Rob 501 - Mathematics for Robotics
Recitation #5

Abhishek Venkataraman (Courtesy: Wubing Qin)

Oct 10, 2017

1 Norms

1. Let $(\mathcal{X}, \mathbb{C})$ be a vector space. A function $||\cdot|| : \mathcal{X} \to \mathbb{R}$ is a norm if:

   - (**Non-negative**): $\forall\, x \in \mathcal{X}, ||x|| \geq 0$ and $||x|| = 0 \Leftrightarrow x = 0$.
   - (**Triangular inequality**): $\forall\, x,\, y \in \mathcal{X},\, ||x+y|| \leq ||x|| + ||y||$.
   - (**Scalability**): $\forall\, \alpha \in \mathbb{C},\, \forall\, x \in \mathcal{X},\, ||\alpha\, x|| = |\alpha|\, ||x||$.

   And, $(\mathcal{X},\, \mathbb{C},\, ||\cdot||)$ is called a normed space.

   Ex:

   (a) In $(\mathbb{R}^n, \mathbb{R})$, prove $||x||_1 = \sum_{i=1}^{n} |x_i|$ and $||x||_\infty = \max_{1\leq i\leq n} |x_i|$ are norms.

   Properties of non-negativity and scalability are easy. Here we only try to prove it satisfies triangular inequality.

   - For $||x||_1 = \sum_{i=1}^{n} |x_i|,\, \forall\, x,\, y \in \mathbb{R}^n$,

   $$\begin{aligned} ||x+y||_1 &= \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} \big(|x_i| + |y_i|\big) \text{ (Equality holds when } \forall\, 1 \leq i \leq n,\, x_i y_i \geq 0) \\ &= \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| = ||x||_1 + ||x||_2 \end{aligned}$$

   - For $||x||_\infty = \max_{1\leq i\leq n} |x_i|,\, \forall\, x,\, y \in \mathbb{R}^n$, suppose

   $$\arg\max_{1\leq i\leq n} |x_i| = I,\ \arg\max_{1\leq i\leq n} |y_i| = J,\ \arg\max_{1\leq i\leq n} |x_i + y_i| = K,$$

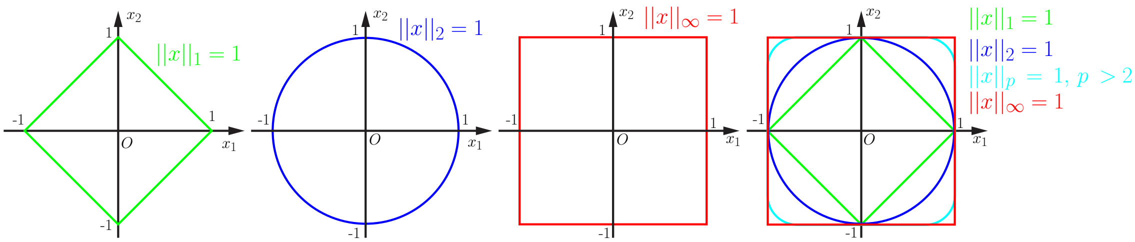
   i.e.,

   $$\forall\, 1 \leq i \leq n,\ |x_i| \leq |x_I|,\ |y_i| \leq |y_J|,\ |x_i + y_i| \leq |x_K + y_K|.$$

   Then we can obtain

   $$\begin{aligned} ||x+y||_\infty &= \max_{1\leq i\leq n} |x_i + y_i| = |x_K + y_K| \leq |x_K| + |y_K| \text{ (Equality holds when } x_K y_K \geq 0) \\ &\leq |x_I| + |y_J| \text{ (Equality holds when } I = J = K) \\ &= \max_{1\leq i\leq n} |x_i| + \max_{1\leq i\leq n} |y_i| = ||x||_\infty + ||x||_\infty. \end{aligned}$$

(b) In $(\mathbb{R}^2, \mathbb{R})$, plot the results of $||x||_1 = 1$, $||x||_2 = 1$, $||x||_\infty = 1$ on the x_1x_2 plane. Then think about $||x||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$, $p \geq 1$.



