MDS codes

Let us start with the definition of *systematic* codes. We will use this notion for linear codes only.

Definition. We say that a (linear) code C of dimensipon k is *systematic* on coordinates $1, 2, \dots k$, if by projecting C onto the first k coordinates we get Q^k. If we specify another set of k coordinates, let us say $\{i_1, \dots, i_k\}$, then the definition is the same, just the projection is onto these coordinates.

In other words, systematicity on coordinates $\{i_1, \dots, i_k\}$ means that there is a generator matrix G' which has the identity matrix on these coordinates. Intuitively we can say that for any generator matrix G of C we can „transform" the identity matrix in positions $\{i_1, \dots, i_k\}$. Transforming means taking linear combinations (and interchanging) rows.

There is an easy way of obtaining a parity check matrix from a generator matrix and the other way round. It says that whenever we have an identity matrix in the generator matrix, then the parity check matrix can have the identity matrix in the remaining positions. In terms of systematicity it gives the following.

Proposition. *If C is systematic on coordinates $\{i_1, \dots, i_k\}$, then $C^\perp$ is systematic on coordinates $\{q, \dots n\} \setminus \{i_1, \dots, i_k\}$.*

The Singleton bound gives an upper bound on the size of a code with minimal distance d

Proposition (Singleton bound)**.** *Let $C \subseteq Q^n$ be a code with minimum distance d. Then $|C| \leq |Q|^{n-d+1}$.*

Let us delete $d-1$ coordinates. Then we gete a mapping $C \to Q^{n-(d-1)}$. This mapping is injective, since the minimum distance is d. This shows $|C| \leq |Q^{n-d+1}| = |Q|^{n-d+1}$.

For linear codes, the Singleton bound is even simpler.

Proposition. *Let C be an $[n, k, d]_q$ code. Then $d \leq n - k + 1$.*

This clearly follows from the non-linear Singleton-bound, since for a k-dimensional linear code $|C| = q^k$, so we get $k \leq n - d + 1$. Rearranging for d gives the result above.

This result follows immediately also from the theorem relating the columns of the parity check matrix H and the minimum distance d of the code. We

have to determine the maximum number s so that every s columns of H are independent. In terms of this s, we have $d = s + 1$. Since H has rank $n - k$, we get $s \leq n - k$, so $d \leq n - k + 1$.

Definition. A code C attaining the Singleton bound is called an *MDS code* (maximum distance separable.

So, we can say that a k-dimensional code is MDS whenever it is systematic on every k coordinates. Having actually the identity matrix at certain coordinates means that we have here the coordinates of the actual message and the remaining coordinates are used for error-correction, so the coordinates can be separated into coordinates of the message and parity check coordinates.

Theorem. *The dual of a linear MDS code is also MDS.*

This follows easily from our remarks on systematicity. A k-dimensional code is MDS if and only if it is systematic on every k coordinates. Then we know that the dual code is systemic on every $n - k$ coordinates, so it also has to be MDS.

Let us prove this directly, because the theorem is very important. We want to prove that $d = n-k+1$ implies that the dual code is also MDS. So, we have to show that every k columns of the original generator matrix G are independent (because that is the parity check matrix of the dual code). Assume to the contrary that, for example, the first k columns of G are not independent. Then, if we restrict our attention to the first k coordinates, the rows (of length k) are also dependent. Therefore a non-trivial linear combinations of them is the zero vector. For the original generator matrix this gives a non-trivial linear combination of the rows, which is 0 in the first k coordinates. This linear combination gives us a codeword of weight at most $n - k$, contradicting $d = n - k + 1$.

Let us see an example of an MDS code. This is a special case of generalized Reed-Solomon codes, which we will study later.

Proposition. *Choose n distinct non-zero elements $\alpha_1, \ldots \alpha_n \in \mathrm{GF}(q)$. Define the parity check matrix $H = (h_{ij})$ with the rule $h_{ij} = \alpha_j^i$, where $i = 1, \ldots n - k, j = 1, \ldots n$. Then the corresponding code is MDS.*

Indeed, if we take any $n-k$ columns, then the corresponding $(n-k)\times(n-k)$ submatrix will be a Vandermonde matrix, hence it is not singular. So, every $n - k$ columns are independent and the code is MDS.

This example also shows a shortcoming of MDS codes, namely that the length can not be (much) larger than q. In some special cases this can be improved by one (or even two) but there are no MDS codes known which have large length over a small alphabet.

Theorem (Ball)**.** *Let q be an odd prime and C be a MDS code. Then $n \leq q+1$.*