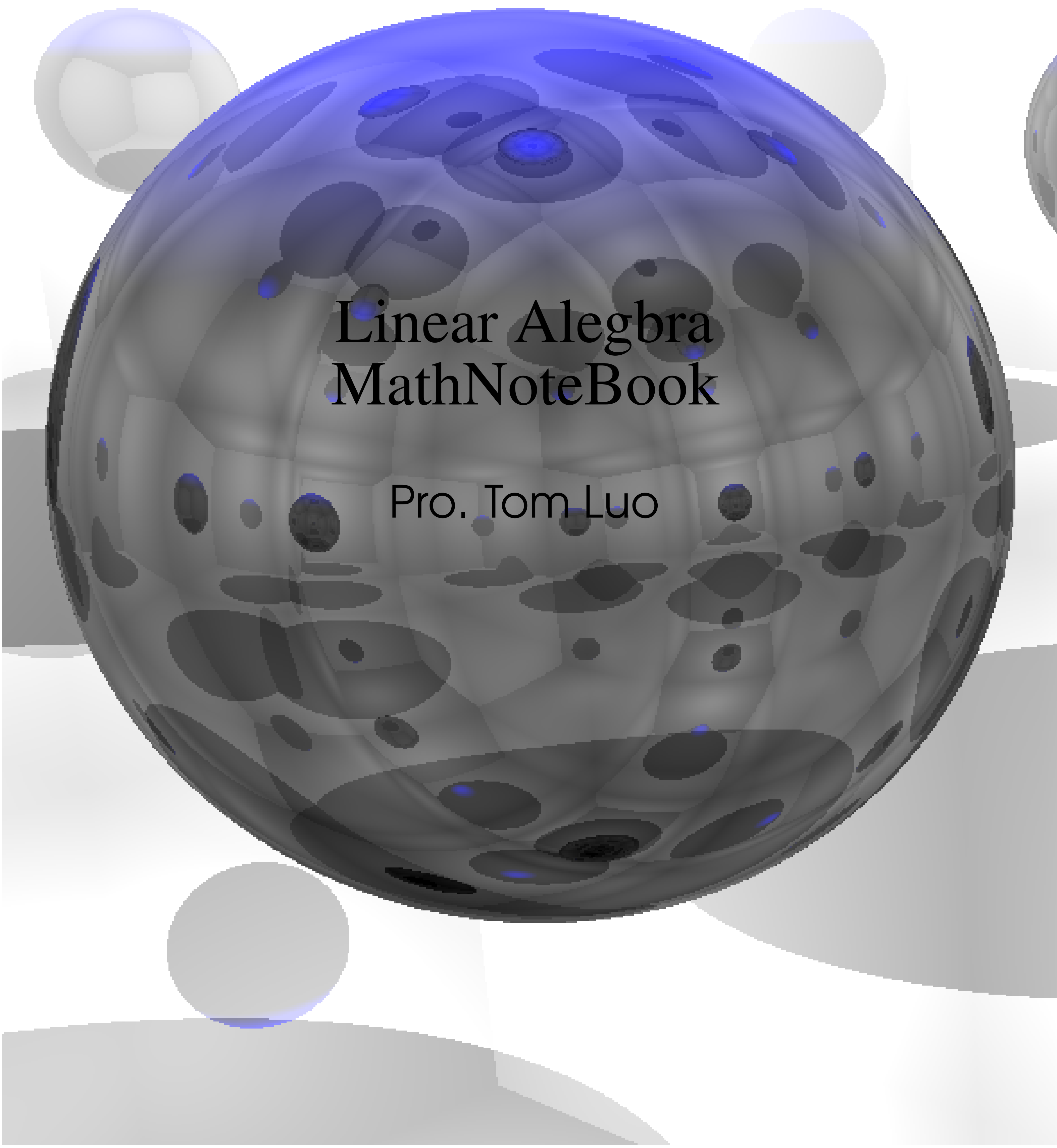