(c) In $(\mathbb{R}^2, \mathbb{R})$, given a symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ and define $f(x) = (x^\top Ax)^{1/2}$. Is it a norm?
Yes, it is an elliptic norm. Proof is not required.
(Hints: To prove this, you need to use the fact that $||x||_2 = (x^\top x)^{1/2}$ is a norm and for any symmetric positive definite matrix A, there exist an invertible matrix L, such that $A = LL^\top$.)

(d) Suppose $(\mathbb{R}^n, \mathbb{R}, ||\cdot||_V)$ is a normed space with some kind of norm $||\cdot||_V$ defined.
In $(\mathbb{R}^{n\times n}, \mathbb{R})$, define $f_V(A) = \sup\limits_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \dfrac{||Ax||_V}{||x||_V}$. Is it a norm? Try to calculate $f_2(A)$.
Yes, it is a norm, called induced norm. Here we only prove that it satisfies triangular inequality since non-negativity and scalability is trivial. $\forall\, A, B \in \mathcal{X}$,

$$
\begin{aligned}
f_V(A+B) &= \sup_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{||(A+B)x||_V}{||x||_V} = \sup_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{||Ax+Bx||_V}{||x||_V} \\
&\leq \sup_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{||Ax||_V + ||Bx||_V}{||x||_V} (\text{triangular inequlity in normed space } (\mathbb{R}^n, \mathbb{R}, ||\cdot||_V)) \\
&\leq \sup_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{||Ax||_V}{||x||_V} + \sup_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \frac{||Bx||_V}{||x||_V} = f_V(A) + f_V(B)
\end{aligned}
$$

For an induced norm, $||A||_V = \sup\limits_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \dfrac{||Ax||_V}{||x||_V} = \sup\limits_{\substack{x\in\mathbb{R}^n\\ x\neq 0}} \left\| A\dfrac{x}{||x||_V} \right\|_V$. If we let $y = \dfrac{x}{||x||_V}$, then $||y||_V = 1$ and $||A||_V = \sup\limits_{\substack{y\in\mathbb{R}^n\\ ||y||_V=1}} ||Ay||_V$. Now let's calculate $||A||_2$.

$$
||A||_2 = \sup_{\substack{y\in\mathbb{R}^n\\ ||y||_2=1}} ||Ay||_2 = \max_{\substack{y\in\mathbb{R}^n\\ y^\top y=1}} \sqrt{(Ay)^\top (Ay)} = \sqrt{\max_{\substack{y\in\mathbb{R}^n\\ y^\top y=1}} y^\top (A^\top A) y}
$$

In homework 2 problem 7, we calculated that for a symmetric matrix M, $\max\limits_{x\in\mathbb{R}^n} x^\top Mx$ subject to $x^\top x = 1$ is equal to $\lambda_{\max}(M)$, and a maximizing x is a normalized e-vector corresponding to $\lambda_{\max}(M)$. Based on this result, we can conclude that

$$
||A||_2 = \sqrt{\lambda_{\max}(A^\top A)}.
$$

2 Inner product

1. Let $(\mathcal{X}, \mathbb{C})$ be a vector space. A function $< \cdot, \cdot >: \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is an inner product if:

   - (**Hermitian symmetry**): $\forall x, y \in \mathbb{C}, < x, y >= \overline{< y, x >}$.
   - (**Linear in the first argument**): $\forall \alpha_1, \alpha_2 \in \mathbb{C},\ x_1, x_2, y \in \mathcal{X},\ < \alpha_1 x_1 + \alpha_2 x_2, y >= \alpha_1 < x_1, y > + \alpha_2 < x_2, y >$.
   - (**Non-negative**): $\forall x \in \mathcal{X}, < x, x > \geq 0$ and $< x, x >= 0 \Leftrightarrow x = 0$.

   And, $(\mathcal{X}, \mathbb{C}, < \cdot, \cdot >)$ is called an inner product space.

2. Let $(\mathcal{X}, \mathbb{C}, < \cdot, \cdot >)$ be an inner product space. Given two vectors $x, y \in \mathcal{X}$, x is orthogonal to y if $< x, y >= 0$.

