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

6.3. Friday

This lecture has two goals. The first is to see **how orthogonality makes it easy to find the projection matrix $\boldsymbol{P}$ and the projection** $\mathrm{Proj}_{\mathcal{C}(\boldsymbol{A})}\boldsymbol{b}$. The key idea is that *Orthogonality makes the product $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ a diagonal matrix*. The second goal is to **show how to construct orthogonal basis of $\mathcal{C}(\boldsymbol{A})$**. For matrix $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$, the columns may not be orthogonal. We intend to convert $a_1, \dots, a_n$ to orthogonal vectors, which will be the columns of a new matrix $\boldsymbol{Q}$.

6.3.1. Orthonormal basis

The vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ are **orthogonal** when their inner product $\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle$ are zero. ($i \neq j$.) With one more step–each vector is just divided by its length, then the collection of vectors become **orthogonal unit vectors**. Their lengths are all 1. Then this basis is called **orthonormal**.

Definition 6.15 [orthonormal] The collection of vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n \in \mathbb{R}^m$ is said to be:

- **orthogonal** if $\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = 0$ for all i, j with $i \neq j$
- **orthonormal** if $\|\boldsymbol{q}_i\|_2 = 1$ for all i and $\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = 0$ for all i, j with $i \neq j$, or equivalently,

$$\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = \begin{cases} 0 & \text{when } i \neq j \qquad (\textbf{orthogonal} \text{ vectors}), \\ 1 & \text{when } i = j \qquad (\textbf{unit} \text{ vectors: } \|\boldsymbol{q}_i\| = 1). \end{cases}$$

Moreover, if $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ are **orthonormal**, then the basis $\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}$ is called **orthonormal basis**. ■

■ **Example 6.6** Given a collection of unit vectors

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad \boldsymbol{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

then $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ forms an *orthonormal basis* for $\mathbb{R}^n$. ■

If we want to express vector $\boldsymbol{b}$ as the linear combination of arbitrary basis (may not be orthogonal) $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n\}$, what should we do?

Answer: Solve the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix}$

What if $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n\}$ is an **orthogonal** basis? How to find solution $\boldsymbol{x}$ s.t.

$$\boldsymbol{b} = x_1\boldsymbol{q}_1 + x_2\boldsymbol{q}_2 + \cdots + x_n\boldsymbol{q}_n? \tag{6.11}$$

Answer: We just do the inner product of each $\boldsymbol{q}_i$ with $\boldsymbol{b}$ to get the coefficient x_i:

$$\begin{aligned} \langle \boldsymbol{q}_i, \boldsymbol{b} \rangle &= x_1\langle \boldsymbol{q}_i, \boldsymbol{q}_1 \rangle + x_2\langle \boldsymbol{q}_i, \boldsymbol{q}_2 \rangle + \cdots + x_n\langle \boldsymbol{q}_i, \boldsymbol{q}_n \rangle \\ &= x_i\langle \boldsymbol{q}_i, \boldsymbol{q}_i \rangle = x_i \end{aligned} \tag{6.12}$$

By substituting Eq.(6.12) into Eq.(6.11), we could express $\boldsymbol{b}$ as:

$$\boldsymbol{b} = \sum_{i=1}^{n} \langle \boldsymbol{q}_i, \boldsymbol{b} \rangle \boldsymbol{q}_i.$$

In this case, from Eq.(6.12) we can see that if columns of $\boldsymbol{A}$ are orthogonal, we could easily obtain the solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:

$$x_i = \langle \boldsymbol{q}_i, \boldsymbol{b} \rangle, \quad i = 1, 2, \dots, n.$$

Definition 6.16 [matrix with orthonormal columns] Given a collection of **orthonormal** vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$, the matrix

$$\boldsymbol{Q} = \begin{bmatrix} q_1 & q_2 & \ldots & q_n \end{bmatrix}$$

is said to be a matrix with **orthonormal** columns.

Note that a matrix with **orthonormal** columns is often denoted as $\boldsymbol{Q}$.

