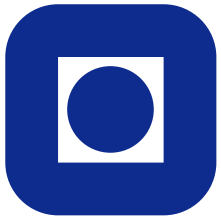

NTNU
Norwegian University of
Science and Technology

TMA4115 - Calculus 3

Lecture 27, April 24

Toke Meier Carlsen
Norwegian University of Science and Technology

Spring 2013

Review of last week's lecture

Last week we introduced and studied

- *the inner product*,
- the *length* of a vector,
- *orthogonality* and *orthogonal* sets in $\mathbb{R}^n$,
- *orthogonal* matrices,
- *orthogonal projections*,
- the *Gram-Schmidt process*,
- *QR factorization*.

NTNU
Norwegian University of
Science and Technology

Today's lecture

Today we shall look at

- *least-squares problems*,
- applications to linear models.

Least-squares problems

- Let A be an $m \times n$ matrix and $\mathbf{b}$ a vector in $\mathbb{R}^m$.
- A *least-square solution* of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

  for all $\mathbf{x}$ in $\mathbb{R}^n$.
- The number $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the *least-squares error* of $A\mathbf{x} = \mathbf{b}$.

NTNU
Norwegian University of
Science and Technology

Solution of the general least-squares problem

Theorem 13

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.

NTNU
Norwegian University of
Science and Technology

Proof of Theorem 13

Let $\hat{\mathbf{b}} = \text{proj}_{\text{Col}(A)}\,\mathbf{b}$. Then $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.

Suppose $\hat{\mathbf{x}}$ is a least-squares solutions of $A\mathbf{x} = \mathbf{b}$. Then $\mathbf{b} - \hat{\mathbf{b}} = \mathbf{b} - A\hat{\mathbf{x}}$ is in $(\text{Col}(A))^{\perp} = \text{Nul}(A^T)$, so $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$ from which it follows that $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

Conversely, suppose $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$. Then $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$, so $\mathbf{b} - A\hat{\mathbf{x}}$ is in $\text{Nul}(A^T) = (\text{Col}(A))^{\perp}$. Since $\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$ and $A\hat{\mathbf{x}}$ is in $\text{Col}(A)$, it follows that $A\hat{\mathbf{x}} = \text{proj}_{\text{Col}(A)}\,\mathbf{b} = \hat{\mathbf{b}}$, so $\hat{\mathbf{x}}$ is a least-squares solutions of $A\mathbf{x} = \mathbf{b}$. Thus $\hat{\mathbf{x}}$ is a least-squares solutions of $A\mathbf{x} = \mathbf{b}$ if and only if $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.

Since $\hat{\mathbf{b}}$ is in $\text{Col}(A)$, there is at least one $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$, so the set of solutions of $A^TA\mathbf{x} = A^T\mathbf{b}$ is nonempty.

NTNU
Norwegian University of
Science and Technology

Example

Let us find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix}.$$

Solution

$$A^TA = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 8 & 10 \end{bmatrix} \text{ and}$$

$$A^T\mathbf{b} = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix} = \begin{bmatrix} -24 \\ -2 \end{bmatrix} \text{ so a least-squares}$$

solution of $A\mathbf{x} = \mathbf{b}$ is the same as a solution of the equation

$$\begin{bmatrix} 4 & 8 \\ 8 & 10 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} -24 \\ -2 \end{bmatrix}.$$

NTNU
Norwegian University of
Science and Technology

Solution (cont.)

$\begin{vmatrix} 4 & 8 \\ 8 & 10 \end{vmatrix} = 40 - 64 = -24$, so $\begin{bmatrix} 4 & 8 \\ 8 & 10 \end{bmatrix}$ is invertible and

$\hat{\mathbf{x}} = \begin{bmatrix} 4 & 8 \\ 8 & 10 \end{bmatrix}^{-1} \begin{bmatrix} -24 \\ -2 \end{bmatrix} = \frac{1}{-24} \begin{bmatrix} 10 & -8 \\ -8 & 4 \end{bmatrix}^{-1} \begin{bmatrix} -24 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{28}{3} \\ \frac{-23}{3} \end{bmatrix}$ is the unique least-squares solution of $A\mathbf{x} = \mathbf{b}$.

