From Singular Values to Canonical Angles

Li Qiu

Hong Kong University of Science and Technology

Sept 2016

Outline

Singular value decomposition (SVD)

- For matrix $A \in \mathbb{C}^{m \times n}$, there are unitary matrices U and V such that

$$A = U\mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{\min\{m,n\}})V^*$$

  where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$.
- We call $\sigma_i, i = 1, \ldots, \min\{m, n\}$, the singular values of A, denoted by $\sigma_i(A)$.

Unitarily invariant norms

- We say a norm $\|\cdot\|$ on $\mathbb{C}^{m\times n}$ is unitarily invariant if $\|U^*AV\| = \|A\|$ for all unitary matrices U and V.
- Clearly a unitarily invariant norm depends only on the singular values.
- We say a function $\Phi : \mathbb{R}^n \to \mathbb{R}$ is a symmetric gauge function if it satisfies the following conditions.
  - Φ is a norm on $\mathbb{R}^n$.
  - $\Phi(Px) = \Phi(x)$ for any permutation matrix P.
  - $\Phi(|x|) = \Phi(x)$.
- There is a one-one correspondence between a symmetric gauge function Φ and a unitarily invariant norm:

$$\|A\| = \Phi(\sigma_1(A), \sigma_2(A), \ldots, \sigma_{\min\{m,n\}}(A)).$$

  (Von Neumannn)

Low rank approximation

- We have

$$\inf_{rank(X)\leq k} \|A - X\| = \Phi(\sigma_{k+1}(A), \sigma_{k+2}(A), \dots, \sigma_{\min\{m,n\}}(A)).$$

- The minimum is achieved at

$$X = U\mathrm{diag}(\sigma_1(A), \dots, \sigma_k(A), 0, \dots, 0)V^*.$$

- Principal Component Analysis (PCA).

Definition

- Let $\mathcal{G}_{m,n}$ denote the set of m dimensional subspaces of $\mathbb{C}^n$.
- The set $\mathcal{G}_{m,n}$ is usually called as a Grassmannian or a Grassmann space.
- For two subspaces $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$, define m canonical angles recursively as

$$
\begin{aligned}
\theta_m(\mathcal{X},\mathcal{Y}) &= \min_{x\in\mathcal{X}, y\in\mathcal{Y}} \angle(x,y) = \angle(x_m, y_m), \\
\theta_{m-1}(\mathcal{X},\mathcal{Y}) &= \min_{x\in\mathcal{X}\ominus\{x_m\}, y\in\mathcal{Y}\ominus\{y_m\}} \angle(x,y) = \angle(x_{m-1}, y_{m-1}), \\
& vdots \\
\theta_1(\mathcal{X},\mathcal{Y}) &= \min_{x\in\mathcal{X}\ominus\{x_m,\ldots,x_2\}, y\in\mathcal{Y}\ominus\{y_m,\ldots,y_2\}} \angle(x,y) = \angle(x_1, y_1),
\end{aligned}
$$

where $\angle(x,y) = \cos^{-1}\frac{|y^*x|}{\|x\|\|y\|}$ represents the angle between two nonzero vectors x and y.

Computation

- From now on we assume $n = 2m$ without loss of generality.
- Let the columns of $X, Y, X_\perp, Y_\perp$ form orthonormal bases of $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{X}^\perp$, $\mathcal{Y}^\perp$ respectively. Then

$$\cos\theta_i(\mathcal{X},\mathcal{Y}) = \sigma_{m-i+1}(X^*Y) = \sigma_{m-i+1}(X_\perp^* Y_\perp),$$
$$\sin\theta_i(\mathcal{X},\mathcal{Y}) = \sigma_i(X^*Y_\perp) = \sigma_i(X_\perp^* Y),$$

  for $i = 1, \ldots, m$.
- Canonical correlation analysis (CCA).
- Clearly $\theta_i(U\mathcal{X}, U\mathcal{Y}) = \theta_i(\mathcal{X},\mathcal{Y})$ for all $U \in \mathcal{U}(n)$.

Unitarily invariant metrics on $\mathcal{G}_{m,n}$

- We say a metric ρ on $\mathcal{G}_{m,n}$ is unitarily invariant if $\rho(U\mathcal{X}, U\mathcal{Y}) = \rho(\mathcal{X}, \mathcal{Y})$ for all $U \in \mathcal{U}(n)$.
- We say a metric on $\mathcal{G}_{m,n}$ is intrinsic if for each $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$, there exists a continuous function $\phi : [0,1] \to \mathcal{G}_{m,n}$ such that $\phi(0) = \mathcal{X}$, $\phi(1) = \mathcal{Y}$, and

$$\rho(\mathcal{X}, \mathcal{Y}) = \rho(\mathcal{X}, \phi(\lambda)) + \rho(\phi(\lambda), \mathcal{Y})$$

  for all $\lambda \in [0,1]$.
- Let Φ be a symmetric gauge function. Then

$$\rho(\mathcal{X}, \mathcal{Y}) = \Phi(\theta_1(\mathcal{X}, \mathcal{Y}), \ldots, \theta_m(\mathcal{X}, \mathcal{Y}))$$

  defines an unitarily invariant intrinsic metric.
- Does this give all unitarily invariant intrinsic metric?
- Conjecture: Yes.