   Ex:

   (a) In $(\mathbb{C}^n, \mathbb{C})$, define $< x, y >= x^\top \bar{y}$.

   (b) In $(\mathbb{R}^n, \mathbb{R})$, define $< x, y >= x^\top y$.

   (c) $\mathcal{X} = \{q(x) \,|\, \text{polynomials in } x \text{ with real coefficients of order } n,\ n \leq 3,\ x \in \mathbb{R}\}$, $\mathcal{F} = \mathbb{R}$.

   (i) Define $< f, g >:= \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x$. Is it an inner product?
   Yes, it is. Hermitian symmetry and non-negativity can be easily checked. So we only prove linearity in the first argument.
   $\forall f_1, f_2, g \in \mathcal{X}, \forall \alpha_1, \alpha_2 \in \mathcal{F}$,

$$\begin{aligned}< \alpha_1 f_1 + \alpha_2 f_2, g >&= \int_{-1}^{1} (\alpha_1 f_1 + \alpha_2 f_2)(x)g(x)\,\mathrm{d}x = \int_{-1}^{1} \big(\alpha_1 f_1(x) + \alpha_2 f_2(x)\big)g(x)\,\mathrm{d}x \\ &= \int_{-1}^{1} \alpha_1 f_1(x)g(x)\,\mathrm{d}x + \int_{-1}^{1} \alpha_2 f_2(x)g(x)\,\mathrm{d}x \\ &= \alpha_1 \int_{-1}^{1} f_1(x)g(x)\,\mathrm{d}x + \alpha_2 \int_{-1}^{1} f_2(x)g(x)\,\mathrm{d}x = \alpha_1 < f_1, g > + \alpha_2 < f_2, g >\end{aligned}$$

   (ii) Given a set of vectors $\{1, x, x^2, x^3\}$, calculate their products defined in (i).

$$< 1, x > = \int_{-1}^{1} 1 \cdot x\mathrm{d}x = 0, \qquad < 1, x^2 >= \int_{-1}^{1} 1 \cdot x^2\mathrm{d}x = 2/3, \qquad < 1, x^3 >= \int_{-1}^{1} 1 \cdot x^3\mathrm{d}x = 0$$
$$< x, x^2 > = \int_{-1}^{1} x \cdot x^2\mathrm{d}x = 0, \qquad < x, x^3 >= \int_{-1}^{1} x \cdot x^3\mathrm{d}x = 2/5, \qquad < x^2, x^3 >= \int_{-1}^{1} x^2 \cdot x^3\mathrm{d}x = 0$$

   (iii) Given another set of vectors $\left\{1,\ x,\ \frac{1}{2}(3x^2 - 1),\ \frac{1}{2}(5x^3 - 3x)\right\}$, calculate their products defined in (i).

$$
\begin{aligned}
&< 1, x >= \int_{-1}^{1} 1 \cdot x \mathrm{d}x = 0, \\
&< 1, \frac{1}{2}(3x^2 - 1) >= \int_{-1}^{1} 1 \cdot \frac{1}{2}(3x^2 - 1)\mathrm{d}x = \frac{1}{2}(x^3 - x)\Big|_{-1}^{1} = 0, \\
&< 1, \frac{1}{2}(5x^3 - 3x) >= \int_{-1}^{1} 1 \cdot \frac{1}{2}(5x^3 - 3x)\mathrm{d}x = 0 \\
&< x, \frac{1}{2}(3x^2 - 1) >= \int_{-1}^{1} x \cdot \frac{1}{2}(3x^2 - 1)\mathrm{d}x = 0, \\
&< x, \frac{1}{2}(5x^3 - 3x) >= \int_{-1}^{1} x \cdot \frac{1}{2}(5x^3 - 3x)\mathrm{d}x = \frac{1}{2}(x^5 - x^3)\Big|_{-1}^{1} = 0, \\
&< \frac{1}{2}(3x^2 - 1), \frac{1}{2}(5x^3 - 3x) >= \int_{-1}^{1} \frac{1}{2}(3x^2 - 1) \cdot \frac{1}{2}(5x^3 - 3x)\mathrm{d}x = 0
\end{aligned}
$$

Thus, those polynomials are all orthogonal. They are the first few vectors of Legendre polynomials.

