Mathematics for Robotics (ROB-GY 6013 Section A)

- **Week 10:**
  - Theme: Least Squares
    - Recursive Least Squares
    - Probability

How did we get here?

$$\hat{\alpha} = (A^T A)^{-1} A^T b$$

How did we get here?

- Projection Theorem

$$\hat{\alpha} = (A^T A)^{-1} A^T b$$

How did we get here?

- Projection Theorem
- What type of problem?

$$\hat{\alpha} = (A^T A)^{-1} A^T b$$

How did we get here?

$A \in \mathbb{R}^{m \times n}$

$x \in \mathbb{R}^{m}$

- Projection Theorem
- What type of problem? $\hat{x} = \arg\min_{m \in M} \|x - m\| = \arg\min_{A\hat{\alpha} \in M} \|b - A\hat{\alpha}\| = \boxed{\arg\min_{\hat{\alpha} \in \mathbb{R}^n} \|A\hat{\alpha} - b\|}$

$$\hat{\alpha} = (A^T A)^{-1} A^T b$$

How did we get here?

$$A \in \mathbb{R}^{m \times n}$$

$$x \in \mathbb{R}^{m}$$

- Projection Theorem
- What type of problem? $\hat{x} = \arg\min_{m \in M} \|x - m\| = \arg\min_{A\hat{\alpha} \in M} \|b - A\hat{\alpha}\| = \boxed{\arg\min_{\hat{\alpha} \in \mathbb{R}^n} \|A\hat{\alpha} - b\|}$

$$\hat{\alpha} = (A^T A)^{-1} A^T b$$

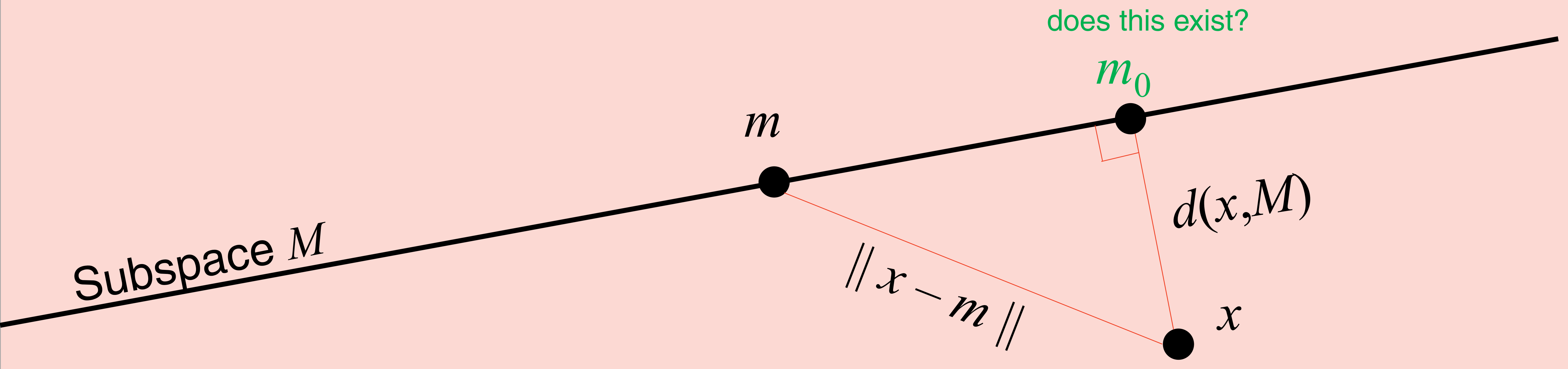
- **Let's review the derivation of the normal equations**

Recap: Without Classical Projection Theorem

- Many problems are minimum-distance problems

Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be a **inner product space**, M be a **subspace** of $\mathcal{X}$, and x be an arbitrary **point** in $\mathcal{X}$.

