112-1
Linear Algebra and its Applications

Review before the final exam

Prof. Argon Chen

TAs

2023/12/18

- Eigenvalues:
  - Why express vector by combination of eigenvectors p.3
  - How many eigenvalues? p.8
  - Projection matrix
- Matrix diagonalization:
  - Schur's lemma
  - P.13 [1 -1], [-1 1] have no difference
- PCA and Classification:
  - p.1-2
  - Rayleigh's quotient
  - P.10 Sum of normalized measurement variance = $\sum_{j=1}^{p} \frac{\sum_{k=1}^{n}\left(x_{kj}-\bar{x}_j\right)^2}{\sum_{k=1}^{n}\left(x_{kj}-\bar{x}_j\right)^2} = \sum_{j=1}^{p} 1 = p = 53$
- SVD:
  - P.4 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ … ? (Q_1, Q_2 order matters)
  - P.7 example 2
  - A^+ and row space component
  - Optimal solution (minimum length)

Eigenvalues

Ex: $A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ has $\lambda_1 = 3$ with $x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\lambda_2 = 2$ with $x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

How about $x = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$**? Not an eigenvector!**

Let $x = x_1 + 5x_2 \Rightarrow$ $Ax = \lambda_1 x_1 + 5\lambda_2 x_2 = \begin{bmatrix} 3 \\ 10 \end{bmatrix}$. **The action of A can still be determined by eigenvalues and eigenvectors!**

Easy!

$Bx = \lambda x$ (matrix times vector = **scalar times vector**)

$B = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}, \lambda_{1,2} = 5, -1, x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

What matrix B does is **scaling** the vectors in the direction of x_1 and x_2.

$x^\star = \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \frac{13}{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \frac{13}{3}x_1 + \frac{2}{3}x_2$

(1) $Bx^\star = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 3 \\ 5 \end{bmatrix} = \cdots = \begin{bmatrix} 23 \\ 21 \end{bmatrix}$

(2) $Bx^\star = B\left(\frac{13}{3}x_1 + \frac{2}{3}x_2\right) = \frac{13}{3} \times (5 \times x_1) + \frac{2}{3} \times (-1 \times x_2) = \frac{65}{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 23 \\ 21 \end{bmatrix}$

Eigenvalues

$B = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$; $\lambda_{1,2} = 5, -1$; $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

$Bx = \lambda x,\ \det(B - \lambda I) = 0, (1-\lambda)(3-\lambda) - 8 = 0,\ \lambda^2 + 4\lambda - 5 = 0$

- How many eigenvalues do $A_{n\times n}$ have?
  - Short answer: ***n*** (Solving the n-th characteristic polynomial)
  - However, they might repeat:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \lambda_{1,2} = 1, 1;\ x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

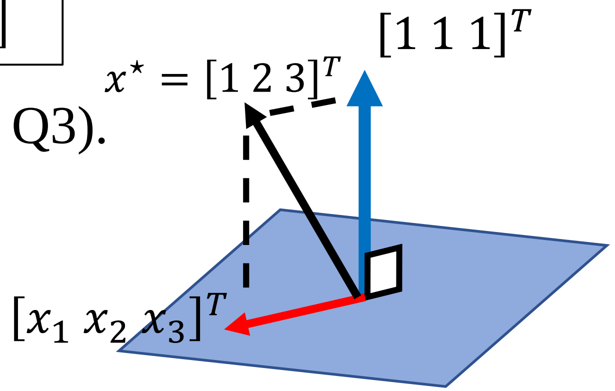
  - and some might cause problems (defective):

$$A_1 = \begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}; \lambda_{1,2} = 0, 0;\ x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{(HW\#9 Q8)}$$

Eigenvalues

Let $u = [1\ 1\ 1]^T$, the projection matrix P is

$$P = I - \frac{uu^T}{u^Tu} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$



- For example, projecting any vector in R^3 onto $x_1 + x_2 + x_3 = 0$ (HW#9 Q3).

