Northeastern University, Department of Mathematics

MATH G5110: Applied Linear Algebra and Matrix Analysis. (Fall 2020)

• Instructor: **He Wang** Email: **he.wang@northeastern.edu**

§6 Least Squares and Data Fitting

Contents

1. Least Squares

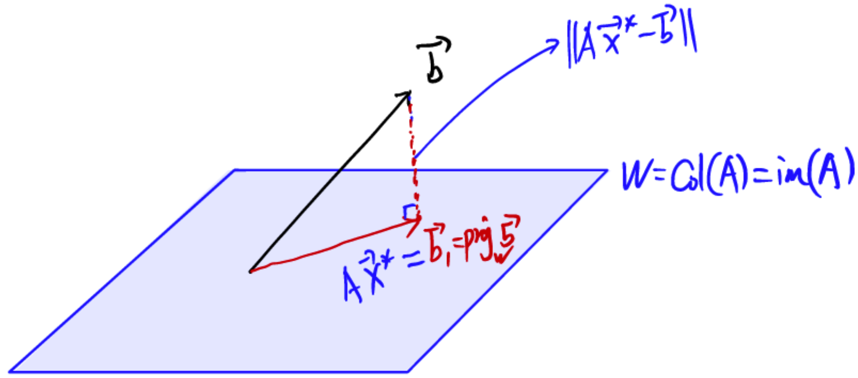
Approximate Solutions to Inconsistent Systems

- Let A be an $n \times m$ matrix and let $\vec{b}$ be an n-dimensional vector such that the system
$$A\vec{x} = \vec{b}$$
is inconsistent (no solution). (if and only if $\vec{b} \notin \text{Col}\, A = \text{im}\, A = \text{Span}(\vec{a}_1, \cdots, \vec{a}_m)$).
- In this case a natural question to ask is which m-dimensional vector(s) $\vec{x}^*$ has/have the property that $A\vec{x}^*$ is closest to $\vec{b}$. Here "closeness" of $A\vec{x}^*$ to $\vec{b}$ is measured by the smallness of
$$||A\vec{x}^* - \vec{b}||$$

[Least-Squares Problem/Solutions]

For an $n \times m$ matrix A and an inconsistent system $A\vec{x} = \vec{b}$, find the vector(s) $\vec{x}^* \in \mathbb{R}^m$ such that
$$||A\vec{x}^* - \vec{b}|| \leq ||A\vec{x} - \vec{b}||$$
for all $x \in \mathbb{R}^m$.

$||A\vec{x}^* - \vec{b}||$ is the **least squares error**.



- To find the Least Square solution(s) $\vec{x}^*$ of an inconsistent system $A\vec{x} = \vec{b}$, we replace the system by the consistent system $A\vec{x} = \vec{b}_1$ with $\vec{b}_1$ the closest vector in Col A to $\vec{b}$, namely $\vec{b}_1 = \text{proj}_{\text{Col}\,A}(\vec{b})$.

Theorem 1 (Solution to the Least-Squares Problem)**.** *Let A be an $n \times m$ matrix. Let $\vec{b} \in \mathbb{R}^n$ and $\vec{b}_1 = \text{proj}_{\text{Col}\,A}(\vec{b})$. Then, any solutions $\vec{x}^*$ of the consistent system $A\vec{x} = \vec{b}_1$ is a least-squares solution.*

Example 2. Find the least-squares solutions for the system $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix}$ and $\begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix}$

$\{\vec{a}_1, \vec{a}_2\}$ is an orthogonal basis for Col(A). since $\vec{a}_1 \cdot \vec{a}_2 = 0$

Step1: $\vec{b}_1 = \text{proj}_{\text{Col}A}(\vec{b}) = \dfrac{\vec{b}\cdot\vec{a}_1}{\vec{a}_1\cdot\vec{a}_1}\vec{a}_1 + \dfrac{\vec{b}\cdot\vec{a}_2}{\vec{a}_2\cdot\vec{a}_2}\vec{a}_2$

$= -\frac{18}{3}\vec{a}_1 + \frac{24}{96}\vec{a}_2 = -6\vec{a}_1 + \frac{1}{4}\vec{a}_2 = \begin{bmatrix} 7 \\ -4 \\ 7 \end{bmatrix}$

