MTAT.05.082: Introduction to Coding Theory **University of Tartu**

Final exam

Instructors: Vitaly Skachek, Yauhen Yakimenka December 14th, 2018

Student name: _____

Student ID: _____

1. This exam contains 9 pages. Check that no pages are missing.
2. It is possible to collect up to 110 points. Try to collect as many points as possible.
3. Justify and prove all your answers.
4. All facts and results that were proved or stated in the class can be used in your solution without a proof. Such results need to be rigorously formulated.
5. Any printed and written material is allowed in the class. No electronic devices are allowed.
6. Exam duration is 3 hours.
7. Good luck!

Question 1	
Question 2	
Question 3	
Question 4	
Total	

Question 1 (20 points).

Consider the following two 3×4 matrices:

$$\mathcal{G}_1 = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix} \qquad \text{and} \qquad \mathcal{G}_2 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 2 & 2 & 0 & 2 \end{pmatrix},$$

that generate codes $\mathcal{C}_1$ and $\mathcal{C}_2$ over $\mathbb{F}_3 = \{0, 1, 2\}$, respectively.

(a) Do these matrices generate the same code or two different codes?

(b) What is the length n, dimension k and minimum distance d of the code $\mathcal{C}_1$?

Prove your answers.

Solution. Let us transform $\mathcal{G}_1$ and $\mathcal{G}_2$ to corresponding systematic generator matrices by applying equivalent transformations on the rows.[1]

$$\mathcal{G}_1 = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix} \overset{(1)\overset{+}{=}(2)}{\sim} \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix} \overset{(1)\overset{+}{=}(3)}{\sim} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix} \overset{(2)\overset{+}{=}(3)}{\sim} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix} \triangleq \mathcal{G}.$$

$$\mathcal{G}_2 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 2 & 2 & 0 & 2 \end{pmatrix} \overset{(1)\overset{-}{=}(2)}{\sim} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 2 & 2 & 0 & 2 \end{pmatrix} \overset{(3)\overset{+}{=}(1)}{\sim} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 0 & 1 \end{pmatrix}$$
$$\overset{(3)\overset{+}{=}(2)}{\sim} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \overset{(2)\overset{-}{=}(3)}{\sim} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix} = \mathcal{G}.$$

Here, the notation $(i) \overset{+}{=} (j)$ denotes addition of the j'th row to the i'th row (analogously for subtraction).

As we can see, both $\mathcal{G}_1$ and $\mathcal{G}_2$ can be transformed into the same systematic form $\mathcal{G}$. Since the systematic matrix is in row-echelon form, it has a full rank. Therefore, the codes generated by $\mathcal{G}_1$ and $\mathcal{G}_2$ are exactly the same.

The length of the code generated by $\mathcal{G}$ is trivially $n = 4$ (number of columns). The rank of $\mathcal{G}$ is 3, hence dimension $k = 3$.

To find the minimum distance, let us first find a parity-check matrix of the code generated by $\mathcal{G}$. It is of size $(n - k) \times n = 1 \times 4$. Let us denote its columns as $\mathcal{H} = (h_1, h_2, h_3, h_4)$. Then it should satisfy the following:

$$\mathcal{H}\mathcal{G}^\intercal = (h_1 + 2h_4, h_2 + 2h_4, h_3 + 2h_4) = (0, 0, 0, 0).$$

[1] From linear algebra point of view, each such transformation is equivalent to changing a basis of the linear subspace.

From this, we obtain that $h_1 = h_2 = h_3 = h_4$. For instance, we can take $\mathcal{H} = (1, 1, 1, 1)$. All 1×1 submatrices of $\mathcal{H}$ are full-rank, therefore the minimum distance $d = 2$. ■

Question 2 (30 points).

Definition: let n be an even integer. A binary (not necessarily linear) code $\mathcal{C}$ is said to be *balanced* if any codeword has equal number of zeros and ones.

Example: for $n = 4$, a code that is formed by *any nonempty subset* of the following set of vectors is balanced:

$$\{(0011), (0101), (1001), (0110), (1010), (1100)\} \ .$$

In what follows, assume that $\mathcal{C}$ is a balanced code.

(a) What is the maximal possible size of $\mathcal{C}$?

(b) Show that the Hamming distance between any two codewords in $\mathcal{C}$ is even.