Or equivalently, a matrix $\boldsymbol{Q}$ is with **orthonormal** columns if and only if

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \\ \cdots \\ \boldsymbol{q}_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \ldots & \boldsymbol{q}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_1 & & \\ & \ddots & \\ & & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{q}_n \end{pmatrix} = \boldsymbol{I}. \tag{6.13}$$

R Note that a matrix $\boldsymbol{Q}$ with orthonormal columns is *not required to be square*! Moreover, $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ in $\boldsymbol{Q}$ is *not required to form a basis.*

Definition 6.17 [orthogonal matrix] A matrix $\boldsymbol{Q}$ is said to be **orthogonal** if it is square and its columns are orthonormal.

Question: Why we call it an orthogonal matrix, but not an orthonormal matrix?

Answer: Orthogonal matrix usually transform an orthogonal basis into another orthogonal basis by matrix multiplication. This special property requires its column to be **orthonormal**.

■ **Example 6.7** If $\boldsymbol{Q}$ is an orthogonal matrix, while $\hat{\boldsymbol{Q}}$ is a matrix with orthonormal columns that is **not square**. Do the products $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$ and $\hat{\boldsymbol{Q}}\hat{\boldsymbol{Q}}^{\mathrm{T}}$ always be *identity matrix*?

Answer:

- $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$ is always *identity matrix*. According to equation (6.13), we have $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$.

Hence $\boldsymbol{Q}^{\mathrm{T}}$ is the left inverse of square matrix $\boldsymbol{Q}$, which implies

$$\boldsymbol{Q}^{-1} = \boldsymbol{Q}^{\mathrm{T}} \implies \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{Q}^{-1} = \boldsymbol{I}.$$

Moreover, solving $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ is equivalent to $\boldsymbol{x} = \boldsymbol{Q}^{-1}\boldsymbol{b} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$, which is *exactly*

$$\boldsymbol{x} = \begin{bmatrix} \langle \boldsymbol{q}_1, \boldsymbol{b} \rangle \\ \langle \boldsymbol{q}_2, \boldsymbol{b} \rangle \\ \vdots \\ \langle \boldsymbol{q}_n, \boldsymbol{b} \rangle \end{bmatrix}.$$

- Although $\hat{Q}^{\mathrm{T}}\hat{Q} = \boldsymbol{I}$, the product $\hat{Q}\hat{Q}^{\mathrm{T}}$ will never be identity matrix for nonsquare $\hat{Q}$. We can verify it by the its rank:

  Assume $\hat{Q} \in \mathbb{R}^{m \times n} (m \neq n)$. Then it's easy to verify that $\operatorname{rank}(\hat{Q}\hat{Q}^{\mathrm{T}}) = \operatorname{rank}(\hat{Q})$. Since $\hat{Q}$ has orthonormal columns, the columns of $\hat{Q}$ are independent, i.e., $\operatorname{rank}(\hat{Q}) = n$. But $\operatorname{rank}(\hat{Q}\hat{Q}^{\mathrm{T}}) = \operatorname{rank}(\hat{Q}) = n \neq m = \operatorname{rank}(\boldsymbol{I}_m)$.

  Moreover, if $\hat{Q}$ has only one column $\hat{q}$, then $\hat{Q}\hat{Q}^{\mathrm{T}} = \hat{q}\hat{q}^{\mathrm{T}} = \operatorname{rank}(1) \neq \operatorname{rank}(\boldsymbol{I}_m)$.

