Slide 20-Orthogonality II

MAT2040 Linear Algebra

Definition 20.1 (**Direct sum**) If U and V are subspaces of a vector space W and each $\mathbf{w} \in W$ can be written uniquely as a sum $\mathbf{u} + \mathbf{v}$, where $\mathbf{u} \in U$ and $\mathbf{v} \in V$, then we say that W is a direct sum of U and V, and we write $W = U \oplus V$.

Theorem 20.2 (**Direct sum of $\mathbb{R}^n$**) If S is a subspace of $\mathbb{R}^n$, then

$$\mathbb{R}^n = S \oplus S^{\perp}$$

Proof. Skipped. See Steven's book P221 or the appendix.

Example 20.3 Let

$$U = \textbf{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}, V = \textbf{Span}\left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

be two subspaces of $\mathbb{R}^3$.

It can be easily checked that $U^\perp = V$ and $V^\perp = U$. Thus

$$\mathbb{R}^3 = U \oplus V.$$

Least Square Solution for the Linear System

Example 20.4 Solve the following linear system:

$$\begin{aligned} x + y &= 3, \\ -2x + 3y &= 1, \\ 2x - y &= 2 \end{aligned}$$

The augmented matrix reduced can be reduced into

$$\left[\begin{array}{cc|c} 1 & 1 & 3 \\ -2 & 3 & 1 \\ 2 & -1 & 2 \end{array}\right] \xrightarrow{elementary\ \ row\ \ operations} \left[\begin{array}{cc|c} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{array}\right]$$

The system is inconsistent, thus does not have a solution.

Question: How can we find a best approximation?

Given an inconsistent linear system $A\mathbf{x} = \mathbf{b}$ ($A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$), we can look at the vector $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is "closest" to $\mathbf{b}$ in the sense of Euclidean length, i.e. find $\hat{\mathbf{x}}$ such that $\| A\hat{\mathbf{x}} - \mathbf{b} \|$ is the smallest.

Definition 20.5 (Residual) For a given linear system $A\mathbf{x} = \mathbf{b}$ ($A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$), then for each $\mathbf{x} \in \mathbb{R}^n$, the residual is defined as

$$r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$$

Definition 20.6 (Least square solution) Given linear system $A\mathbf{x} = \mathbf{b}(A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m)$, a vector $\hat{\mathbf{x}}(\mathbf{x} \in \mathbb{R}^n)$ that satisfies the minimum residual condition

$$\| r(\hat{\mathbf{x}}) \| = \min_{\mathbf{x}} \| r(\mathbf{x}) \|$$

is called the least square solution for $A\mathbf{x} = \mathbf{b}$.

Theorem 20.7 (Projection onto a Subspace) Let S be a subspace of $\mathbb{R}^m$, for each $\mathbf{b} \in \mathbb{R}^m$, there exists a unique $\mathbf{p} \in S$ such that
(1) $\mathbf{b} - \mathbf{p} \in S^{\perp}$
(2) $\| \mathbf{b} - \mathbf{y} \| \geq \| \mathbf{b} - \mathbf{p} \|, \forall\ \mathbf{y} \in S$.
$\mathbf{p}$ is called the projection of $\mathbf{b}$ on the subspace S.

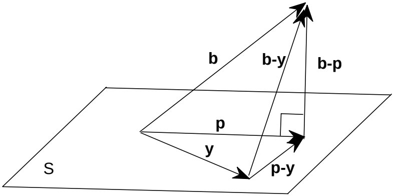

Figure: Projection of $\mathbf{b} \in \mathbb{R}^m$ onto subspace S.

Proof. Since $\mathbb{R}^m = S \oplus S^{\perp}$, each element $\mathbf{b} \in \mathbb{R}^m$ can be expressed uniquely as a sum

$$\mathbf{b} = \mathbf{p} + \mathbf{z}$$

where $\mathbf{p} \in S$ and $\mathbf{z} \in S^{\perp}$, thus $\mathbf{b} - \mathbf{p} \in S^{\perp}$.
Then, for any $\mathbf{y} \in S$, we have

$$\begin{aligned}
& \| \mathbf{b} - \mathbf{y} \|^2 \\
= & \| \mathbf{b} - \mathbf{p} + \mathbf{p} - \mathbf{y} \|^2 \\
= & \| \mathbf{b} - \mathbf{p} \|^2 + \| \mathbf{p} - \mathbf{y} \|^2 \\
\geq & \| \mathbf{b} - \mathbf{p} \|^2
\end{aligned}$$

since $\mathbf{b} - \mathbf{p} \in S^{\perp}$ and $\mathbf{p} - \mathbf{y} \in S$ $(\mathbf{p}, \mathbf{y} \in S)$, where the Pythagorean's Law is used.