Inner product space $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$

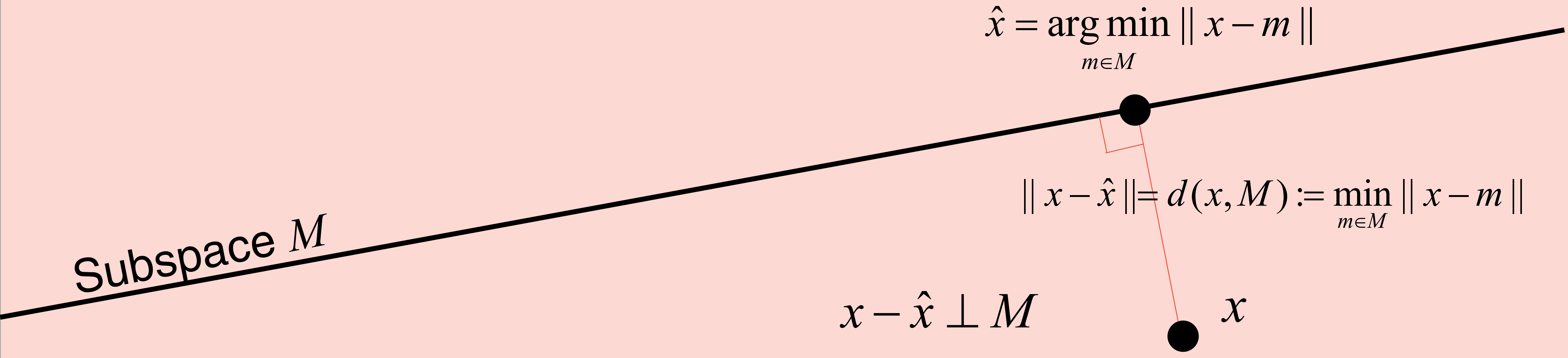


Recap: Classical Projection Theorem

- Many problems are minimum-distance problems

Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be a **finite-dimensional** (real) **inner product space**, M be a **subspace** of $\mathcal{X}$, and x be an arbitrary **point** in $\mathcal{X}$.

Inner product space $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$



Recap: Classical Projection Theorem

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be a finite-dimensional (real) **inner product space** and M a **subspace** of $\mathcal{X}$. Then, $\forall x \in \mathcal{X}$, $\exists$ **unique** $\hat{x} \in M$ such that

$$\|x - \hat{x}\| = d(x, M) := \inf_{m \in M} \|x - m\| = \min_{m \in M} \|x - m\|$$

  where we can write **min**imum instead of **inf**imum, because the infimum is achieved. Moreover $\hat{x} \in M$ is characterized by $x - \hat{x} \perp M$

- Use of $\hat{x}$ (best approximation) instead of m_0 because this is the standard notation for estimation problems.

Find $\hat{x}$ with Projection Theorem

- By Projection Theorem, $\hat{x}$ **exists** and is **unique** (characterized by $x - \hat{x} \perp M$)

Find $\hat{x}$ with Projection Theorem

- By Projection Theorem, $\hat{x}$ **exists** and is **unique** (characterized by $x - \hat{x} \perp M$)
- Let $M = \mathrm{span}\{y^1, \ldots, y^k\}$: $\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + \ldots + \alpha_k y^k$

Find $\hat{x}$ with Projection Theorem

- By Projection Theorem, $\hat{x}$ **exists** and is **unique** (characterized by $x - \hat{x} \perp M$)
- Let $M = \text{span}\{y^1, \ldots, y^k\}$: $\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + \ldots + \alpha_k y^k$
- Then: $(x - \hat{x} \perp M)$

$$\Updownarrow$$

$$x - \hat{x} \perp y^i \quad 1 \leq i \leq k$$

Find $\hat{x}$ with Projection Theorem

- By Projection Theorem, $\hat{x}$ **exists** and is **unique** (characterized by $x - \hat{x} \perp M$)
- Let $M = \mathrm{span}\{y^1, \dots, y^k\}$: $\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + \ldots + \alpha_k y^k$
- Then: $(x - \hat{x} \perp M)$

$$\Updownarrow$$

$$x - \hat{x} \perp y^i \quad 1 \le i \le k$$

$$\Updownarrow$$

$$\left\langle x - \hat{x}, y^i \right\rangle = 0 \quad 1 \le i \le k$$

Find $\hat{x}$ with Projection Theorem

- By Projection Theorem, $\hat{x}$ **exists** and is **unique** (characterized by $x - \hat{x} \perp M$)
- Let $M = \text{span}\{y^1, \dots, y^k\}$: $\hat{x} = \alpha_1 y^1 + \alpha_2 y^2 + \dots + \alpha_k y^k$
- Then: $(x - \hat{x} \perp M)$

$$\Updownarrow$$

$$x - \hat{x} \perp y^i \quad 1 \le i \le k$$

$$\Updownarrow$$

$$\langle x - \hat{x}, y^i \rangle = 0 \quad 1 \le i \le k$$

$$\langle x - \hat{x}, y^i \rangle = \langle \hat{x}, y^i \rangle - \langle x, y^i \rangle$$

$$\Updownarrow$$

$$\langle \hat{x}, y^i \rangle = \langle x, y^i \rangle \quad 1 \le i \le k$$

Find $\hat{x}$ with Projection Theorem

- By Projection Theorem, $\hat{x}$ **exists** and is **unique** (characterized by $x-\hat{x}\perp M$)
- Let $M = \operatorname{span}\{y^1, \ldots, y^k\}$: $\hat{x}=\alpha_1 y^1+\alpha_2 y^2+\ldots+\alpha_k y^k$
- Then: $(x-\hat{x}\perp M)$

$$\Updownarrow$$

$$x-\hat{x}\perp y^i \quad 1\le i\le k$$

$$\Updownarrow$$

$$\langle x-\hat{x}, y^i\rangle=0 \quad 1\le i\le k \qquad\qquad \langle x-\hat{x}, y^i\rangle=\langle \hat{x}, y^i\rangle-\langle x, y^i\rangle$$

$$\Updownarrow$$

$$\boxed{\langle \hat{x}, y^i\rangle=\langle x, y^i\rangle \quad 1\le i\le k}$$

$$\langle \widehat{x}, y^i \rangle = \langle x, y^i \rangle \; 1 \le i \le k$$

$$\Updownarrow$$

$$\langle \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k, y^i \rangle = \langle x, y^i \rangle \; 1 \le i \le k$$

$$\Updownarrow$$

$$\alpha_1 \langle y^1, y^i \rangle + \alpha_2 \langle y^2, y^i \rangle + \cdots + \alpha_k \langle y^k, y^i \rangle = \langle x, y^i \rangle \; 1 \le i \le k$$

