COM521500
Math. Methods for SP I
Lecture 4: Singular Value Decomposition & Orthogonal Projection

Singular Value Decomposition (SVD)

Theorem 4.1 Every $\mathbf{A} \in \mathbb{C}^{m \times n}$ can be decomposed as

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$$

where $\mathbf{U} \in \mathbb{C}^{m \times m}$ and $\mathbf{V} \in \mathbb{C}^{n \times n}$ are unitary, and

$$\mathbf{\Sigma} = \mathrm{Diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m \times n}, \qquad p = \min(m, n),$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$.

The values σ_i are called the **singular values** of $\mathbf{A}$. The columns $\mathbf{u}_i$ & $\mathbf{v}_i$ of $\mathbf{U}$ & $\mathbf{V}$ are called the **left and right singular vectors** of $\mathbf{A}$.

Outer product representation of SVD:

$$\mathbf{A} = \sum_{i=1}^{p} \sigma_i \mathbf{u}_i \mathbf{v}_i^H$$

Relationship with the 2-norm:

Recall $\|\mathbf{A}\|_2 = \sqrt{\lambda_{max}}$, where λ_{max} is the max. eigenvalue of $\mathbf{A}^H\mathbf{A}$.

By SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$,

$$\mathbf{A}^H\mathbf{A} = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^H.$$

It follows that the eigenvalues of $\mathbf{A}^H\mathbf{A}$ are σ_i^2, and that the eigenvector matrix of $\mathbf{A}^H\mathbf{A}$ is $\mathbf{V}$. Thus,

$$\|\mathbf{A}\|_2 = \sigma_1$$

Relationship with eigendecomposition:

Consider a Hermitian $\mathbf{A} \in \mathbb{C}^{n\times n}$. Eigendcomposition:

$$\begin{aligned}\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H &\Longleftrightarrow \mathbf{A}\mathbf{Q} = \mathbf{Q}\mathbf{\Lambda} \\ &\Longleftrightarrow \mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i, \quad i = 1, \ldots, n\end{aligned}$$

SVD:

$$\begin{aligned}\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H &\Longleftrightarrow \mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \\ &\Longleftrightarrow \mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad i = 1, \ldots, n\end{aligned}$$

Hence, for Hermitian $\mathbf{A}$ we have $\mathbf{U} = \mathbf{V} = \mathbf{Q}$ & $\mathbf{\Lambda} = \mathbf{\Sigma}$.

Partitioning the SVD

Suppose that the number of nonzero singular values is $r \leq p$; i.e., $\sigma_{r+1} = \sigma_{r+2} = \ldots \sigma_p = 0$.

The SVD can be rewritten as

$$\mathbf{A} = \begin{bmatrix}\mathbf{U}_1 & \mathbf{U}_2\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^H \\ \mathbf{V}_2^H\end{bmatrix}$$

where $\tilde{\mathbf{\Sigma}} = \mathrm{Diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$, $\mathbf{U}_1 \in \mathbb{C}^{m\times r}$, $\mathbf{U}_2 \in \mathbb{C}^{m\times m-r}$, $\mathbf{V}_1 \in \mathbb{C}^{n\times r}$, and $\mathbf{V}_2 \in \mathbb{C}^{n\times m-r}$.

Property 4.1 $\text{rank}(\mathbf{A}) = r$.

Property 4.2 $N(\mathbf{A}) = R(\mathbf{V}_2)$.

Property 4.3 $R(\mathbf{A}) = R(\mathbf{U}_1)$.

Property 4.4 $R(\mathbf{A}^H) = R(\mathbf{V}_1)$.

Property 4.5 $R_\perp(\mathbf{A}) = R(\mathbf{U}_2)$.

Inverse

Consider a square, nonsingular $\mathbf{A}$.

$$\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^H$$

An alternate form of the inverse:

$$\mathbf{A}^{-1} = \sum_{i=1}^{p} \frac{1}{\sigma_i}\mathbf{v}_i\mathbf{u}_i^H$$

Linear System of Equations

Given $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{b} \in \mathbb{C}^n$, the problem of the linear system of eqns. is find an $\mathbf{x} \in \mathbb{C}^m$ (or multiple $\mathbf{x}$'s) such that

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

We have learnt that for $m = n$, $\mathbf{A}\mathbf{x} = \mathbf{b}$ is always satisfied if $\mathbf{A}$ is nonsingular.