$$x_1 + x_2 + x_3 = [1\ 1\ 1]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$$

- Eigenvalues and eigenvectors are

$$\lambda_{1,2} = 1, 1; v_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}. \qquad \lambda_3 = 0; v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- Suppose we have $x^\star = [1\ 2\ 3]^T$, what $Px^\star$ does is
  - $x^\star = -v_2 + 2v_3$
  - $Px^\star = P(-v_2 + 2v_3) = -(1 \times v_2) + 2(0 \times v_3) = -v_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

Properties of projection matrix P:
- $P^T = P$ (symmetric)
- $P^2 = P$ (idempotent)

$Px = \lambda x, P^2x = \lambda Px$
$\lambda x = \lambda^2 x, \lambda(\lambda - 1)x = 0 \Rightarrow \lambda = 0 \vee 1$

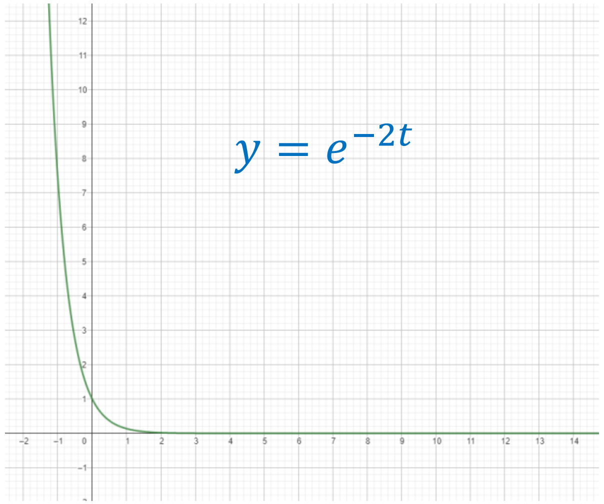
- Why its eigenspace is composed of column space and null space? $\boldsymbol{Ax = \lambda x}$

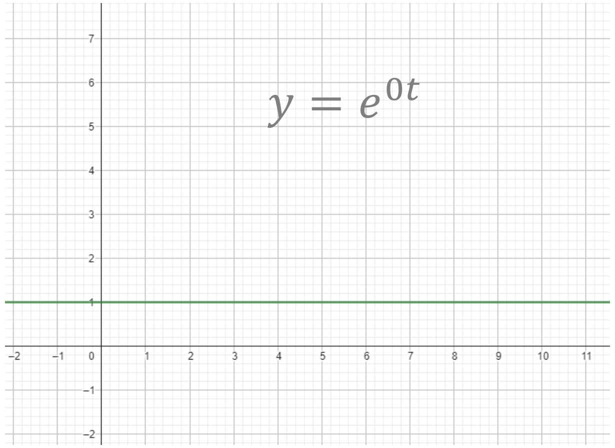
Eigenvalues

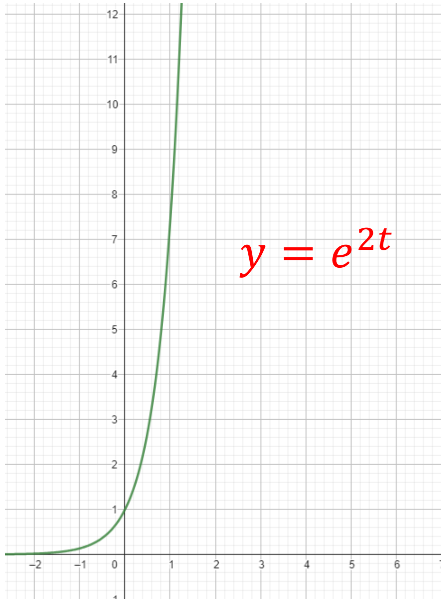
- How to know $u(t) = e^{At} \cdot u_0$ will blow up to infinity or stay stable?