- Find α's

$$\langle \widehat{x}, y^i \rangle = \langle x, y^i \rangle \; 1 \leq i \leq k$$

$$\Updownarrow$$

$$\langle \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k, y^i \rangle = \langle x, y^i \rangle \; 1 \leq i \leq k$$

$$\Updownarrow$$

$$\alpha_1 \langle y^1, y^i \rangle + \alpha_2 \langle y^2, y^i \rangle + \cdots + \alpha_k \langle y^k, y^i \rangle = \langle x, y^i \rangle \; 1 \leq i \leq k$$

- Find α's

$$\langle \widehat{x}, y^i \rangle = \langle x, y^i \rangle \; 1 \leq i \leq k$$
$$\Updownarrow$$
$$\langle \alpha_1 y^1 + \alpha_2 y^2 + \cdots + \alpha_k y^k, y^i \rangle = \langle x, y^i \rangle \; 1 \leq i \leq k$$
$$\Updownarrow$$
$$\alpha_1 \langle y^1, y^i \rangle + \alpha_2 \langle y^2, y^i \rangle + \cdots + \alpha_k \langle y^k, y^i \rangle = \langle x, y^i \rangle \; 1 \leq i \leq k$$

$$\boxed{\begin{array}{ll} i = 1 & \alpha_1 \langle y^1, y^1 \rangle + \alpha_2 \langle y^2, y^1 \rangle + \cdots + \alpha_k \langle y^k, y^1 \rangle = \langle x, y^1 \rangle \\ i = 2 & \alpha_1 \langle y^1, y^2 \rangle + \alpha_2 \langle y^2, y^2 \rangle + \cdots + \alpha_k \langle y^k, y^2 \rangle = \langle x, y^2 \rangle \\ \quad \vdots & \qquad\qquad\qquad\qquad \vdots \\ i = k & \alpha_1 \langle y^1, y^k \rangle + \alpha_2 \langle y^2, y^k \rangle + \cdots + \alpha_k \langle y^k, y^k \rangle = \langle x, y^k \rangle \end{array}}$$

Normal Equations

$$G(y^1, \cdots, y^k) := \begin{bmatrix} \langle y^1, y^1\rangle & \langle y^1, y^2\rangle & \cdots & \langle y^1, y^k\rangle \\ \langle y^2, y^1\rangle & \langle y^2, y^2\rangle & \cdots & \langle y^2, y^k\rangle \\ \vdots & \vdots & & \vdots \\ \langle y^k, y^1\rangle & \langle y^k, y^2\rangle & \cdots & \langle y^k, y^k\rangle \end{bmatrix} \quad \alpha := \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}, \ \beta := \begin{bmatrix} \langle x, y^1\rangle \\ \langle x, y^2\rangle \\ \vdots \\ \langle x, y^k\rangle \end{bmatrix} =: \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

Gram Matrix

$G_{ij} := \langle y^i, y^j \rangle$

$$\boxed{\begin{array}{ll} i = 1 & \alpha_1\langle y^1, y^1\rangle + \alpha_2\langle y^2, y^1\rangle + \cdots + \alpha_k\langle y^k, y^1\rangle = \langle x, y^1\rangle \\ i = 2 & \alpha_1\langle y^1, y^2\rangle + \alpha_2\langle y^2, y^2\rangle + \cdots + \alpha_k\langle y^k, y^2\rangle = \langle x, y^2\rangle \\ \vdots & \qquad\qquad\qquad\qquad \vdots \\ i = k & \alpha_1\langle y^1, y^k\rangle + \alpha_2\langle y^2, y^k\rangle + \cdots + \alpha_k\langle y^k, y^k\rangle = \langle x, y^k\rangle \end{array}}$$

Normal Equations $\quad G^T\alpha = \beta$

Linear Regression (Overdetermined Equations)

- **Seek** $\hat{x} = \arg\min_{m \in M} \|x - m\| = \arg\min_{A\hat{\alpha} \in M} \|b - A\hat{\alpha}\|$

- **Pick basis (columns of A)**
  - Let $M = \text{span}\{y^1, \ldots, y^k\}$:

    $\hat{x} = \alpha_1 y^1 + \alpha_1 y^1 + \ldots + \alpha_k y^k$

- **Use normal equations**

$$G^T \alpha = \beta$$

- **Get α's**

Inner Product

- Picked dot product as the inner product for linear regression
    - Connection between 2-norm and sum of squared errors (regular least squares)

