TMA 4115 Matematikk 3

Lecture for KJ & NANO

Alexander Schmeding

NTNU

20. April 2017

In today's lecture we will discuss

- again least square problems,
- symmetric and orthogonal matrices,
- the Spectral Theorem for symmetric matrices

Least square solutions

A $m \times n$ matrix, $\overrightarrow{b} \in \mathbb{R}^m$. $\overrightarrow{y} \in \mathbb{R}^n$ is called **least square solution** of $A\overrightarrow{x} = \overrightarrow{b}$ if

$$\|\overrightarrow{b} - A\overrightarrow{y}\| \leq \|\overrightarrow{b} - A\overrightarrow{x}\|, \quad \text{for all } \overrightarrow{x} \in \mathbb{R}^n.$$

Note: $\overrightarrow{y}$ is a least square solution if and only if

- $A\overrightarrow{y} = \text{proj}_{\text{Col}(A)}(\overrightarrow{b})$
- It solves the **normal equation** $A^T A\overrightarrow{x} = A^T\overrightarrow{b}$

For every linear system there is a least square solution and it is the best (approximate) solution.

Spring 2011 Problem 5

Let $A = \begin{bmatrix} 1 & 3 & 0 & 1 \\ 2 & 1 & 5 & -3 \\ -1 & -1 & -2 & 1 \end{bmatrix}$ and $\overrightarrow{b} = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix}$. Find the nearest point in $\mathrm{Col}(A)$ to $\overrightarrow{b}$.

Avoid Gram-Schmidt and compute a least square solution $\overrightarrow{y}$ to $A\overrightarrow{x} = \overrightarrow{b}$ the nearest point is then $A\overrightarrow{y}$. Normal equation: $A^T A\overrightarrow{x} = A^T\overrightarrow{b}$ yields

$$A^T A = \begin{bmatrix} 6 & 6 & 12 & -6 \\ 6 & 11 & 7 & -1 \\ 12 & 7 & 29 & -17 \\ -6 & -1 & -17 & 11 \end{bmatrix}.$$

Can we simplify the problem so that we do not need $A^T A$?

Revisiting an example

Idea: Col(A) is important, but A is not!
Use the method with simpler matrix B with Col(B) =Col(A).

A basis for Col(A) is $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$. Set $B = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ -1 & -1 \end{bmatrix}$ and compute:

$$B^T B = \begin{bmatrix} 6 & 6 \\ 6 & 5 \end{bmatrix}, B^T \overrightarrow{b} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$$

Solving $B^T B \overrightarrow{x} = B^T \overrightarrow{b}$ yields the least square solution $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$

Revisiting an example

The nearest point to $\overrightarrow{b}$ in Col(A) =Col(B) is

$$B\begin{bmatrix}3\\-1\end{bmatrix}=\begin{bmatrix}0\\5\\-2\end{bmatrix}$$

If we know a basis of Col(A), we can replace A by a simpler matrix!

Remark

Can compute nearest points in a subspace by solving least square problems.
$\rightarrow$ avoid Gram-Schmidt ($=$ messy computation).

An observation

In setting up the normal equation we computed matrices of the form $A^T A$. Here are some examples:

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, \quad \begin{bmatrix} 4 & 17 \\ 17 & 81 \end{bmatrix}, \quad \begin{bmatrix} 6 & 6 & 12 & -6 \\ 6 & 11 & 7 & -1 \\ 12 & 7 & 29 & -17 \\ -6 & -1 & -17 & 11 \end{bmatrix}$$

These matrices have an interesting structure:
Their transpose coincides with the matrix![1]

These special matrices are called **symmetric** and we will now investigate them.

[1] By the rules of the transpose $(A^T A)^T = A^T (A^T)^T = A^T A$.

Spectral Theorem for symmetric matrices

A symmetric $n \times n$ matrix, then the following holds:

(a) A has n real eigenvalues counting multiplicities[a]

(b) dim E_λ equals the multiplicity of the eigenvalue λ,
Recall: E_λ = eigenspace to eigenvalue λ

(c) $\lambda \neq \mu$ eigenvalues of A, then $E_\lambda \subseteq E_\mu^\perp$,

(d) A is orthogonally diagonalisable.

[a] **Multiplicity of eigenvalue**: Multiplicity of the root of the characteristic polynomial.
Example: If $(\lambda - 1)^n(\lambda + 3)$ is the characteristic polynomial, $\lambda = 1$ is an eigenvalue of multiplicity n, $\lambda = -3$ is eigenvalue of multiplicity 1.

Spectrum (of a matrix) = set of eigenvalues.

$A = PDP^{-1}$ with $P = \begin{bmatrix} \overrightarrow{u}_1 & \dots & \overrightarrow{u}_n \end{bmatrix}$ orthogonal, D a diagonal matrix (diagonal entries $\lambda_1, \dots, \lambda_n$ the eigenvalues of A.) Since $P^{-1} = P^T$ we have

$$A = PDP^T = \begin{bmatrix} \overrightarrow{u}_1 & \dots & \overrightarrow{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \overrightarrow{u}_1^T \\ \vdots \\ \overrightarrow{u}_n^T \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 \overrightarrow{u}_1 & \dots & \lambda_n \overrightarrow{u}_n \end{bmatrix} \begin{bmatrix} \overrightarrow{u}_1^T \\ \vdots \\ \overrightarrow{u}_n^T \end{bmatrix}$$

We obtain the so called **spectral decomposition** of A

$$A = \lambda_1 \overrightarrow{u}_1 \overrightarrow{u}_1^T + \lambda_2 \overrightarrow{u}_2 \overrightarrow{u}_1^T + \dots + \lambda_n \overrightarrow{u}_n \overrightarrow{u}_n^T$$

Note: $(\overrightarrow{u}_k \overrightarrow{u}_k^T) \overrightarrow{x} = \mathrm{proj}_{\mathrm{span}\{\overrightarrow{u}_k\}}(x)$