Step2: Solve $A\vec{x} = \vec{b}_1$

$\begin{bmatrix} \vec{a}_1 & \vec{a}_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{b}_1$

$x_1\vec{a}_1 + x_2\vec{a}_2 = \vec{b}_1$

So $\vec{x} = \begin{bmatrix} -6 \\ \frac{1}{4} \end{bmatrix}$ is a least squares solution

Theorem 3. *(Normal Equation) The set of Least-Square solutions of the inconsistent system $A\vec{x} = \vec{b}$ coincides with the solution set of the consistent system of **normal equations***

$$(A^T A)\vec{x} = A^T\vec{b}.$$

Proof. Proof: Let $V = \text{im}\,A$.
$\vec{x}_*$ is a least-squares solution for $A\vec{x} = \vec{b} \iff A\vec{x}_* = \text{proj}_V\,\vec{b}$
$\iff \vec{b} - A\vec{x}_* = \vec{b}^\perp \in (\text{im}\,A)^\perp = \ker A^T$
$\iff A^T(\vec{b} - A\vec{x}_*) = \vec{0}$
$\iff A^T\vec{b} - A^T A\vec{x}_* = \vec{0}$
$\iff A^T A\vec{x}_* = A^T\vec{b}$ □

Example 4. Find the least-squares solutions for the system $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix}$

$$A^TA = \begin{bmatrix} -1 & 1 & -1 \\ 4 & 8 & 4 \end{bmatrix} \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 96 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} -1 & 1 & -1 \\ 4 & 8 & 4 \end{bmatrix} \begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix} = \begin{bmatrix} -18 \\ 24 \end{bmatrix}$$

Solve the normal equation $A^TA\,\vec{x} = A^T\vec{b}$

$$\left[\begin{array}{cc|c} 3 & 0 & -18 \\ 0 & 96 & 24 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 1 & 0 & -6 \\ 0 & 1 & 4 \end{array}\right]$$

So $\vec{x} = \begin{bmatrix} -6 \\ 4 \end{bmatrix}$ is the least-squares solution

(2) The image im(A) is a plane in $\mathbb{R}^3$ passing the origin. Find the distance from the vector $\vec{b}$ (or the point $(14, -4, 0)$) to the plane im(A). (Hint: Use the geometric meaning of the least-squares solution in (1))

The distance is given be the norm of $\vec{b}^{\perp} = \vec{b} - \text{proj}_{\text{im}(A)}\,\vec{b}$.

We know that $\text{proj}_{\text{im}(A)}\,\vec{b} = Ax^* = \begin{bmatrix} -1 & 4 \\ 1 & 8 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} -6 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 7 \\ -4 \\ 7 \end{bmatrix}$.

So, $\vec{b}^{\perp} = \begin{bmatrix} 14 \\ -4 \\ 0 \end{bmatrix} - \begin{bmatrix} 7 \\ -4 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ 0 \\ -7 \end{bmatrix}$. So the distance is $||\vec{b}^{\perp}|| = 7\sqrt{2}$.

Example 5. Find the least-squares solutions for the system $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}$

Step1. Construct the normal equation $A^TA\vec{x}=A^T\vec{b}$

$$A^TA=\begin{bmatrix}1&1&1&1\\1&1&0&0\\0&0&1&1\end{bmatrix}\begin{bmatrix}1&1&0\\1&1&0\\1&0&1\\1&0&1\end{bmatrix}=\begin{bmatrix}4&2&2\\2&2&0\\2&0&2\end{bmatrix}$$

$$A^T\vec{b}=\begin{bmatrix}1&1&1&1\\1&1&0&0\\0&0&1&1\end{bmatrix}\begin{bmatrix}1\\3\\2\\4\end{bmatrix}=\begin{bmatrix}10\\4\\6\end{bmatrix}$$

Solve the normal equation

$$\left[\begin{array}{ccc|c}4&2&2&10\\2&2&0&4\\2&0&2&6\end{array}\right]\rightarrow\cdots\rightarrow \text{rref}=\left[\begin{array}{ccc|c}1&0&1&3\\0&1&-1&-1\\0&0&0&0\end{array}\right]$$