Inner Product

- Picked dot product as the inner product for linear regression
  - Connection between 2-norm and sum of squared errors (regular least squares)
- Weighted least squares
  - Norm as quadratic form with positive definite matrix S

Inner Product

- Picked dot product as the inner product for linear regression
    - Connection between 2-norm and sum of squared errors (regular least squares)
- Weighted least squares
    - Norm as quadratic form with positive definite matrix S
    - The world is an uncertain place, S can help

Inner Product

- Picked dot product as the inner product for linear regression
  - Connection between 2-norm and sum of squared errors (regular least squares)
- Weighted least squares
  - Norm as quadratic form with positive definite matrix S
  - The world is an uncertain place, S can help
- Let S be an $n \times n$ positive definite matrix ($S > 0$) and let the inner product on $\mathbb{R}^n$ be

$$\langle x, y \rangle := x^T S y$$

  so that $\|x\|^2 = \langle x, x \rangle = x^T S x$.

Inner Product

- Picked dot product as the inner product for linear regression
  - Connection between 2-norm and sum of squared errors (regular least squares)
- Weighted least squares
  - Norm as quadratic form with positive definite matrix S
  - The world is an uncertain place, S can help
- Let S be an $n \times n$ positive definite matrix ($S > 0$) and let the inner product on $\mathbb{R}^n$ be

$$\langle x, y \rangle := x^T S y$$

so that $\|x\|^2 = \langle x, x \rangle = x^T S x$.

$$\hat{\alpha} = (A^T A)^{-1} A^T b$$

Inner Product

- Picked dot product as the inner product for linear regression
  - Connection between 2-norm and sum of squared errors (regular least squares)
- Weighted least squares
  - Norm as quadratic form with positive definite matrix S
  - The world is an uncertain place, S can help
- Let S be an $n \times n$ positive definite matrix ($S > 0$) and let the inner product on $\mathbb{R}^n$ be

$$\langle x, y \rangle := x^T S y$$

so that $\|x\|^2 = \langle x, x \rangle = x^T S x$.

$$\hat{\alpha} = (A^T S A)^{-1} A^T S b$$

Regular Least Squares

- Let the inner product on $\mathbb{R}^n$ be

$$\langle x, y \rangle := x^T y$$

so that $\|x\|^2 = \langle x, x \rangle = x^T x$.

$$G_{ij} =< A_i, A_j >= A_i^T A_j = [A^T A]_{ij}$$

$$\beta_i =< A_i, b >= A_i^T b = [A^T b]_i$$

Weighted Least Squares

- Let S be an $n \times n$ positive definite matrix ($S > 0$) and let the inner product on $\mathbb{R}^n$ be

$$\langle x, y \rangle := x^T S y$$

so that $\|x\|^2 = \langle x, x \rangle = x^T S x$.

$$G_{ij} =< A_i, A_j >= A_i^T S A_j = [A^T S A]_{ij}$$

$$\beta_i =< A_i, b >= A_i^T S b = [A^T S b]_i$$

Weighted Least Squares

- Let S be an $n \times n$ positive definite matrix ($S > 0$) and let the inner product on $\mathbb{R}^n$ be

$$\langle x, y \rangle := x^T S y$$

  so that $\|x\|^2 = \langle x, x \rangle = x^T S x$.

$$G_{ij} = < A_i, A_j > = A_i^T S A_j = [A^T S A]_{ij}$$

$$\beta_i = < A_i, b > = A_i^T S b = [A^T S b]_i$$

- **One last thing to verify**

$$G(y^1, \cdots, y^k) := \begin{bmatrix} \langle y^1, y^1 \rangle & \langle y^1, y^2 \rangle & \cdots & \langle y^1, y^k \rangle \\ \langle y^2, y^1 \rangle & \langle y^2, y^2 \rangle & \cdots & \langle y^2, y^k \rangle \\ \vdots & & & \vdots \\ \langle y^k, y^1 \rangle & \langle y^k, y^2 \rangle & \cdots & \langle y^k, y^k \rangle \end{bmatrix} \quad \alpha := \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}, \ \beta := \begin{bmatrix} \langle x, y^1 \rangle \\ \langle x, y^2 \rangle \\ \vdots \\ \langle x, y^k \rangle \end{bmatrix} =: \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

Is this invertible?

Gram Matrix

$G_{ij} := \langle y^i, y^j \rangle$

If real: $G^T = G$

Normal Equations $$G^T \alpha = \beta$$

Proposition 3.41 (Invertibility of the Gram Matrix)

- Let $g(y^1, \ldots, y^k) := \det G(y^1, \ldots, y^k)$ be the **determinant** of the **Gram Matrix**.