$u(t) = e^{At} \cdot u_0 = \sum_{i=1}^{n} c_i e^{\lambda_i t} x_i$ where $c = S^{-1} u_0$

- If all $Re(\lambda_i) < 0$, then the system is stable.
- If all $Re(\lambda_i) \leq 0$ and some $Re(\lambda_i) = 0$, then the system is neutrally stable.
- If any $Re(\lambda_i) > 0$, then the system is unstable.







$U^H U = UU^H = I$ (*complex* version of orthogonal matrix)

Matrix diagonalization

Schur's Lemma – Trianglizing by a Unitary M

- **For any matrix A, there is a unitary matrix $M=U$ such that $U^{-1}AU=T$ is upper triangular. The eigenvalues of A, shared by the similar matrix T, appear along its diagonal.**

***Proof*: Take 4 by 4 matrix A as an example. There is at least one eigenvalue λ_1. Unitary matrix U_1 can be constructed by the corresponding eigenvector x_1 as the first vector and by Gram-Schmidt process for the subsequent 3 vectors. Then,**

$$AU_1 = U_1\begin{bmatrix} \lambda_1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix} \text{ or } U_1^{-1}AU_1 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix}$$

$$\boxed{Ax_1 = \lambda_1 x_1} \Rightarrow U_1 = \begin{bmatrix} | & * & * & * \\ x_1 & * & * & * \\ | & * & * & * \\ & * & * & * \end{bmatrix}$$

Gram-Schmidt

Second step: look at the lower right 3 by 3 matrix and there exists at least one eigenvalue λ_2 so that its corresponding eigenvector and two perpendicular vectors (Gram-Schmidt) can form M_2 and

$$U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & M_2 & \\ 0 & & & \end{bmatrix} \text{ such that } U_2^{-1}\left(U_1^{-1}AU_1\right)U_2 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{bmatrix}$$

1. Remain first column of $U_1^{-1}AU_1$
2. Not involve the first column of $U_1^{-1}AU_1$

Last step: $U_3^{-1}\left(U_2^{-1}U_1^{-1}AU_1U_2\right)U_3 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & \lambda_3 & * \\ 0 & 0 & 0 & * \end{bmatrix} = T$

- **The product $U=U_1U_2U_3$ is still unitary (exercise)**

Matrix diagonalization

$$\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \to \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{bmatrix} \to \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{bmatrix} \to \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & * & * \\ 0 & 0 & \lambda_3 & * \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}$$

- Give a 4×4 example:

$$A = \begin{bmatrix} 4 & 0 & -2 & 0 \\ 0 & 7 & 0 & -12 \\ 4 & 0 & -2 & 0 \\ 0 & 2 & 0 & -3 \end{bmatrix}; \ \lambda_1 = 0; x_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix} \to \begin{bmatrix} \frac{1}{\sqrt{5}} \\ 0 \\ \frac{2}{\sqrt{5}} \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & -6 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 7 & -12 \\ 0 & 0 & 2 & -3 \end{bmatrix}$$

$$U_1 = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ U_1^T A U_1 = \begin{bmatrix} 0 & -6 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 7 & -12 \\ 0 & 0 & 2 & -3 \end{bmatrix} \Rightarrow U_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, U_1^T A U_1 = U_2^T (U_1^T A U_1) U_2$$

//

$$(7-\lambda)(-3-\lambda) + 24 = 0, \lambda_{1,2} = 3, 1; x_3 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \to \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix}, U_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{3}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \\ 0 & 0 & \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \end{bmatrix}, U^T A U = \begin{bmatrix} 0 & -6 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & -14 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $U = U_1 U_2 U_3$

Matrix diagonalization

- (p.13)

Example: $A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$ **has the eigenvalue λ=1 (twice)**

$$U^{-1}AU = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

If $U = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$, then

$$U^{-1}AU = U^{T}AU = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$$

Matrix diagonalization

Diagonalizing General Matrices – Jordan Form

- **Goal: make $M^{-1}AM$ as nearly diagonal as possible**
- **If A has s independent eigenvectors, it is similar to a matrix with s blocks:**

M is any matrix

$$J = M^{-1}AM = \begin{bmatrix} J_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ & & & & J_s \end{bmatrix}$$

Each Jordan block J_i is a triangular matrix with only a single eigenvalue λ_i and only one eigenvector.

