Decoding error-correcting codes with Gröbner bases

Stanislav Bulygin   Ruud Pellikaan

WIC, May 24, 2007

Outline

- Introduction
- Unknown syndromes and MDS bases
- Decoding up to half the minimum distance
- Complexity of the algorithm

The decoding of cyclic codes up to half the BCH distance is well-known by Peterson, Arimoto and Gorenstein-Zierler,

by means of the syndromes s_i of a received word and the error-locator polynomial with coefficients σ_i.

Suppose that the defining set of the cyclic code contains $2t$ consecutive elements.
The generalized Newton identities

$$s_1 + \sigma_1 s_{i-1} + \cdots + \sigma_t s_{i-t} = 0, \quad i = t+1, \ldots, 2t.$$

are t linear equations in the variables $\sigma_1, \ldots, \sigma_t$ with the known syndromes $s_1, \ldots, s_{2t}$ as coefficients.

Gaussian elimination solves this system of linear equations with complexity $\mathcal{O}(n^3)$.

This complexity was improved by the algorithm of Berlekamp-Massey and a variant of the Euclidean algorithm due to Sugiyama et al.

Both these algorithms are more efficient
and are basically equivalent,
but they decode up to the BCH error-correcting capacity,
which is often strictly smaller than the true capacity.

They do not correct up to the true error-correcting capacity.

Gröbner bases techniques were addressed to remedy this problem.

These methods can be divided into the following categories:

- Unknown syndromes by Berlekamp and Tzeng-Hartmann-Chien,
- Power sums by Cooper and Chen-Reed-Helleseth-Truong,
- Newton identities by Augot-Charpin-Sendrier.

Our method is a generalization of the first one.

The theory of Gröbner basis is about
solving systems of polynomial equations in several variables

It is as a common generalization of

- Linear Algebra,
  linear systems of equations in several variables,
- Euclidean Algorithm,
  polynomial equations of arbitrary degree in one variable.

The polynomial equations are linearized by treating the monomials as new variables. The number of variables grows exponentially in the degree of the polynomials.

The complexity of computing a Gröbner basis is
doubly exponential in general,
and exponential in our case of a finite set of solutions.

The complexity of our algorithm is exponential.
The complexity coefficient is measured under the assumption that the over-determined system of quadratic equations is semi-regular using the results of Bardet et al. applied to algorithm F5 of Faugère.

Let $\mathbf{b}_1, \ldots, \mathbf{b}_n$ be a basis of $\mathbb{F}_q^n$.
B is the $n \times n$ matrix with $\mathbf{b}_1, \ldots, \mathbf{b}_n$ as rows.

The (unknown) syndrome of a word $\mathbf{e}$ with respect to B
is the column vector $\mathbf{u}(\mathbf{e}) = \mathbf{u}(B, \mathbf{e}) = B\mathbf{e}^T$.
with entries $u_i(\mathbf{e}) = u_i(B, \mathbf{e}) = \mathbf{b}_i \cdot \mathbf{e}$ for $i = 1, \ldots, n$.

The matrix B is invertible.
So the syndrome $\mathbf{u}(B, \mathbf{e})$ determines the error vector $\mathbf{e}$ uniquely:

$$B^{-1}\mathbf{u}(B, \mathbf{e}) = B^{-1}B\mathbf{e}^T = \mathbf{e}^T.$$

The coordinatewise star product of $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$ by

$$\mathbf{x} * \mathbf{y} = (x_1 y_1, \ldots, x_n y_n).$$

Then $\mathbf{b}_i * \mathbf{b}_j$ is a linear combination of the basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$.
There are structure constants $\mu_{ijl} \in \mathbb{F}_q$ such that

$$\mathbf{b}_i * \mathbf{b}_j = \sum_{l=1}^{n} \mu_{ijl} \mathbf{b}_l.$$

$\mathcal{U}(\mathrm{e})$ is the $n \times n$ matrix of (unknown) syndromes of a word e with entries

$$u_{ij}(\mathbf{e}) = (\mathbf{b}_i * \mathbf{b}_j) \cdot \mathbf{e}.$$

The entries of $\mathcal{U}(\mathrm{e})$ and $\mathbf{u}(\mathrm{e})$ are related by

$$u_{ij}(\mathbf{e}) = \sum_{l=1}^{n} \mu_{ijl} u_l(\mathbf{e}).$$

Lemma

The rank of $\mathcal{U}(\mathrm{e})$ is equal to the weight of e.