Can $\mathbf{A}\mathbf{x} = \mathbf{b}$ be satisfied when $m \neq n$, and/or when $\mathbf{A}$ is rank deficient?

$$\begin{aligned} & \mathbf{A}\mathbf{x} = \mathbf{b} \\ \Longleftrightarrow & \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H\mathbf{x} = \mathbf{b} \\ \Longleftrightarrow & \mathbf{\Sigma}\mathbf{d} = \mathbf{c} \end{aligned}$$

where

$$\mathbf{d} = \mathbf{V}^H\mathbf{x} = \begin{bmatrix} \mathbf{V}_1^H\mathbf{x} \\ \mathbf{V}_2^H\mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}$$

$$\mathbf{c} = \mathbf{U}^H\mathbf{b} = \begin{bmatrix} \mathbf{U}_1^H\mathbf{x} \\ \mathbf{U}_2^H\mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$$

Case A: $m > n$, and $r = n$.

In this case $\mathbf{V} = \mathbf{V}_1$, $\mathbf{d}_1 = \mathbf{d}$, &

$$\begin{aligned} \mathbf{\Sigma d} &= \mathbf{c} \\ \Longleftrightarrow \begin{bmatrix} \tilde{\mathbf{\Sigma}}\mathbf{d} \\ \mathbf{0} \end{bmatrix} &= \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} \end{aligned}$$

$\mathbf{Ax} = \mathbf{b}$ can only be satisfied if $\mathbf{b} \in R_\perp(\mathbf{U}_2) = R(\mathbf{U}_1)$.

Case B: $m > n$, and $r = n$.

In this case $\mathbf{U} = \mathbf{U}_1$, $\mathbf{c}_1 = \mathbf{c}$, &

$$\begin{aligned} \mathbf{\Sigma d} &= \mathbf{c} \\ \Longleftrightarrow \begin{bmatrix} \tilde{\mathbf{\Sigma}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{c} \end{bmatrix} \\ \Longleftrightarrow \tilde{\mathbf{\Sigma}}\mathbf{d}_1 &= \mathbf{c} \end{aligned}$$

$\mathbf{Ax} = \mathbf{b}$ can always be satisfied, but $\mathbf{x}$ is not unique.

If $\mathbf{x}_o$ is a solution to $\mathbf{Ax} = \mathbf{b}$, then $\mathbf{x}_o + \mathbf{V}_2\mathbf{c}_2$, for any $\mathbf{c}_2 \in \mathbb{C}^{n-r}$ is also a solution.

Case C: $r < \min(m, n)$.

$$\boldsymbol{\Sigma}\mathbf{d} = \mathbf{c}$$

$$\Longleftrightarrow \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$$

$\mathbf{Ax} = \mathbf{b}$ can only be satisfied if $\mathbf{b} \in R(\mathbf{U}_1)$.

If $\mathbf{x}_o$ is a solution to $\mathbf{Ax} = \mathbf{b}$, then $\mathbf{x}_o + \mathbf{V}_2\mathbf{c}_2$, for any $\mathbf{c}_2 \in \mathbb{C}^{n-r}$ is also a solution.

Low Rank Approximation

Theorem 4.2 Let $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ be the SVD of $\mathbf{A}$. For $k < r = \text{rank}(\mathbf{A})$, the solution to the problem

$$\min_{\substack{\mathbf{B} \in \mathbb{C}^{m\times n}, \\ \text{rank}(\mathbf{B})=k}} \|\mathbf{A} - \mathbf{B}\|_2$$

is

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^H.$$

Moreover, the minimal objective function value is

$$\|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}$$

Theorem 4.3 Let $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ be the SVD of $\mathbf{A}$. For $k < r = \text{rank}(\mathbf{A})$, the solution to the problem

$$\min_{\substack{\mathbf{B}\in\mathbb{C}^{m\times n},\\ \text{rank}(\mathbf{B})=k}} \|\mathbf{A}-\mathbf{B}\|_F^2$$

is

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^H.$$

Recall the KL transform in Lecture 3.

The vector $\hat{\mathbf{x}}_n$, formed from truncating $N-r$ KL coefficients, has the covariance matrix given by

$$\mathbf{R}_{\hat{x}} = \mathbf{V}\text{Diag}(\lambda_1,\ldots,\lambda_r,0,\ldots,0)\mathbf{V}^H$$

From Theorems 4.2 & 4.3 we know that $\mathbf{R}_{\hat{x}}$ is the closest rank-r matrix to the true signal covariance matrix $\mathbf{R}_x$, in the 2-norm and Frobenius-norm senses.

Orthogonal Projection

The idea: An arbitrary vector $\mathbf{y}$ can be expressed as

$$\mathbf{y} = \mathbf{y}_s + \mathbf{y}_c$$

where $\mathbf{y}_s \in \mathcal{S}$, & $\mathbf{y}_c \in \mathcal{S}_\perp$.

We are interested in obtaining a matrix $\mathbf{P}$, called the **orthogonal projection**, such that

$$\mathbf{P}\mathbf{y} = \mathbf{y}_s$$

Application: noise reduction

Consider a received signal that consists of a signal vector $\mathbf{s} \in \mathcal{S}$ and noise $\mathbf{w}$:

$$\mathbf{y} = \mathbf{s} + \mathbf{w}$$

We don't know $\mathbf{s}$, but we do know $\mathcal{S}$.