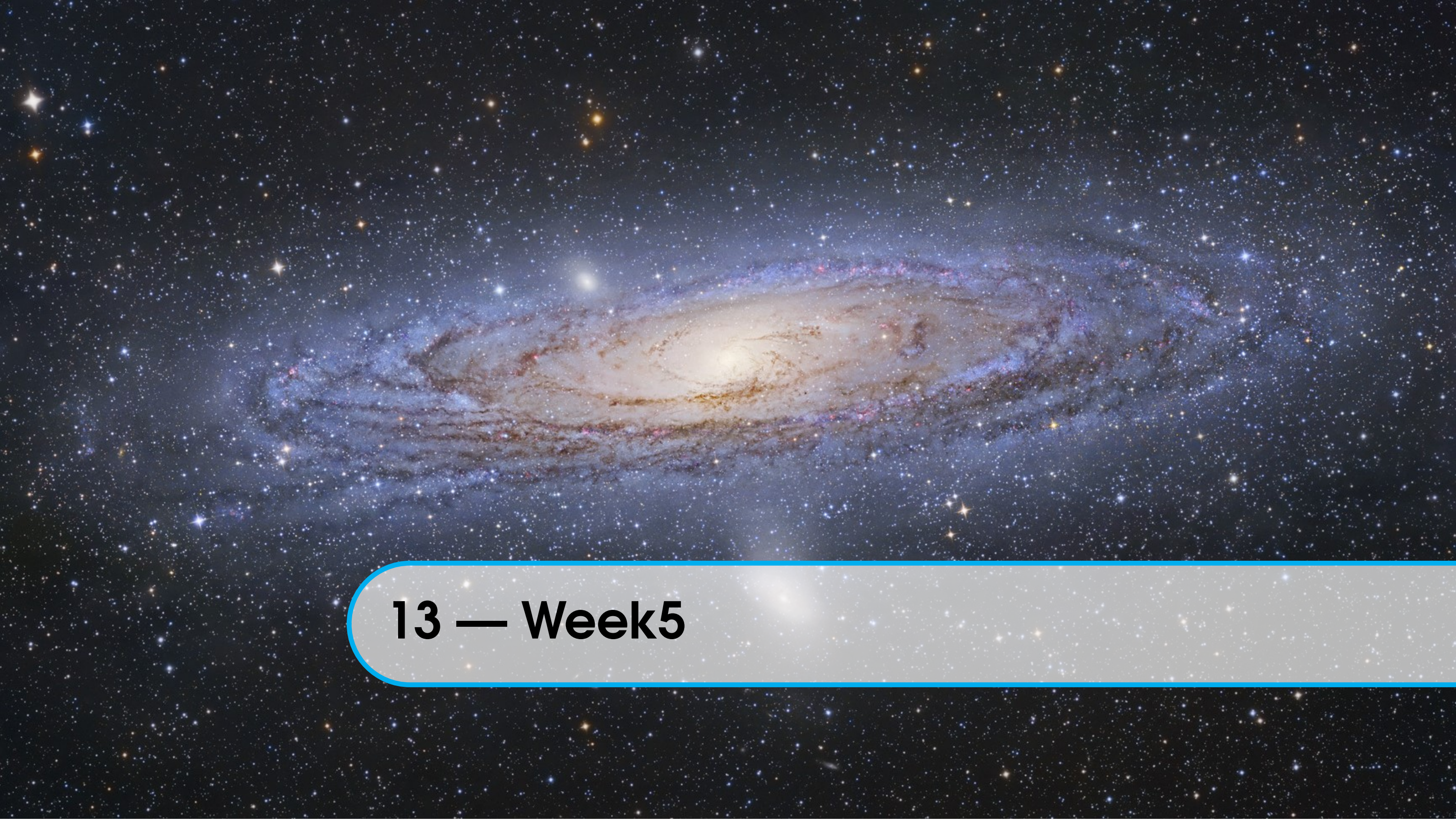