$x_1=3-x_3$
$x_2=-1+x_3$
x_3 free

$$\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}3-x_3\\-1+x_3\\x_3\end{bmatrix}=\begin{bmatrix}3\\-1\\0\end{bmatrix}+x_3\begin{bmatrix}-1\\1\\1\end{bmatrix}$$

A technical property:

Proposition 6. *Let A be an $n\times m$ matrix.*

- $\ker(A)=\ker(A^TA)$
- *If* $\ker(A)=\{0\}$, *then A^TA is an invertible matrix.*

Proof. $\square$

Corollary 7. *If* $\operatorname{rank}A=m$, *then* $\ker(A)=\{0\}$, *then A^TA is an $m\times m$ invertible matrix. In this case, the normal equation $(A^TA)\vec{x}=A^T\vec{b}$ has a unique solution:*

$$\vec{x}=(A^TA)^{-1}A^T\vec{b}$$

QR factorization method Suppose A is $n\times m$ matrix with full column rank. Solve the least squares solution using QR factorization $A=QR$ where Q is an orthogonal matrix $n\times m$ and R is an $m\times m$ upper triangular matrix with rank m.

$$\begin{aligned}
\vec{x} &= (A^T A)^{-1} A^T \vec{b} \\
&= ((QR)^T QR)^{-1} (QR)^T \vec{b} \\
&= (R^T Q^T QR)^{-1} R^T Q^T \vec{b} \\
&= (R^T R)^{-1} R^T Q^T \vec{b} \\
&= (R^T R)^{-1} R^T Q^T \vec{b} \\
&= R^{-1} Q^T \vec{b}
\end{aligned}$$

Example 8.

2. **Data Fitting**

Problem: Fitting a function of a certain type of data. We use the following three example to illustrate this application.

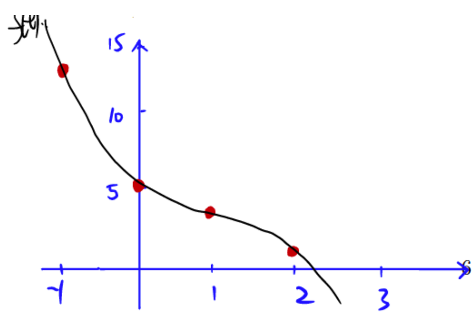
Example 9. Find a cubic polynomial $f(t) = c_0 + c_1 t + c_2 t^2 + c_3 t^3$ whose graph passes through the points $(0, 5)$, $(1, 3)$, $(-1, 13)$, $(2, 1)$

Solution:

We need to solve the linear system $\begin{cases} c_0 & = 5 \\ c_0 + c_1 + c_2 + c_3 & = 3 \\ c_0 - c_1 + c_2 - c_3 & = 13 \\ c_0 + 2c_1 + 4c_2 + 8c_3 & = 1 \end{cases}$

$$[A|\vec{b}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 5 \\ 1 & 1 & 1 & 1 & 3 \\ 1 & -1 & 1 & -1 & 13 \\ 1 & 2 & 4 & 8 & 1 \end{bmatrix} \to \cdots \to \mathbf{rref}[A|\vec{b}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

So, the linear system has the unique solution $\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \\ 3 \\ -1 \end{bmatrix}$ So, the cubic polynomial is $f(t) = 5 - 4t + 3t^2 - t^3$.



perfect fit, but calculation is hard.

Example 10. Fit a quadratic function $g(t) = c_0 + c_1 t + c_2 t^2$ to the four data points $(0, 5)$, $(1, 3)$, $(-1, 13)$, $(2, 1)$

We need to solve the linear system

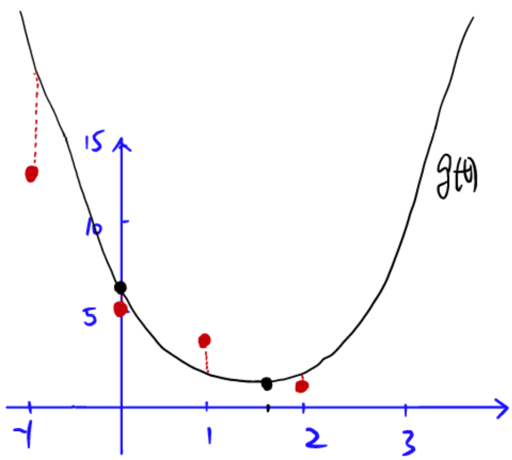
$$\begin{cases} c_0 & = 5 \\ c_0 + c_1 + c_2 & = 3 \\ c_0 - c_1 + c_2 & = 13 \\ c_0 + 2c_1 + 4c_2 & = 1 \end{cases}$$