Proposition 6.2

If $\boldsymbol{Q}$ has orthonormal columns, then it *leaves lengths unchanged*, in other words,

$$\textbf{Same length} \quad \|\boldsymbol{Q}\boldsymbol{x}\| = \|\boldsymbol{x}\| \text{ for every vector } \boldsymbol{x}.$$

Also, $\boldsymbol{Q}$ preserves inner products for vectors, i.e., :

$$\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \quad \text{for every vectors } \boldsymbol{x} \text{ and } \boldsymbol{y}.$$

Proofoutline. $\|\boldsymbol{Q}\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$ because

$$\begin{aligned} \langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{x} \rangle &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})\boldsymbol{x} \\ &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \end{aligned}$$

Hence we have $\|\boldsymbol{Qx}\| = \|\boldsymbol{x}\|$. Just using $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$, we can derive $\langle \boldsymbol{Qx}, \boldsymbol{Qy} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$. ■

Orthogonal matrices are excellent for computations, since the inverse of matrices could usually be converted into transpose.

When Least Squares Meet Orthogonality. In particular, if $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ has orthonormal columns, the least square problem is easy:

Although $\boldsymbol{Qx} = \boldsymbol{b}$ may not have a solution, but the normal equation

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$$

must have the unique solution $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$. Why? Since $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$, we derive

$$\hat{x} = (\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

6.3.1.1. Summary

Hence the **least squares solution** to $\boldsymbol{Qx} = \boldsymbol{b}$ is $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$. In other words, $\boldsymbol{QQ}^{\mathrm{T}}\boldsymbol{b} \approx \boldsymbol{b}$. **The projection matrix is $\boldsymbol{P} = \boldsymbol{QQ}^{\mathrm{T}}$.** Note that the projection $\mathrm{Proj}_{C(\boldsymbol{Q})}(\boldsymbol{b}) = \boldsymbol{QQ}^{\mathrm{T}}\boldsymbol{b}$ doesn't equal to $\boldsymbol{b}$ in general.

For general matrix $\boldsymbol{A}$, the projection matrix is more complicated:

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}.$$

6.3.2. Gram-Schmidt Process

"Orthogonal is good". So our goal for this section is: *Given a collection of independent vectors, how to make them orthonormal?*

We start with three independent vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ in $\mathbb{R}^3$. In order to construct orthonormal vectors, firstly we construct three **orthogonal** vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Secongly we divide $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ by their lengths to get three **orthonormal** vectors $\boldsymbol{q}_1 = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|}, \boldsymbol{q}_2 = \frac{\boldsymbol{B}}{\|\boldsymbol{B}\|}, \boldsymbol{q}_3 = \frac{\boldsymbol{C}}{\|\boldsymbol{C}\|}$.

- Firstly we set $\boldsymbol{A} = \boldsymbol{a}$.

- The next vector $\boldsymbol{B}$ must be perpendicular to $\boldsymbol{A}$. Look at the figure (6.4) below, We find that $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}}(\boldsymbol{b})$. Or equivalently,

$$\textbf{First Gram-Schmidt step} \qquad \boldsymbol{B} = \boldsymbol{b} - \frac{\langle \boldsymbol{A}, \boldsymbol{b} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A}.$$

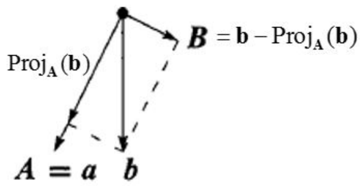

Figure 6.4: Subtract projection to get $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}}\, \boldsymbol{b}$.

You can take inner product between $\boldsymbol{A}$ and $\boldsymbol{B}$ to verify that $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal in Figure (6.4). Note that $\boldsymbol{B}$ is not zero (otherwise $\boldsymbol{a}$ and $\boldsymbol{b}$ would be dependent. We will show it later.)

- Then we want to construct another vector $\boldsymbol{C}$. Most likely $\boldsymbol{c}$ is **not** perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. What we do is to **subtract $\boldsymbol{c}$ off its projections onto the column space of $\boldsymbol{A}$ and $\boldsymbol{B}$ to get $\boldsymbol{C}$:**

$$\textbf{Next Gram-Schmidt step} \qquad \begin{aligned} \boldsymbol{C} &= \boldsymbol{c} - \text{Proj}_{\text{span}\{\boldsymbol{A},\boldsymbol{B}\}}(\boldsymbol{c}) \\ &= \boldsymbol{c} - \text{Proj}_{\boldsymbol{A}}(\boldsymbol{c}) - \text{Proj}_{\boldsymbol{B}}(\boldsymbol{c}) \\ &= \boldsymbol{c} - \frac{\langle \boldsymbol{A}, \boldsymbol{c} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A} - \frac{\langle \boldsymbol{B}, \boldsymbol{c} \rangle}{\langle \boldsymbol{B}, \boldsymbol{B} \rangle} \boldsymbol{B}. \end{aligned}$$