  Then $g(y^1, \ldots, y^k) \neq 0$ if and only if $\{y^1, \ldots, y^k\}$ is **linearly independent**.

$$G(y^1, \cdots, y^k) := \begin{bmatrix} \langle y^1, y^1 \rangle & \langle y^1, y^2 \rangle & \cdots & \langle y^1, y^k \rangle \\ \langle y^2, y^1 \rangle & \langle y^2, y^2 \rangle & \cdots & \langle y^2, y^k \rangle \\ \vdots & & & \vdots \\ \langle y^k, y^1 \rangle & \langle y^k, y^2 \rangle & \cdots & \langle y^k, y^k \rangle \end{bmatrix} \quad \alpha := \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}, \ \beta := \begin{bmatrix} \langle x, y^1 \rangle \\ \langle x, y^2 \rangle \\ \vdots \\ \langle x, y^k \rangle \end{bmatrix} =: \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

Yes. A^TSA is invertible

Gram Matrix

$G_{ij} := \langle y^i, y^j \rangle$

If real: $G^T = G$

Normal Equations

$$G^T \alpha = \beta$$

Proposition 3.41 (Invertibility of the Gram Matrix)

- Let $g(y^1, \dots, y^k) := \det G(y^1, \dots, y^k)$ be the **determinant** of the **Gram Matrix**.

  Then $g(y^1, \dots, y^k) \neq 0$ if and only if $\{y^1, \dots, y^k\}$ is **linearly independent**.

Recursive Least Squares

- Consider the model

$$\textbf{\textit{Model:}} \begin{cases} y_i = C_i x + e_i, \ i = 1, 2, 3, \cdots \\ \\ C_i \in \mathbb{R}^{m \times n} \\ i = \text{time index} \\ x = \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\ y_i = \text{measurements} \in \mathbb{R}^m \\ e_i = \text{model ``mismatch''} \in \mathbb{R}^m \end{cases}$$

Recursive Least Squares

- Consider the model
- At a particular time ($i = 1, 2, 3, \ldots$) this is regular least squares

$$
\textbf{\textit{Model:}}
\begin{cases}
y_i = C_i x + e_i, \ i = 1, 2, 3, \cdots \\
\\
C_i \in \mathbb{R}^{m \times n} \\
i = \text{time index} \\
x = \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\
y_i = \text{measurements} \in \mathbb{R}^m \\
e_i = \text{model ``mismatch''} \in \mathbb{R}^m
\end{cases}
$$

Recursive Least Squares

- Consider the model
- At a particular time ($i = 1, 2, 3, \ldots$) this is regular least squares $\quad C_i x \approx y_i$

Model:
$$
\begin{cases}
y_i = C_i x + e_i, \;\; i = 1, 2, 3, \cdots \\
\\
C_i \in \mathbb{R}^{m \times n} \\
i = \text{time index} \\
x = \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\
y_i = \text{measurements} \in \mathbb{R}^m \\
e_i = \text{model ``mismatch''} \in \mathbb{R}^m
\end{cases}
$$

Recursive Least Squares

- Consider the model
- At a particular time ($i = 1, 2, 3, \ldots$) this is regular least squares $\quad C_i x \approx y_i$

Model:
$$\begin{cases} y_i = C_i x + e_i, \ i = 1, 2, 3, \cdots \\ \\ C_i \in \mathbb{R}^{m \times n} \\ i = \text{time index} \\ x = \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\ y_i = \text{measurements} \in \mathbb{R}^m \\ e_i = \text{model ``mismatch''} \in \mathbb{R}^m \end{cases}$$

- Choose inner product for weighted least squares

$$\begin{aligned} \widehat{x}_k :&= \underset{x \in \mathbb{R}^n}{\arg\min} \left(\quad (y_i - C_i x)^\top S_i (y_i - C_i x) \right) \\ &= \underset{x \in \mathbb{R}^n}{\arg\min} \left(\quad e_i^\top S_i e_i \right) \end{aligned}$$

Recursive Least Squares

- Consider the model
- At a particular time ($i = 1, 2, 3, \ldots$) this is regular least squares $\quad C_i x \approx y_i$

Model:
$$\begin{cases} y_i = C_i x + e_i, \; i = 1, 2, 3, \cdots \\ \\ C_i \in \mathbb{R}^{m \times n} \\ i = \text{time index} \\ x = \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\ y_i = \text{measurements} \in \mathbb{R}^m \\ e_i = \text{model ``mismatch''} \in \mathbb{R}^m \end{cases}$$

- Choose inner product for weighted least squares

$$\widehat{x}_k := \underset{x \in \mathbb{R}^n}{\arg\min} \left(\quad (y_i - C_i x)^\top S_i (y_i - C_i x) \right)$$
$$= \underset{x \in \mathbb{R}^n}{\arg\min} \left(\quad e_i^\top S_i e_i \right)$$

- **Now do it for all time indices at the same time**

Recursive Least Squares

- Consider the model
- At a particular time ($i = 1, 2, 3, \ldots$) this is regular least squares $\quad C_i x \approx y_i$