Let B_r be the $r \times n$ sub matrix of B with $\mathbf{b}_1, \ldots, \mathbf{b}_r$ as rows.

$\mathbf{b}_1, \ldots, \mathbf{b}_n$ is called an MDS basis and B an MDS matrix
if all the $t \times t$ sub matrices of B_t have rank t for all $t = 1, \ldots, n$.

Let C_t be the code with B_t as parity check matrix.

Proposition
B is an MDS matrix if and only if C_t is an [n,n-t,t+1] code for all t.

MDS bases are known to exist if $n \leq q$.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be n mutually distinct elements in $\mathbb{F}_q$.
Define

$$\mathbf{b}_i = (x_1^{i-1}, \ldots, x_n^{i-1}).$$

Then $\mathbf{b}_1, \ldots, \mathbf{b}_n$ with matrix $B(\mathbf{x})$ are MDS and are called a Vandermonde basis and matrix, resp.

If $\alpha \in \mathbb{F}_q^*$ is an element of order n and $x_j = \alpha^{j-1}$, then we get a Reed-Solomon (RS) basis and matrix with $b_i * b_j = b_{i+j-1}$ and $u_{ij}(\mathbf{e}) = u_{i+j-1}(\mathbf{e})$.

Proposition

Suppose that B is an MDS matrix.

Let $\mathcal{U}_{u,v}(\mathbf{e})$ be the $u \times v$ sub matrix of $\mathcal{U}(\mathbf{e})$ consisting of the first u rows and v columns.

Then

$$\text{rank}(\mathcal{U}_{nv}(\mathbf{e})) = \begin{cases} v & \text{if } \ v \leq \text{wt}(\mathbf{e}), \\ \text{wt}(\mathbf{e}) & \text{if } \ v > \text{wt}(\mathbf{e}). \end{cases}$$

Let C be an $\mathbb{F}_q$-linear code of
length n, dimension k, minimum distance d,
and redundancy $r = n - k$.

Choose a parity check matrix H of C.
Let $\mathbf{h}_1, \ldots, \mathbf{h}_r$ be the rows of H.
There are constants $a_{ij} \in \mathbb{F}_q$ such that

$$\mathbf{h}_i = \sum_{j=1}^{n} a_{ij}\mathbf{b}_j .$$

Let A be the $r \times n$ matrix with entries a_{ij}. Then $H = AB$.

Let $\mathbf{y} = \mathbf{c} + \mathbf{e}$ be a received word
with $\mathbf{c} \in C$ a code word and $\mathbf{e}$ an error vector.

The syndromes of $\mathbf{y}$ and $\mathbf{e}$ with respect to H are equal and known

$$s_i(\mathbf{y}) := \mathbf{h}_i \cdot \mathbf{y} = \mathbf{h}_i \cdot \mathbf{e} = s_i(\mathbf{e})$$

Expressed in the unknown syndromes of $\mathbf{e}$ with respect to B:

$$s_i(\mathbf{y}) = \sum_{j=1}^{n} a_{ij} u_j(\mathbf{e}).$$

The system $E(\mathbf{y})$ of equations in the variables $U_1, \ldots, U_n$ is given by:

$$\sum_{l=1}^{n} a_{jl} U_l \;=\; s_j(\mathbf{y}) \;\text{ for }\; j = 1, \ldots, r.$$

It consists of $n-k$ independent linear equations in n variables

The system $E(t)$ in the variables $U_1, \ldots, U_n, V_1, \ldots, V_t$ is given by:

$$\sum_{j=1}^{t} \sum_{l=1}^{n} \mu_{ijl} U_l V_j \;=\; \sum_{l=1}^{n} \mu_{it+1l} U_l \;\text{ for }\; i = 1, \ldots, n.$$

It consists of n quadratic equations in $n+t$ variables.

The system of equations $E(t, \mathbf{y})$ is the union of $E(t)$ and $E(\mathbf{y})$.

It consists of $n-k$ linear equations in n variables and n quadratic equations in $n+t$ variables.

The linear equations are independent and used to eliminate $n-k$ variables.

Thus we get a system of n quadratic equations in $k+t$ variables.