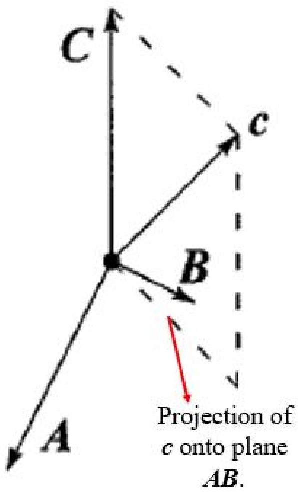

Figure 6.5: Subtract $\boldsymbol{c}$ off its projections onto the column space of $\boldsymbol{A}$ and $\boldsymbol{B}$ to get $\boldsymbol{C}$

- Finally we get orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ are obtained by dividing their lengths (shown in Figure (6.6)):

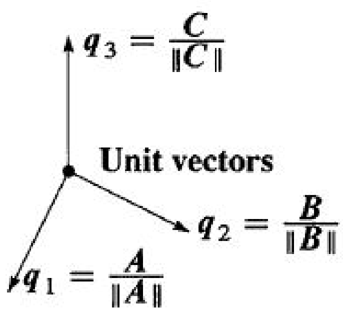

Figure 6.6: Final Gram-Schmidt step

Next we show an example of Gram-Schmidt step:

■ **Example 6.8** How to construct orthonormal vectors from

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}?$$

- Firstly we set $\boldsymbol{A} = \boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

- $$\begin{aligned} \boldsymbol{B} &= \boldsymbol{b} - \mathrm{Proj}_{\boldsymbol{A}}(\boldsymbol{b}) = \boldsymbol{b} - \frac{\langle \boldsymbol{A}, \boldsymbol{b} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle}\boldsymbol{A} \\ &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} 2^{-1} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{pmatrix} \end{aligned}$$

- $$\begin{aligned} \boldsymbol{C} &= \boldsymbol{c} - \mathrm{Proj}_{\boldsymbol{A}}(\boldsymbol{c}) - \mathrm{Proj}_{\boldsymbol{B}}(\boldsymbol{c}) = \boldsymbol{c} - \frac{\langle \boldsymbol{A}, \boldsymbol{c} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle}\boldsymbol{A} - \frac{\langle \boldsymbol{B}, \boldsymbol{c} \rangle}{\langle \boldsymbol{B}, \boldsymbol{B} \rangle}\boldsymbol{B} \\ &= \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} 2^{-1} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} (\frac{1}{2})^{-1} \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \end{aligned}$$

  Hence we obtain our orthonormal vectors:

$$\boldsymbol{q}_1 = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, , \boldsymbol{q}_2 = \frac{\boldsymbol{B}}{\|\boldsymbol{B}\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \boldsymbol{q}_3 = \frac{\boldsymbol{C}}{\|\boldsymbol{C}\|} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

And we derive the orthogonal matrix $\boldsymbol{Q}$:

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

When will the Gram-Schmidt process "fail"? Let's describe this process in general case first, then we answer this question.

6.3.2.1. Gram-Schmidt process in general case

Algorithm: Gram-Schmidt Process

Input: a collection of vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$, presumably linear independent.

Firstly construct orthogonal vectors $\boldsymbol{A}_1, \dots, \boldsymbol{A}_n$.

$\boldsymbol{A}_1 = \boldsymbol{a}_1$.