(d) $\mathcal{X} = \{f \,|\, f : \mathbb{R} \to \mathbb{R},\ f(x + T) = f(x),\ T \text{ is a given constant}\}$, $\mathcal{F} = \mathbb{R}$.

(i) Define $< f,\, g >:= \frac{1}{T} \int_0^T f(x)g(x)\,\mathrm{d}x$. Is it an inner product?
Yes, it is an inner product. Proof is similar to (c). Try it out on your own.

(ii) Given a set of vectors $\{1,\ \sin(\frac{2\pi}{T}x),\ \cos(\frac{2\pi}{T}x),\ \sin(\frac{4\pi}{T}x),\ \sin(\frac{4\pi}{T}x)\}$, calculate their products defined in (i).
Please try to calculate it on your own, and you will figure out that this is actually an orthogonal set, i.e., each vector is orthogonal to any other vectors in this set. Maybe you have noticed that this is the first few vectors in the Fourier series. Yes, that is exactly what we are doing in Fourier series. We are projecting an infinite-dimensional periodic function to an orthogonal basis. The coefficients in front of those sines and cosines are actually coordinates in these directions. One caveat is that this is not a finite dimensional space.

3 Gram Schmidt process

1. In an inner product space $(\mathcal{X},\, \mathcal{F},\, < \cdot >)$, given $x, y \in \mathcal{X}$. *x is orthogonal to y* if $< x,\, y >= 0$, denoted as $x \perp y$.

2. A set S is *orthogonal* if $\forall\, (x, y \in S,\ x \neq y), x \perp y$.

3. A set S is *orthonormal* if S is orthogonal and $\forall\, x \in S,\ \|x\| = 1$.

4. In a finite dimensional vector space, any set of linear independent vectors can be completed to a basis.

5. In an inner product space $(\mathcal{X},\, \mathcal{F},\, < \cdot >)$, given $S \subset \mathcal{X}$ a subset, the *orthogonal complement of S* is

$$S^{\perp} := \{x \in \mathcal{X} \,|\, < x,\, y >= 0,\ \forall\, y \in S\}.$$

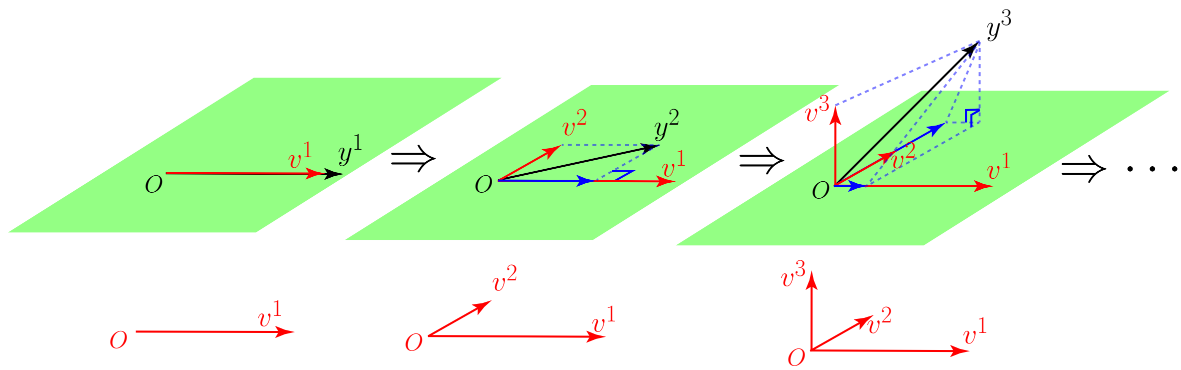
6. In an inner product space $(\mathcal{X},\, \mathcal{F},\, < \cdot >)$, given a set of linear independent vectors $Y = \{y^1,\, y^2,\, \ldots,\, y^n\}$. There exists an orthogonal set $V = \{v^1,\, v^2,\, \ldots,\, v^n\}$ such that $\mathrm{span}\,\{V\} = \mathrm{span}\,\{Y\}$. V can be otained

by

$$k = 1, \qquad v^1 = y^1,$$
$$k \geq 2, \qquad v^k = y^k - \sum_{i=1}^{k-1} \frac{< y^k, v^i >}{< v^i, v^i >} v^i.$$

7. Ex:

- In $(\mathbb{R}^3, \mathbb{R})$, demonstrate how Gram Schmidt process works.