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \cdot & \cdot & \\ & & \cdot & 1 \\ & & & \lambda_i \end{bmatrix}$$

Jordan Form - Examples

Example 1: $T = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ **and** $A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$ **and** $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ **share**

$$J = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$M^{-1}TM = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = J \quad \mathbf{(2 \to 1)}$$

$U^{-1}AU = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = T$ and then $M^{-1}TM = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = J$ **(make triangular and then $2 \to 1$)**

$$P^{-1}BP = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = J \quad \textbf{(permutations)}$$

The three matrices are similar because they share the same Jordan form.

- A is positive definite implies $A^T = A$; $\lambda_i > 0 \ \forall i = 1, \ldots, n$; Invertible

PCA and Classification (p.1-2)

Minimizing a quadratic function $P(x)$ is equivalent to solving $Ax = b$.

6H If A is symmetric positive definite, then $P(x) = \frac{1}{2}x^{\mathrm{T}}Ax - x^{\mathrm{T}}b$ reaches its minimum at the point where $Ax = b$. At that point $P_{\min} = -\frac{1}{2}b^{\mathrm{T}}A^{-1}b$.

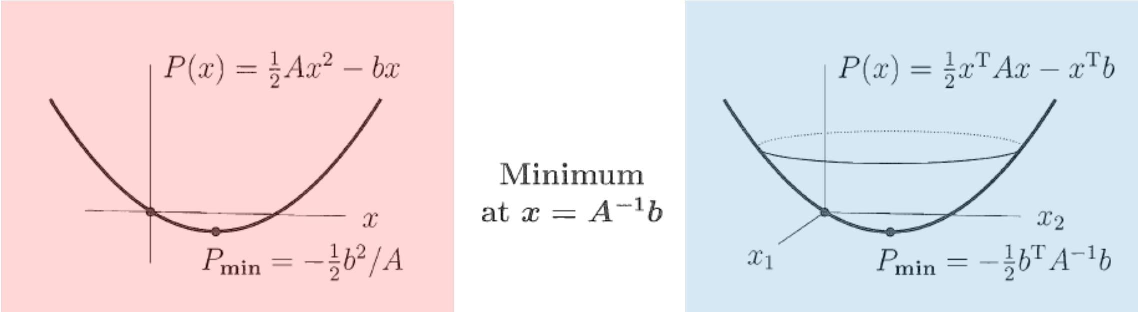
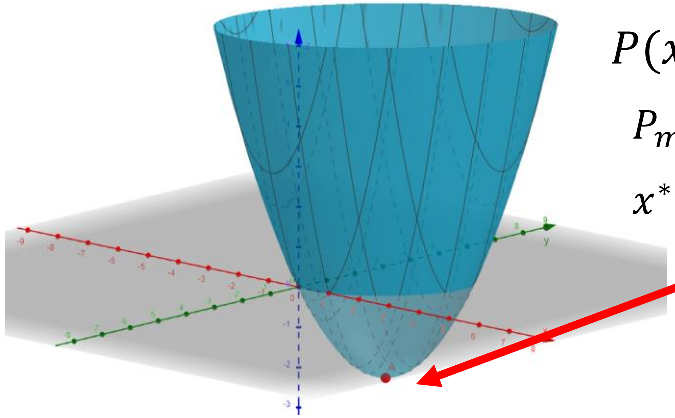


Figure 6.4: The graph of a positive quadratic $P(x)$ is a parabolic bowl.