We can enhance the signal by performing a projection

$$\mathbf{P}\mathbf{y} = \mathbf{s} + \mathbf{w}_s$$

where $\mathbf{w}_s = \mathbf{P}\mathbf{w}$ is a residual noise vector.

A matrix $\mathbf{P} \in \mathbb{C}^{n \times n}$ is an **orthogonal projection** onto $\mathcal{S}$ if

1. $R(\mathbf{P}) = \mathcal{S}$,
2. $\mathbf{P}^2 = \mathbf{P}$, and
3. $\mathbf{P}^H = \mathbf{P}$.

Note that a matrix having the property $\mathbf{P}^2 = \mathbf{P}$ is called an idempotent matrix.

We have learnt that for a subspace $\mathcal{S}$ with a dimension m, there is a full rank matrix $\mathbf{X} \in \mathbb{C}^{n \times m}$, such that $\mathcal{S} = R(\mathbf{X})$.

An orthogonal projection onto $\mathcal{S} = R(\mathbf{X})$ is

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H \qquad (*)$$

Exercise: Verify that $(*)$ satisfies the 3 properties for an orthogonal projection matrix.

Theorem 4.4 The orthogonal projection matrix in $(*)$ is unique (i.e., there does not exist $\mathbf{P}_1$ such that $\mathbf{P}_1$ is an orthogonal projection onto $\mathcal{S}$ and $\mathbf{P}_1 \neq \mathbf{P}$).

The orthogonal complement projection:

By observing that

$$\mathbf{y} = \mathbf{y}_s + \mathbf{y}_c = \mathbf{P}\mathbf{y} + \mathbf{y}_c,$$

we obtain

$$\mathbf{y}_c = (\mathbf{I} - \mathbf{P})\mathbf{y}$$

and that $(\mathbf{I} - \mathbf{P})$ is the orthogonal projection onto the orthogonal complement subspace $\mathcal{S}_\perp$.

Property 4.6

- $\mathbf{V}_1\mathbf{V}_1^H$ is the orthogonal projection onto $R(\mathbf{A}^H)$.
- $\mathbf{V}_2\mathbf{V}_2^H$ is the orthogonal projection onto $N(\mathbf{A})$.
- $\mathbf{U}_1\mathbf{U}_1^H$ is the orthogonal projection onto $R(\mathbf{A})$.
- $\mathbf{U}_2\mathbf{U}_2^H$ is the orthogonal projection onto $R_\perp(\mathbf{A})$.

Property 4.7 The eigenvalues of a projection matrix is either 1 or 0. The number of nonzero eigenvalues is the dimension of the associated subspace.

Distance between subspaces:

Let $\mathcal{S}_1$ & $\mathcal{S}_2$ be two subspaces with $\dim \mathcal{S}_1 = \dim \mathcal{S}_2$.

Let $\mathbf{P}_1$ & $\mathbf{P}_2$ be the orthogonal projection matrices of $\mathcal{S}_1$ & $\mathcal{S}_2$, respectively.

The distance between $\mathcal{S}_1$ & $\mathcal{S}_2$ is defined as

$$\begin{aligned}\mathrm{dist}(\mathcal{S}_1, \mathcal{S}_2) &= \|\mathbf{P}_1 - \mathbf{P}_2\|_2 \\ &= \max_{\|\mathbf{x}\|_2 = 1} \|\mathbf{P}_1\mathbf{x} - \mathbf{P}_2\mathbf{x}\|_2\end{aligned}$$

Theorem 4.5 Suppose

$$\mathbf{W} = [\ \mathbf{W}_1\ \mathbf{W}_2\ ], \qquad \mathbf{Z} = [\ \mathbf{Z}_1\ \mathbf{Z}_2\ ]$$

are unitary, where $\mathbf{W}_1, \mathbf{Z}_1 \in \mathbb{C}^{n\times k}$. If $\mathcal{S}_1 = R(\mathbf{W}_1)$ & $\mathcal{S}_2 = R(\mathbf{Z}_1)$, then

$$\text{dist}(\mathcal{S}_1, \mathcal{S}_2) = \|\mathbf{W}_1^H \mathbf{Z}_2\|_2 = \|\mathbf{Z}_1^H \mathbf{W}_2\|_2$$

Property 4.8 $0 \leq \text{dist}(\mathcal{S}_1, \mathcal{S}_2) \leq 1.$

Property 4.9 If $\mathcal{S}_1 = \mathcal{S}_2$, then $\text{dist}(\mathcal{S}_1, \mathcal{S}_2) = 0.$

Property 4.10 If $\mathcal{S}_1 \cup \mathcal{S}_2^{\perp} \neq \{\mathbf{0}\}$, then $\text{dist}(\mathcal{S}_1, \mathcal{S}_2) = 1.$