EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 4

Overview

- Basic definition: orthogonality, orthogonal projection, distance between subspaces, matrix inverse
- Matrix decomposition: singular value decomposition

Reading

- Chapter 4 of *Numerical Linear Algebra* by Lloyd Trefethen and David Bau
- Chapters 2 and 3 of *Matrix Computations* by Gene Golub and Charles Van Loan
- Chapter 3 of *Mathematical Modeling of Continuous Systems* by Carlo Tomasi

Matrix multiplication

- Let $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, $A \in \mathbb{R}^{m \times n}$, then

$$A\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{a}_j$$

  The output vector is a linear combination of matrix columns with coefficients given by the entries of $\mathbf{x}$

- Example

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 2 \end{bmatrix} + 3 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \begin{bmatrix} 4 \\ 4 \end{bmatrix} + \begin{bmatrix} 9 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}$$

Orthogonality

- A set vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in $\mathbb{R}^m$ is orthogonal if $\mathbf{x}_i^\top \mathbf{x}_j = 0$ when $i \neq j$, and orthonormal if $\mathbf{x}_i^\top \mathbf{x}_j = \delta_{ij}$
- Orthogonal vectors are maximally independent for they point in totally different directions
- Subspace: A collection of subspaces $S_1, \ldots, S_p$ in $\mathbb{R}^m$ is mutually orthogonal if $\mathbf{x}^\top \mathbf{y} = 0$ whenever $\mathbf{x} \in S_i$ and $\mathbf{y} \in S_j$ for $i \neq j$
- The orthogonal complement of a subspace $S \subset \mathbb{R}^m$ is

$$S^\perp = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y}^\top \mathbf{x} = 0 \ \forall \mathbf{x} \in S\}$$

- It can be shown that $\text{ran}(A)^\perp = \text{null}(A^\top)$
- The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ form an orthonormal basis for a subspace $S \subset \mathbb{R}^m$ if they are orthonormal and span S

Orthogonality (cont'd)

- A matrix $Q \in \mathbb{R}^{m \times m}$ is said to be orthogonal if $Q^\top Q = I$
- If $Q = [\mathbf{q}_1, \ldots, \mathbf{q}_m]$ is orthogonal, then the $\mathbf{q}_i$ form an orthonormal basis for $\mathbb{R}^m$
- It is always possible to extend such a basis to a full orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ for $\mathbb{R}^m$

Theorem

If $V_1 \in \mathbb{R}^{m \times r}$ has orthonormal columns, then there exists $V_2 \in \mathbb{R}^{m \times (m-r)}$ such that

$$V = [V_1 \; V_2] \in \mathbb{R}^{m \times m}$$

is orthogonal. Note that $ran(V_1)^\perp = ran(V_2)$

Proof.

This is a standard result from introductory linear algebra □

Norms and orthogonal transformations

- The vector 2-norm is invariant under orthogonal transformation Q

$$\|Q\mathbf{x}\|_2^2 = \mathbf{x}^\top Q^\top Q\mathbf{x} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|^2$$

- Likewise, matrix 2-norm and Frobenius norm are invariant with respect to orthogonal transformations Q and Z

$$\begin{aligned} \|QAZ\|_F &= \|A\|_F \\ \|QAZ\|_2 &= \|A\|_2 \end{aligned}$$

Singular values and singular vectors

- Let S be the unit sphere in $\mathbb{R}^n$ and any $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and $\text{ran}(A) = n$. The image AS is a hyperellipse in $\mathbb{R}^m$
- The n singular values of A, $\sigma_1, \sigma_2, \ldots, \sigma_n$, are the lengths of the n principal semi-axes of AS, where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$
- The n left singular vectors of A are the unit vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ oriented in the directors of the principal semi-axes of AS
- The n right singular vectors of A are the unit vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \in S$ that are the preimages of the principal semi-axes of AS, numbered so that $A\mathbf{v}_j = \sigma_j \mathbf{u}_j$

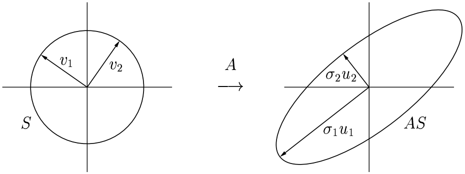


Figure 4.1. *SVD of a* 2×2 *matrix.*

Singular value decomposition (SVD)

Theorem

If A is a real m-by-n matrix, then there exists orthogonal matrices

$$U = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}, \ \textit{and } V = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$$

such that

$$U^\top A V = \Sigma = \textit{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m \times n}, \ \ p = \min(m, n),$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$, *or*

$$A = U \Sigma V^\top$$

- The σ_i are the singular values of A and the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th left singular vector and the i-th right singular vector respectively
- It follows that $AV = U\Sigma$, and $A^\top U = V\Sigma^\top$

Existence of SVD

Proof.

Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ be unit 2-norm vectors that satisfy $A\mathbf{x} = \sigma\mathbf{y}$ with $\sigma = \|A\|_2$. There exists $V_2 \in \mathbb{R}^{n\times(n-1)}$ and $U_2 \in \mathbb{R}^{m\times(m-1)}$ so $V = [\mathbf{x} \quad V_2] \in \mathbb{R}^{n\times n}$ and $U = [\mathbf{y} \quad U_2] \in \mathbb{R}^{m\times m}$ are orthogonal. It follows that

$$U^\top AV = U^\top \left[\ \sigma\mathbf{y} \quad AV_2 \ \right] = \begin{bmatrix} \sigma & \mathbf{w}^\top \\ \mathbf{0} & B \end{bmatrix} \equiv A_1$$

for some $\mathbf{w}$ and B. Since

$$\left\| \ A_1 \begin{bmatrix} \sigma \\ \mathbf{w} \end{bmatrix} \ \right\|_2^2 \geq (\sigma^2 + \mathbf{w}^\top\mathbf{w})^2$$

we have $\|A_1\|_2^2 \geq (\sigma^2 + \mathbf{w}^\top\mathbf{w})$. But $\sigma^2 = \|A\|_2^2 = \|A_1\|_2^2$, and thus $\mathbf{w}$ must be $\mathbf{0}$. By induction, we complete this proof. □