- Visualize the positive quadratic function
- In 2-D, it's a parabola
  - $P(x) = \frac{1}{2}Ax^2 - bx, x^* = b/A$
  - $P_{min} = -\frac{1}{2}b^2/A$
- In 3-D, it's a bowl
  - $P(x) = \frac{1}{2}x^TAx - x^Tb, x^* = A^{-1}b$
  - $P_{min} = -\frac{1}{2}b^TA^{-1}b$

Proof. Suppose $Ax = b$. For any vector y, we show that $P(y) \geq P(x)$:

$$\begin{aligned} P(y) - P(x) &= \frac{1}{2}y^{\mathrm{T}}Ay - y^{\mathrm{T}}b - \frac{1}{2}x^{\mathrm{T}}Ax + x^{\mathrm{T}}b \\ &= \frac{1}{2}y^{\mathrm{T}}Ay - y^{\mathrm{T}}Ax + \frac{1}{2}x^{\mathrm{T}}Ax \quad (\text{set } b = Ax) \qquad (1) \\ &= \frac{1}{2}(y-x)^{\mathrm{T}}A(y-x). \end{aligned}$$



$P(x)$

$P_{min} = -\frac{1}{2}b^TA^{-1}b = -\frac{7}{3}$

$x^* = \left(\frac{4}{3}, \frac{5}{3}\right)$

Example: Minimize $P(x) = \boxed{x_1^2 - x_1x_2 + x_2^2 - x_1 - 2x_2}$

Calculus:
$\partial P/\partial x_1 = 2x_1 - x_2 - 1 = 0$
$\partial P/\partial x_2 = -x_1 + 2x_2 - 2 = 0$

Linear algebra: solve $Ax=b$ where $P(x) = \frac{1}{2}x^{\mathrm{T}}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}x - x^{\mathrm{T}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

PCA and Classification (p.1-2)

Minimum/Maximum and Solving $Ax=\lambda x$

- ***Rayleigh's quotient*:**

$$\boxed{R(x) = \frac{x^T A x}{x^T x}}$$

- ***Rayleigh's Principle:***

The quotient $R(x)$ is maximized by the first eigenvector $x=x_1$ of A corresponding to the largest eigenvalue λ_1 and its maximum value is λ_1:

$$\boxed{R(x_1) = \frac{x_1^T A x_1}{x_1^T x_1} = \frac{x_1^T \lambda_1 x_1}{x_1^T x_1} = \lambda_1}$$

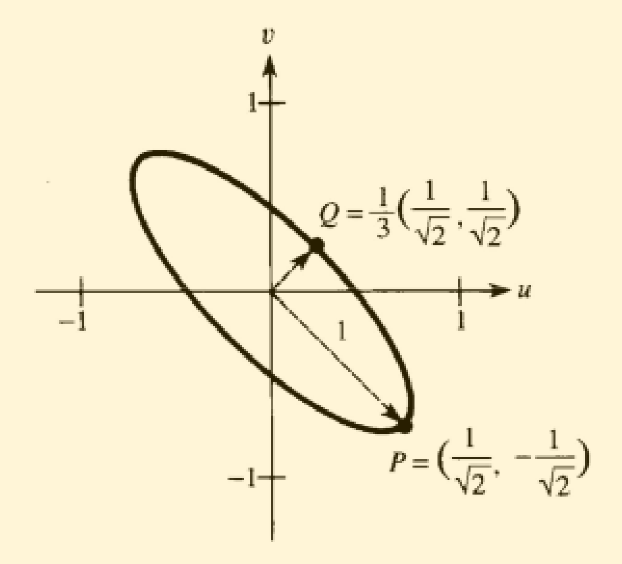
Geometrically:

Fix numerator at 1: $x^T Ax=1$ ellipsoid

$\Rightarrow$ denominator $x^T x=||x||^2$ as small as possible $\Rightarrow$ shortest axis $\Rightarrow$ smallest $1/\sqrt{\lambda_i}$ $\Rightarrow$ largest eigenvalue λ_1

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1 \text{ (ellipsoid)}$$
$$= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1/a^2 & 0 \\ 0 & 1/b^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Minimizing a Rayleigh's quotient $R(x)$ is equivalent to solving $Ax = \lambda x$.