Model:
$$\begin{cases} y_i = C_i x + e_i, \ i = 1, 2, 3, \cdots \\ \\ C_i \in \mathbb{R}^{m \times n} \\ i = \text{time index} \\ x = \text{an unknown } \underline{\text{constant}} \text{ vector} \in \mathbb{R}^n \\ y_i = \text{measurements} \in \mathbb{R}^m \\ e_i = \text{model ``mismatch''} \in \mathbb{R}^m \end{cases}$$

- Choose inner product for weighted least squares

$$\begin{aligned} \widehat{x}_k :&= \underset{x \in \mathbb{R}^n}{\arg\min} \left(\sum_{i=1}^{k} (y_i - C_i x)^\top S_i (y_i - C_i x) \right) \\ &= \underset{x \in \mathbb{R}^n}{\arg\min} \left(\sum_{i=1}^{k} e_i^\top S_i e_i \right) \end{aligned}$$

- **Now do it for all time indices at the same time**

Recursive Least Squares: Batch Solution

- First rewrite $y_i = C_i x + e_i,\ i = 1, 2, 3, \ldots, k$ in matrix form: $Y_k = A_k x + E_k$

$$Y_k = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, A_k = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix}, E_k = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

$$R_k = \begin{bmatrix} S_1 & & & \mathbf{0} \\ & S_2 & & \\ & \mathbf{0} & \ddots & \\ & & & S_k \end{bmatrix} = \mathrm{diag}(S_1, S_2, \cdots, S_k) > 0$$

Recursive Least Squares: Batch Solution

- First rewrite $y_i = C_i x + e_i,\ i = 1, 2, 3, \ldots, k$ in matrix form: $Y_k = A_k x + E_k$

$$Y_k = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}, A_k = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix}, E_k = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

$$R_k = \begin{bmatrix} S_1 & & \mathbf{0} & \\ & S_2 & & \\ & \mathbf{0} & \ddots & \\ & & & S_k \end{bmatrix} = \mathrm{diag}(S_1, S_2, \cdots, S_k) > 0$$

Batch Solution $\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$

Recursive Least Squares: Batch Solution

- A_k too "small": Cannot use normal equations if $\mathrm{rank}(A_k) < n$
  - There is some minimum k_0 where $\mathrm{rank}(A_{k0}) = n$

Recursive Least Squares: Batch Solution

- A_k too “small”: Cannot use normal equations if $\mathrm{rank}(A_k) < n$
  - There is some minimum k_0 where $\mathrm{rank}(A_{k0}) = n$
- A_k too “big”: Memory is an issue

Recursive Least Squares: Batch Solution

- A_k too "small": Cannot use normal equations if $\text{rank}(A_k) < n$
    - There is some minimum k_0 where $\text{rank}(A_{k0}) = n$
- A_k too "big": Memory is an issue
- Can we start with the smallest ($Y_{k0} = A_{k0}x + E_{k0}$) and keep growing to $Y_k = A_k x + E_k$ for greater and greater k?

Recursive Least Squares: Batch Solution

- A_k too "small": Cannot use normal equations if $\mathrm{rank}(A_k) < n$
  - There is some minimum k_0 where $\mathrm{rank}(A_{k0}) = n$
- A_k too "big": Memory is an issue
- Can we start with the smallest ($Y_{k0} = A_{k0}x + E_{k0}$) and keep growing to $Y_k = A_k x + E_k$ for greater and greater k?
  - Is there a recursive relationship between $\hat{x}_k$ and $\hat{x}_{k+1}$

    when going from $Y_k = A_k x + E_k$ to $Y_{k+1} = A_{k+1} x + E_{k+1}$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

- **Normal equations** for k

$$\left(\sum_{i=1}^{k} C_i^\top S_i C_i\right)\widehat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

- **Normal equations** for k

$$\left(\sum_{i=1}^{k} C_i^\top S_i C_i\right)\widehat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

- **Normal equations** for $k+1$

$$(\sum_{i=1}^{k+1} C_i^\top S_i C_i)\,\widehat{x}_{k+1} = \sum_{i=1}^{k+1} C_i^\top S_i y_i$$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

- **Normal equations** for k

$$\underbrace{\left(\sum_{i=1}^{k} C_i^\top S_i C_i\right)}_{Q_k} \widehat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

- **Normal equations** for $k+1$

$$\left(\sum_{i=1}^{k+1} C_i^\top S_i C_i\right) \widehat{x}_{k+1} = \sum_{i=1}^{k+1} C_i^\top S_i y_i$$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

- **Normal equations** for k

$$\underbrace{\left(\sum_{i=1}^{k} C_i^\top S_i C_i\right)}_{Q_k} \widehat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

- **Normal equations** for $k+1$

$$\underbrace{\left(\sum_{i=1}^{k+1} C_i^\top S_i C_i\right)}_{Q_{k+1} = Q_k + C_{k+1}^\top S_{k+1} C_{k+1}} \widehat{x}_{k+1} = \sum_{i=1}^{k+1} C_i^\top S_i y_i$$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