Remark 1:
If $\mathbf{b} \in S$, then the projection of $\mathbf{p}$ onto S is just $\mathbf{b}$.

Remark 2: Let $A\mathbf{x} = \mathbf{b}$ ($A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$) is a linear system, then the residual $\mathbf{b} - A\mathbf{x}$ will reach its minimum when $\mathbf{x} = \hat{\mathbf{x}}$, where $\mathbf{A}\hat{\mathbf{x}}$ is the projection of $\mathbf{b}$ onto $Col(A)$ (the column space of A). Moreover,
$\mathbf{b} - A\hat{\mathbf{x}} \perp A\hat{\mathbf{x}} - A\mathbf{y} \in Col(A)$ and
$\| \mathbf{b} - A\mathbf{y} \|^2 = \| \mathbf{b} - A\hat{\mathbf{x}} \|^2 + \| A\hat{\mathbf{x}} - A\mathbf{y} \|^2 \geq \| \mathbf{b} - A\hat{\mathbf{x}} \|^2$

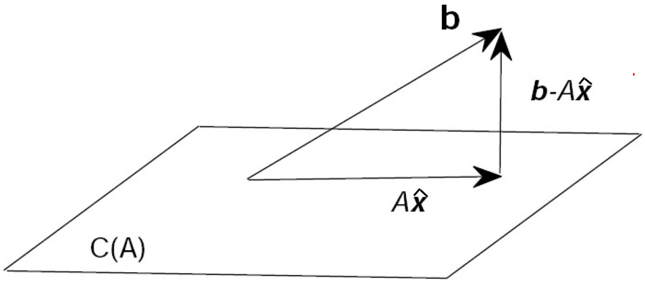

Figure: Projection of $\mathbf{b} \in V$ onto column space $Col(A)$.

Theorem 20.8 (**Normal equations for the linear system**) Given the linear system $A\mathbf{x} = \mathbf{b}$ ($A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$), let the projection of $\mathbf{b}$ onto the subspace $Col(A)$ is $\mathbf{p}$, then there exists a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$, s.t.
$\mathbf{p} = A\hat{\mathbf{x}} \in Col(A), \mathbf{b} - A\hat{\mathbf{x}} \in Col(A)^{\perp} = Null(A^T)$ and
$\| \mathbf{b} - A\mathbf{x} \| \geq \| \mathbf{b} - A\hat{\mathbf{x}} \|$ for any $\mathbf{x} \in \mathbb{R}^n$.
$\mathbf{b} - A\hat{\mathbf{x}} \in Col(A)^{\perp} = Null(A^T)$ gives the condition

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

i.e.,

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

which is called the **normal equation**, and it is a $n \times n$ linear system.

Remark
The normal equations may not have a unique solution, but the **projection vector p** of **b** onto $Col(A)$ is unique, i.e., there are possible two vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ satisfies $A\hat{\mathbf{x}} = A\hat{\mathbf{y}} = \mathbf{p}$.

Theorem 20.9 (**Unique Solution Condition for the Normal Equations**) If A is a $m \times n$ matrix of rank n, the normal equations

$$A^T A\hat{\mathbf{x}} = A^T \mathbf{b}$$

have a unique solution

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

and $\hat{\mathbf{x}}$ is the unique least square solution for the linear system $A\hat{\mathbf{x}} = \mathbf{b}$.

Proof. Only need to show that $A^T A$ is nonsingular. Only need to show that the linear system $A^T A\mathbf{x} = \mathbf{0}$ has only a trivial solution. Suppose $\mathbf{s}$ is the solution of $A^T A\mathbf{s} = \mathbf{0}$, then

$$A^T A\mathbf{s} = \mathbf{0}$$

Multiplying the above equation both sides from the left by $\mathbf{s}^T$, then one can reach

$$\mathbf{s}^T A^T A\mathbf{s} = 0$$

which means

$$(A\mathbf{s})^T A\mathbf{s} = \| A\mathbf{s} \|^2 = 0.$$

Thus

$$A\mathbf{s} = \mathbf{0}$$

Since the rank of A is n=the number of columns, the columns are linearly independent, thus the linear system $A\mathbf{s} = \mathbf{0}$ only has a trivial solution. Thus A^TA is nonsingular.