- In $(\mathbb{R}^3, \mathbb{R})$, given a set $Y = \left\{ y^1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, y^2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \right\}$ and $< x, y >= x^\top y$, find an orthogonal set V such that span $\{V\}$ = span $\{Y\}$. Complete set V to a basis. What if you are asked to complete set V to an orthogonal basis.
  Apply Gram Schmidt process, we get

$$v^1 = y^1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$
$$v^2 = y^2 - \frac{< y^2, v^1 >}{< v^1, v^1 >} v^1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

If we want to complete V to a basis, suppose we choose $v^3 = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$ such that $\{v^1, v^2, v^3\}$ is a linearly independent set. Then $\exists\, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$, $\alpha_1 v^1 + \alpha_2 v^2 + \alpha_3 v^3 = 0$ has non-trivial solution. This can be written into

$$\underbrace{\begin{bmatrix} v^1 & v^2 & v^3 \end{bmatrix}}_{V} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = 0.$$

Choose v^3 such that $\det V \neq 0$, e.g., $v^3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

If we need to find an orthogonal basis, we can either complete set Y to a basis and then apply Gram Schmidt process, or do the following. We have an orthogonal set $\{v^1, v^2\}$ after apply Gram Schmidt process to the given vectors $\{y^1, y^2\}$, we only need to find another vector v^4 such that $v^4 \perp v^1$ and $v^4 \perp v^2$, i.e., $< v^1, v^4 >= 0$, $< v^2, v^4 >= 0$.

$$\begin{bmatrix} (v^1)^\top \\ (v^2)^\top \end{bmatrix} v^4 = 0 \Rightarrow \begin{bmatrix} 1 & 1 & 0 \\ 1/2 & -1/2 & 1 \end{bmatrix} v^4 = 0 \Rightarrow v^4 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Then, $\{v^1, v^2, v^4\}$ is an orthogonal basis.

- $\mathcal{X} = \{q(x) \,|\, \text{polynomials in } x \text{ with real coefficients of order } n,\ n \leq 3,\ x \in \mathbb{R}\}$, $\mathcal{F} = \mathbb{R}$. The inner product is defined as $< f, g >:= \int_{-1}^{1} f(x)g(x)\,\mathrm{d}x$. Given a set of vectors $\{1, x, x^2\}$, apply Gram Schmidt process to find an orthogonal set V, and then find $V^\perp$.
  Apply Gram Schmidt process, we have

$$v^1 = 1,$$

$$v^2 = y^2 - \frac{< y^2, v^1 >}{< v^1, v^1 >} v^1 = x - \frac{\int_{-1}^{1} x\,\mathrm{d}x}{2} \cdot 1 = x,$$

$$v^3 = y^3 - \frac{< y^3, v^1 >}{< v^1, v^1 >} v^1 - \frac{< y^3, v^2 >}{< v^2, v^2 >} v^2 = x^2 - \frac{\int_{-1}^{1} x^2\,\mathrm{d}x}{2} \cdot 1 - \frac{\int_{-1}^{1} x^3\,\mathrm{d}x}{2/3} \cdot x = x^2 - \frac{1}{3},$$

We want to find $V^\perp$, then we need to find vector v^4 first such that $v^4 \perp v^1$, $v^4 \perp v^2$, $v^4 \perp v^3$. Suppose $v^4 = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$, then

$$\begin{cases} < v^1, v^4 >= 0 \\ < v^2, v^4 >= 0 \\ < v^3, v^4 >= 0 \end{cases} \Rightarrow \begin{cases} \beta_0 + \frac{2}{3}\beta_2 = 0, \\ \frac{2}{3}\beta_1 + \frac{2}{5}\beta_3 = 0, \\ \frac{2}{3}\beta_0 + \frac{2}{5}\beta_2 = 0, \end{cases} \Rightarrow \beta = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 5/3 \end{bmatrix}$$

Thus, $v^4 = \frac{5}{3}x^3 - x$ and $V^\perp = \text{span}\left\{\frac{5}{3}x^3 - x\right\}$.
Here we have used the results below.