As matrix equation $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 5 \\ 3 \\ 13 \\ 1 \end{bmatrix}$

$A^TA = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix}$ $\quad A^T\vec{b} = \begin{bmatrix} 22 \\ -8 \\ 20 \end{bmatrix}$

Solve the normal equation $A^TA\vec{x} = A^T\vec{b}$ $\quad \vec{x} = \begin{bmatrix} 5.9 \\ -5.3 \\ 1.5 \end{bmatrix} = \vec{c}^*$

So, the quadratic function $g(t) = 5.9 - 5.3t + 1.5t^2$



$A\vec{c}^* = \begin{bmatrix} g(a_1) \\ g(a_2) \\ g(a_3) \\ g(a_4) \end{bmatrix}$ $\quad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$

$\|\vec{b} - A\vec{c}^*\|^2 = (b_1 - g(a_1))^2 + (b_2 - g(a_2))^2 + (b_3 - g(a_3))^2 + (b_4 - g(a_4))^2$

The sum of the vertical distances between graph and data points is minimal.

Example 11. Fit a linear function $h(t) = c_0 + c_1 t$ to the four data points $(0, 5)$, $(1, 3)$, $(-1, 13)$, $(2, 1)$

We need to solve the linear system

$$\begin{cases} c_0 & = 5 \\ c_0 + c_1 & = 3 \\ c_0 - c_1 & = 13 \\ c_0 + 2c_1 & = 1 \end{cases}$$

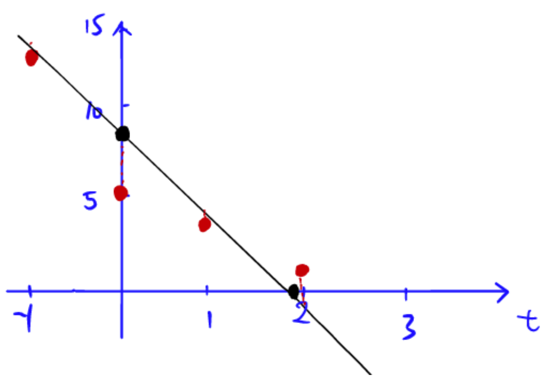
As matrix equation $A\vec{x} = \vec{b}$, where $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 5 \\ 3 \\ 13 \\ 1 \end{bmatrix}$

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 13 \\ 1 \end{bmatrix} = \begin{bmatrix} 22 \\ -8 \end{bmatrix}$$

Solve the normal equation $A^TA\vec{x} = A^T\vec{b}$. $\vec{x} = \begin{bmatrix} 7.4 \\ -3.8 \end{bmatrix}$

So the linear function is $h(t) = 7.4 - 3.8t$

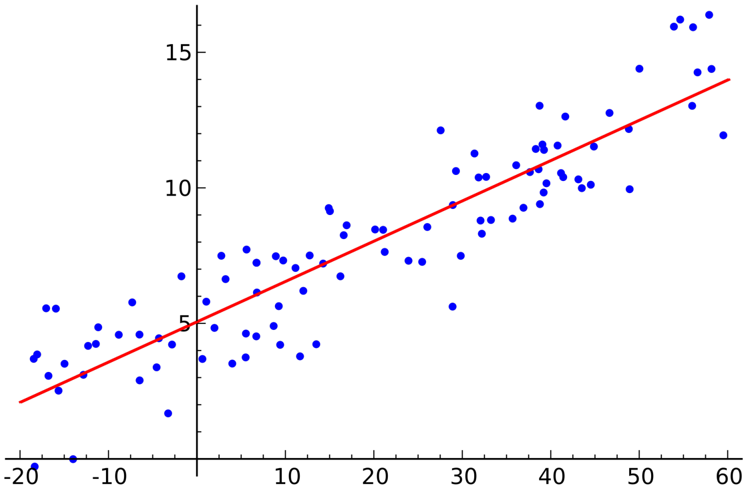


Remark: More generally, we can consider n-points $(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)$.