Linear Alegbra MathNoteBook

Pro. Tom Luo



13 — Week5

13.1 Friday

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

13.1.1 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–just divide each vector by its length, then the vectors become **orthogonal unit vectors**. Their lengths are all 1. Then its basis is called **orthonormal**.

Definition 13.1 — orthonormal. The vectors $\boldsymbol{q}_1,\dots,\boldsymbol{q}_n$ are **orthonormal** if

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

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

■ **Example 13.1** Unit vectors $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \boldsymbol{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$ is an *orthonormal basis* for $\mathbb{R}^n$. ■

If we want to express vector $\boldsymbol{b}$ as a linear combination of arbitrary basis $\{\boldsymbol{q}_1,\boldsymbol{q}_2,\dots,\boldsymbol{q}_n\}$, what should you do?
Answer: To solve the system $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{bmatrix}$.

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

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

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

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

Since $x_i = \langle \boldsymbol{q}_i,\boldsymbol{b}\rangle$, 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, since $\{\boldsymbol{q}_1,\boldsymbol{q}_2,\dots,\boldsymbol{q}_n\}$ forms a basis, the columns of $\boldsymbol{A}$ must be ind. Hence $\boldsymbol{A}$ is invertible, then we get the solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{13.1}$$

Definition 13.2 — matrix with orthonormal columns. Define $\boldsymbol{Q} = \begin{bmatrix} q_1 & q_2 & \dots & q_n\end{bmatrix}$. If vectors $\boldsymbol{q}_1,\dots,\boldsymbol{q}_n$ are **orthonormal**, then we say $\boldsymbol{Q}$ is a matrix with **orthonormal** columns.

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

Such matrix **is easy to work with** because we have:

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \\ \dots \\ \boldsymbol{q}_n^{\mathrm{T}}\end{pmatrix}\begin{pmatrix} q_1 & q_2 & \dots & q_n\end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_1 & & \\ & \ddots & \\ & & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{q}_n\end{pmatrix} = \boldsymbol{I}. \tag{13.2}$$

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

Definition 13.3 — orthogonal matrix. An **square** that is a *matrix with orthonormal columns* is called **othogonal matrix**.

■ Example 13.2
If $\boldsymbol{Q}$ is a 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 (13.2), we have $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$.
  Hence $\boldsymbol{Q}^{\mathrm{T}}$ is the left inverse of square matrix $\boldsymbol{Q}$.
  Hence $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^{\mathrm{T}} \implies \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{Q}^{-1} = \boldsymbol{I}$.

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*

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

- But the product $\hat{Q}\hat{Q}^{\mathrm{T}}$ will never be identity matrix. Assume $\hat{Q}$ is a $m \times n$ matrix. ($m \neq n$.) 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 ind. Hence $\operatorname{rank}(\hat{Q}) = n$.
  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 $\hat{Q}\hat{Q}^{\mathrm{T}} = \hat{q}\hat{q}^{\mathrm{T}} = \operatorname{rank}(1) \neq \boldsymbol{I}_m$.

Proposition 13.1
If $\boldsymbol{Q}$ has orthonormal columns, then it *leaves lengths unchanged*, in other words,

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

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

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

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

Orthogonal matrices are excellent for computations, since numbers can never grow too large when lengths of vectors are fixed.
In particular, if $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ has orthonormal columns, the least square problem is easy:
Although $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ 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 a unique solution $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$. Why? Since $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$, we derive $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$.

Summary:
Hence the **least squares solution** to $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ is $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$. In other words, $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} \approx \boldsymbol{b}$. **The projection matrix is** $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$. Note that the projection $\operatorname{Proj}_{\operatorname{col}(\boldsymbol{Q})}(\boldsymbol{b}) = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$ doesn't equal to $\boldsymbol{b}$ in general.
For general $\boldsymbol{A}$, the projection matrix is $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$.