Perturbation of subspaces

- Let $\mathcal{X}, \mathcal{Y} \in \mathcal{G}_{m,n}$ and $\mathcal{X} \cap \mathcal{Y} = \{0\}$, i.e., $\theta_m(\mathcal{X}, \mathcal{Y}) > 0$. The perturbed versions $\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}$ satisfies

$$\rho(\tilde{\mathcal{X}}, \mathcal{X}) \leq \alpha \quad \text{and} \quad \rho(\tilde{\mathcal{Y}}, \mathcal{Y}) \leq \beta.$$

  How can we ensure $\tilde{\mathcal{X}} \cap \tilde{\mathcal{Y}} = \{0\}$?
- $\tilde{\mathcal{X}} \cap \tilde{\mathcal{Y}} = \{0\}$ if (and only if)

$$\alpha + \beta < \Phi(0, \ldots, 0, \theta_m(\mathcal{X}, \mathcal{Y})).$$

- In general, we may ask how to ensure

$$\text{nullity}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}) := \dim \tilde{\mathcal{X}} \cap \tilde{\mathcal{Y}} < k.$$

C-S decomposition

- Let $W \in \mathcal{U}(n)$. Then there exist $U_1, U_2, V_1, V_2 \in \mathcal{U}(m)$ such that

$$W = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} C & -S \\ S & C \end{bmatrix} \begin{bmatrix} V_1^* & 0 \\ 0 & V_2^* \end{bmatrix}$$

  where

$$C = \mathrm{diag}\{c_1, c_2, \dots, c_m\}$$
$$S = \mathrm{diag}\{s_1, s_2, \dots, s_m\}.$$

- Clearly $c_i^2 + s_i^2 = 1$ and $C^2 + S^2 = I$.

Direct rotation

- Let the columns of $X, Y, X_\perp, Y_\perp$ form orthonormal bases of $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{X}^\perp$, $\mathcal{Y}^\perp$ respectively.
- The unitary matrix

$$M = [Y \;\; Y_\perp][X \;\; X_\perp]^*$$

has the property

$$M[X \;\; X_\perp] = [Y \;\; Y_\perp].$$

In particular, $M\mathcal{X} = \mathcal{Y}$.

- Apply C-S decomposition to

$$W = [X \;\; X_\perp]^*[Y \;\; Y_\perp] = \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} C & -S \\ S & C \end{bmatrix} \begin{bmatrix} V_1^* & 0 \\ 0 & V_2^* \end{bmatrix}.$$

- Define

$$[\hat{X} \;\; \hat{X}_\perp] = [XU_1 \;\; X_\perp U_2] \quad \text{and} \quad [\hat{Y} \;\; \hat{Y}_\perp] = [YV_1 \;\; Y_\perp V_2]$$

$$\hat{W} = [\hat{X} \;\; \hat{X}_\perp]^*[\hat{Y} \;\; \hat{Y}_\perp] = \begin{bmatrix} C & -S \\ S & C \end{bmatrix}$$

$$\hat{M} = [\hat{Y} \;\; \hat{Y}_\perp][\hat{X} \;\; \hat{X}_\perp]^* = [\hat{X} \;\; \hat{X}_\perp] \begin{bmatrix} C & -S \\ S & C \end{bmatrix} [\hat{X} \;\; \hat{X}_\perp]^*.$$

- The matrix $\hat{M}$ is called the direct rotation from $\mathcal{X}$ to $\mathcal{Y}$.

Direct rotation squared

- Let $P_{\mathcal{X}}$ and $P_{\mathcal{X}^\perp}$ denote the orthogonal projection onto $\mathcal{X}$ and $\mathcal{X}^\perp$ respectively. Then

$$P_{\mathcal{X}} - P_{\mathcal{X}^\perp} = \hat{X}\hat{X}^* - \hat{X}_\perp \hat{X}_\perp^* = [\hat{X} \ \hat{X}_\perp] \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} [\hat{X} \ \hat{X}_\perp]^*$$

  which is called the reflexion with respect to $\mathcal{X}$.
- Similarly we have

$$P_{\mathcal{Y}} - P_{\mathcal{Y}^\perp} = [\hat{Y} \ \hat{Y}_\perp] \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} [\hat{Y} \ \hat{Y}_\perp]^*.$$

- Direct computation gives

$$(P_{\mathcal{Y}} - P_{\mathcal{Y}^\perp})(P_{\mathcal{X}} - P_{\mathcal{X}^\perp}) = \hat{M}^2.$$

A "Pythagorean theorem"

- Let
  - $\hat{M}$ be the direct rotation from $\mathcal{X}$ to $\mathcal{Y}$,
  - $\hat{N}$ be the direct rotation from $\mathcal{Y}$ to $\mathcal{Z}$,
  - $\hat{L}$ be the direct rotation from $\mathcal{X}$ to $\mathcal{Z}$.

  Then
  $$\hat{L}^2 = \hat{N}^2 \hat{M}^2.$$

- A four-line proof:

$$\begin{aligned}
\hat{N}^2 \hat{M}^2 &= (P_{\mathcal{Z}} - P_{\mathcal{Z}^\perp})(P_{\mathcal{Y}} - P_{\mathcal{Y}^\perp})(P_{\mathcal{Y}} - P_{\mathcal{Y}^\perp})(P_{\mathcal{X}} - P_{\mathcal{X}^\perp}) \\
&= (P_{\mathcal{Z}} - P_{\mathcal{Z}^\perp})(P_{\mathcal{Y}} + P_{\mathcal{Y}^\perp})(P_{\mathcal{X}} - P_{\mathcal{X}^\perp}) \\
&= (P_{\mathcal{Z}} - P_{\mathcal{Z}^\perp})(P_{\mathcal{X}} - P_{\mathcal{X}^\perp}) \\
&= \hat{L}^2.
\end{aligned}$$

Recap

- Matrices vs pairs of subspaces.
- Singular values vs canonical angles.
- Unitarily invariant norms vs unitarily invariant intrinsic metrics.
- Rank of a matrix (with perturbation) vs nullity of a pair of subspaces (with perturbation).
- Some key tools: C-S decomposition, direct rotation, multiplicative Pythagorean theorem, ...

Application

- Secure robust control through networks.
- An architecture: