A FIRST COURSE

IN

LINEAR ALGEBRA

A FIRST COURSE

IN

LINEAR ALGEBRA

MAT2040 Notebook

Prof. Tom Luo

The Chinese University of Hongkong, Shenzhen

Prof. Ruoyu Sun

University of Illinois Urbana-Champaign

香港中文大學(深圳)

The Chinese University of Hong Kong, Shenzhen

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Survey Methodology / Robert M. Groves . . . [et al.].
p. cm.—(Wiley series in survey methodology)
"Wiley-Interscience."
Includes bibliographical references and index.
ISBN 0-471-48348-6 (pbk.)
1. Surveys—Methodology. 2. Social sciences—Research—Statistical methods. I. Groves, Robert M. II. Series.

HA31.2.S873 2004
001.4'33—dc22 2004044064
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Contributors

Zhi-quan Luo, Shenzhen Research Institute of Big Data, Lecturer

Ruoyu Sun, Industrial and Enterprise Systems Engineering, Lecturer

Jie Wang, The Chinese University of Hongkong, Shenzhen, Typer

Foreword

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Preface

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

place
date

Acknowledgments

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

I. R. S.

Acronyms

ASTA	Arrivals See Time Averages
BHCA	Busy Hour Call Attempts
BR	Bandwidth Reservation
b.u.	bandwidth unit(s)
CAC	Call / Connection Admission Control
CBP	Call Blocking Probability(-ies)
CCS	Centum Call Seconds
CDTM	Connection Dependent Threshold Model
CS	Complete Sharing
DiffServ	Differentiated Services
EMLM	Erlang Multirate Loss Model
erl	The Erlang unit of traffic-load
FIFO	First in - First out
GB	Global balance
GoS	Grade of Service
ICT	Information and Communication Technology
IntServ	Integrated Services
IP	Internet Protocol
ITU-T	International Telecommunication Unit – Standardization sector
LB	Local balance
LHS	Left hand side

LIFO	Last in - First out
MMPP	Markov Modulated Poisson Process
MPLS	Multiple Protocol Labeling Switching
MRM	Multi-Retry Model
MTM	Multi-Threshold Model
PASTA	Poisson Arrivals See Time Averages
PDF	Probability Distribution Function
pdf	probability density function
PFS	Product Form Solution
QoS	Quality of Service
r.v.	random variable(s)
RED	random early detection
RHS	Right hand side
RLA	Reduced Load Approximation
SIRO	service in random order
SRM	Single-Retry Model
STM	Single-Threshold Model
TCP	Transport Control Protocol
TH	Threshold(s)
UDP	User Datagram Protocol

6.2. Thursday

6.2.1. Orthogonality

Recall that two vectors are orthogonal if their inner product is zero:

$$\boldsymbol{u} \perp \boldsymbol{v} \iff \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$$

Orthogonality among vectors has an important property:

Proposition 6.1 If **nonzero** vectors $v_1, \dots, v_k$ are mutually orthogonal, i.e., $v_i \perp v_j$ for any $i \neq j$, then $\{v_1, \dots, v_k\}$ must be ind.

Proof. It suffices to show that

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

- 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 6.10 [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 $\boldsymbol{V}$:

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

■ **Example 6.3** 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. Hence, two planes (both with dimension 2 in $\mathbb{R}^3$) cannot be orthogonal subspaces.

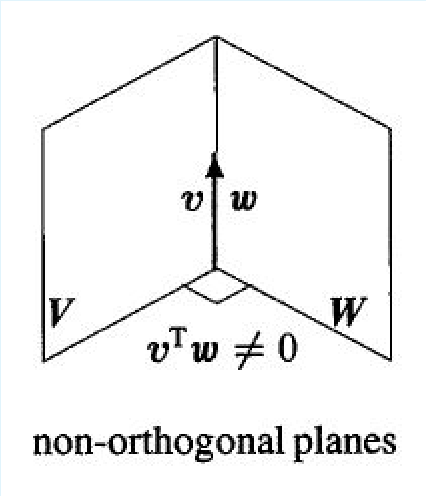

Figure 6.1: Orthogonality is impossible when $\dim \boldsymbol{U} + \dim \boldsymbol{V} > \dim(\boldsymbol{U} \cup \boldsymbol{V})$

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

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

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

then equivalently,

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

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 6.1 For subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$, we obtain

$$\dim(\boldsymbol{U} \cup \boldsymbol{V}) \leq \dim(\boldsymbol{U}) + \dim(\boldsymbol{V}).$$

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 6.11 [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 $\boldsymbol{U}$. We denote $\boldsymbol{V}$ as $\boldsymbol{U}^{\perp}$.