NTNU
Norwegian University of
Science and Technology

Example

Let us find all least-squares solutions of $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 7 \\ 2 \\ 3 \\ 6 \\ 5 \\ 4 \end{bmatrix}.$$

NTNU
Norwegian University of
Science and Technology

Solution

$$A^TA = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \text{ and}$$

$$A^T\mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 2 \\ 3 \\ 6 \\ 5 \\ 4 \end{bmatrix} = \begin{bmatrix} 27 \\ 12 \\ 15 \end{bmatrix},$$

NTNU
Norwegian University of
Science and Technology

Solution (cont.)

so a least-squares solution of $A\mathbf{x} = \mathbf{b}$ is the same as a solution of the equation $\begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 27 \\ 12 \\ 15 \end{bmatrix}$.

We reduce the augmented matrix of the above system to its reduced echelon form.

$$\begin{bmatrix} 6 & 3 & 3 & 27 \\ 3 & 3 & 0 & 12 \\ 3 & 0 & 3 & 15 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{9}{2} \\ 1 & 1 & 0 & 4 \\ 1 & 0 & 1 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{9}{2} \\ 0 & \frac{1}{2} & \frac{-1}{2} & \frac{-1}{2} \\ 0 & \frac{-1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 0 & 1 & 5 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

NTNU
Norwegian University of
Science and Technology

Solution

So the least-squares solutions of $A\mathbf{x} = \mathbf{b}$ are

$$\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ -1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$ where t is a free parameter.

Uniqueness of least-squares solutions

Theorem 14

Let A be an $m \times n$ matrix. Then the following statements are logically equivalent.

1. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.
2. The columns of A are linearly independent.
3. The matrix $A^T A$ is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$.

NTNU
Norwegian University of
Science and Technology

Example

Let us find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}.$$

Solution

Let $\mathbf{v}_1, \mathbf{v}_2$ be the columns of A. Notice that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. Recall that $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $A\hat{\mathbf{x}} = \text{proj}_{\text{Col}(A)}\,\mathbf{b}$.

We have that $\text{proj}_{\text{Col}(A)}\,\mathbf{b} = \frac{\mathbf{b}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{b}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 = A\begin{bmatrix} \frac{\mathbf{b}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1} \\ \frac{\mathbf{b}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2} \end{bmatrix}$, so

$$\hat{\mathbf{x}} = \begin{bmatrix} \frac{\mathbf{b}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1} \\ \frac{\mathbf{b}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2} \end{bmatrix} = \begin{bmatrix} 3 \\ \frac{1}{2} \end{bmatrix}.$$

NTNU
Norwegian University of
Science and Technology

QR-factorization and least-squares solutions

Theorem 15

Let A be an $m \times n$ matrix with linearly independent columns, let $A = QR$ be a QR-factorization of A and let $\mathbf{b}$ be in $\mathbb{R}^m$. Then $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$ is the unique least-squares solution of $A\mathbf{x} = \mathbf{b}$.

NTNU
Norwegian University of
Science and Technology

Proof of Theorem 15

Since the columns of Q forms an orthonormal basis of $\text{Col}(A)$, it follows that $\text{proj}_{\text{Col}(A)}\,\mathbf{b} = QQ^T\mathbf{b}$.

Let $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$. Then

$A\hat{\mathbf{x}} = QRR^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b} = \text{proj}_{\text{Col}(A)}\,\mathbf{b}$, so $\hat{\mathbf{x}}$ is the unique least-squares solution of $A\mathbf{x} = \mathbf{b}$.

NTNU
Norwegian University of
Science and Technology

Example

Suppose that $A = \begin{bmatrix} 1 & -1 \\ 1 & 4 \\ 1 & -1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}$.

Let us find the least-squares solution of $A\mathbf{x} = \mathbf{b}$ where

$\mathbf{b} = \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix}$.

NTNU
Norwegian University of
Science and Technology

Solution

The least-squares solution of $A\mathbf{x} = \mathbf{b}$ is

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b} = \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}^T \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 5 & -3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{-1}{2} & \frac{1}{2} & \frac{-1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 5 & -3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 17 \\ 9 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 58 \\ 18 \end{bmatrix} = \begin{bmatrix} 2.9 \\ 0.9 \end{bmatrix}$$

NTNU
Norwegian University of
Science and Technology

Least-squares lines

- Suppose we are given a finite number of points $(x_1, y_1), \ldots, (x_n, y_n)$.
- We want to determine the line $y = \beta_0 + \beta_1 x$ which is as "close" to the points as possible.
- For each point (x_j, y_j) we call
  - y_j for the *observed* value of y,
  - $\beta_0 + \beta_1 x_j$ for the *predicted* value of y,
  - $y_j - (\beta_0 + \beta_1 x_j)$ for the *residual*.
- The *least-squares line* is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals.