Therefore,

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

will be the unique solution for the normal equations. Consequently, $\hat{\mathbf{x}}$ is the unique least square solution for $A\mathbf{x} = \mathbf{b}$.

Then projection vector is given by $\mathbf{p} = A\hat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$ where $P = A(A^TA)^{-1}A^T$ is called the **projection matrix**.

Definition 20.10 (**Idempotent**) Let A be a square matrix that satisfies $A = A^2$, then A is called an idempotent matrix.

Remark. The projection matrix $P = A(A^TA)^{-1}A^T$ is an idempotent matrix.

It can be easily checked that
$P^2 = A(A^TA)^{-1}A^TA(A^TA)^{-1}A^T = A(A^TA)^{-1}A^T = P$

Example 20.11 Find the least square solution for the system:

$$\begin{aligned} x + y &= 3, \\ -2x + 3y &= 1, \\ 2x - y &= 2 \end{aligned}$$

where

$$A = \begin{bmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

The normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$ for this system are

$$\begin{bmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & -2 & 2 \\ 1 & 3 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}.$$

It can be simplified as

$$\begin{bmatrix} 9 & -7 \\ -7 & 11 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$$

Thus, the least square solution is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{83}{50} \\ \frac{71}{50} \end{bmatrix}$$

Applications of Least Square Solution

A collected data is usually trying to find a functional relation among variables. For example, the data may involve the temperature $T_1, \cdots, T_n$ of a liquid measured at times $t_1, \cdots, t_n$ respectively. If the temperature T can be represented by a function of time t, then one can use the function to predict the future temperature.

Applications of Least Square Solution

If the data set is as follows:

x	x_1	$\cdots$	x_n
y	y_1	$\cdots$	y_n

there are n data points, it is possible to find a polynomial of degree $n-1$ such that all the data satisfies the polynomial, such polynomial is called the **interpolation polynomial**. However, the data usually collected from the experiment involves experimental errors, it is **unreasonable** to require the function pass through all the points. In reality, finding a polynomial with lower-order degree is more reasonable and truer than the higher order polynomial passing all the points.

Applications of Least Square Solution: Example 1

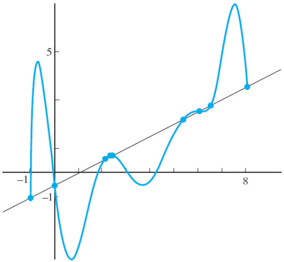

Figure: Line fitting

As one example shown in the above figure, the collected data roughly follows a linear relation, it is not good to find the interpolation polynomial, which will have oscillations (**Runge's phenomenon**).

Applications of Least Square Solution: Example 1

Instead find the interpolation polynomial, we are trying to find a linear function

$$y = c_0 + c_1 x$$

that best fits the data in the least square sense.
Now if we require

$$y_i = c_0 + c_1 x_i, \;\; i = 1, \cdots, n$$

then we get a linear system of n equations and two unknowns.

Applications of Least Square Solution: Example 1

The matrix-vector form is

$$A\mathbf{c} = \mathbf{y}$$

where $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

The above linear system may not have a solution. But we can find the least square solution $\hat{\mathbf{c}}$. The least square solution $\hat{\mathbf{c}} = [\hat{c}_0, \hat{c}_1]^T$ satisfies following property

$$\| r(\hat{\mathbf{c}}) \|^2 = \min_{\mathbf{c}} \| r(\mathbf{c}) \|^2 = \min_{\mathbf{c}} \| \mathbf{y} - A\mathbf{c} \|^2 = \min_{c_0, c_1} \sum_{i=1}^{n} (y_i - (c_0 + c_1 x_i))^2$$

The normal equations

$$A^T A\hat{\mathbf{c}} = A^T \mathbf{y}$$

will provide the least square solution for $\hat{\mathbf{c}} = [\hat{c}_0, \hat{c}_1]^T$. And this gives the best linear fitting function in the sense of least square.

Example Find the best line fitting to the data using the least square method

x	0	3	6
y	1	4	5

In this case, the linear system is $A\mathbf{c} = \mathbf{y}$, where $A = \begin{bmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}$,

$\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix}$.