Eigenvalues/vectors: 1, $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ and 9, $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

PCA and Classification (p.1-2)

Algebraically: **Diagonalize $A \Rightarrow A = Q\Lambda Q^T$ ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$)**

$$R(x) = \frac{(Q^T x)^T \Lambda (Q^T x)}{(Q^T x)^T (Q^T x)} = \frac{y^T \Lambda y}{y^T y} = \frac{\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2}{y_1^2 + \cdots + y_n^2} \boxed{\leq} \lambda_1 \ (\geq \lambda_n)$$

since $\lambda_1 (y_1^2 + y_2^2 + \cdots + y_n^2) \geq (\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2)$

The maximum R must be at $y_1=1$ and $y_2=y_3=\ldots=y_n=0$

- **Rayleigh quotient is never above λ_1 and never below λ_n**

- When $y = [1\ 0\ \ldots 0]^T$, $x = Qy$
  $=$ the first eigenvector corresponding to λ_1 (max)
- When $y = [0\ 0\ \ldots 1]^T$, $x = Qy$
  $=$ the last eigenvector corresponding to λ_n (min)

PCA and Classification

$$y_{ki} = \frac{x_{ki} - \bar{x}_i}{\sigma_i}, \text{ where } \sigma_i = \sqrt{\frac{1}{65-1}\sum_{k=1}^{65}(x_{ki} - \bar{x}_i)^2}$$

- (p.10) Sum of normalized measurement variances

  (1) = Trace(ρ) = Sum of λ_i

  (2) = Sum of all principal components variances = 53

(2)
- All diagonal elements are 1 in ρ.
- In addition, the first principal component has the largest variance λ_1
- Why the variances are additive?
  - All principal components are orthogonal to each other.

    $e_i^T e_j^T = 0 \ \forall i \neq j$

(1)
- **Let** $\mathbf{B} = \begin{bmatrix} y_{11} & \cdots & y_{1,53} \\ \vdots & \vdots & \vdots \\ y_{65,1} & \cdots & y_{65,53} \end{bmatrix}$

Then,

$$\rho = \begin{bmatrix} \rho_{11} = \dfrac{\sum_{k=1}^{65}(x_{k1}-\bar{x}_1)(x_{k1}-\bar{x}_1)}{\sqrt{\sum_{k=1}^{65}(x_{k1}-\bar{x}_1)^2}\sqrt{\sum_{k=1}^{65}(x_{k1}-\bar{x}_1)^2}} = 1 & \rho_{12} & \cdots & \rho_{1,53} \\ \rho_{21} = \dfrac{\sum_{k=1}^{65}(x_{k2}-\bar{x}_2)(x_{k1}-\bar{x}_1)}{\sqrt{\sum_{k=1}^{65}(x_{k2}-\bar{x}_2)^2}\sqrt{\sum_{k=1}^{65}(x_{k1}-\bar{x}_1)^2}} & \rho_{22} & & \rho_{2,53} \\ \vdots & & \ddots & \vdots \\ \rho_{53,1} & \rho_{53,2} & \cdots & \rho_{53,53} \end{bmatrix}$$

$$= \frac{1}{65-1}\mathbf{B}^T\mathbf{B}$$

(1) λ_i is the i-th eigenvalue of ρ, and $trace(\rho) = \sum_{i=1}^{53}\lambda_i$

SVD

$$A_{m\times n} = Q_1 \Sigma Q_2^T = \underbrace{\begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix}}_{(m\times m)} \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix}}_{(m\times n)} \underbrace{\begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}^T}_{(n\times n)}$$

