TMA 4115 Matematikk 3

Lecture 27 for MBIOT5, MTKJ, MTNANO

Alexander Schmeding

NTNU

8. April 2014

Least square solutions

Let A be an $m \times n$ matrix and $\overrightarrow{b} \in \mathbb{R}^m$. A vector $\overrightarrow{y} \in \mathbb{R}^n$ 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: A vector $\overrightarrow{y}$ is a least square solution if and only if

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

For every linear system there is a least square solution. If the system is inconsistent a least square solution is the best approximate solution we can get.

Revisiting example 21.9

22.7 Example, (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 $A\overrightarrow{y}$. Try to solve the normal equation $A^T A\overrightarrow{x} = A^T\overrightarrow{b}$. However, we obtain

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

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

Revisiting example 21.9

Idea: The column space Col(A) is important not the matrix A! Use the method with an easier matrix B with Col(B) =Col(A).

In 21.9 we saw that $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$ is a basis of Col(A). Thus we set $B = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ -1 & -1 \end{bmatrix}$ and compute with B:

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

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

Revisiting example 21.9

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

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

as before!

Remark We can compute nearest points in a subspace by solving least square problems.
This avoids Gram-Schmidt, which will in general be quite a messy computation.

An observation

In setting up the normal equation we computed matrices of the form A^TA. 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:
Taking their transpose we get back the same matrix![1]

We call these special matrices **symmetric** and will now investigate their properties.

[1]Note that the rules for the transpose imply $(A^TA)^T = A^T(A^T)^T = A^TA$.

23.13 Definition

A quadratic form $Q(\overrightarrow{x}) = \overrightarrow{x}^T A \overrightarrow{x}$ is called

- **positive definite**, if $Q(\overrightarrow{x}) > 0, \ \forall \overrightarrow{x} \neq \overrightarrow{0}$,
- **negative definite**, if $Q(\overrightarrow{x}) < 0, \ \forall \overrightarrow{x} \neq \overrightarrow{0}$,
- **indefinite**, if $Q(\overrightarrow{x})$ assumes both negative and positive values.

These properties are quite important in applications for quadratic forms. We shall discuss one after the next theorem.

23.14 Theorem

Let A be a symmetrix matrix and $Q(\overrightarrow{x}) = \overrightarrow{x}^T A \overrightarrow{x}$ be the associated quadratic form. Then Q is

- positive definite, if all eigenvalues of A are positive
- negative definite, if all eigenvalues of A are negative
- indefinite, if there are positive and negative eigenvalues

Application (in Calculus) Recall from calculus that for a (sufficiently) differentiable function $f : \mathbb{R} \to \mathbb{R}$ we know: If $f'(x) = 0$ and $f''(x) \neq 0$ then

if $f''(x) < 0$ we have a local maximum at x

if $f''(x) > 0$ we have a local minimum at x

Let now $f : \mathbb{R}^n \to \mathbb{R}$ be a (sufficiently) differentiable function. It turns out that a similar criterion holds in this case, but we need symmetric matrices!

Example in two dimensions

In calculus you learn how to compute for $\overrightarrow{x} \in \mathbb{R}^n$ the derivatives $f'(\overrightarrow{x})$ and $f''(\overrightarrow{x})$.

Furthermore, $f''(\overrightarrow{x})$ is determined by the **Hessian**

$$Hf(\overrightarrow{x}) = \left[\frac{\partial^2 f}{\partial_{x_i}\partial_{x_j}}(\overrightarrow{x})\right]_{1\leq i,j\leq n}$$

Example Consider the function $f: \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto \cos(x)\sin(y)$. Then the Hessian computes as:

$$Hf\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = -\begin{bmatrix} \cos(x)\sin(y) & \sin(x)\cos(y) \\ \sin(x)\cos(y) & \cos(x)\sin(y) \end{bmatrix}$$

It is a symmetric matrix[2].

Study maxima and minima via the Hessian!

[2] This is no coincidence, by Schwartz law the Hessian of each sufficiently differentiable function will be symmetric!

Example in two dimensions

The generalized criterion is: If $f'(\overrightarrow{x}) = \overrightarrow{0}$ and the Hessian $Hf(\overrightarrow{x})$ is

- positive definite, then f has a local minimum at $\overrightarrow{x}$,
- negative definite, then f has a local maximum at $\overrightarrow{x}$,
- indefinite, then f has a saddle point at $\overrightarrow{x}$,

Using this technique one can compute the critical points of for example $f(x, y) = \cos(x)\sin(y)$

The graph of $\cos(x)\sin(y)$: