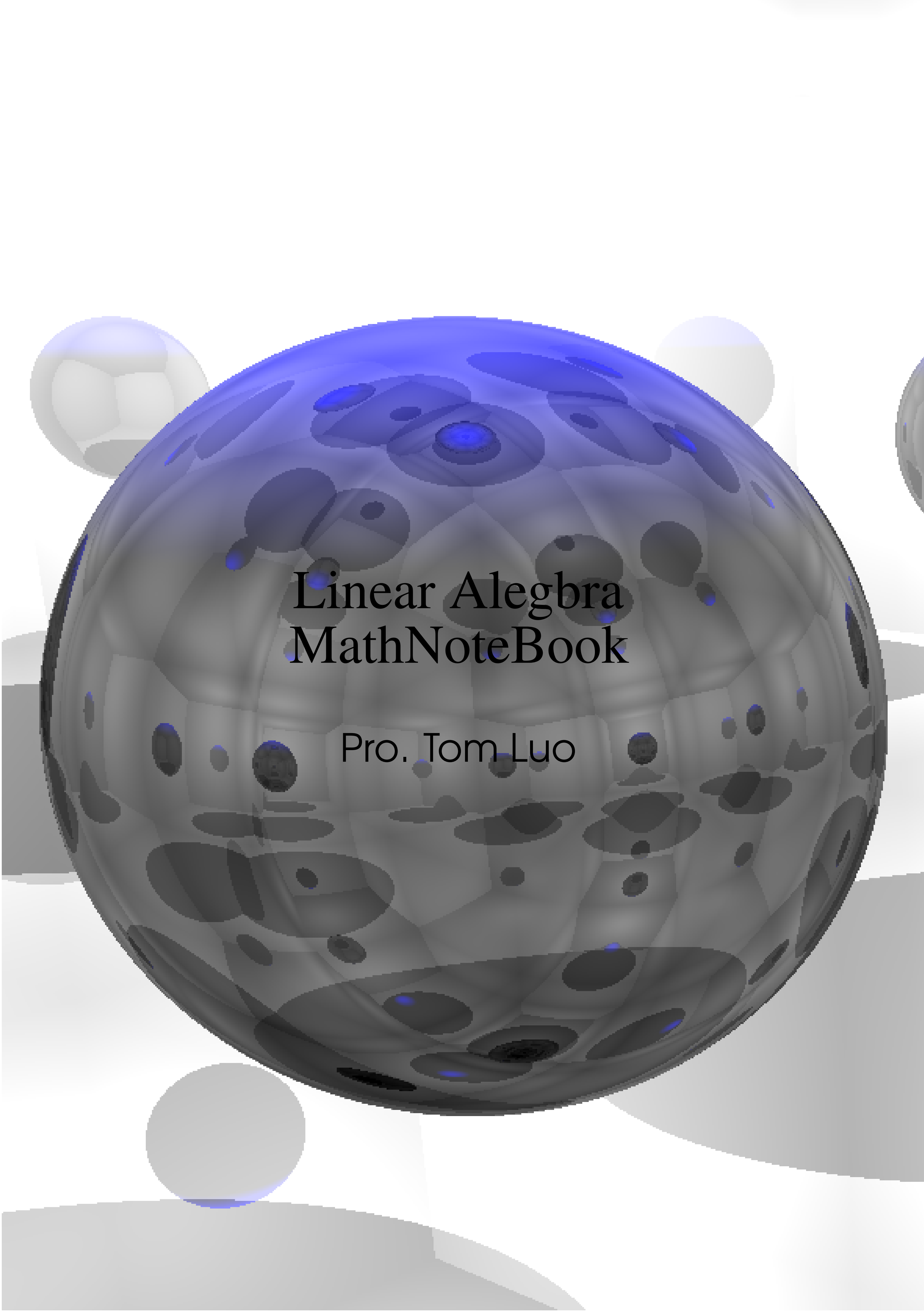

Linear Alegbra MathNoteBook

Pro. Tom Luo

12 — Week5

12.1 Thursday

12.1.1 *Orthogonality and Projection*

Two vectors are orthogonal if their inner product is zero:

$$\boldsymbol{u} \perp \boldsymbol{v} \iff \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0 \qquad (\text{if } \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n, \text{ then } \boldsymbol{u}^{\mathrm{T}} \boldsymbol{v} = 0.)$$

And orthogonality among vectors has an important property:

Proposition 12.1
If **nonzero** vectors $v_1, \dots, v_k$ are mutually orthogonal (mutulally means $v_i \perp v_j$ for any $i \neq j$), then $\{v_1, \dots, v_k\}$ must be ind.

Proof. We only need to show that

$$\text{if } \alpha_1 v_1 + \cdots + \alpha_k v_k = \mathbf{0}, \qquad \text{then } \alpha_i = 0 \text{ for any } i \in \{1, 2, \dots, k\}.$$

- We do inner product to show α_1 must be zero:

$$\begin{aligned}
\langle v_1, \alpha_1 v_1 + \cdots + \alpha_k v_k \rangle &= \langle v_1, \mathbf{0} \rangle = 0 \\
&= \alpha_1 \langle v_1, v_1 \rangle + \alpha_2 \langle v_1, v_2 \rangle + \cdots + \alpha_k \langle v_1, v_k \rangle \\
&= \alpha_1 \langle v_1, v_1 \rangle = \alpha_1 \|v_1\|_2^2 \\
&= 0
\end{aligned}$$

  Since $v_1 \neq \mathbf{0}$, we have $\alpha_1 = 0$.
- Similarly, we have $\alpha_i = 0$ for $i = 1, \dots, k$.

■

Now we can also talk about orthogonality among spaces:

Definition 12.1 — subspace orthogonality. Two subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$ of a vector space are **orthogonal** if every vector $\boldsymbol{u}$ in $\boldsymbol{U}$ is *perpendicular* to every vector $\boldsymbol{v}$ in $\boldsymbol{V}$:

$$\textbf{Orthogonal subspaces} \qquad \boldsymbol{u} \perp \boldsymbol{v} \quad \forall \boldsymbol{u} \in \boldsymbol{U}, \boldsymbol{v} \in \boldsymbol{V}.$$

■ **Example 12.1** Two walls look *perpendicular* but they are not orthogonal subspaces! The meeting line is in both $\boldsymbol{U}$ and $\boldsymbol{V}$-and this line is not perpendicular to itself. Two planes (dimensions 2 and 2 in $\mathbb{R}^3$) cannot be orthogonal subspaces.

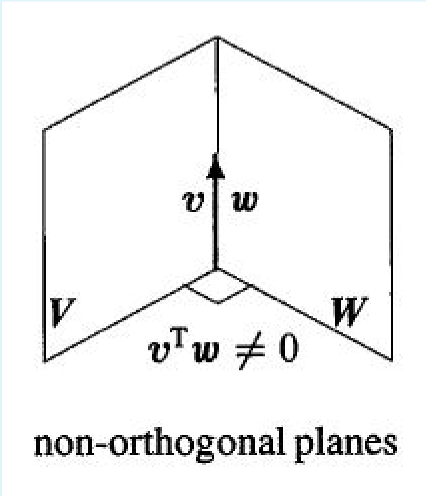

Figure 12.1: Orthogonality is impossible when $\dim \boldsymbol{U} + \dim \boldsymbol{V} > \dim(\boldsymbol{U} \cup \boldsymbol{V})$

■

R When a vector is in two orthogonal subspaces, it *must* be zero. It is **perpendicular** to itself.
The reason is clear: this vector $\boldsymbol{u} \in \boldsymbol{U}$ and $\boldsymbol{u} \in \boldsymbol{V}$, so $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$. It has to be zero vector.

If two subspaces are perpendicular, their basis must be ind.

Theorem 12.1 Assume $\{u_1, \dots, u_k\}$ is the basis for $\boldsymbol{U}$, $\{v_1, \dots, v_l\}$ is the basis for $\boldsymbol{V}$. If $\boldsymbol{U} \perp \boldsymbol{V}$ ($u_i \perp v_j$ for $\forall i, j$), then $u_1, u_2, \dots, u_k, v_1, v_2, \dots, v_l$ must be ind.

Proof. Suppose there exists $\{\alpha_1, \dots, \alpha_k\}$ and $\{\beta_1, \dots, \beta_l\}$ such that

$$\alpha_1 u_1 + \cdots + \alpha_k u_k + \beta_1 v_1 + \cdots + \beta_l v_l = \mathbf{0}$$

then equibalently,

$$\alpha_1 u_1 + \cdots + \alpha_k u_k = -(\beta_1 v_1 + \cdots + \beta_l v_l)$$

Then we set $\boldsymbol{w} = \alpha_1 u_1 + \cdots + \alpha_k u_k$, obviously, $\boldsymbol{w} \in \boldsymbol{U}$ and $\boldsymbol{w} \in \boldsymbol{V}$.
Hence it must be zero (This is due to remark above). Thus we have

$$\alpha_1 u_1 + \cdots + \alpha_k u_k = \mathbf{0}$$
$$\beta_1 v_1 + \cdots + \beta_l v_l = \mathbf{0}.$$

Due to the independence, we have $\alpha_i = 0$ and $\beta_j = 0$ for $\forall i, j$. ■

Corollary 12.1 If $\{u_1, u_2, \dots, u_k, v_1, v_2, \dots, v_l\} \in \boldsymbol{W}$, then $\dim(\boldsymbol{W}) \geq \dim(\boldsymbol{U}) + \dim(\boldsymbol{V})$. Note that $\boldsymbol{U} \cup \boldsymbol{V} \subset \boldsymbol{W}$.

For subspaces $\boldsymbol{U}$ and $\boldsymbol{V} \in \mathbb{R}^n$, if $\mathbb{R}^n = \boldsymbol{U} \cup \boldsymbol{V}$, and moreover, $n = \dim(\boldsymbol{U}) + \dim(\boldsymbol{V})$, then we say $\boldsymbol{V}$ is the **orthogonal complement** of $\boldsymbol{U}$.

Definition 12.2 — orthogonal complement. For subspaces $\boldsymbol{U}$ and $\boldsymbol{V} \in \mathbb{R}^n$, if $\dim(\boldsymbol{U}) + \dim(\boldsymbol{V}) = n$ and $\boldsymbol{U} \perp \boldsymbol{V}$, then we say $\boldsymbol{V}$ is the **orthogonal complement** of $\boldsymbol{U}$. And we denote $\boldsymbol{V}$ as $\boldsymbol{U}^{\perp}$.
Moreover, $\boldsymbol{V} = \boldsymbol{U}^{\perp} \iff \boldsymbol{V}^{\perp} = \boldsymbol{U}$.

Example 12.2 Suppose $\boldsymbol{U} \cup \boldsymbol{V} = \mathbb{R}^3$, $\boldsymbol{U} = \operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$. If $\boldsymbol{V}$ is the orthogonal complement of $\boldsymbol{U}$, then $\boldsymbol{V} = \operatorname{span}\{\boldsymbol{e}_3\}$.

Moreover, $\boldsymbol{U}$ could also be expressed as $\operatorname{span}\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$.

Example:
Next let's show the nullspace is the orthogonal complement of the row space. (In $\mathbb{R}^n$). Suppose $\boldsymbol{A}$ is a $m \times n$ matrix.

- Firstly, we show $\dim(N(\boldsymbol{A})) + \dim(C(\boldsymbol{A}^{\mathrm{T}})) = \dim(N(\boldsymbol{A}) \cup C(\boldsymbol{A}^{\mathrm{T}})) = \dim(\mathbb{R}^n) = n$:
  We know $\dim(N(\boldsymbol{A})) = n - r$, where $r = \operatorname{rank}(\boldsymbol{A})$. And $r = C(\boldsymbol{A}^{\mathrm{T}})$).
  Hence $\dim(N(\boldsymbol{A})) + \dim(C(\boldsymbol{A}^{\mathrm{T}})) = n$.
- Then we show $N(\boldsymbol{A}) \perp C(\boldsymbol{A}^{\mathrm{T}})$:
  For any $x \in N(\boldsymbol{A})$, if we set $\boldsymbol{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$, then we obtain:

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  Hence *every row has a zero product with* x. In other words, $\langle a_i, x \rangle = 0$ for $\forall i \in \{1, 2, \dots, m\}$. Hence for any $y = \sum_{i=1}^{m} \alpha_i a_i \in C(\boldsymbol{A}^{\mathrm{T}})$, we obtain:

$$\begin{aligned} \langle x, y \rangle = \langle y, x \rangle &= \langle \sum_{i=1}^{m} \alpha_i a_i, x \rangle \\ &= \sum_{i=1}^{m} \alpha_i \langle a_i, x \rangle = 0. \end{aligned}$$

  Hence $x \perp y$ for $\forall x \in N(\boldsymbol{A})$ and $y \in C(\boldsymbol{A}^{\mathrm{T}})$.