NTNU
Norwegian University of
Science and Technology

Finding the least-squares line

Finding the least-squares line is equivalent to computing the least-squares solution of the equation $X\beta = \mathbf{y}$ where

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

NTNU
Norwegian University of
Science and Technology

Example

Let us determine the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points (1,0), (2,1), (4,2) and (5,3).

Solution

Let $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$. Then

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ 12 & 46 \end{bmatrix} \text{ and}$$

$$X^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 25 \end{bmatrix}.$$

NTNU
Norwegian University of
Science and Technology

Solution (cont.)

The columns of X are linearly independent so the equation $X\beta = \mathbf{y}$ has a unique least-squares solution which is

$$\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = (X^TX)^{-1}X^T\mathbf{y} = \begin{bmatrix}4 & 12\\ 12 & 46\end{bmatrix}^{-1}\begin{bmatrix}6\\ 25\end{bmatrix}$$
$$= \frac{1}{186-144}\begin{bmatrix}46 & -12\\ -12 & 4\end{bmatrix}\begin{bmatrix}6\\ 25\end{bmatrix}$$
$$= \frac{1}{40}\begin{bmatrix}-24\\ 28\end{bmatrix} = \begin{bmatrix}-0.6\\ 0.7\end{bmatrix}$$

so $y = -0.6 + 0.7x$ is the equation of the least-squares line that best fits the data points.

NTNU
Norwegian University of
Science and Technology

The general linear model

- Any equation of the form $X\beta = \mathbf{y} + \epsilon$ where X and $\mathbf{y}$ are given, is referred to as a *linear model*.
- The goal is to minimize the length of the *residual vector* ϵ.
- This amounts to finding a least-squares solution of $X\beta = \mathbf{y}$.

NTNU
Norwegian University of
Science and Technology

Example

Suppose data points $(x_1, y_1), \ldots, (x_n, y_n)$ appear to lie along a parabola.
Suppose we wish to approximate the data by an equation of the form $y = \beta_0 + \beta_1 x + \beta_2 x^2$.
Let us describe the linear model that produces a least-square fit of the data point to the above equation.

NTNU
Norwegian University of
Science and Technology

Solution

Let $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ and

$$\epsilon = \begin{bmatrix} y_1 - (\beta_0 + \beta_1 x_1 + \beta_2 x_1^2) \\ \vdots \\ y_n - (\beta_0 + \beta_1 x_n + \beta_2 x_n^2) \end{bmatrix}.$$

Then $X\beta = \mathbf{y} + \epsilon$ is the linear model that produces a least-square fit of the data point to the equation $y = \beta_0 + \beta_1 x + \beta_2 x^2$.

NTNU
Norwegian University of
Science and Technology

Example

Suppose we are given data points $(u_1, v_1, y_1), \ldots, (u_n, v_n, y_n)$ that we expect to satisfy an equation of the form $y = \beta_0 + \beta_1 u + \beta_2 v$.
Let us describe the linear model that produces a least-square fit of the data point to the above equation.

NTNU
Norwegian University of Science and Technology

Solution

Let $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & u_1 & v_1 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ and

$$\epsilon = \begin{bmatrix} y_1 - (\beta_0 + \beta_1 u_1 + \beta_2 v_1) \\ \vdots \\ y_n - (\beta_0 + \beta_1 u_n + \beta_2 v_n) \end{bmatrix}.$$

Then $X\beta = \mathbf{y} + \epsilon$ is the linear model that produces a least-square fit of the data point to the equation $y = \beta_0 + \beta_1 u + \beta_2 v$.

NTNU
Norwegian University of
Science and Technology

Tomorrow's lecture

Tomorrow we shall introduce and study

- *symmetric matrices*,
- *quadratic forms*.

Sections 7.1–7.2 in "Linear Algebras and Its Applications" (pages 393–407).

NTNU
Norwegian University of
Science and Technology