- **Normal equations** for k

$$\underbrace{\left(\sum_{i=1}^{k} C_i^\top S_i C_i\right)}_{Q_k} \widehat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

- **Normal equations** for $k+1$

$$\underbrace{\left(\sum_{i=1}^{k+1} C_i^\top S_i C_i\right)}_{Q_{k+1} = Q_k + C_{k+1}^\top S_{k+1} C_{k+1}} \widehat{x}_{k+1} = \underbrace{\sum_{i=1}^{k+1} C_i^\top S_i y_i}_{\sum_{i=1}^{k} C_i^\top S_i y_i + C_{k+1}^\top S_{k+1} y_{k+1}}$$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

- **Normal equations** for k

$$\underbrace{\left(\sum_{i=1}^{k} C_i^\top S_i C_i\right)}_{Q_k} \widehat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

- **Normal equations** for $k+1$

$$\underbrace{\left(\sum_{i=1}^{k+1} C_i^\top S_i C_i\right)}_{Q_{k+1} = Q_k + C_{k+1}^\top S_{k+1} C_{k+1}} \widehat{x}_{k+1} = \underbrace{\sum_{i=1}^{k+1} C_i^\top S_i y_i}_{\underbrace{\sum_{i=1}^{k} C_i^\top S_i y_i}_{Q_k \widehat{x}_k} + C_{k+1}^\top S_{k+1} y_{k+1}}$$

Recursive Least Squares: Recursion

- Normal equations for $k \geq k_0$ $\quad (A_k^\top R_k A_k)\widehat{x}_k = A_k^\top R_k Y_k$

Batch Solution

$$\widehat{x}_k = (A_k^\top R_k A_k)^{-1} A_k^\top R_k Y_k$$

- **Normal equations** for k

$$\underbrace{\left(\sum_{i=1}^{k} C_i^\top S_i C_i\right)}_{Q_k} \widehat{x}_k = \sum_{i=1}^{k} C_i^\top S_i y_i$$

- **Normal equations** for $k+1$

$$\underbrace{\left(\sum_{i=1}^{k+1} C_i^\top S_i C_i\right)}_{Q_{k+1} = Q_k + C_{k+1}^\top S_{k+1} C_{k+1}} \widehat{x}_{k+1} = \underbrace{\sum_{i=1}^{k+1} C_i^\top S_i y_i}_{\underbrace{\sum_{i=1}^{k} C_i^\top S_i y_i}_{Q_k \widehat{x}_k} + C_{k+1}^\top S_{k+1} y_{k+1}}$$

$$\boxed{\begin{aligned} Q_{k+1} &= Q_k + C_{k+1}^\top S_{k+1} C_{k+1} \\ Q_{k+1}\widehat{x}_{k+1} &= Q_k \widehat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}. \end{aligned}}$$

Recursive Least Squares: Recursion

$$\boxed{\begin{aligned} Q_{k+1} &= Q_k + C_{k+1}^\top S_{k+1} C_{k+1} \\ Q_{k+1}\widehat{x}_{k+1} &= Q_k \widehat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}. \end{aligned}}$$

$$\widehat{x}_{k+1} = Q_{k+1}^{-1}\left[Q_k \widehat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}\right]$$

$$\widehat{x}_{k+1} = \widehat{x}_k + \underbrace{Q_{k+1}^{-1} C_{k+1}^\top S_{k+1}}_{\text{Kalman gain}} \underbrace{\left(y_{k+1} - C_{k+1}\widehat{x}_k\right)}_{\text{Innovations}}$$

Recursive Least Squares: Recursion

$$\boxed{\begin{aligned} Q_{k+1} &= Q_k + C_{k+1}^\top S_{k+1} C_{k+1} \\ Q_{k+1}\widehat{x}_{k+1} &= Q_k \widehat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1}. \end{aligned}}$$

$$\widehat{x}_{k+1} = Q_{k+1}^{-1} \left[Q_k \widehat{x}_k + C_{k+1}^\top S_{k+1} y_{k+1} \right]$$

$$\widehat{x}_{k+1} = \widehat{x}_k + \underbrace{Q_{k+1}^{-1} C_{k+1}^\top S_{k+1}}_{\text{Kalman gain}} \underbrace{\left(y_{k+1} - C_{k+1} \widehat{x}_k \right)}_{\text{Innovations}}$$

- Sneak peek at Kalman Filter

Kalman Filter: Sneak Peek

- **Model:** Linear time-varying discrete-time system with “white” Gaussian noise

$$
\begin{aligned}
x_{k+1} &= A_k x_k + G_k w_k, \quad x_0 \text{ initial condition} \\
y_k &= C_k x_k + v_k
\end{aligned}
$$

where $x \in \mathbb{R}^n$, $w \in \mathbb{R}^p$, $y \in \mathbb{R}^m$, $v \in \mathbb{R}^m$. Moreover, the random vectors x_0, and, for $k \geq 0$, w_k, v_k are all independent Gaussian (normal) random vectors.