Hence $N(\boldsymbol{A})^{\perp} = C(\boldsymbol{A}^{\mathrm{T}})$.
If we applying this equation to $\boldsymbol{A}^{\mathrm{T}}$, then we have $N(\boldsymbol{A}^{\mathrm{T}})^{\perp} = C(\boldsymbol{A})$.

Theorem 12.2 — Fundamental theorem for linear alegbra, part 2.
$N(\boldsymbol{A})$ is the orthogonal complement of the row space $C(\boldsymbol{A}^{\mathrm{T}})$ (in $\mathbb{R}^n$).
$N(\boldsymbol{A}^{\mathrm{T}})$ is the orthogonal complement of the row space $C(\boldsymbol{A})$ (in $\mathbb{R}^m$).

Corollary 12.2 $\boldsymbol{Ax}=\boldsymbol{b}$ is solvable if and only if $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$ implies $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$=0.

Proof.
$\boldsymbol{Ax}=\boldsymbol{b}$ is solvable. $\iff \boldsymbol{b}\in C(\boldsymbol{A})$. $\iff \boldsymbol{b}\in N(\boldsymbol{A}^{\mathrm{T}})^{\perp}$
$\iff \boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}=0$ for $\forall y\in N(\boldsymbol{A}^{\mathrm{T}})$ $\iff \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$ implies $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$=0. ■

The Inverse Negative Propositions is more important:

Corollary 12.3 $\boldsymbol{Ax}=\boldsymbol{b}$ has no solution if and and only if $\exists\boldsymbol{y}$ s.t. $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=0$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}\neq 0$.

R

Theorem 12.3 $\boldsymbol{Ax}\geq\boldsymbol{b}$ has no solution if and only if $\exists\boldsymbol{y}\geq\boldsymbol{0}$ such that $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}\geq\boldsymbol{0}$.

$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=0$ requires exists one linear combination of the row space to be zero.

Necessity case. Suppose $\exists\boldsymbol{y}\geq\boldsymbol{0}$ such that $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}\geq\boldsymbol{0}$. And we assume there exists x^* such that $\boldsymbol{A}x^*\geq\boldsymbol{b}$. By postmultiplying $\boldsymbol{y}^{\mathrm{T}}$ we have

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}x^*\geq\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}>\boldsymbol{0}\implies\boldsymbol{0}>\boldsymbol{0}.$$

which is a contradiction! ■

The complete proof for this theorem is not required in this course.

■ **Example 12.3** Given the system

$$\begin{aligned} x_1+x_2&\geq 1\\ -x_1&\geq -1\\ -x_2&\geq 2 \end{aligned}\tag{12.1}$$

Eq(1)×1+Eq(2)×1+Eq(3)×1 gives

$$0\geq 2$$

which is a contradiction!
So the key idea of theorem (12.3) is to construct a linear combination of row space to let it become zero. Then if the right hand is larger than zero, then this system has no solution. ■

R

Corollary 12.4 If $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}$, then $N(\boldsymbol{A}^{\mathrm{T}})^{\perp}=C(A)=C(\boldsymbol{A}^{\mathrm{T}})=N(\boldsymbol{A})$.

Corollary 12.5 The system $\boldsymbol{Ax}=\boldsymbol{b}$ may not have a solution, but $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ always have at least one solution for $\forall\boldsymbol{b}$.

Proof. Since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is symmetric, we have $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=C(\boldsymbol{AA}^{\mathrm{T}})$. You can check by yourself that $C(\boldsymbol{AA}^{\mathrm{T}})=C(\boldsymbol{A}^{\mathrm{T}})$. Hence $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})=C(\boldsymbol{A}^{\mathrm{T}})$.

For any vector $\boldsymbol{b}$ we have $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \in C(\boldsymbol{A}^{\mathrm{T}}) \implies \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \in C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$, which means there exists a linear combination of the columns of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ that equals to $\boldsymbol{b}$.
Equivalently, there exists a solution to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$. ■

Corollary 12.6 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible if and only if columns of $\boldsymbol{A}$ are ind.