To construct $\boldsymbol{A}_j$, $j \in \{2, \dots, n\}$, we compute $\boldsymbol{a}_j$ minus its projection in the column space spanned by $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_{j-1}\}$:

$$\begin{aligned} \boldsymbol{A}_j &= \boldsymbol{a}_j - \text{Proj}_{\text{span}\{\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_{j-1}\}}(\boldsymbol{a}_j) \\ &= \boldsymbol{a}_j - \text{Proj}_{\boldsymbol{A}_1}(\boldsymbol{a}_j) - \text{Proj}_{\boldsymbol{A}_2}(\boldsymbol{a}_j) - \cdots - \text{Proj}_{\boldsymbol{A}_{j-1}}(\boldsymbol{a}_j) \\ &= \boldsymbol{a}_j - \frac{\langle \boldsymbol{A}_1, \boldsymbol{a}_j \rangle}{\langle \boldsymbol{A}_1, \boldsymbol{A}_1 \rangle} \boldsymbol{A}_1 - \frac{\langle \boldsymbol{A}_2, \boldsymbol{a}_j \rangle}{\langle \boldsymbol{A}_2, \boldsymbol{A}_2 \rangle} \boldsymbol{A}_2 - \cdots - \frac{\langle \boldsymbol{A}_{j-1}, \boldsymbol{a}_j \rangle}{\langle \boldsymbol{A}_{j-1}, \boldsymbol{A}_{j-1} \rangle} \boldsymbol{A}_{j-1} \end{aligned}$$

Secondly, after getting $\boldsymbol{A}_1, \dots, \boldsymbol{A}_n$, we can construct orthonormal vectors:

$$\boldsymbol{q}_j = \frac{\boldsymbol{A}_j}{\|\boldsymbol{A}_j\|} \quad \text{for } j = 1, 2, \dots, n.$$

So when do this process fail? When $\exists j$ such that $\boldsymbol{A}_j = \boldsymbol{0}$, we cannot continue this process anymore:

Proposition 6.3 $\boldsymbol{A}_j \neq \boldsymbol{0}$ for $\forall j$ if and only if $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are indendent.

Proofoutline. $\boldsymbol{A}_j = \boldsymbol{0} \iff \boldsymbol{a}_j = \text{Proj}_{\text{span}\{\boldsymbol{A}_1, \dots, \boldsymbol{A}_{j-1}\}}(\boldsymbol{a}_j)$. It suffices to prove $\exists j$ s.t. $\boldsymbol{A}_j = \boldsymbol{0}$ if and only if $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are depependent.

Sufficiency. Given $\boldsymbol{A}_j = \mathbf{0}$, then $\boldsymbol{a}_j = \text{Proj}_{\text{span}\,\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}}(\boldsymbol{a}_j) \in \text{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}\}$. It's easy to verify that $\text{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}\} = \text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{j-1}\}$. Hence $\boldsymbol{a}_j \in \text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{j-1}\}$. Hence $\boldsymbol{a}_1,\dots,\boldsymbol{a}_j$ are dependent. Thus $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$ are dependent.

Necessity. Given dependent vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$, obviously, $\boldsymbol{a}_n \in \text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{n-1}\}$. It's easy to verify that $\boldsymbol{a}_n = \text{Proj}_{\text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{n-1}\}}(\boldsymbol{a}_n)$. Thus $\boldsymbol{a}_n = \text{Proj}_{\text{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{n-1}\}}(\boldsymbol{a}_n) \implies \boldsymbol{A}_n = \mathbf{0}$. ■

6.3.3. The Factorization $\boldsymbol{A} = \boldsymbol{QR}$

We know that Gaussian Elimination leads to *LU decomposition*; in fact, Gram-Schmidt process leads to *QR factorization*. These two decomposition methods are quite important in Linear Algebra, let's discuss QR factorization briefly:

Given a matrix $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix}$, we finally end with a matrix $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix}$. How are these two matrices related?

Answer: Since the linear combination of $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ leads to $\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3$ (vice versa), there must be a third matrix connecting $\boldsymbol{A}$ to $\boldsymbol{Q}$. This third matrix is the triangular $\boldsymbol{R}$ such taht $\boldsymbol{A} = \boldsymbol{QR}$.