Moreover, $\boldsymbol{V} = \boldsymbol{U}^{\perp}$ iff $\boldsymbol{V}^{\perp} = \boldsymbol{U}$.

■ **Example 6.4** Suppose $\boldsymbol{U} \cup \boldsymbol{V} = \mathbb{R}^3$, $\boldsymbol{U} = \text{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$. If $\boldsymbol{V}$ is the orthogonal complement of $\boldsymbol{U}$, then $\boldsymbol{V} = \text{span}\{\boldsymbol{e}_3\}$.

Next we study the relationship between the null space and the row space in $\mathbb{R}^n$.

Theorem 6.3 — Fundamental theorem for linear alegbra, part 2. Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$,

$N(\boldsymbol{A})$ is the orthogonal complement of the row space of $\boldsymbol{A}$, $\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$ (in $\mathbb{R}^n$).

$N(\boldsymbol{A}^{\mathrm{T}})$ is the orthogonal complement of the column space $\mathcal{C}(\boldsymbol{A})$ (in $\mathbb{R}^m$).

Proof.

- Firstly, we show $\dim(N(\boldsymbol{A})) + \dim(\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})) = n$:

  We know that $\dim(N(\boldsymbol{A})) = n - r$ and $\dim(\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})) = r$, where $r = \operatorname{rank}(\boldsymbol{A})$.

  Hence $\dim(N(\boldsymbol{A})) + \dim(\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})) = n$.

- Then we show $N(\boldsymbol{A}) \perp \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$:

  For any $x \in N(\boldsymbol{A})$, if we set $\boldsymbol{A} = \begin{bmatrix} a_1^{\mathrm{T}} \\ a_2^{\mathrm{T}} \\ \vdots \\ a_m^{\mathrm{T}} \end{bmatrix}$, then we obtain:

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

  Hence *every row has a zero product with* $\boldsymbol{x}$, i.e., $\langle a_i, \boldsymbol{x} \rangle = 0$ for $\forall i \in \{1, 2, \dots, m\}$.

  For any $y = \sum_{i=1}^{m} \alpha_i a_i \in \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$, we obtain:

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

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

Hence $N(\boldsymbol{A})^{\perp} = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$. Similarly, we have $N(\boldsymbol{A}^{\mathrm{T}})^{\perp} = \mathcal{C}(\boldsymbol{A})$. ■

Corollary 6.2 $\quad \boldsymbol{Ax}=\boldsymbol{b}$ is solvable if and only if $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$ implies $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$=0.

Proof. The following statements are equivalent:

- $\boldsymbol{Ax}=\boldsymbol{b}$ is solvable.
- $\boldsymbol{b}\in\mathcal{C}(\boldsymbol{A})$.
- $\boldsymbol{b}\in N(\boldsymbol{A}^{\mathrm{T}})^{\perp}$
- $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}=0$ for $\forall y\in N(\boldsymbol{A}^{\mathrm{T}})$
- Given $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$, i.e., $y\in N(\boldsymbol{A}^{\mathrm{T}})$, it implies $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}=0$.

■

The **Inverse Negative Proposition** is more commonly useful:

Corollary 6.3 $\quad \boldsymbol{Ax}=\boldsymbol{b}$ has no solution if and and only if $\exists\boldsymbol{y}$ s.t. $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=0$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}\neq 0$.

We could extend this corollary into general case:

R

Theorem 6.4 $\quad \boldsymbol{Ax}\ge\boldsymbol{b}$ has no solution if and only if $\exists\boldsymbol{y}\ge\boldsymbol{0}$ such that $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{0}$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}\ge\boldsymbol{0}$.

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

The complete proof for this theorem is not required in this course. We only show the necessity case.

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

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

which is a contradiction! ■

■ **Example 6.5** Given the system

$$x_1 + x_2 \geq 1 \tag{6.3}$$

$$-x_1 \geq -1 \tag{6.4}$$

$$-x_2 \geq 2 \tag{6.5}$$

Eq.(6.3)×1+Eq(6.4)×1+Eq.(6.5)×1 gives

$$0 \geq 2$$

which is a contradiction!

So the key idea of theorem (6.4) is to construct a linear combination of row space to let it become zero. If the right hand is larger than zero, then this system has no solution. ■

R

Corollary 6.4 If $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$, then $N(\boldsymbol{A}^{\mathrm{T}})^{\perp} = \mathcal{C}(A) = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}}) = N(\boldsymbol{A})$.

Corollary 6.5 The system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ may not have a solution, but $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ always have at least one solution for $\forall \boldsymbol{b}$.

Proof. Since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is symmetric, we have $\mathcal{C}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$. Show by yourself that $\mathcal{C}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$, hence $\mathcal{C}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$.

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

Or equivalently, there exists a solution to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$. ■

Corollary 6.6 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible if and only if $\boldsymbol{A}$ is full column rank, i.e., columns of $\boldsymbol{A}$ are ind.

Proof. We have shown that $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = C(\boldsymbol{A}^{\mathrm{T}})$.

Hence $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{\perp} = C(\boldsymbol{A}^{\mathrm{T}})^{\perp} \implies N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = N(\boldsymbol{A})$.

Thus, the following statements are equivalent:

- $\boldsymbol{A}$ has ind. columns
- $N(\boldsymbol{A}) = \{\boldsymbol{0}\}$
- $N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \{\boldsymbol{0}\}$
- $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible.

■

6.2.2. Least Squares Approximations

The linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 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

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

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. The minimizer $\boldsymbol{x}$ is called the **linear least squares solution**.

6.2.2.1. Least Squares by Convex Optimization

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

The Chian Rule. Given two vectors $f(x)$,$g(x)$ of appropriate size,

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

Examples of Matrix Derivative.

$$\frac{\partial(a^{\mathrm{T}}\boldsymbol{x})}{\partial\boldsymbol{x}}=a \tag{6.6}$$

$$\frac{\partial(a^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}}=\frac{\partial((\boldsymbol{A}^{\mathrm{T}}a)^{\mathrm{T}}\boldsymbol{x})}{\partial\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}a \tag{6.7}$$

$$\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}} \tag{6.8}$$

$$\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \tag{6.9}$$

Thus, in order to minimize $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2=(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})$, it suffices to let its **derivative** with respect to $\boldsymbol{x}$ to be **zero.** (Since $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2$ is convex, which will be discussed 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 (6.5), this equation always exists a solution. This equation is called the **normal equation**.

Theorem 6.5 A vector $\boldsymbol{x}_{\mathrm{LS}}$ is an optimal solution to the least squares problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_2^2 \tag{6.10a}$$

if and only if it satisfies

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_{\mathrm{LS}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \tag{6.10b}$$

6.2.2.2. Fit a stright line

Given a collection of data $(\boldsymbol{x}_i, y_i)$ for $i = 1, \ldots, m$, we can use a stright line to fit these points:

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

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

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

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 $\|\hat{\boldsymbol{y}} - \boldsymbol{y}\|^2 = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2$. Then by theorem (6.5), it suffices to sovle $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$.

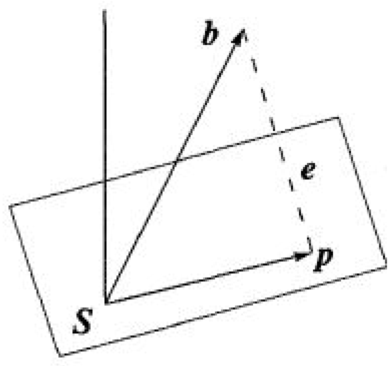

Figure 6.2: The projection of $\boldsymbol{b}$ onto a subspace $\boldsymbol{S} := \mathcal{C}(\boldsymbol{A})$.

6.2.3. Projections

In corollary (6.6), we know that if $\boldsymbol{A}$ has ind. columns, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible. On this condition, the normal equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ has the unique solution $\boldsymbol{x}^* = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$, which follows that the error $\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*$ is minimized. Note that $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ is **approximately** equal to $\boldsymbol{b}$.

- If $\boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{x}^*$ are exactly in the same space, i.e., $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$, then $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$. The error is equal to zero.
- Otherwise, just as the Figure (6.2) shown, $\boldsymbol{A}\boldsymbol{x}^*$ is the projection of $\boldsymbol{b}$ to subspace $\mathcal{C}(\boldsymbol{A})$.

Definition 6.12 [Projection] Let $\boldsymbol{S} \in \mathbb{R}^m$ be a non-empty closed set and $\boldsymbol{b} \in \mathbb{R}^m$ be given. Then the projection of $\boldsymbol{b}$ onto the set $\boldsymbol{S}$ is the solution to

$$\min_{\boldsymbol{z} \in \boldsymbol{S}} \|\boldsymbol{z} - \boldsymbol{b}\|_2^2,$$

where we use notation $\mathrm{Proj}_{\boldsymbol{S}}(\boldsymbol{b})$ to denote the projection of $\boldsymbol{b}$ onto $\boldsymbol{S}$. ■

By definition, the projection of $\boldsymbol{b}$ onto the subspace $\mathcal{C}(\boldsymbol{A})$ is given by

$$\mathrm{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b}) := \boldsymbol{A}\boldsymbol{x}^*, \quad \text{where } \boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|.$$

Definition 6.13 [Projection matrix] Given the projection

$$\text{Proj}_{\mathcal{C}(A)}(\boldsymbol{b}) := \boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b},$$

since $[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{b}$, we call the projection operator $\boldsymbol{P} := \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ as the **projection matrix** of $\boldsymbol{A}$. ■

Definition 6.14 [Idempotent] Let $\boldsymbol{A}$ be a **square** matrix that satisfies $\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}$, then $\boldsymbol{A}$ is called an **idempotent** matrix. ■

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

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

6.2.3.1. Observations

- Suppose $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$, i.e., $\exists \boldsymbol{x}$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Then 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 span$\{\boldsymbol{a}\}$. (Shown in figure (6.3).)

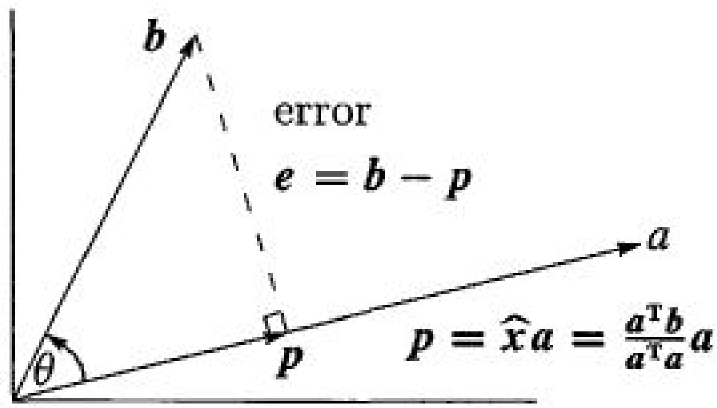

Figure 6.3: The projection of $\boldsymbol{b}$ onto a line $\boldsymbol{a}$.

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

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

Changing an Orthogonal Basis. Note that the error $\boldsymbol{b} - \mathrm{Proj}_{\mathrm{span}\{\boldsymbol{a}\}}(\boldsymbol{b})$ is perpendicular to $\boldsymbol{a}$, and $\boldsymbol{b} - \mathrm{Proj}_{\mathrm{span}\{\boldsymbol{a}\}}(\boldsymbol{b}) \in \mathrm{span}\{\boldsymbol{a},\boldsymbol{b}\}$.

If we define $\boldsymbol{b}' = \boldsymbol{b} - \mathrm{Proj}_{\mathrm{span}\{\boldsymbol{a}\}}(\boldsymbol{b})$, then it's easy to check that $\mathrm{span}\{\boldsymbol{a},\boldsymbol{b}'\} = \mathrm{span}\{\boldsymbol{a},\boldsymbol{b}\}$ and $\boldsymbol{a} \perp \boldsymbol{b}'$.

Hence, we convert the basis $\{\boldsymbol{a},\boldsymbol{b}\}$ into another basis $\{\boldsymbol{a},\boldsymbol{b}'\}$ such that the elements are orthogonal to each other. For general subspace we could also use this approach to obtain an orthogonal basis, which will be discussed in next lecture.