The normal equations $A^T A\hat{\mathbf{c}} = A^T\mathbf{y}$ simplify into

$$\begin{bmatrix} 3 & 9 \\ 9 & 45 \end{bmatrix} \begin{bmatrix} \hat{c}_0 \\ \hat{c}_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 42 \end{bmatrix}$$

The solution is $\begin{bmatrix} \hat{c}_0 \\ \hat{c}_1 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} \\ \frac{2}{3} \end{bmatrix}$. Thus, the best fitted line in the sense of least square is

$$y = \frac{4}{3} + \frac{2}{3}x$$

Applications of Least Square Solution: Example 2

Example Find the best quadratic fitting to the data using the least square method

x	0	1	2	3
y	3	2	4	4

Let

$$y = c_0 + c_1 x + c_2 x^2$$

and requires

$$y_i = c_0 + c_1 x_i + c_2 x_i^2, \ \ i = 1, \cdots, 4$$

Then

$$A\mathbf{c} = \mathbf{y}$$

where $A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_3^2 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$

The normal equations are

$$A^T A\hat{\mathbf{c}} = A^T\mathbf{y},$$

which satisfies (Here $\hat{\mathbf{c}} = [\hat{c}_0, \hat{c}_1, \hat{c}_2]^T$)

$$\begin{aligned} \| r(\hat{\mathbf{c}}) \|^2 &= \min_{\mathbf{c}} \| r(\mathbf{c}) \|^2 \\ &= \min_{\mathbf{c}} \| \mathbf{y} - A\mathbf{c} \|^2 \\ &= \min_{c_0, c_1, c_2} \sum_{i=1}^{n} \left(y_i - (c_0 + c_1 x_i + c_2 x_i^2) \right)^2 \end{aligned}$$

Substituting the data into the above normal equation, one has

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} \hat{c}_0 \\ \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 4 \\ 4 \end{bmatrix}$$

These simplify into

$$\begin{bmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{bmatrix} \begin{bmatrix} \hat{c}_0 \\ \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} 13 \\ 22 \\ 54 \end{bmatrix}$$

The solution is $[2.75, -0.25, 0.25]^T$. Thus,

$$p(x) = 2.75 - 0.25x + 0.25x^2$$

is the best quadratic fitting polynomial in sense of the least square.

Appendix: Proof of Theorem 20.2

Theorem 20.2 (**Direct sum of $\mathbb{R}^n$**) If S is a subspace of $\mathbb{R}^n$, then

$$\mathbb{R}^n = S \oplus S^\perp$$

Recall: **Theorem 19.19**
If S is a subspace of $\mathbb{R}^n$, then $\dim S + \dim S^\perp = n$. Furthermore, if $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ is a basis for S and $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_n\}$ is a basis for $S^\perp$, then $\{\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

Proof. The result is trivial if $S = \{\mathbf{0}\}$ or $S = \{\mathbb{R}^n\}$. From theorem 19.19, one can see that each vector $\mathbf{x} \in \mathbb{R}^n$ can be uniquely expressed in the form:

Appendix: Proof of Theorem 20.2

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_r \mathbf{u}_r + \alpha_{r+1} \mathbf{u}_{r+1} + \cdots + \alpha_n \mathbf{u}_n$$

where $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ is a basis for S and $\{\mathbf{u}_{r+1}, \cdots, \mathbf{u}_n\}$ is a basis for $S^\perp$.

Let $\mathbf{u} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_r \mathbf{u}_r$ and $\mathbf{v} = \alpha_{r+1} \mathbf{u}_{r+1} + \cdots + \alpha_n \mathbf{u}_n$, then $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$, $\mathbf{x} = \mathbf{u} + \mathbf{v}$. To show the uniqueness, suppose that $\mathbf{x}$ can also be written as $\mathbf{x} = \mathbf{y} + \mathbf{z}$, where $y \in S$ and $z \in S^\perp$, then

$$\mathbf{u} + \mathbf{v} = \mathbf{y} + \mathbf{z}$$

and

$$\mathbf{u} - \mathbf{y} = \mathbf{z} - \mathbf{v}$$

LHS $\in S$ and RHS $\in S^\perp$ but $S \cap S^\perp = \{\mathbf{0}\}$ Thus

$$\mathbf{u} = \mathbf{y}, \quad \mathbf{z} = \mathbf{v}$$