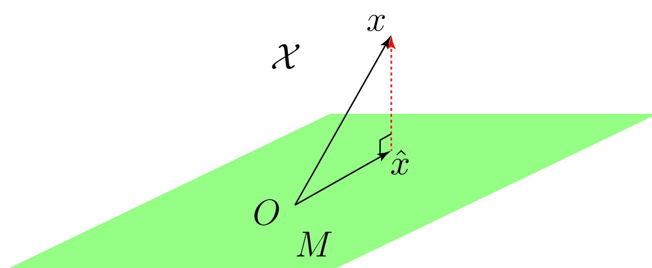
$$< 1, 1 > = \int_{-1}^{1} 1 \cdot 1\mathrm{d}x = 2, \quad < 1, x >= \int_{-1}^{1} 1 \cdot x\mathrm{d}x = 0, \quad < 1, x^2 >= \int_{-1}^{1} 1 \cdot x^2\mathrm{d}x = 2/3,$$

$$< 1, x^3 > = \int_{-1}^{1} 1 \cdot x^3\mathrm{d}x = 0, \quad < x, x >= \int_{-1}^{1} x \cdot x^2\mathrm{d}x = 2/3, \quad < x, x^2 >= \int_{-1}^{1} x \cdot x^2\mathrm{d}x = 0,$$

$$< x, x^3 > = \int_{-1}^{1} x \cdot x^3\mathrm{d}x = 2/5, \quad < x^2, x^2 >= \int_{-1}^{1} x^2 \cdot x^2\mathrm{d}x = 2/5, \quad < x^2, x^3 >= \int_{-1}^{1} x^2 \cdot x^3\mathrm{d}x = 0$$

4 Projection theorem

1. Projection Theorem: $(\mathcal{X}, \mathbb{R}, < \cdot, \cdot >)$ is a **finite dimensional inner product space**, and $M \subset \mathcal{X}$ is a subspace of $\mathcal{X}$. Then $\forall x \in \mathcal{X}$, there exists a unique $\hat{x} \in M$ such that $\|x - \hat{x}\| = \inf_{y \in M} \|x - y\|$. Moreover, $\hat{x}$ is characterized by $(x - \hat{x}) \perp M$.



2. Normal equation:
In a finite dimensional inner product space $(\mathcal{X}, \mathbb{R}, < \cdot >)$, $M = \text{span}\left\{y^1, y^2, \ldots, y^k\right\}$ is a subspace of $\mathcal{X}$ where $\{y^1, y^2, \ldots, y^k\}$ is a linear independent set. We need to find $\hat{x} = \arg\min_{m \in M} \|x - m\|$.
Suppose $\hat{x}$ takes the form
$$\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k,$$
then the coordinates of $\hat{x}$ expressed in the basis $\{y^1, y^2, \ldots, y^k\}$, i.e., $\alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix}$ is given by the normal equation
$$\begin{bmatrix} < y^1, y^1 > & \cdots & < y^k, y^1 > \\ \vdots & \ddots & \vdots \\ < y^1, y^k > & \cdots & < y^k, y^k > \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} < x, y^1 > \\ \vdots \\ < x, y^k > \end{bmatrix}$$

3. A function $P : \mathcal{X} \to M$ is an orthogonal projection operator if $P(x) = \arg\min_{m \in M} \|x - m\|$.

4. Ex:

   - In $(\mathbb{R}^{2\times 2, \mathbb{R}})$, the inner product is defined as $< A, B >= \text{trace}(A^\top Q B)$, where $Q = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, and $M = \text{span}\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right\}$. Given $x = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, find $\hat{x} = \arg\min_{y \in M} \|x - y\|$.
   Suppose $\hat{x}$ takes the form $\hat{x} = \alpha_1 y^1 + \alpha_2 y^2$, we need to solve α_1, α_2 using normal equation such

that $\hat{x} = \arg\min_{y \in M} \|x - y\|$.