13.1.2 Gram-Schmidt Process

"Orthogonal is good". So our goal for this section is: *Given ind. vectors, how to make them orthonormal?*
We start with three ind. vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ in $\mathbb{R}^3$. In order to construct orthonormal vectors, firstly we construct three **orthogonal** vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Then 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 (13.1) below, We find that $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}}(\boldsymbol{b})$. Hence

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

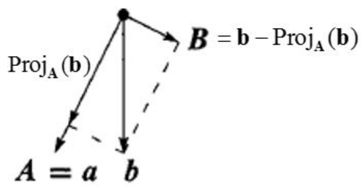
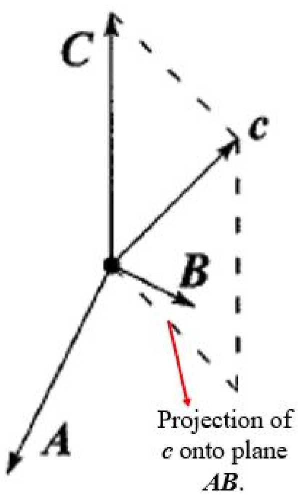


Figure 13.1: Subtract projection to get $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}}\,\boldsymbol{b}$.

You can take inner product between $\boldsymbol{A}$ and $\boldsymbol{B}$ to verify that $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal in Figure (13.1). Note that $\boldsymbol{B}$ is not zero (otherwise $\boldsymbol{a}$ and $\boldsymbol{b}$ would be dep. We will show it later.)

Then we want to construct vector $\boldsymbol{C}$. $\boldsymbol{C}$ is not a linear combination of $\boldsymbol{A}$ and $\boldsymbol{B}$. (Because $\boldsymbol{c}$ is not a linear combination of $\boldsymbol{a}$ and $\boldsymbol{b}$.) But most likely $\boldsymbol{c}$ is **not** perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. Hence we *subtract* $\boldsymbol{c}$ *off its projections onto the space of* $\boldsymbol{A}$ *and* $\boldsymbol{B}$. to get $\boldsymbol{C}$:

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



Finally we get $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ are obtained by dividing their lengths (shown in Figure (13.2)):

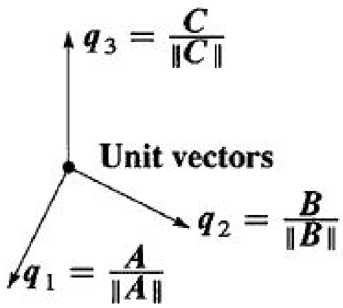


Figure 13.2: Final Gram-Schmidt step

Next we show an example of Gram-Schmidt step:

■ **Example 13.3** How to construct orthonormal vectors for $\boldsymbol{a} = \begin{pmatrix}1\\0\\1\end{pmatrix}, \boldsymbol{b} = \begin{pmatrix}1\\0\\0\end{pmatrix}, \boldsymbol{c} = \begin{pmatrix}2\\1\\0\end{pmatrix}$?

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

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

-

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

Hence we obtain our orthonormal vectors:

$$q_1 = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|} = \begin{pmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\end{pmatrix}, , q_2 = \frac{\boldsymbol{B}}{\|\boldsymbol{B}\|} = \begin{pmatrix}\frac{1}{\sqrt{2}}\\0\\-\frac{1}{\sqrt{2}}\end{pmatrix}, q_3 = \frac{\boldsymbol{C}}{\|\boldsymbol{C}\|} = \begin{pmatrix}0\\1\\0\end{pmatrix}$$

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

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

■

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

Gram-Schmidt process in general case

Input: Ind. vectors $a_1,\dots,a_n$.
Firstly we want to construct orthogonal vectors $\boldsymbol{A}_1,\dots,\boldsymbol{A}_n$.
In step $j \in \{1,\dots,n\}$, we want to compute a_j minus its projection in the space spanned by $\{\boldsymbol{A}_1,\boldsymbol{A}_2,\dots,\boldsymbol{A}_{j-1}\}$:

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

After we get $\boldsymbol{A}_1,\dots,\boldsymbol{A}_n$, 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 13.2 $\boldsymbol{A}_j \neq \boldsymbol{0}$ for $\forall j$ if and only if $a_1,a_2,\dots,a_n$ are ind.

