TMA 4115 Matematikk 3

MTEL, MTENERG, MTIØT, MTTK

Alexander Schmeding

NTNU

7. April 2014

Least square solutions

Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. A vector $\mathbf{y} \in \mathbb{R}^n$ is called **least square solution** of $A\mathbf{x} = \mathbf{b}$ if

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

Note: A vector $\mathbf{y}$ is a least square solution if and only if

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

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.

Weight of baggage

We want to know the weight x of our suitcase.
If it is too heavy we have to pay extra on the airplane and we don't want that.

On a scale

weight $x = x_1$kg

Second scale

weight $x = x_2$kg

Third scale

weight $x = x_3$kg

What is the weight of our suitcase?

Weight of baggage II

From our scales we have the information:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leftrightarrow \begin{cases} x & = x_1 \\ x & = x_2 \\ x & = x_3 \end{cases}$$

Unless $x_1 = x_2 = x_3$ holds (by chance), the linear system is **inconsistent**, i.e. there is no exact solution.

Problem: We want an "approximate" solution. This means a number y such that the errors $|x_i - y|$ are as small as possible.

Weight of baggage III

Idea: If $A\mathbf{x} = \mathbf{b}$ is inconsistent, approximate solution means an $\mathbf{x}$ such that $A\mathbf{x}$ is as near as possible to the target vector $\mathbf{b}$.

$A\mathbf{x}$ minimizes the distance to $\mathbf{b}$ if it is $\mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$!
This is in the column space, and we can then solve
$A\mathbf{x} = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$.

In our example with $A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, we set $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and obtain

$$\mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b}) = \frac{\mathbf{b} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v_1} = \begin{bmatrix} \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \\ \frac{x_1+x_2+x_3}{3} \end{bmatrix}$$

Thus $Ax = \mathrm{proj}_{\mathrm{Col}(A)}(\mathbf{b})$ shows: $x = \frac{x_1+x_2+x_3}{3}$.

Application for Least squares I

Spring 2011 Problem 5

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

Recall: If $\mathbf{y}$ is a least square solution, then $A\mathbf{y}$ is the nearest point to $\mathbf{b}$! Try to solve the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$. 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 $A^T A$?

Application for least squares II

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).

One computes that $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$ is a basis of Col(A).Set
$B = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ -1 & -1 \end{bmatrix}$ and compute with B:

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

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

Application for least squares III

Now the nearest point to $\mathbf{b}$ 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.

An easy way to solve the last example?

If we are given an orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ for a subspace W the nearest point to $\mathbf{b} \in \mathbb{R}^n$ is given by

$$\text{proj}_W(\mathbf{b}) = \frac{\mathbf{b} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u_1} + \cdots + \frac{\mathbf{b} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u_p} = \sum_{i=1}^{p} \frac{\mathbf{b} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}\mathbf{u_i}$$

So if we know an orthogonal basis (for the column space) its easy to solve the problem:

Spring 2011 Problem 5

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

We will now investigate how such a basis can be constructed.

The Gram-Schmidt Process

Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ be a basis for a non-zero subspace $W \subseteq \mathbb{R}^n$.Define

$$
\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
\vdots &= \quad \vdots \qquad \quad \vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \sum_{i=1}^{p-1} \frac{\mathbf{x}_p \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}\mathbf{v}_i
\end{aligned}
$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W and in addition

$$\text{span } \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \text{span } \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \quad \text{for } 1 \leq k \leq p$$