Let's discuss a specific example to show how to do QR factorization.

■ **Example 6.9** Given $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix}$, whose columns are independent, then we can use Gram-Schmidt process to obtain the corresponding orthonormal vectors $\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3$ from $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$. As a result, we can write $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ 0 & 0 & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix}$$

We define $\boldsymbol{R} \triangleq \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ 0 & 0 & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix}, \boldsymbol{Q} \triangleq \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix}$.
Hence $\boldsymbol{A}$ could be factorized into:

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

where $\boldsymbol{R}$ is upper triangular, $\boldsymbol{Q}$ is a matrix with orthonormal columns.

QR factorization holds for every matrix with independent columns:

Theorem 6.6 Every $m \times n$ matrix $\boldsymbol{A}$ with ind. columns can be factorized as

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

where $\boldsymbol{Q}$ is a matrix with *orthonormal columns*, $\boldsymbol{R}$ is an upper triangular matrix (always square).

We omit the proof of this theorem. Now we show that the inverse of $\boldsymbol{R}$ always exists:

Proof. suppose $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{bmatrix}$. Thus we derive

$$\boldsymbol{R} = \boldsymbol{Q}^{-1}\boldsymbol{A} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_1 & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_2 & \dots & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_n \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{a}_2 & \dots & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{a}_n \end{bmatrix}$$

For every step j we have

$$\boldsymbol{A}_j = \boldsymbol{a}_j - \mathrm{Proj}_{\mathrm{span}\{a_1,\dots,a_{j-1}\}}(\boldsymbol{a}_j), \qquad \boldsymbol{q}_j = \frac{\boldsymbol{A}_j}{\|\boldsymbol{A}_j\|}.$$

Since $\langle \boldsymbol{A}_j, \boldsymbol{a}_j \rangle = \langle \boldsymbol{a}_j, \boldsymbol{a}_j \rangle - \langle \mathrm{Proj}_{\mathrm{span}\{a_1,\dots,a_{j-1}\}}(\boldsymbol{a}_j), \boldsymbol{a}_j \rangle = \|a_j\|^2 - \|\mathrm{Proj}_{\mathrm{span}\{a_1,\dots,a_{j-1}\}}(\boldsymbol{a}_j)\|^2 > 0$, we have $\langle \boldsymbol{q}_j, \boldsymbol{a}_j \rangle = \frac{\langle \boldsymbol{A}_j, \boldsymbol{a}_j \rangle}{\|\boldsymbol{A}_j\|} > 0$. Hence the diagonal of $\boldsymbol{R}$ are all positive. Hence this triangular matrix is *invertible*. ■

Proposition 6.4 If $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, then the least squares solution is given by:

$$\boldsymbol{x} = (\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

Explain: Since we have

$$\begin{aligned}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} \\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} &= \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}\end{aligned}$$

it's equivalent to solve $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$.

Sicne $\boldsymbol{R}$ is *invertible*, we solve by back substitution to get

$$\boldsymbol{x} = (\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

■

6.3.4. Function Space

Sometimes we may also discuss orthonormal basis and Gram-Schmidt process on function space. There is a simple example:

■ **Example 6.10** For subspace span$\{1, x, x^2\} \subset \mathcal{C}[-1,1]$, firstly, how to define orthogonal for the basis $\{1, x, x^2\}$?