- (p.4) Normally, we will let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ $(\sigma_i = \sqrt{\lambda_i}$, where $A^T A v_i = \lambda_i v_i)$
- But, as long as the “paired” relationship is established, the SVD should work

$$\begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} \cdot \begin{pmatrix} \boxed{\sqrt{30}} & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{\sqrt{30}}{30} & \frac{-2\sqrt{5}}{5} & \frac{-3\sqrt{70}}{70} & \frac{-2\sqrt{105}}{105} \\ \frac{\sqrt{30}}{15} & \frac{\sqrt{5}}{5} & \frac{-3\sqrt{70}}{35} & \frac{-4\sqrt{105}}{105} \\ \frac{\sqrt{30}}{10} & 0 & \frac{\sqrt{70}}{14} & \frac{-2\sqrt{105}}{35} \\ \frac{2\sqrt{30}}{15} & 0 & 0 & \frac{\sqrt{105}}{15} \end{pmatrix}^T$$

$$\begin{pmatrix} 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 & \boxed{30^{0.5}} \end{pmatrix} \cdot \begin{pmatrix} \frac{-2\cdot 105^{0.5}}{105} & \frac{-4\cdot 105^{0.5}}{105} & \frac{-2\cdot 105^{0.5}}{35} & \frac{105^{0.5}}{15} \\ \frac{-2\cdot 5^{0.5}}{5} & \frac{5^{0.5}}{5} & 0 & 0 \\ \frac{-3\cdot 70^{0.5}}{70} & \frac{-3\cdot 70^{0.5}}{35} & \frac{70^{0.5}}{14} & 0 \\ \frac{30^{0.5}}{30} & \frac{30^{0.5}}{15} & \frac{30^{0.5}}{10} & \frac{2\cdot 30^{0.5}}{15} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}$$

SVD

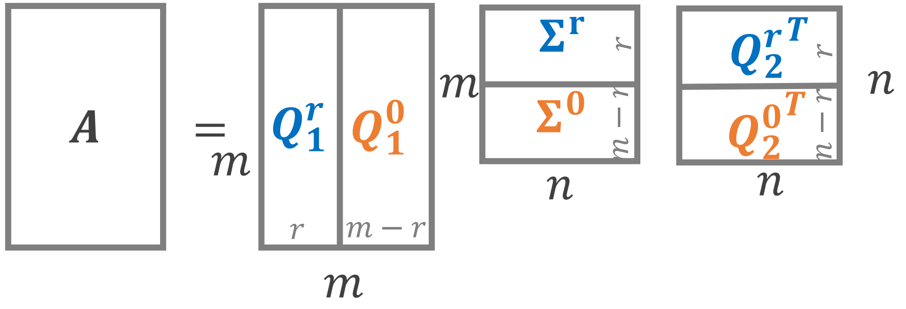
$$A = Q_1 \Sigma Q_2^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

$$Av_k = \sum_{i=1}^{r} \sigma_i u_i v_i^T v_k = \sigma_k u_k \Rightarrow u_k = \frac{1}{\sigma_k} A v_k \text{ (columns combination)}$$

$$A^T u_k = \sum_{i=1}^{r} \sigma_i v_i u_i^T u_k = \sigma_k v_k \Rightarrow v_k = \frac{1}{\sigma_k} A^T u_k \text{ (rows combination)}$$

Therefore, we know that

- u_k are in $R(A)$ for $k = 1, \ldots, r$
- v_k are in $R(A^T)$ for $k = 1, \ldots, r$



SVD

$A = Q_1 \Sigma Q_2^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$, since $v_i^T v_j = 0$ and $u_i^T u_j = 0 \ \forall\ i \neq j$, we have:

- $Av_i = \sigma_i u_i \Rightarrow u_i = \frac{1}{\sigma_i} A v_i$
- $A^T u_i = \sigma_i v_i \Rightarrow v_i = \frac{1}{\sigma_i} A^T u_i$

$$A^T = \sum_{i=1}^{r} \sigma_i v_i u_i^T$$

- (p.7) example 2:

$A = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$, since $m = 3 > 1 = n$, we start from $A^T A = 9 \Rightarrow \lambda_1 = 9, \sigma_1 = 3; \lambda_{2,3} = 0$ and $Q_2 = [1]$

Also, we by $AQ_2 = Q_1 \Sigma$, $\quad Av_1 = u_1 \cdot \sigma_1 \Rightarrow u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{3} \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$

u_2 and u_3 are in $N(A^T)$, solve $-x_1 + 2x_2 + x_3 = 0$ and we have $Q_1 = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$ (and probably with Gram-Schmidt to and them orthonormal)