$$< y^1, y^1 > = \text{trace}\big((y^1)^\top Q y^1\big) = \text{trace}\left(\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) = \text{trace}\left(\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\right) = 3$$
$$< y^1, y^2 > = \text{trace}\big((y^1)^\top Q y^2\big) = \text{trace}\left(\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\right) = \text{trace}\left(\begin{bmatrix}1 & 2\\1 & 1\end{bmatrix}\right) = 2$$
$$< y^2, y^2 > = \text{trace}\big((y^2)^\top Q y^2\big) = \text{trace}\left(\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\right) = \text{trace}\left(\begin{bmatrix}1 & 1\\1 & 2\end{bmatrix}\right) = 3$$
$$< x, y^1 > = \text{trace}\big(x^\top Q y^1\big) = \text{trace}\left(\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) = \text{trace}\left(\begin{bmatrix}4 & 3\\10 & 7\end{bmatrix}\right) = 11$$
$$< x, y^2 > = \text{trace}\big(x^\top Q y^2\big) = \text{trace}\left(\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}\right) = \text{trace}\left(\begin{bmatrix}3 & 4\\7 & 10\end{bmatrix}\right) = 13$$

Thus,

$$\begin{bmatrix}< y^1, y^1 > & < y^2, y^1 >\\< y^1, y^2 > & < y^2, y^2 >\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix} = \begin{bmatrix}< x, y^1 >\\< x, y^2 >\end{bmatrix} \Rightarrow \begin{bmatrix}3 & 2\\2 & 3\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix} = \begin{bmatrix}11\\13\end{bmatrix} \Rightarrow \begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix} = \begin{bmatrix}7/5\\17/5\end{bmatrix}$$

The minimizing vector is

$$\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 = \frac{7}{5}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} + \frac{17}{5}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix} = \begin{bmatrix}7/5 & 17/5\\17/5 & 7/5\end{bmatrix}$$

- $\mathcal{X} = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$, $\mathcal{F} = \mathbb{R}$. Define inner product $< f, g >= \int_{-1}^{1} f(t)g(t)\mathrm{d}t$. $M = \text{span}\left\{1, t, t^2\right\}$, $x = \mathrm{e}^t$, find $\hat{x} = \arg\min_{y \in M} \|x - y\|$.
  Suppose $\hat{x}$ takes the form $\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + +\alpha_3 y^3$, we need to solve α_1, α_2, α_3 using normal equation such that $\hat{x} = \arg\min_{y \in M} \|x - y\|$.
  In page (3), we have the inner product table, now we only need to calculate

$$\begin{cases}< \mathrm{e}^t, 1 >= \int_{-1}^{1} \mathrm{e}^t \mathrm{d}t = \mathrm{e} - \mathrm{e}^{-1}\\ < \mathrm{e}^t, t >= \int_{-1}^{1} t\mathrm{e}^t \mathrm{d}t = t\mathrm{e}^t\Big|_{-1}^{1} - \int_{-1}^{1} \mathrm{e}^t \mathrm{d}t = 2\mathrm{e}^{-1}\\ < \mathrm{e}^t, t^2 >= \int_{-1}^{1} t^2\mathrm{e}^t \mathrm{d}t = t^2\mathrm{e}^t\Big|_{-1}^{1} - \int_{-1}^{1} 2t\mathrm{e}^t \mathrm{d}t = \mathrm{e} - 5\mathrm{e}^{-1}\end{cases} \Rightarrow \begin{bmatrix}2 & 0 & 2/3\\0 & 2/3 & 0\\2/3 & 0 & 2/5\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\\\alpha_3\end{bmatrix} = \begin{bmatrix}\mathrm{e} - \mathrm{e}^{-1}\\2\mathrm{e}^{-1}\\\mathrm{e} - 5\mathrm{e}^{-1}\end{bmatrix}$$

  Solve α_1, α_2, α_3, then you get $\hat{x}$.