In a real-time implementation, computing the inverse of Q_{k+1} can be time consuming. An attractive alternative can be obtained by applying the **Matrix Inversion Lemma**,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(DA^{-1}B + C^{-1}\right)^{-1}DA^{-1}$$

Now, following the substitution rule

$$A \leftrightarrow Q_k \quad B \leftrightarrow C_{k+1}^\top \quad C \leftrightarrow S_{k+1} \quad D \leftrightarrow C_{k+1},$$

yields, after some tedious calculations,

$$\begin{aligned} Q_{k+1}^{-1} &= \left(Q_k + C_k^\top S_{k+1}C_{k+1}\right)^{-1} \\ &= Q_k^{-1} - Q_k^{-1}C_{k+1}^\top\left[C_{k+1}Q_k^{-1}C_{k+1}^\top + S_{k+1}^{-1}\right]^{-1}C_{k+1}Q_k^{-1}, \end{aligned}$$

which is a recursion for Q_k^{-1}. Upon defining

$$P_k = Q_k^{-1},$$

we have

$$P_{k+1} = P_k - P_kC_{k+1}^\top\left[C_{k+1}P_kC_{k+1}^\top + S_{k+1}^{-1}\right]^{-1}C_{k+1}P_k$$

We note that we are now inverting a matrix that is $m \times m$, instead of one that is $n \times n$. Typically, $n >> m$ (means very much greater than), and thus the savings can be important.

■

Probability Space: The *RIGHT* way to begin

- Disclaimer: This part is just for fun.

Definition: Probability Space

- $(\Omega, \mathcal{F}, P)$ is called a **probability space**.
  - Ω is the sample space. Think of it as the set of all possible outcomes of an experiment.
  - $E \subset \Omega$ is an event
  - $\mathcal{F}$ is the collection of allowed events. It must at least contain $\emptyset$ and Ω. It is closed with respect to set complement, countable unions, and countable intersections. Such sets are called sigma algebras.
  - $P: \mathcal{F} \to [0, 1]$ is a probability measure. It has to satisfy a few basic operations:

1. $P(\emptyset) = 0$ *and* $P(\Omega) = 1$.
2. *For each* $E \in \mathscr{F}$, $0 \leq P(E) \leq 1$
3. *If the sets* $E_1, E_2, \ldots$ *are disjoint (i.e.,* $E_i \cap E_j = \emptyset$ *for* $i \neq j$*), then* $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

Definition: Probability Space

- $(\Omega, \mathcal{F}, P)$ is called a **probability space**.
    - Ω is the sample space. Think of it as the set of all possible outcomes of an experiment.
    - $E \subset \Omega$ is an event
    - $\mathcal{F}$ is the collection of allowed events. It must at least contain $\emptyset$ and Ω. It is closed with respect to set complement, countable unions, and countable intersections. Such sets are called sigma algebras.
    - $P: \mathcal{F} \to [0, 1]$ is a probability measure. It has to satisfy a few basic operations:

These are typically called the **Axioms of Probability**

1. $P(\emptyset) = 0$ *and* $P(\Omega) = 1$.
2. *For each* $E \in \mathscr{F}$, $0 \leq P(E) \leq 1$
3. *If the sets* $E_1, E_2, \ldots$ *are disjoint (i.e.,* $E_i \cap E_j = \emptyset$ *for* $i \neq j$*), then* $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

Why go through this?

- Rolling a die is easy. You can count outcomes and assign probabilities to each outcome.
- Many questions don't lead themselves to outcomes that are easy to count.
  - What is the probability of thermonuclear war? (Neither 0% nor 100%)
    - How do we find a way to **reasonably** assign a number between 0 and 1?

Minimal Introduction to Probability

- **Overall objective:** minimal set of knowledge such that the rest of the course makes sense
- This week is an extremely brief taste of probability
- Moving from counting (think about rolling a 6-sided die) to integrating

Rolling a Die (Discrete)

- **Sample space** Ω :
  - {“rolling a 1”, “rolling a 2”, “rolling a 3”, “rolling a 4”, “rolling a 5”, “rolling a 6”}
- **Probability** of an **event**
  - P(“rolling a 1”) $= 1/6$
  - P(“rolling an even number”) $= 3/6$
- **Expected value**: e.g., average points rolled
  - Assign point value to each event
  - Sum of point value × event probability

$$1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=\frac{21}{6}$$

Probability Distribution (Continuous)

- **Sample space** Ω :
  - Real numbers $x \in [-\infty, \infty]$

- **Probability** of an **event**
  - Integral of a **probability density function** $f(x)$

$$P(x \in [c, d]) = \int_c^d f(x)dx$$

$$P(x \in [-\infty, \infty]) = 1 = \int_{-\infty}^{\infty} f(x)dx$$

- **Expectation operator**: e.g., average function value $g(x)$
  - Assign $g(x)$ to each $x \in [-\infty, \infty]$
  - Integral of $g(x) \times f(x)$

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f(x)dx$$