime power q.*

Theorem 3.4 actually implies that any two representation of $\mathbb{F}_q$ for the same q are isomorphic. It provides a justification for speaking of *the* finite field with q elements, or of *the* finite field of order q. Next we study polynomials which will in turn define a canonical representation of $\mathbb{F}_q$ for any prime power q.

3.2 Polynomials

In elementary algebra one regards a polynomial as an expression of the form $a_0 + a_1 X + \cdots + a_n X^n$. The a_0's are called coefficients and are usually real or complex numbers; X is viewed as a variable: that is, substituting an arbitrary number α for X, a well defined number $a_0 + a_1\alpha + \cdots + a_n\alpha^n$ is obtained. The arithmetic of polynomials is governed by familiar rules of addition, subtraction, multiplication, and division over the relevant field.

More formally, $\mathbf{P}(\mathbf{X})$ of *degree* d over $\mathbb{F}_q$ is of the form $\sum_{i=0}^{d} C_i X^i$ where $C_i \in \mathbb{F}_q$ and $C_d \neq 0$. If $\alpha \in \mathbb{F}_q$ is a root of $\mathbf{P}(\mathbf{X})$, then $\mathbf{P}(\alpha) = 0$. The set of all polynomials over $\mathbb{F}_q$ is denoted by $\mathbb{F}_q[X]$. The definitions of addition and multiplication of polynomials are the natural *extension* of the ones mentioned above (all the operations are over $\mathbb{F}_q$). For polynomial $P(X)$ of degree d, if there exists $Q_1(X), Q_2(X) \in \mathbb{F}_q[X]$ such that $0 < deg(Q_1) < d$, and $0 < deg(Q_2) < d$ and $P(X) = Q_1(x).Q_2(x)$, it implies $P(X)$ is *not* irreducible.
Now we present the important concept of *irreducible* polynomials.

Definition 3.5 (Irreducible polynomial)**.** *A polynomial $\mathbf{P}(\mathbf{X}) \in \mathbb{F}_q[X]$ is said to be irreducible over $\mathbb{F}_q$ (or irreducible in $\mathbb{F}_q$, or prime in $\mathbb{F}_q$) if $\mathbf{P}(\mathbf{X})$ has positive degree and $\mathbf{P}(\mathbf{X}) = \mathbf{B}(\mathbf{X})\mathbf{C}(\mathbf{X})$ with $\mathbf{B}(\mathbf{X})$, $\mathbf{C}(\mathbf{X}) \in \mathbb{F}_q[X]$ implies either $\mathbf{B}(\mathbf{X})$ or $\mathbf{C}(\mathbf{X})$ is a constant polynomial.*

In other words, $\mathbf{P}(\mathbf{X})$ over $\mathbb{F}_q$ is irreducible if there are no non-trivial factors of $\mathbf{P}(\mathbf{X})$ in $\mathbb{F}_q$. A polynomial in $\mathbb{F}_q$ of positive degree that is not irreducible over $\mathbb{F}_q$ is called *reducible over* $\mathbb{F}_q$. The reducibility or irreducibility of a given polynomial depends heavily on the field under consideration.

As for example, the polynomial $x^2 + 1$ is *not* irreducible over the field $\mathbb{F}_2$. Since we have
$x^2 + 1 = x^2 + (x + x) + 1 = (x + 1)(x + 1)$

Definition 3.6. *If $E[X]$ is an irreducible polynomial of degree e, $\mathbb{F}_p[X]/E[X]$ is a field where:*

- *The elements of $\mathbb{F}_p[X]/E[X]$ are polynomials of degree less than e*
- *The addition is the usual polynomial addition*
- *The multiplication is the usual polynomial multiplication mod $E[X]$*
- *The additive and multiplicative inverses are the same as the additive and multiplicative inverses in $\mathbb{F}_p$*

It can be proved that the additive and multiplicative inverses do exists.

Theorem 3.7. *$\mathbb{F}_p[X]/E[X]$ is a field.*

since, in modulo 2 addition, $x+x = 0$. But the polynomial $x^2 + x + 1$ is irreducible over $\mathbb{F}_2$. This can be easily verified by checking that this polynomial has no degree 1 polynomial as a factor. Note that all non-trivial factors of a degree 2 polynomial needs to be of degree 1. For the proof of the next theorem and more details on it, the reader is again asked to refer to [2].

Theorem 3.8. *For every finite field $\mathbb{F}_q$ and every positive integer e, there exists an irreducible polynomial $\mathbf{E}(\mathbf{x})$ over $\mathbb{F}_q$ of degree e.*

Theorems 3.4, 3.7 and, 3.8 imply a canonical representation of every finite field $\mathbb{F}_{p^e}$. When $e = 1$, $\mathbb{F}_p = \mathbb{Z}_p$ otherwise the field $\mathbb{F}_{p^e}$ is equivalent to $\mathbb{F}_p[X]/E[X]$, where $E[X]$ is an irreducible polynomial of degree e and p is a prime.

References

[1] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to Algorithms, Second Edition*. McGraw-Hill Book Company, Boston Burr Ridge , IL Dubuque , IA Madison , WI New York San Francisco St. Louis, Montral Toronto, 2001.

[2] Rudolf Lidl and Harald Niederreiter. *Finite Fields And Its Applications*. Addison-Wesley, Reading, Massachusetts, 1983.