Proof. We have shown that $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = C(\boldsymbol{A}^{\mathrm{T}})$.
Hence $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{\perp} = C(\boldsymbol{A}^{\mathrm{T}})^{\perp} \implies N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = N(\boldsymbol{A})$.
$\boldsymbol{A}$ has ind. columns $\iff N(\boldsymbol{A}) = \{\boldsymbol{0}\} \iff N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \{\boldsymbol{0}\} \iff \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible. ■

12.1.2 Least Squares Approximations

$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ often has no solution, if so, what should we do?
We cannot always get the error $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$ down to zero, so we want to use least square method to minimize the error. In other words, our goal is to

$$\min_{\boldsymbol{x}} \boldsymbol{e}^2 = \min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \sum_{i=1}^{m} (a_i^{\mathrm{T}}\boldsymbol{x} - b_i)^2$$

where $\boldsymbol{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$.
The minimizer $\boldsymbol{x}$ is called **linear least squares solution**.

Matrix Calculus

Firstly, you should know some basic calculus knowledge for matrix:

- $\frac{\partial (f^{\mathrm{T}}g)}{\partial x} = \frac{\partial f(x)}{\partial x} g(x) + \frac{\partial g(x)}{\partial x} f(x)$

Example:

- $\frac{\partial (a^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = a$
- $\frac{\partial (a^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial ((\boldsymbol{A}^{\mathrm{T}}a)^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}a$
- $\frac{\partial (\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}$
- $\frac{\partial (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$

Thus, in order to minimize $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$, we only need to let its **partial derivative** with respect to $\boldsymbol{x}$ to be **zero.** (Since its second derivative is non-negative, we will talk about it in detail in other courses.) Hence we have

$$\begin{aligned}
\frac{\partial (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})}{\partial \boldsymbol{x}} &= \frac{\partial (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})}{\partial \boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \frac{\partial (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})}{\partial \boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = 2\frac{\partial (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})}{\partial \boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\
&= 2(\frac{\partial (\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} - \frac{\partial (\boldsymbol{b})}{\partial \boldsymbol{x}})(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\
&= 2\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}.
\end{aligned}$$

Or equivalently,

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

According to corollary (12.5), this equation always exists a solution. And this equation is called **normal equation**.

Theorem 12.4 The partial derivatives of $||\boldsymbol{Ax}-\boldsymbol{b}||^2$ are **zero** when $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Fit a stright line

Given a collection of data $(\boldsymbol{x}_i, y_i)$ for $i=1,\ldots,m$, we can fit the model parameters:

$$\begin{cases} y_1 = a_0 + a_1 x_{1,1} + a_2 x_{1,2} + \cdots + a_n x_{1,n} + \varepsilon_1 \\ y_2 = a_0 + a_1 x_{2,1} + a_2 x_{2,2} + \cdots + a_n x_{2,n} + \varepsilon_2 \\ \quad \vdots \\ y_m = a_0 + a_1 x_{m,1} + a_2 x_{m,2} + \cdots + a_n x_{m,n} + \varepsilon_m \end{cases}$$

Our fit line is

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

In *compact matrix form*, we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,n} \\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \ldots & x_{m,n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Or equivalently, we have

$$\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{A} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,n} \\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \ldots & x_{m,n} \end{bmatrix}_{m\times(n+1)}$, $\boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{(n+1)\times 1}$, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}_{m\times 1}$.

Our goal is to minimize $||\hat{\boldsymbol{y}}-\boldsymbol{y}||^2 = ||\boldsymbol{Ax}-\boldsymbol{y}||^2$. Then by theorem (12.4), we only need to sovle $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$.

12.1.3 Projeections

In corollary (12.6), we know that if $\boldsymbol{A}$ has ind. columns, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible. On this condition, the normal equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ has unique solution $\boldsymbol{x}^* = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.
Thus the error $\boldsymbol{b}-\boldsymbol{Ax}^*$ is minimum. And $\boldsymbol{Ax}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ **approximately** equals to $\boldsymbol{b}$.

- If $\boldsymbol{b}$ and $\boldsymbol{Ax}^*$ are exactly in the same space, then $\boldsymbol{Ax}^* = \boldsymbol{b}$.
- Otherwise, just as the Figure (12.2) shown, $\boldsymbol{Ax}^*$ is the projection of $\boldsymbol{b}$ to subspace $C(\boldsymbol{A})$.

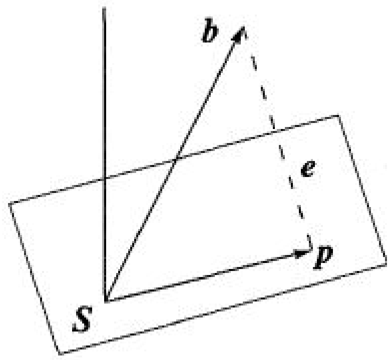

Figure 12.2: The projection of $\boldsymbol{b}$ onto a subspace $C(\boldsymbol{A})$.