(c) Define a "balanced sphere" of radius $2r$, $0 \leq r \leq n/2$, around the codeword $\bar{\boldsymbol{x}} \in \mathcal{C}$ as

$$\mathbb{S}_{2r}(\bar{\boldsymbol{x}}) = \{\bar{\boldsymbol{y}} \ : \ \bar{\boldsymbol{y}} \text{ has equal number of zeros and ones, and } \mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \leq 2r\} \ ,$$

where $\mathsf{d}_H(\cdot, \cdot)$ denotes the Hamming distance. What is the size of $\mathbb{S}_{2r}(\bar{\boldsymbol{x}})$?

(d) Let d be an integer, $1 \leq d \leq n/2$. By using the results in (a) and (c), formulate and prove a variation of a sphere-packing bound on the size of a balanced code $\mathcal{C}$ of length n with an additional property that for any two codewords $\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \in \mathcal{C}$, $\mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \geq 2d$.

(e) Check that your result in part (a) is a special case of the bound in part (d) (for $r = 0$).

Solution.

(a) Assume that a binary code $\mathcal{C}$ of even length n is balanced. Therefore, there are $n/2$ zeroes and $n/2$ ones in each codeword in $\mathcal{C}$. The maximal possible size of such a code is achieved when *all* the binary vectors of length n with $n/2$ ones are in the code. The number of such vectors is equal to the number of choices of $n/2$ positions out of n:

$$|\mathcal{C}| \leq \binom{n}{n/2}.$$

(b) The Hamming distance between any two codewords $\bar{\boldsymbol{c}}_1$ and $\bar{\boldsymbol{c}}_2$ in $\mathcal{C}$ is

$$\begin{aligned} \mathsf{d}_H(\bar{\boldsymbol{c}}_1, \bar{\boldsymbol{c}}_2) = |\operatorname{supp}(\bar{\boldsymbol{c}}_1 + \bar{\boldsymbol{c}}_2)| &= |\operatorname{supp}(\bar{\boldsymbol{c}}_1)| + |\operatorname{supp}(\bar{\boldsymbol{c}}_2)| - 2 \cdot |\operatorname{supp}(\bar{\boldsymbol{c}}_1) \cap \operatorname{supp}(\bar{\boldsymbol{c}}_2)| \\ &= n - \underbrace{2 \cdot |\operatorname{supp}(\bar{\boldsymbol{c}}_1) \cap \operatorname{supp}(\bar{\boldsymbol{c}}_2)|}_{even} \rightarrow \text{even}. \end{aligned}$$

Here $\operatorname{supp}(\cdot)$ denotes the support of a vector, i.e. the set of positions with non-zero entries. See Fig. 1 for illustration.

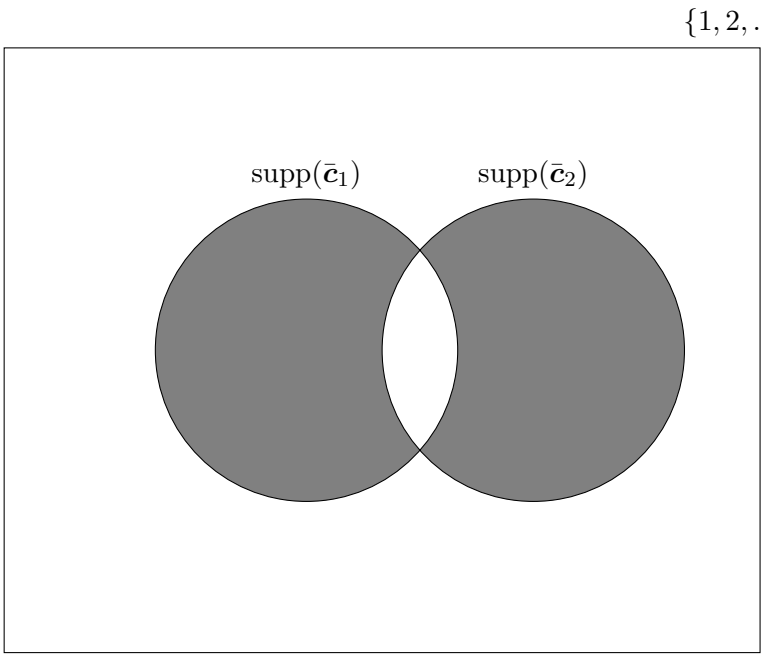

Figure 1: Illustration for Question 2.(b).

(c) If $\mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) = 2t$ and both $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ have Hamming weight $n/2$, then one of them can be obtained from another by changing exactly t ones to zeroes and t zeroes to ones. For a fixed $\bar{\boldsymbol{x}}$, there are $\binom{n/2}{t}$ ways to choose additional positions of ones in $\bar{\boldsymbol{x}}$ and, independently, $\binom{n/2}{t}$ ways to choose additional positions of zeroes. Therefore, the total number of such $\bar{\boldsymbol{y}}$ is

$$\binom{n/2}{t}^2.$$

Hence, the size of a "balanced sphere" is

$$|\mathbb{S}_{2r}(\bar{\boldsymbol{x}})| = \sum_{t=0}^{r} \binom{n/2}{t}^2.$$

(d) Since for any two codewords $\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \in \mathcal{C}$, $\mathsf{d}_H(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \geq 2d$, then no two balanced spheres of radius $2(d-1)$ around codewords of $\mathcal{C}$ intersect. In particular, this requires that

$$\sum_{\bar{\boldsymbol{c}} \in \mathcal{C}} \left|\mathbb{S}_{2(d-1)}(\bar{\boldsymbol{c}})\right| \leq \binom{n}{n/2}.$$

Since the size of each balanced sphere is the same, we obtain:

$$|\mathcal{C}| \leq \frac{\binom{n}{n/2}}{\sum_{t=0}^{d-1} \binom{n/2}{t}^2} \tag{1}$$

(e) By setting $r = d - 1 = 0$ in (1), we obtain that

$$|\mathcal{C}| \leq \frac{\binom{n}{n/2}}{\sum_{t=0}^{0} \binom{n/2}{t}^2} = \frac{\binom{n}{n/2}}{\binom{n/2}{0}^2} = \frac{\binom{n}{n/2}}{1} = \binom{n}{n/2}.$$

■

Question 3 (30 points).

(a) Consider a linear $[n, k = n - r, d]$ code over a finite field $\mathbb{F}$ defined by a parity-check matrix

$$H = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 \\ \alpha_1 & \alpha_2 & \dots & \alpha_{n-1} & 0 \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_{n-1}^2 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_1^{r-2} & \alpha_2^{r-2} & \dots & \alpha_{n-1}^{r-2} & 0 \\ \alpha_1^{r-1} & \alpha_2^{r-1} & \dots & \alpha_{n-1}^{r-1} & 1 \end{pmatrix},$$

where $\alpha_1, \alpha_2, \cdots, \alpha_{n-1}$ are distinct nonzero elements of $\mathbb{F}$.

Prove that the code $\mathcal{C}$ is MDS.

(b) Prove that there exists a generator matrix of the code $\mathcal{C}$ of the following form:

$$G = \begin{pmatrix} 1 & 1 & \dots & 1 & 0 \\ \alpha_1 & \alpha_2 & \dots & \alpha_{n-1} & 0 \\ \alpha_1^2 & \alpha_2^2 & \dots & \alpha_{n-1}^2 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_1^{k-2} & \alpha_2^{k-2} & \dots & \alpha_{n-1}^{k-2} & 0 \\ \alpha_1^{k-1} & \alpha_2^{k-1} & \dots & \alpha_{n-1}^{k-1} & 1 \end{pmatrix} \cdot \begin{pmatrix} u_1 & & & 0 \\ & u_2 & & \\ & & \ddots & \\ 0 & & & u_n \end{pmatrix},$$

where $u_1, u_2, \cdots, u_n$ are nonzero elements of $\mathbb{F}$.

Hint: it was shown in the class that the dual of a GRS code is a GRS code. You can use that fact.

Solution.

(a) The code $\mathcal{C}$ is MDS if and only if $d = n - k + 1 = r + 1$. The latter is equivalent to the fact that any r columns of H are linearly-independent. In other words, any r columns form an $r \times r$ invertible (full-rank) matrix.

If these r columns do not include the last columns, the matrix has Vandermonde form and, hence, full rank.