Finally, we have $A = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} [1]$

SVD

Optimal Solution of $Ax=b$

- $Ax=b$

  (1) **Rows of A are dependent $\Rightarrow$ very likely no solution (b is not in the column space of A) $\Rightarrow A\hat{x}=p \Rightarrow A^T A\hat{x} = A^T b$**

  (2) **Columns of A are dependent? $A^T A$ not invertible with null space $\Rightarrow$ No unique Solution!**

- Optimal solution of $Ax = b$ under (1) and (2) is defined as $\hat{x}$ (with the **minimum length**)
- Recall: $x_c = x_p + x_n$ (but x_p also could have null space component)
- $\hat{x} = x_r + x_n$, $||\hat{x}||^2 = ||x_r||^2 + ||x_n||$^2
- How to make x have the shortest $||x||^2 = x^T x$?
  - Get rid of the **null space** component!

How to find x^+? Let us begin from a simple case:

Example 5. A is diagonal, with dependent rows and dependent columns:

$$A\hat{x} = p \quad \text{is} \quad \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \\ \hat{x}_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ 0 \end{bmatrix}.$$

- The optimal solution should be $\hat{x} = \begin{bmatrix} \frac{b_1}{\sigma_1} & \frac{b_2}{\sigma_2} & 0 & 0 \end{bmatrix}^T$
- And also, we find that

$$\hat{x} = x^+ = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{4\times 3} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{\Sigma}^+ b,$$

where $\Sigma = \boldsymbol{A} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{3\times 4}$

SVD

- In general, we can say $x^+ = A^+ b$, where $\boxed{A^+ = Q_2 \Sigma^+ Q_1^T}$

Proof:

$Ax = b \Rightarrow (Q_1 \Sigma Q_2^T)x = b$

$\Sigma(Q_2^T x) = Q_1^T b$, (let $y = Q_2^T x$)

$\boxed{\Sigma y = Q_1^T b}$

From the simple case, we know $x^+ = \Sigma^+ b$ if we want to solve $\Sigma x = b$

Therefore, we have $y^+ = \Sigma^+ Q_1^T b$, and thus $x^+ = Q_2 \Sigma^+ Q_1^T b = A^+ b$

SVD

- **Example 2: A=[-1 2 2]**

  Q_1= [1] with singular value=3:

$$A = \begin{bmatrix}-1 & 2 & 2\end{bmatrix} = [1]\begin{bmatrix}3 & 0 & 0\end{bmatrix}\begin{bmatrix}-\frac{1}{3} & \frac{2}{3} & \frac{2}{3}\\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}\\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3}\end{bmatrix}$$

$$A^+ = \begin{bmatrix}-1 & 2 & 2\end{bmatrix}^+ = \begin{bmatrix}-\frac{1}{3} & \frac{2}{3} & \frac{2}{3}\\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}\\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3}\end{bmatrix}\begin{bmatrix}\frac{1}{3}\\ 0\\ 0\end{bmatrix}[1] = \begin{bmatrix}-\frac{1}{9}\\ \frac{2}{9}\\ \frac{2}{9}\end{bmatrix}$$

For example, $b = [9]^T$ and we have to solve $Ax = b$ and find the optimal solution:

$$x^+ = A^+b = \begin{bmatrix}-1\\ 2\\ 2\end{bmatrix},$$

and it is orthogonal to $N(A) \in span\left(\begin{bmatrix}2\\ 1\\ 0\end{bmatrix}, \begin{bmatrix}0\\ 1\\ -1\end{bmatrix}\right)$

$$[-1\ \ 2\ \ 2]^T \begin{bmatrix}2\\ 1\\ 0\end{bmatrix} = 0$$

$$[-1\ \ 2\ \ 2]^T \begin{bmatrix}0\\ 1\\ -1\end{bmatrix} = 0.$$

Therefore, we confirm that

$$x^+ \in R(A^T)$$