Algebra and wireless communication

Felipe Voloch

In this talk we will explain what are low-correlation sequences and how they are used in cellular telephony and other wireless systems. We will talk about linear feedback shift register sequences and how to get low-correlation sequences from them.

More information available in my home page:

`http://www.ma.utexas.edu/users/voloch`

Low-correlation sequences are used in CDMA (Code Division Multiple Access) communications systems, which are used in applications such as cellular telephones. Typically one is faced with the problem of assigning to each user a signature (in the form of a binary string). Signatures for different users have to be not only different, but very different, so that errors in transmission can be recognized. Plus, it is also important to have the sequences very different even after shifts because of syncronization problems and because it is often hard to pinpoint the beginning of a transmission. The correlation is a measure of similarity between sequences so low-correlation is what we will strive for. We need as many sequences as there will be users and we want them as short as possible to optimize their transmission.

We consider infinite sequences of period n with symbols in $\{0, 1\}$. The idea is to form a large family $\mathbf{F}$ of such sequences which are pairwise cyclically distinct (i.e. no sequence is a shift of any other) and which have small correlation. Here the correlation $c_{st}(\Delta)$ between sequences $s = \{s(j)\}$ and $t = \{t(j)\}$ of period n for shift Δ is

$$c_{st}(\Delta) = \left| \sum_{j=0}^{n-1} (-1)^{s(j+\Delta)-t(j)} \right|$$

Note that $c_{st}(\Delta)$ is the absolute value of the inner product of the vectors

$$((-1)^{s(\Delta)}, (-1)^{s(\Delta+1)}, \ldots, (-1)^{s(\Delta+n-1)})$$

and

$$((-1)^{t(0)}, (-1)^{t(1)}, \ldots, (-1)^{t(n-1)}).$$

So to make the correlation small is to make these two vectors almost orthogonal. In practice, the signal will be transmitted by a waveform closely associated to such vectors and the inner product of the vectors is a bound on the convolution of the waveforms.

Another way of expressing the correlation is through the Hamming distance. The Hamming distance of two strings of same length is the number of positions in which they differ. Thus, the Hamming distance $d(s_\Delta, t)$ between $s_\Delta = \{s(j+\Delta)\}$ and $t = \{t(j)\}$ is the number of j's with $s(j+\Delta) \neq t(j), j = 0, \ldots, n-1$. Hence, $c_{st}(\Delta) = |(-1)d(s_\Delta, t) + (-1)^0(n - d(s_\Delta, t))| = |n - 2d(s_\Delta, t)|$.

One measures whether or not a family $\mathbf{F}$ of sequences is good by considering the maximum correlation parameter

$$C_{\max}(\mathbf{F}) = \max\{c_{st}(\Delta) \mid s, t \in \mathbf{F};\ s \neq t \,\mathrm{or}\, \Delta \neq 0\}.$$

In applications, the large family size allows for a large number of users and a small correlation parameter translates to little interference from one user to another.

Linear feedback shift register sequences

These sequences, abbreviated LFSR sequences, are defined as follows. We fix a string $c_0c_1\dots c_m$ of zeros and ones and given an initial segment $x_0x_1\dots x_m$ we generate the infinite string $x_0x_1\dots$, where the subsequent digits are determined by the formula

$$x_{n+1} = c_0x_{n-m} \oplus c_1x_{n-m+1} \oplus \cdots \oplus c_mx_n.$$

where $\oplus$ stands for addition modulo two (aka XOR). The set of all squences that we obtain from the fixed $c_0c_1\dots c_m$ by varying the initial segment $x_0x_1\dots x_m$, is a good family with low-correlation (as long as we exclude the all-zero initial input).

For example, if we fix the string 011 then we generate the following sequences (the initial input consists of the first three digits) of period 7. (This is $x_n = x_{n-2} + x_{n-3}$).

$$1001011, 0101110, 0010111, 1100101, 1011100, 0111001, 1110010$$

Another example with $x_n = x_{n-3} + x_{n-5}$.

$$1000010010110011111000110111010$$

A LFSR sequence of degree m has period at most $2^m - 1$. This can be seen as follows. Consider the vectors $v_k = (x_k, x_{k+1}, \ldots, x_{k+m-1})$. If some $v_k = (0, \ldots, 0)$ then the whole sequence consists of zeros. Otherwise, there are at most $2^m - 1$ possibilities for a length m binary vector. Thus among $v_0, v_1, \ldots, v_{2^m-1}$, there has to be a repetition, but when $v_k = v_r$ we must also have $v_{k+1} = v_{r+1}$ etc..., by the recurrence defining the sequence x. Hence the period is $r - k \leq 2^m - 1$. LFSR sequences of period $2^m - 1$ exist but it is a bit harder to prove this in general. We will give some examples shortly. Note that for such a sequence $v_0, v_1, \ldots, v_{2^m-2}$ are all distinct and non-zero, so they run through all possible non-zero binary vectors of length m. A consequence of this is that a maximal period LFSR sequence has $2^{m-1} - 1$ zeros and 2^{m-1} ones. Indeed, among all possible non-zero binary vectors of length m, $2^{m-1} - 1$ start with zero and 2^{m-1} start with one.

A shift of a LFSR sequence is another sequence generated by the same LFSR. Indeed, the shift of x by k starts by $x_k, x_{k+1}, \ldots, x_{k+m-1}, \ldots$ and is therefore the result of starting the LFSR sequence with this initial input.

For this reason, LFSR sequences are not by themselves the solution to our problem of making families sequences which are cyclically distinct. We remedy this problem by combining two or more different LFSR's.

A possibility, that works well in practice, is the following. Fix two maximal length LFSR's of the same degree m. Generate a sequence x with the first one and initial input $(1, 0, \ldots, 0)$. For each initial input $a \neq (0, 0, \ldots, 0)$, generate the sequence $y(a)$ with the second LFSR, so that all $y(a)$'s are shifts of one another. Consider the family formed by adding x to all possible $y(a)$'s. That is $\mathbf{F} = \{x + y(a) \mid a \in \{0,1\}^m, a \neq (0, 0, \ldots, 0)\}$. Usually the sequence x is obtained from that of $y = y(1, 0, \ldots, 0)$ by *decimation*. The decimation of y by d is, by definition, the sequence $x_n = y_{dn}, n = 0, 1, \ldots$. It can be shown that a decimation of y is a LFSR sequence of the same degree m as y and is maximal if d is relatively prime with $2^m - 1$. When m is odd we can take $d = 2^k + 1, (k, m) = 1, k \leq (m-1)/2$, and this family $\mathbf{F}$ is known as the family of Gold sequences. The correlations in this case are at most $2^{\frac{m+1}{2}}$. Other values of d also give correlations at most $2^{\frac{m+1}{2}}$, for instance $d = 2^{2k} - 2^k + 1$ (Kasami sequences), $2^{\frac{m-1}{2}} + 3$ (Welsh sequences) and $2^{\frac{m-1}{2}} + 2^r - 1$, where $r = (m-1)/4$ if $(m-1)/2$ is even and $r = (3m-1)/4$ if $(m-1)/2$ is odd (Niho sequences). Are there any others? The answer to this question is not known!

References:

Fan, Darnell “Sequence design for Communications Applications” Wiley 1996.

Golomb “Shift register sequences”, 2nd edition, Aegean Press 1982

Canteaut, Charpin, Dobbertin, “ Weight divisibility of cyclic codes, highly nonlinear functions on F_{2^m}, and crosscorrelation of maximum-length sequences”. SIAM J. Discrete Math. 13 (2000), no. 1, 105–138.