- In $(\mathbb{R}^n, \mathbb{R})$ with inner product defined in the standard way, i.e., $< x, y >= x^\top y$.
  $M = \text{span}\left\{v^1, v^2, \ldots, v^k\right\}$, $k \le n$. Find the matrix representation of the othogonal projector that projects $x \in \mathcal{X}$ onto $\hat{x} \in M$. Use the standard basis $E = \{e^1, e^2, \ldots, e^n\}$ for $(\mathbb{R}^n, \mathbb{R})$, and basis $V = \{v^1, v^2, \ldots, v^k\}$ for subspace M. What if we also express the projection $\hat{x}$ in standard basis $E = \{e^1, e^2, \ldots, e^n\}$? What if V is an orthonormal set?

  Assume that $P(x) = \hat{x} = \alpha_1 v^1 + \alpha_2 v^2 + \cdots + \alpha_k v^k$. Thus, $[P(x)]_V = \begin{bmatrix}\alpha_1\\\vdots\\\alpha_k\end{bmatrix} := \alpha$.

  Based on normal equation, we obtain that the coefficient α satisfies

$$G^\top \alpha = \beta,$$

where

$$G = \begin{bmatrix} < v^1,\, v^1 > & \cdots & < v^k,\, v^1 > \\ \vdots & \ddots & \vdots \\ < v^1,\, v^k > & \cdots & < v^k,\, v^k > \end{bmatrix} = \begin{bmatrix} (v^1)^\top v^1 & \cdots & (v^1)^\top v^k \\ \vdots & \ddots & \vdots \\ (v^k)^\top v^1 & \cdots & (v^k)^\top v^k \end{bmatrix} = \begin{bmatrix} (v^1)^\top \\ \vdots \\ (v^k)^\top \end{bmatrix} \begin{bmatrix} v^1 & \cdots & v^k \end{bmatrix} = V^\top V,$$

$$\beta = \begin{bmatrix} < x,\, v^1 > \\ \vdots \\ < x,\, v^k > \end{bmatrix} = \begin{bmatrix} x^\top v^1 \\ \vdots \\ x^\top v^k \end{bmatrix} = \begin{bmatrix} (v^1)^\top x \\ \vdots \\ (v^k)^\top x \end{bmatrix} = \begin{bmatrix} (v^1)^\top \\ \vdots \\ (v^k)^\top \end{bmatrix} x = \begin{bmatrix} v^1 & \cdots & v^k \end{bmatrix}^\top x = V^\top x.$$

Thus, $\alpha = \left(G^\top\right)^{-1} \beta = \left(V^\top V\right)^{-1} V^\top x$. The matrix representation of the orthogonal projector from standard basis E to basis V is $P = \left(V^\top V\right)^{-1} V^\top$.
If we express $\hat{x}$ in standard basis $E = \{e^1, e^2, \ldots, e^n\}$, then we need to find the matrix representation of linear operator $\mathcal{L} : \mathbb{R}^k \to \mathbb{R}^n$ that maps vector expressed in V to the same vector expressed in E. Notice that $[x]_E = V[x]_V$, then the resulting matrix representation of the orthogonal projector becomes $\hat{P} = VP = V\left(V^\top V\right)^{-1} V^\top$.
If V is an orthonormal set, then $\left(V^\top V\right) = I$,

$$\hat{P}(x) = V\left(V^\top V\right)^{-1} V^\top x = VV^\top x.$$

What we have derived in class for an orthonormal set is that $\hat{P}(x) = \sum_{i=1}^{k} < x,\, v^i > v^i$, are they equivalent?

$$\begin{aligned} \hat{P}(x) &= \sum_{i=1}^{k} < x,\, v^i > v^i = \sum_{i=1}^{k} v^i < v^i,\, x >= \sum_{i=1}^{k} v^i \left((v^i)^\top x\right) = \sum_{i=1}^{k} \left(v^i (v^i)^\top\right) x \\ &= \left(\sum_{i=1}^{k} v^i (v^i)^\top\right) x = \begin{bmatrix} v^1 & \cdots & v^k \end{bmatrix} \begin{bmatrix} (v^1)^\top \\ \vdots \\ (v^k)^\top \end{bmatrix} x = VV^\top x. \end{aligned}$$