- Find a linear function $h(t) = c_0 + c_1 t$ fits the data by the least squares.

More generally, the following question is very standard in statistics.

Example 12. Consider the data with n points (a_1, b_1), (a_2, b_2), $\ldots$, (a_n, b_n). Find a linear function $h(t) = c_0 + c_1 t$ fits the data by the least squares. (Suppose $a_1 \neq a_2$)



We need to solve the least-squares problem for $A\vec{x} = \vec{b}$, for $A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \\ 1 & a_n \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

$$A^T A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \\ 1 & a_n \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n a_i \\ \sum_{i=1}^n a_i & \sum_{i=1}^n a_i^2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \\ 1 & a_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n b_i \\ \sum_{i=1}^n a_i a_i \end{bmatrix}$$

Since $a_1 \neq a_2$, we know that rank $A = 2$.
The normal equation $A^T A\vec{x} = A^T \vec{b}$ has a unique solution

$$\vec{x}_* = (A^T A)^{-1} A^T \vec{b} = \frac{1}{n \sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2} \begin{bmatrix} \sum_{i=1}^n a_i^2 & -\sum_{i=1}^n a_i \\ -\sum_{i=1}^n a_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n b_i \\ \sum_{i=1}^n a_i a_i \end{bmatrix}$$

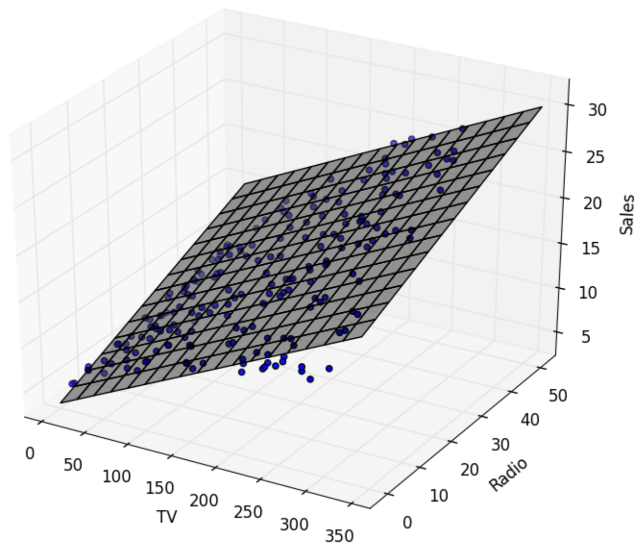
$$= \frac{1}{n \sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2} \begin{bmatrix} (\sum_{i=1}^n a_i^2)(\sum_{i=1}^n b_i) - (\sum_{i=1}^n a_i)(\sum_{i=1}^n a_i a_i) \\ -(\sum_{i=1}^n a_i)(\sum_{i=1}^n b_i) + n \sum_{i=1}^n a_i a_i \end{bmatrix}$$

Example 13. Consider the data with m inputs and 1 output:

$(a_{11}, a_{12}, ..., a_{1m}, b_1)$, $(a_{21}, a_{22}, ..., a_{2m}, b_2)$, $\ldots$, $(a_{n1}, a_{n2}, ..., a_{nm}, b_n)$.

Find a linear function $h(t_1, t_2, ..., t_n) = c_0 + c_1 t_1 + c_2 t_2 + \cdots + c_n t_n$ fits the data by the least squares.

For example, when $m = 2$,



We need to solve the least-squares problem for $A\vec{x} = \vec{b}$, for $A = \begin{bmatrix} 1 & a_{11} & a_{12} & \ldots & a_{1m} \\ 1 & a_{21} & a_{22} & \ldots & a_{2m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & a_{n1} & a_{n2} & \ldots & a_{nm} \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$

Example 14. Consider the data with m inputs and s outputs:

$(a_{11}, a_{12}, ..., a_{1m}, b_{11}, ..., b_{1s})$, $(a_{21}, a_{22}, ..., a_{2m}, b_{21}, ..., b_{2s})$, $\ldots$, $(a_{n1}, a_{n2}, ..., a_{nm}, b_{n1}, ..., b_{ns})$.

Find a linear function $H(\vec{t}) = \vec{c}_0 + C\vec{t}$ fits the data by the least squares.