Proofoutline. $\boldsymbol{A}_j = \boldsymbol{0} \Longleftrightarrow a_j = \mathrm{Proj}_{\mathrm{span}\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}(a_j)}$
Hence we only need to prove $\exists j$ s.t. $\boldsymbol{A}_j = \boldsymbol{0}$ if and only if $a_1,a_2,\dots,a_n$ are dep.
Sufficiency. Given $\boldsymbol{A}_j = \boldsymbol{0}$, then $a_j = \mathrm{Proj}_{\mathrm{span}\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}(a_j)\in\mathrm{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}\}}$. It's easy to verify that $\mathrm{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}\} = \mathrm{span}\{a_1,\dots,a_{j-1}\}$. Hence $a_j \in \mathrm{span}\{a_1,\dots,a_{j-1}\}$.
Hence $a_1,\dots,a_j$ are dep. Thus $a_1,\dots,a_n$ are dep.

Necessity. Given $a_1,a_2,\dots,a_n$ are dep. Then obviously, $a_n \in \mathrm{span}\{a_1,\dots,a_{n-1}\}$. It's easy to verify that $a_n = \mathrm{Proj}_{\mathrm{span}\{a_1,\dots,a_{n-1}\}}(\boldsymbol{a}_n)$. Thus $a_n = \mathrm{Proj}_{\mathrm{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{n-1}\}}(\boldsymbol{a}_n) \implies \boldsymbol{A}_n = \boldsymbol{0}$. ■

13.1.3 *The Factorization $A = QR$*

We know Gaussian Elimination leads to *LU decomposition*; in fact, Gram-Schmidt process leads to *QR factorization*. These two decomposition methods are quite important in LA, 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 matrix 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 $\boldsymbol{A} = \boldsymbol{QR}$.
In general case, $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k$ are combinations of $\boldsymbol{q}_1,\dots,\boldsymbol{q}_k$ at every step.
(In general suppose $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n\end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n\end{bmatrix}$)

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

■ **Example 13.4** Given $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c}\end{bmatrix}$, whose columns are ind. We can write $\boldsymbol{A}$ as:

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

where $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ are **orthonormal**.
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. ■

We have a theorem about QR faactorization (without proof):

Theorem 13.1 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 a upper triangular matrix (always square).

We postmultiply $\boldsymbol{Q}^{\mathrm{T}}$ both sides for $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ to obtain $\boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}$. In fact, 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

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

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 13.3 If $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, then we have a simple way to solve $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

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 substitution to get

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

■

13.1.4 Function Space

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

Example 13.5 For subspace $\operatorname{span}\{1,x,x^2\} \subset C[-1,1]$, firstly, how to define orthogonal for the basis $\{1,x,x^2\}$?

Pre-requisite: 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, we obtain:

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

$\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$ 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 $\operatorname{span}\{1,x,x^2\}$. ■

Example 13.6 Consider the collection $\mathscr{F}$ of functions defined on $[0,2\pi]$, where

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

Using various trigonometric identities, we can show that if f and g are **distinct**(different) functions in $\mathscr{F}$, 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 $\{1,\sin mx,\cos mx\}$ for $k = 1,2,\dots$ are *orthogonal* in $C[0,2\pi]$. Note that this set is **not orthonormal**! ■

This example motivates the fourier transformation:

13.1.5 Fourier Series

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 $f(x) \in C[0, 2\pi]$. We have an orthogonal basis! But what kind of function could be expressed in this way? There is a theorem for this condition (without proof):

Theorem 13.2 If a function f have the finite length in its function space $C[a,b]$, 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 13.3: Enjoy fourier series!

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

Similarly we derive

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