Now consider the case when these r columns *include* the last column. Without loss of generality, assume that the matrix is formed by columns $1, 2, \ldots, r-2, r-1$ and the last one. Let us calculate the determinant of the following matrix:

$$\det \left[\begin{array}{cccc|c} 1 & 1 & \ldots & 1 & 0 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_{r-1} & 0 \\ \alpha_1^2 & \alpha_2^2 & \ldots & \alpha_{r-1}^2 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_1^{r-2} & \alpha_2^{r-2} & \ldots & \alpha_{r-1}^{r-2} & 0 \\ \hline \alpha_1^{r-1} & \alpha_2^{r-1} & \ldots & \alpha_{r-1}^{r-1} & 1 \end{array}\right] = \det \left[\begin{array}{cccc} 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_{r-1} \\ \alpha_1^2 & \alpha_2^2 & \ldots & \alpha_{r-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{r-2} & \alpha_2^{r-2} & \ldots & \alpha_{r-1}^{r-2} \end{array}\right] \neq 0,$$

because the latter matrix is also of the Vandermonde form.

(b) Note that $\operatorname{rank} G = k$ (compare with (a)). Therefore, we will prove the required statement if we find such non-zero $u_1, u_2, \ldots, u_n \in \mathbb{F}$, that satisfy:

$$GH^\intercal = O_{k\times r},$$

where $O_{k\times r}$ denotes all-zero matrix of the size $k \times r$. Let us consider multiplication of $(i+1)$'th row of G by $(j+1)$'th column of $H^\intercal$ for $0 \leq i + j \leq (k-1) + (r-1) - 1 = n - 3$:

$$\sum_{t=1}^{n-1} \alpha_t^i u_t \alpha_t^j = \sum_{t=1}^{n-1} \alpha_t^{i+j} u_t = 0. \tag{2}$$

Note that these equations depends only on the sum $s = i + j$ ($s = 0, 1, \ldots, n-3$) and not particular values of i and j.

However, for $i = k-1$ and $j = r-1$ (i.e. $i + j = n - 2$) we obtain the following equation:

$$\sum_{t=1}^{n-1} \alpha_t^{n-2} u_t + u_n = 0. \tag{3}$$

We see that u_n appears only in this last equation. Hence, let us first show that it is possible to choose the required $u_1, u_2, \ldots, u_{n-1}$. They should satisfy (2), which can be re-written in matrix form as follows:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_{n-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{n-3} & \alpha_2^{n-3} & \cdots & \alpha_{n-1}^{n-3} \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{n-1} \end{pmatrix} = \bar{\mathbf{0}}^\intercal. \tag{4}$$

Observe that the matrix here is a parity-check matrix of $[n-1, 1, n-1]$ Reed-Solomon code. Therefore $(u_1, u_2, \ldots, u_{n-1})$ is any codeword of this code and—if we require it to be non-zero—it has the Hamming weight $n-1$, i.e. all $u_1, u_2, \ldots, u_{n-1}$ are non-zero. Fix some particular $u_1, u_2, \ldots, u_{n-1}$.

Now turn to the equation (3). In a straightforward manner, we obtain:

$$u_n = -\sum_{t=1}^{n-1} \alpha_t^{n-2} u_t.$$

The only thing we still need to check is that such $u_n \neq 0$. Assume, to the contrary that

$$\sum_{t=1}^{n-1} \alpha_t^{n-2} u_t = 0.$$

Together with (4) this gives us:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_{n-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{n-2} & \alpha_2^{n-2} & \cdots & \alpha_{n-1}^{n-2} \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix} = \bar{\mathbf{0}}^{\mathsf{T}}.$$

However, the matrix of this system is invertible and, therefore, the only solution is $u_1 = u_2 = \cdots = u_n = 0$, which contradicts the fact that all $u_1, u_2, \ldots, u_{n-1}$ are non-zero. This contradiction proves that $u_n \neq 0$.

Therefore, we have proved that there is a choice of non-zero $u_1, u_2, \ldots, u_n$ for which G is indeed a generator matrix of $\mathcal{C}$.

■

Question 4 (30 points).

Let $\mathbb{F} = \mathbb{F}_7$ be a field of integer residues modulo 7. Suppose that $\mathcal{C}$ is a $[6, 2, 5]$ Reed-Solomon code over $\mathbb{F}$, with a parity-check matrix of the code given by

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 6 & 5 & 4 & 3 & 2 & 1 \\ 6^2 & 5^2 & 4^2 & 3^2 & 2^2 & 1^2 \\ 6^3 & 5^3 & 4^3 & 3^3 & 2^3 & 1^3 \end{pmatrix}.$$

Assume that $\bar{\boldsymbol{c}} \in \mathcal{C}$ is transmitted, and $\bar{\boldsymbol{y}} = (2, 6, 2, 6, 2, 6) \in \mathbb{F}^6$ is received. In this question, you will decode $\bar{\boldsymbol{y}}$.