Pre-requisite Knowledge: Inner product.

$$\langle f, g\rangle = \int_a^b fg \, \mathrm{d}x \text{ for } f, g \in C[a,b]. \qquad \|f\|^2 = \int_a^b f^2 \, \mathrm{d}x$$

If we have defined inner product, then we can talk about *orthogonality* for $\{1, x, x^2\}$. It's easy to verify that

$$\langle 1, x\rangle = 0 \quad \langle x, x^2\rangle = 0 \quad \langle 1, x^2\rangle = \frac{2}{3}.$$

If we do the Gram-Schmidt Process similarly, we obtain:

$$\boldsymbol{A}=1,\qquad \boldsymbol{B}=x,\qquad \boldsymbol{C}=x^2-\frac{\langle 1,x^2\rangle}{\langle 1,1\rangle}1-\frac{\langle x,x^2\rangle}{\langle x,x\rangle}x=x^2-\frac{1}{3}$$

where $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$ are *orthogonal*. We can divide their length to obtain orthonormal basis:

$$\boldsymbol{q}_1=\frac{\boldsymbol{A}}{\|\boldsymbol{A}\|}=\frac{1}{\sqrt{\int_{-1}^{1}1^2\,\mathrm{d}x}}=\frac{1}{2}$$
$$\boldsymbol{q}_2=\frac{\boldsymbol{B}}{\|\boldsymbol{B}\|}=\frac{x}{\sqrt{\int_{-1}^{1}x^2\,\mathrm{d}x}}=\frac{x}{2/3}=\frac{3}{2}x$$
$$\boldsymbol{q}_3=\frac{\boldsymbol{C}}{\|\boldsymbol{C}\|}=\frac{x^2-\frac{1}{3}}{\sqrt{\int_{-1}^{1}(x^2-\frac{1}{3})^2\,\mathrm{d}x}}=\frac{x^2-\frac{1}{3}}{\frac{8}{45}}=\frac{45x^2-15}{8}$$

Hence $\{\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3\}$ is the orthonormal basis for $\{1,x,x^2\}$. ■

■ **Example 6.11** Consider the collection $\mathcal{F}$ of functions defined on $[0,2\pi]$, where

$$\mathcal{F}:=\{1,\cos x,\sin x,\cos 2x,\sin 2x,\dots,\cos mx,\sin mx,\dots\}$$

Using various trigonometric identities, we can show that if f and g are **distinct**(different) functions in $\mathcal{F}$, we have $\int_0^{2\pi}fg\,\mathrm{d}x=0$. For example,

$$\langle\sin x,\sin 2x\rangle=\int_0^{2\pi}\sin x\sin 2x\,\mathrm{d}x=\int_0^{2\pi}\frac{1}{2}(\cos x-\cos 3x)\,\mathrm{d}x=0.$$

And moreover, if $f=g$, we have $\int_0^{2\pi}f^2\,\mathrm{d}x=\pi$. For example,

$$\langle\sin 5x,\sin 5x\rangle=\int_0^{2\pi}\sin^2 5x\,\mathrm{d}x=\int_0^{2\pi}\frac{1}{2}(1+\cos 10x)\,\mathrm{d}x=\pi.$$

In conclusion, the collection of functions $\{1,\sin mx,\cos mx\}$ for $k=1,2,\dots$ are *orthogonal* in $\mathcal{C}[0,2\pi]$. Note that this set is **not orthonormal**. ■

This example gives a motivation of the fourier transformation:

6.3.5. Fourier Series

Since we have shown the orthogonality of $\mathcal{F}$ in Example.(6.11), our question is that what kind of function can be written as the linear combination of functions from $\mathcal{F}$.

The Fourier series of a function is its expansion into sines and cosines:

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots$$

where $f(x) \in \mathcal{C}[0,2\pi]$. So our question turns into what kind of function could be expressed as fuourier series?

Theorem 6.7 If a function f have the finite length in its function space $\mathcal{C}[a,b]$, then it could be expressed as *fourier series*.

But how to compute the coefficients $a_i's$ and $b_j's$? The key is orthogonality! For example, in order to get a_1, we just do the inner product between $f(x)$ and $\cos x$:

Figure 6.7: Enjoy fourier series!

$$\langle f(x), \cos x\rangle = a_1\langle \cos x, \cos x\rangle + 0 \implies a_1 = \frac{\langle f(x), \cos x\rangle}{\langle \cos x, \cos x\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x \, dx$$

Similarly we derive

$$a_m = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos mx \, dx \qquad b_m = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin mx \, dx.$$