NDMI011: Combinatorics and Graph Theory 1

Lecture #14
Linear codes

Irena Penev

1 Some Linear Algebra preliminaries

In what follows, for a field $\mathbb{F}$ and a positive integer n, we denote by $\mathbb{F}^n$ the set of all row vectors of length n whose entries are all in $\mathbb{F}$. For vectors $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ in $\mathbb{F}^n$, we define $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i$, where the summation and multiplication denote the operations from the field $\mathbb{F}$; note that $\langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{F}$. If $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, then $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal*.

Instead of multiplying matrices by column vectors on the right ($A\mathbf{x}$), we will multiply matrices by row vectors on the left ($\mathbf{x}A$). If A is an $n \times m$ matrix with entries in $\mathbb{F}$, and $\mathbf{x} \in \mathbb{F}^n$,[1] then we can think of $\mathbf{x}$ as a $1 \times n$ matrix, and we can compute $\mathbf{x}A$ according to the usual rules of matrix multiplication.[2] Note that if $\mathbf{x} = (x_1, \dots, x_n)$ and $A = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix}$ (i.e. $\mathbf{r}_1, \dots, \mathbf{r}_n$ are the rows of A, from top to bottom), then $\mathbf{x}A = \sum\limits_{i=1}^{n} x_i \mathbf{r}_i$. Furthermore, if $\mathbf{e}_i$ is the i-th standard basis vector of $\mathbb{F}^n$, i.e. the row vector whose i-th entry is 1, and all of whose other entries are 0, then $\mathbf{e}_i A$ is equal to the i-th row of A.

With these adjustments, all familiar theorems of Linear Algebra still hold, but with rows and columns reversed. For instance, Gaussian elimination is performed on columns, not rows.[3]

[1] So, A has n rows and m columns, and $\mathbf{x}$ is a row vector of length n.

[2] Indeed, we multiply a $1 \times n$ matrix by an $n \times m$ matrix, and we obtain a $1 \times m$ matrix, i.e. a row vector of length m.

[3] Alternatively, given a matrix A, we can perform Gaussian elimination as follows: we first form the transpose A^T, then we perform the familial Gaussian elimination on rows to obtain a matrix B, and then we take the transpose of B. The result is the same as if we performed Gaussian elimination on the columns of A directly.

For a field $\mathbb{F}$ and a subspace C of $\mathbb{F}^n$, we define $C^\perp = \{\mathbf{y} \in \mathbb{F}^n \mid \langle \mathbf{x}, \mathbf{y} \rangle = 0 \text{ for all } \mathbf{x} \in C\}$. It is easy to check that $C^\perp$ is a subspace of $\mathbb{F}^n$.[4]

Theorem 1.1. *Let $\mathbb{F}$ be a field, and let C be a subspace of $\mathbb{F}^n$. Then $\dim C + \dim C^\perp = n$.*

Proof. Set $k = \dim C$; we must show that $\dim C^\perp = n - k$. If $k = 0$, then $C = \{\mathbf{0}\}$ and $C^\perp = \mathbb{F}^n$, and it follows that $\dim C^\perp = n = n - k$. From now on, we assume that $k \geq 1$. Let $\{\mathbf{c}_1, \dots, \mathbf{c}_k\}$ be some basis for C, and let $G = \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_k \end{bmatrix}$. Then $C^\perp = \{\mathbf{y} \in \mathbb{F}^n \mid \mathbf{y}G^T = \mathbf{0}\} = \text{Ker}(G^T)$.[5] By the Rank-nullity theorem, we have that $\text{rank}(G^T) + \dim \text{Ker}(G^T) = n$. But $\text{rank}(G^T) = \text{rank}(G) = k$ (because G has k rows, and they are linearly independent), and as we saw $C^\perp = \text{Ker}(G^T)$. It follows that $k + \dim C^\perp = n$, i.e. $\dim C^\perp = n - k$. □

Proposition 1.2. *Let $\mathbb{F}$ be a field, and let C be a subspace of $\mathbb{F}^n$. Then $(C^\perp)^\perp = C$.*

Proof. Obviously, $C \subseteq (C^\perp)^\perp$;[6] since C and $(C^\perp)^\perp$ are both subspaces of $\mathbb{F}^n$, it follows that C is a subspace of $(C^\perp)^\perp$. On the other hand, by Theorem 1.1, we have that

$$\dim(C^\perp)^\perp \quad = \quad n - \dim C^\perp \quad = \quad n - (n - \dim C) \quad = \quad \dim C,$$

and we deduce that $C = (C^\perp)^\perp$. □

2 Linear codes

A *linear code* is a subspace C of a vector space $\mathbb{F}_q^n$, where $\mathbb{F}_q$ is a finite field of size q (here, q is a prime power).[7] Note that every linear code contains the zero vector.