(a) Find the syndrome polynomial $S(x)$.

(b) Find the error-locator and the error-evaluator polynomials by either Peterson-Gorenstein-Zierler method or by Euclid's algorithm.

(c) What are the error locations and error values?

(d) What is $\bar{\boldsymbol{c}}$ if we assume that there were at most $\lfloor (d-1)/2 \rfloor$ errors?

(e) Compute $\boldsymbol{H} \cdot \bar{\boldsymbol{c}}^T$ and show that indeed $\bar{\boldsymbol{c}} \in \mathcal{C}$.

Solution. Before proceeding further, let us re-write $\boldsymbol{H}$ modulo 7:

$$\boldsymbol{H} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 4 & 2 & 2 & 4 & 1 \\ 6 & 6 & 1 & 6 & 1 & 1 \end{pmatrix}$$

(a) Syndrome vector:

$$\bar{\boldsymbol{s}} = \boldsymbol{H}\bar{\boldsymbol{y}}^{\intercal} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 4 & 2 & 2 & 4 & 1 \\ 6 & 6 & 1 & 6 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 6 \\ 2 \\ 6 \\ 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 24 \\ 78 \\ 56 \\ 94 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 3 \end{pmatrix}$$

Therefore, the syndrome polynomial is

$$S(x) = 3x^3 + x + 3$$

(b) We use the extended Euclid's algorithm. Initialisation:

$$\begin{aligned} r_{-1}(x) = a(x) = x^{d-1} = x^4 \qquad & t_{-1}(x) = 0 \\ r_0(x) = b(x) = S(x) = 3x^3 + x + 3 \qquad & t_0(x) = 1. \end{aligned}$$

Iterations of the extended Euclid's algorithm are as follows.

Iteration 1:

$$\begin{aligned} r_{-1}(x) = x^4 &= \underbrace{(5x)}_{q_1(x)} \cdot (3x^3 + x + 3) + \underbrace{(2x^2 + 6x)}_{r_1(x)} \\ t_{-1}(x) = 0 &= (5x) \cdot (1) + \underbrace{(2x)}_{t_1(x)} \end{aligned}$$

Iteration: 2

$$\begin{aligned} r_0(x) = 3x^3 + x + 3 &= \underbrace{(5x+6)}_{q_2(x)} \cdot (2x^2 + 6x) + \underbrace{(3)}_{r_2(x)} \\ t_0(x) = 1 &= (5x+6) * (2x) + \underbrace{(4x^2 + 2x + 1)}_{t_2(x)} \end{aligned}$$

We observe that $\deg r_2 = 0 < \frac{d-1}{2} = \frac{5-1}{2} = 2$. Thus, we stop iterations.

We set $\Lambda(x) = t_2(x) = 4x^2 + 2x + 1$ and $\Gamma(x) = r_2(x) = 3$. We verify that $\Lambda(0) = 1$.

(c) To find locations of errors, we need to find roots of $\Lambda(x)$. By simply try elements of $\mathbb{F}_7$, we find the roots: $\{1, 2\}$. Therefore, there are errors in positions j where $\alpha_j^{-1} \in \{1, 2\}$. In other words, $\alpha_j \in \{1, 4\}$ and $j \in \{6, 3\}$. That is, the errors are in positions 3 and 6.

Since $\Lambda'(x) = x + 2$, we obtain values of errors:

$$e_3 = -\alpha_3 \cdot \frac{\Gamma(\alpha_3^{-1})}{\Lambda'(\alpha_3^{-1})} = -4 \cdot \frac{3}{2+2} = -3 = 4,$$
$$e_6 = -\alpha_6 \cdot \frac{\Gamma(\alpha_6^{-1})}{\Lambda'(\alpha_6^{-1})} = -1 \cdot \frac{3}{1+2} = -1 = 6.$$

(d) If we assume the aforementioned number of errors then

$$\bar{\boldsymbol{c}} = \bar{\boldsymbol{y}} - \bar{\boldsymbol{e}} = (2, 6, 2, 6, 2, 6) - (0, 0, 4, 0, 0, 6) = (2, 6, 5, 6, 2, 0).$$

(e) Verification:

$$\boldsymbol{H}\bar{\boldsymbol{c}}^{\mathsf{T}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 4 & 2 & 2 & 4 & 1 \\ 6 & 6 & 1 & 6 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 6 \\ 5 \\ 6 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 21 \\ 84 \\ 56 \\ 91 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

■