Definition 12.3 — Projection. The projection of $\boldsymbol{b}$ onto the subspace $C(\boldsymbol{A})$ is denoted as $\text{Proj}_{C(\boldsymbol{A})}(\boldsymbol{b})$.

Definition 12.4 — Projection matrix. Given $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \text{Proj}_{C(\boldsymbol{A})}(\boldsymbol{b})$. Since $[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{b}$ is the projection of $\boldsymbol{b}$, we call $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ as **projection matrix**.

Definition 12.5 — Idempotent. Let $\boldsymbol{A}$ be a **square** matrix that satisfies $\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}$, then $\boldsymbol{A}$ is called a **idempotent** matrix.

Let's show the projection matrix is *idempotent*:

$$\begin{aligned}
\boldsymbol{P}^2 &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} \\
&= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} \\
&= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{P}.
\end{aligned}$$

Observations

- If $\boldsymbol{b} \in C(\boldsymbol{A})$, then $\exists \boldsymbol{x}$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Moreover, the projection of $\boldsymbol{b}$ is exactly $\boldsymbol{b}$:

$$\begin{aligned}
\boldsymbol{P}\boldsymbol{b} &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}) \\
&= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}) \\
&= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x} \\
&= \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.
\end{aligned}$$

- Assume $\boldsymbol{A}$ has only one column, say, $\boldsymbol{a}$. Then we have

$$\begin{aligned}
\boldsymbol{x}^* &= (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \\
\boldsymbol{A}\boldsymbol{x}^* &= \boldsymbol{P}\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}) = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \times \boldsymbol{a} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\|\boldsymbol{a}\|^2} \times \boldsymbol{a}
\end{aligned}$$

  More interestingly,

$$\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\|\boldsymbol{a}\|^2} \times \boldsymbol{a} = \frac{\|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\theta}{\|\boldsymbol{a}\|^2} \times \boldsymbol{a} = \|\boldsymbol{b}\|\cos\theta \times \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}$$

which is the projection of $\boldsymbol{b}$ onto a line $\boldsymbol{a}$. (Shown in figure below.)

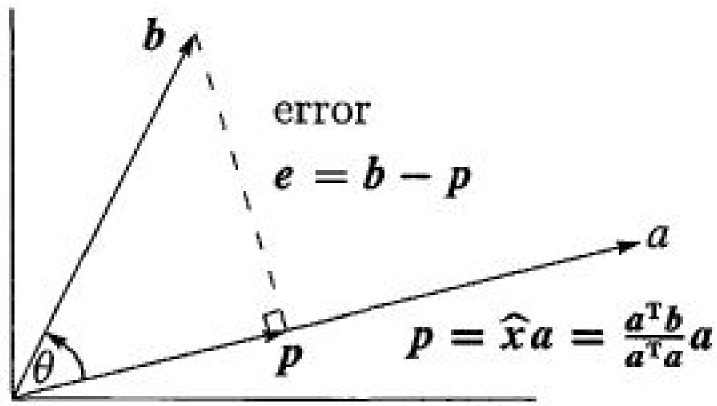

Figure 12.3: The projection of $\boldsymbol{b}$ onto a line $\boldsymbol{a}$.

More generally, we can write the projection of $\boldsymbol{b}$ as:

$$\text{Proj}_{\boldsymbol{a}}(\boldsymbol{b}) = \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle} \boldsymbol{a}$$

Look at the figure above! The error is $\boldsymbol{b} - \text{Proj}_{\boldsymbol{a}}(\boldsymbol{b})$, which is obviously perpendicular to $\boldsymbol{a}$. And $\boldsymbol{b} - \text{Proj}_{\boldsymbol{a}}(\boldsymbol{b}) \in \text{span}\{\boldsymbol{a}, \boldsymbol{b}\}$.
If we define $\boldsymbol{b}' = \boldsymbol{b} - \text{Proj}_{\boldsymbol{a}}(\boldsymbol{b})$, then it's easy to check $\text{span}\{\boldsymbol{a}, \boldsymbol{b}'\} = \text{span}\{\boldsymbol{a}, \boldsymbol{b}\}$ and $\boldsymbol{a} \perp \boldsymbol{b}'$. Hence we convert a basis to another basis such that the elements are orthogonal to each other. We will discuss it in detail in next lecture.