Notationally, if the linear code C is an $(n, k, d)_q$-code, then we write that C is an $[n, k, d]_q$-code (here, square brackets indicate that C is a linear code). Clearly, an $[n, k, d]_q$-code is a subspace of $\mathbb{F}_q^n$.[8] Furthermore, as our next proposition shows, the (vector space) dimension of an $[n, k, d]_q$-code is k.

[4] Check this!

[5] $\text{Ker}(G^T) = \{\mathbf{y} \in \mathbb{F}^n \mid \mathbf{y}G^T = \mathbf{0}\}$ is simply the definition of $\text{Ker}(G^T)$.

[6] Indeed, every vector in C is orthogonal to every vector in $C^\perp$. On the other hand, $(C^\perp)^\perp$ is the set of all vectors in $\mathbb{F}$ that are orthogonal to every vector in $C^\perp$. It follows that $C \subseteq (C^\perp)^\perp$.

[7] So, elements of $\mathbb{F}_q$ are row vectors of length n, all of whose entries are in the field $\mathbb{F}_q$.

[8] This is because the alphabet over which C is a code must be of size q, and since C is a linear code, it is a subspace of $\mathbb{F}^n$, where $\mathbb{F}$ is some finite field. So, $\mathbb{F}$ is a field of size q, and so it is equal (technically, isomorphic) to $\mathbb{F}_q$ (because all fields of the same size are isomorphic.

Proposition 2.1. *Let C be an $[n, k, d]_q$-code. Then $\dim C = k$, i.e. the dimension of C as a vector space is k.*

Proof. Since C is an $[n, k, d]_q$-code, we know that C is a subspace of $\mathbb{F}_q^n$; set $\ell = \dim C$. We must show that $\ell = k$. Let $\{\mathbf{c}_1, \dots, \mathbf{c}_\ell\}$ be a basis for C. Then C is the set of all vectors of the form $\sum_{i=1}^{\ell} \alpha_i \mathbf{c}_i$, where $\alpha_1, \dots, \alpha_\ell \in \mathbb{F}_q$. There are q choices for each α_i,[9] and so there are q^ℓ choices for the ℓ-tuple $(\alpha_1, \dots, \alpha_\ell)$. On the other hand, since $\{\mathbf{c}_1, \dots, \mathbf{c}_\ell\}$ is linearly independent (because it is a basis), we know that $\sum_{i=1}^{\ell} \alpha_i \mathbf{c}_i = \sum_{i=1}^{\ell} \beta_i \mathbf{c}_i$ (where $\alpha_1, \dots, \alpha_\ell, \beta_1, \dots, \beta_\ell \in \mathbb{F}_q$) if and only if $(\alpha_1, \dots, \alpha_\ell) = (\beta_1, \dots, \beta_\ell)$. It follows that $|C| = q^\ell$, and consequently, $\ell = \log_q |C|$. Since $k = \log_q |C|$ (by definition), it follows that $\ell = k$, which is what we needed to show. □

Now, suppose that $C \subseteq \mathbb{F}_q^n$ be an $[n, k, d]_q$-code, with $0 < k < n$. By Proposition 2.1, we have that $\dim C = k$, and so C is a non-null proper subspace of $\mathbb{F}_q^n$. Let G be any matrix whose rows form a basis for C (in particular, $G \in \mathbb{F}_q^{k \times n}$); then G is called the *generator matrix* of the linear code C. Note that this implies that $C^\perp = \{\mathbf{y} \in \mathbb{F}_q^n \mid \mathbf{y}G^T = \mathbf{0}\}$. Next, suppose H is any matrix such that the rows of H^T form a basis for $C^\perp$ (so, H^T is a generator matrix for $C^\perp$). The matrix H is called a *parity check matrix* for C, and by Proposition 1.2, it satisfies $C = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}H = \mathbf{0}\}$,[10] i.e. $C = \mathrm{Ker}(H)$. Note that the parity check matrix H can be used to check whether a vector $\mathbf{x} \in \mathbb{F}_q^n$ is a codeword of C. Indeed, if $\mathbf{x}H = \mathbf{0}$, then $\mathbf{x} \in C$, and otherwise, $\mathbf{x} \notin C$. Note that, given a generator matrix for C, one can easily compute a parity check matrix for C, and vice versa.

Given a vector $\mathbf{x} \in \mathbb{F}_q^n$, the *Hamming weight* of $\mathbf{x}$, denoted by $\mathrm{wt}(\mathbf{x})$, is the number of non-zero coordinates in $\mathbf{x}$.