Singular vectors

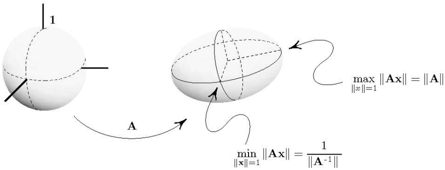
- As $A = U\Sigma V^\top$, by comparing the columns in the equations $AV = U\Sigma$ and $A^\top U = V\Sigma^\top$, it is easy to show

$$\begin{aligned} A\mathbf{v}_i &= \sigma\mathbf{u}_i \\ A^\top\mathbf{u}_i &= \sigma\mathbf{v}_i \end{aligned}$$

  where $i = 1, \ldots, \min(m, n)$
- The σ_i are the singular values of A and the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the i-th left singular vector and i-th right singular vector respectively
- U is a set of eigenvectors of $AA^\top \in \mathbb{R}^{m\times m}$
- Σ is a diagonal matrix whose values are the square root of eigenvalues of $AA^\top \in \mathbb{R}^{m\times m}$
- V is a set of eigenvectors of $A^\top A \in \mathbb{R}^{n\times n}$
- It can be shown that singular values σ_i are the square roots of eigenvalues, λ_i, i.e., $\sigma_i = \sqrt{\lambda_i}$

Singular values

- The singular values of a matrix A are precisely the lengths of the semi-axes of the hyperellipsoid E defined by $E = \{A\mathbf{x} : \|\mathbf{x}\|_2 = 1\}$
- The semi-axes are described by the singular vectors



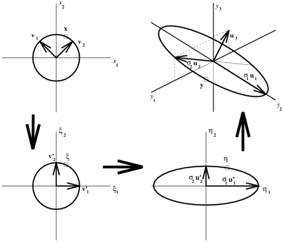
- The SVD reveals the structure of a matrix. Let

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$$

then

$$\begin{aligned} \text{rank}(A) &= r \\ \text{null}(A) &= \text{span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\} \\ \text{ran}(A) &= \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \end{aligned}$$

Geometric interpretation of SVD



- Consider $A\mathbf{x} = \mathbf{b}$ where $A \in \mathbb{R}^{3\times 2}, \mathbf{x} \in \mathbb{R}^{2\times 1}$ and $\mathbf{b} \in \mathbb{R}^{3\times 1}$, and $A = U\Sigma V^\top$
- Apply left rotation to $\mathbf{x}$ using right singular vectors V, $\xi = V^\top \mathbf{x}$
- Scale with Σ, i.e., $\eta = \Sigma\xi = \Sigma V^\top \mathbf{x}$
- Apply right rotation using left singular vectors U, $\mathbf{b} = U\eta = U\Sigma V^\top \mathbf{x}$
- Best approximation with r eigenvectors in 2-norm

SVD expansion

- We can decompose A in terms of singular values and vectors

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \otimes \mathbf{v}_i$$

  where $\otimes$ is the Kronecker product

- Matrix 2-norm and Frobenius norm

$$\begin{aligned} \|A\|_F &= \sqrt{\sigma_1^2 + \cdots + \sigma_p^2}, \quad p = \min(m, n) \\ \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} &= \|A\|_2 = \sigma_1 \\ \min_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} &= \sigma_n, \quad m \geq n \end{aligned}$$

  and $|\det(A)| = \prod_{i=1}^{n} \sigma_i$

- Closely related to eigenvalues, eigen-decomposition and principal component analysis

Application of SVD

- Matrix algebra: pseudo inverse, solving homogeneous linear equation, least squares minimization, rank, null space, etc.
- Computer vision: denoise, eigenface, eigentexture, eigen-X, structure from motion, etc.
- Pattern recognition: principal component analysis, dimensionality reduction, multivariate Gaussian, etc.
- Application: image/data analysis, document retrieval, etc.

Covariance matrix

- Let $\mathbf{x}_i \in \mathbb{R}^m$ and $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, covariance matrix $C \in \mathbb{R}^{m \times m}$,

$$\begin{aligned} C &= E[(\mathbf{x} - E[\mathbf{x}])(\mathbf{x} - E[\mathbf{x}])^\top] = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top \\ &= \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}) \otimes (\mathbf{x}_i - \boldsymbol{\mu}) = E[\mathbf{x}\mathbf{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top = \frac{1}{n}\overline{X}\,\overline{X}^\top \end{aligned}$$

  where $\boldsymbol{\mu} = E[\mathbf{x}] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$, and $\overline{X} = X - \mathbf{1}\boldsymbol{\mu}$
- Covariance C is positive semi-definite (i.e., $\mathbf{x}^\top C\mathbf{x} \geq 0$)
- Second order statistics of $\mathbf{x}$
- The variation can be compactly modeled with principal component analysis
- Related to multivariate Gaussian distribution, principal component analysis, SVD, and others

Gram matrix

- Let $\mathbf{x}_i \in \mathbb{R}^m$ and $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, Gram matrix, $G \in \mathbb{R}^{n \times n}$ is defined by
$$G = X^\top X$$
where $G_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j =< \mathbf{x}_i | \mathbf{x}_j >$
- Compute the pairwise similarities or correlations between two points
- Related to kernel methods (e.g., kernel PCA, support vector machine), regression, spectral clustering, SVD, and others