Theorem

Let B be an MDS matrix with structure constants μ_{ijl}.
Let H be a parity check matrix of the code C such that $H = AB$.
Let $\mathbf{y} = \mathbf{c} + \mathbf{e}$ be a received word
with $\mathbf{c}$ in C the codeword sent and $\mathbf{e}$ the error vector.

Suppose that the weight of $\mathbf{e}$ is not zero and at most $(d-1)/2$.

Let t be the smallest positive integer such that $E(t, \mathbf{y})$
has a solution $(\mathbf{u}, \mathbf{v})$ over some extension $\mathbb{F}_{q^m}$ of $\mathbb{F}_q$.

Then $\mathrm{wt}(\mathbf{e}) = t$ and the solution is unique satisfying $\mathbf{u} = \mathbf{u}(\mathbf{e})$.

Experiments were done on an
AMD Athlon 64 Processor 2800+ (1.8MHz), 512MB RAM under Linux.

The computations of Gröbner bases were realized in SINGULAR 3-0-1.

Code	err. cap.	mindist.	GB dec.	no. of rec.	average
[25,11,4]	1	2.99	1.10	300	0.0037
[25,11,5]	2	21.58	2.89	300	0.0096
[25,8,5]	2	0.99	1.84	300	0.0061
[25,8,6]	2	3.38	1.79	300	0.0060
[25,8,7]	3	12.26	6.94	300	0.0231
[31,15]	2	-	10.76	300	0.0359
[31,15]	3	-	11.19	10	1.119

no. of err.	[120,40]	[120,30]	[120,20]	[120,10]	[150,10]
2	1	1	1	1	1
3	13	1	1	1	1
4	313	9	1	1	1
5	-	62	1	1	1
6	-	200	5	1	3
7	-	933	14	1	4
8	-	-	32	1	4
9	-	-	74	1	4
10	-	-	183	2	6
11	-	-	633	3	6
12	-	-	-	4	6
13	-	-	-	5	8
14	-	-	-	6	8
15	-	-	-	14	10
16	-	-	-	20	11
17	-	-	-	29	16
18	-	-	-	71	16
19	-	-	-	139	34
20	-	-	-	327	53
21	-	-	-	483	84
22	-	-	-	-	133
23	-	-	-	-	241
24	-	-	-	-	513

Given a decoding algorithm for a code C of rate R over $\mathbb{F}_q$ of complexity $Compl(C)$,

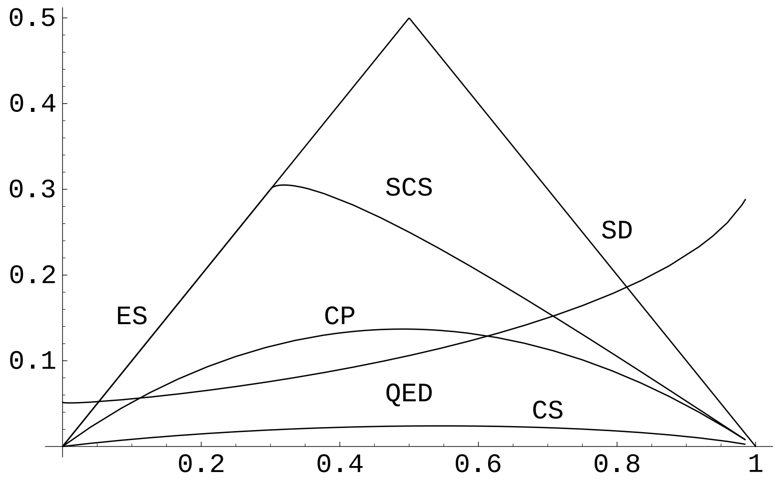
the complexity coefficient $CC(R)$ is defined as

$$CC(R) = \frac{1}{n} \log_q(Compl(C)).$$

In the binary case the complexity of our method is worse than exhaustive search.

But with increasing alphabet our method is better.
The following figure compares the complexity coefficients for $q = 2^{10}$ of

- exhaustive search (ES),
- syndrome decoding (SD),
- systematic coset search (SCS),
- covering polynomials (CP),
- covering sets (CD) and
- our method using quadratic equations (QED).


0.5
0.4
0.3
0.2
0.1
0.2
0.4
0.6
0.8
1
SCS
SD
ES
CP
QED
CS

Questions?