Proposition 2.2. *Let $C \subsetneqq \mathbb{F}_q^n$ be an $[n, k, d]_q$-code, with $0 < k < n$, and let H be a parity check matrix for C. Then $d = \min\{wt(\mathbf{x}) \mid \mathbf{x} \in C, \mathbf{x} \neq \mathbf{0}\}$.*

Proof. Fix $\mathbf{x} \in C \setminus \{\mathbf{0}\}$ with minimum Hamming weight. We must show that $d = \mathrm{wt}(\mathbf{x})$.

First, since C is a linear code, we know that $\mathbf{0} \in C$, and so (since $\mathbf{x}$ and $\mathbf{0}$ are distinct codewords in C) we have that $d(\mathbf{x}, \mathbf{0}) \geq d$. But obviously, $d(\mathbf{x}, \mathbf{0}) = \mathrm{wt}(\mathbf{x})$, and it follows that $\mathrm{wt}(\mathbf{x}) \geq d$.

It remains to show that $\mathrm{wt}(\mathbf{x}) \leq d$. Fix distinct $\mathbf{y}, \mathbf{z} \in C$ such that $d(\mathbf{y}, \mathbf{z}) = d$.[11] Since C is a vector space, we know that $\mathbf{y} - \mathbf{z} \in C$, and so by the choice of $\mathbf{x}$, we have that $\mathrm{wt}(\mathbf{x}) \leq \mathrm{wt}(\mathbf{y} - \mathbf{z})$. But now

$$d \;=\; d(\mathbf{y}, \mathbf{z}) \;=\; \mathrm{wt}(\mathbf{y} - \mathbf{z}) \;\geq\; \mathrm{wt}(\mathbf{x}),$$

[9] This is because $|\mathbb{F}_q| = q$.

[10] Let us check this. Clearly, $(C^\perp)^\perp = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}(H^T)^T = \mathbf{0}\} = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}H = \mathbf{0}\}$. Since $(C^\perp)^\perp = C$ (by Proposition 1.2), it follows that $C = \{\mathbf{x} \in \mathbb{F}_q^n \mid \mathbf{x}H = \mathbf{0}\}$.

[11] The minimum distance between codewords in C is d. So, there exists distinct vectors in C (say, $\mathbf{y}$ and $\mathbf{z}$) whose distance is precisely d.

which is what we needed to show. □

3 Hamming codes

Fix an integer $\ell \geq 2$, and set $n = 2^\ell - 1$, $k = 2^\ell - \ell - 1$, and $d = 3$. Our goal in this section is to construct an $[n, k, d]_2$-code, called a *Hamming code*.[12] We do this by constructing its parity check matrix H; then the code in question will simply be the subspace $C = \{\mathbf{x} \in \mathbb{F}_2^n \mid \mathbf{x}H = \mathbf{0}\}$.

Note that the binary representation of the integer $n = 2^\ell - 1$ is $\underbrace{1\dots1}_{\ell}$. More generally, the binary representation of any integer in $\{1, \dots, n\}$ has at most ℓ digits. Now, for all $i \in \{1, \dots, n\}$, let $\mathbf{h}_i \in \mathbb{F}_2^\ell$ be the vector giving the binary representation of i, with zeros added to the front if necessary (so that the length of the representation is ℓ).[13] Let

$$H = \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_n \end{bmatrix}.$$

Note that $H \in \mathbb{F}_2^{n \times \ell}$. We now define the code C by setting

$$C = \{\mathbf{x} \in \mathbb{F}_2^n \mid \mathbf{x}H = \mathbf{0}\}.$$

Let us show that C is an $[n, k, d]_2$-code. Obviously, C is a subspace of $\mathbb{F}_2^n$.[14] Let us show that $\dim C = k$.[15] As usual, for all $i \in \{1, \dots, \ell\}$, let $\mathbf{e}_i^\ell$ be the vector in $\mathbb{F}_2^\ell$ whose i-th coordinate is 1, and all of whose other coordinates are 0. Then each of $\mathbf{e}_1^\ell, \dots, \mathbf{e}_\ell^\ell$ is a row of H, and furthermore, the set $\{\mathbf{e}_1^\ell, \dots, \mathbf{e}_\ell^\ell\}$ is a basis for $\mathbb{F}_2^\ell$; so, $\text{rank}(H) = \ell$. The Rank-nullity theorem guarantees that $\text{rank}(H) + \dim \text{Ker}(H) = n$, and we deduce that $\dim \text{Ker}(H) = n - \ell = k$. But $C = \text{Ker}(H)$, and so $\dim C = k$.

It remains to show that the minimum distance of words in C is $d = 3$. We will use Proposition 2.2. As usual, for all $i \in \{1, \dots, n\}$, let $\mathbf{e}_i^n$ be the vector in $\mathbb{F}_2^n$ whose i-th coordinate is 1, and all of whose other coordinates are 0. Note that the vectors of $\mathbb{F}_2^n$ of Hamming weight 1 are precisely the vectors $\mathbf{e}_1^n, \dots, \mathbf{e}_n^n$. But note that, for all $i \in \{1, \dots, n\}$, we have that $\mathbf{e}_i H = \mathbf{h}_i \neq \mathbf{0}$, and so $\mathbf{e}_i \notin C$. Next, vectors of $\mathbb{F}_2^n$ of Hamming weight 2 are precisely the

---

[12] It is also possible to construct "q-ary Hamming codes," which are over the (more general) field $\mathbb{F}_q$. For the sake of simplicity, though, we consider only binary Hamming codes, i.e. those over the field $\mathbb{F}_2$.

[13] For example, if $\ell = 2$, then $n = 3$, and we have that $\mathbf{h}_1 = (0, 1)$, $\mathbf{h}_2 = (1, 0)$, and $\mathbf{h}_3 = (1, 1)$.

[14] So, C is a linear code, and furthermore, the first coordinate (i.e. the n-part) and the subscript (i.e. 2) of $[n, k, d]_2$ are correct.

[15] In view of Proposition 2.1, this will guarantee that second coordinate (i.e. the k-part) of $[n, k, d]_2$ is correct.

vectors of the form $\mathbf{e}_i^n + \mathbf{e}_j^n$, with $i \neq j$. Now, for distinct $i, j \in \{1, \ldots, n\}$, we have that $(\mathbf{e}_i^n + \mathbf{e}_j^n)H = \mathbf{h}_i + \mathbf{h}_j$; since $\mathbf{h}_i \neq \mathbf{h}_j$ (and our field is $\mathbb{F}_2$), we have that $\mathbf{h}_i + \mathbf{h}_j \neq \mathbf{0}$, and it follows that $\mathbf{e}_i^n + \mathbf{e}_j^n \notin C$. We have now shown that C does not contain any non-zero vectors of Hamming weight at most two. On the other hand, C does contain a vector of Hamming weight at most three, e.g. the vector $\mathbf{e}_1^n + \mathbf{e}_2^n + \mathbf{e}_3^n$.[16] So, $\min\{\text{wt}(\mathbf{x}) \mid \mathbf{x} \in C, \mathbf{x} \neq \mathbf{0}\} = 3 = d$, and so by Proposition 2.2, we see that the minimum distance in C is d.

We have now shown that C is indeed an $[n, k, d]_2$-code, that is, C is a $[2^\ell - 1, 2^\ell - \ell - 1, 3]_2$-code. The code that we just constructed is called a *Hamming code*.

Finally, let us explain how error checking works for the Hamming code C that we just constructed. Suppose $\mathbf{w} \in \mathbb{F}_2^n$. Then by construction, $\mathbf{w} \in C$ if and only if $\mathbf{w}H = \mathbf{0}$. Suppose now that $\mathbf{w}$ differs in exactly one coordinate from some codeword in C, that is, that $\mathbf{w}$ can be obtained from a codeword in C by introducing one error (i.e. by changing exactly one 1 into 0, or vice versa, in some codeword of C). This means that there exist some $\mathbf{x} \in C$ and $i \in \{1, \ldots, n\}$ such that $\mathbf{w} = \mathbf{x} + \mathbf{e}_i^n$, and so

$$\begin{aligned} \mathbf{w}H &= (\mathbf{x} + \mathbf{e}_i^n)H \\ &= \underbrace{\mathbf{x}H}_{=\mathbf{0}} + \underbrace{\mathbf{e}_i^n H}_{=\mathbf{h}_i} \\ &= \mathbf{h}_i. \end{aligned}$$

But $\mathbf{h}_i$ is simply the integer i written in binary code! This means that if $\mathbf{w}$ was obtained from a codeword in C by introducing exactly one error, then the coordinate of that error is the integer whose binary representation is given by the vector $\mathbf{w}H$; we can correct the error by altering the entry (from 1 to 0, or vice versa) in that one coordinate of $\mathbf{w}$.

[16] Indeed,

$$\begin{aligned} (\mathbf{e}_1^n + \mathbf{e}_2^n + \mathbf{e}_3^n)H &= \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 \\ &= (\underbrace{0, \ldots, 0}_{n-2}, 0, 1) + (\underbrace{0, \ldots, 0}_{n-2}, 0, 1) + (\underbrace{0, \ldots, 0}_{n-2}, 1, 1) \\ &= \mathbf{0}, \end{aligned}$$

and so $\mathbf{e}_1^n + \mathbf{e}_2^n + \mathbf{e}_3^n \in C$.