ME 491 Lecture 3.

Ch.1 Singular Value Decomposition (SVD)

* Brief recap of eigen decomposition.

Let A be an $n \times n$ square matrix that has a non-zero determinant, so we can perform eigen decomposition for A

$$A = Q \Lambda Q^{-1}, \quad \text{where}$$

$$Q = \begin{bmatrix} | & | & & | \\ x_1, & x_2, & \cdots, & x_n \\ | & | & & | \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

x_i: eigenvectors of A, λ_i: eigenvalues of A

$$A x_i = \lambda_i x_i$$

Eigen decomposition is very useful and convenient for square matrices, but a lot of datasets we encounter in engineering will not be square, so what do we do?

Ch.1.1 Overview

For a large data set $X \in C^{n \times m}$

$C^{n \times m}$: $n \times m$ matrix with complex values

$$X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & & | \end{bmatrix}$$

column $x_k \in C^n$: may be measurements from simulations or experiments k: k^{th} distinct set of measurements

Many example data in this course will consist of time series, so

$$x_k = x(k\Delta t)$$

Often $n \gg m$, resulting in tall-skinny matrix.

SVD is a unique matrix decomposition:

$$X = U\Sigma V^*$$

$$U \in \mathbb{C}^{n\times n}, \quad V \in \mathbb{C}^{m\times m}, \quad \Sigma \in \mathbb{R}^{n\times m}$$

U, V: unitary matrices, meaning

$$U^*U = U^*U = I$$

$*$: complex conjugate transpose.

For a complex number $z = a+bi$, $z^* = a-bi$

Σ: real, non-negative entries on the diagonal, zeros of the diagonal.

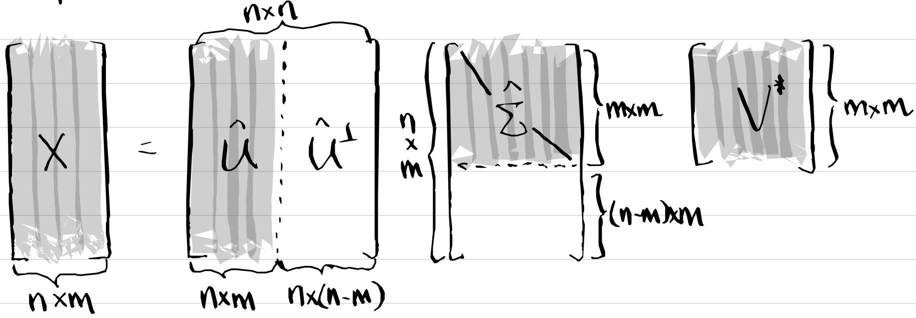
When $n \geq m$, Σ has at most m non-zero elements on the diagonal:

$$\Sigma = \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix}$$

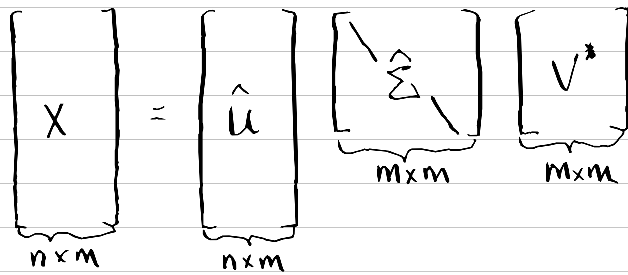
Because of this, we can use a more memory efficient way to store SVD.

$$X = U\Sigma V^* = \begin{bmatrix} \hat{U} & \hat{U}^\perp \end{bmatrix} \begin{bmatrix} \hat{\Sigma} \\ 0 \end{bmatrix} V^* = \hat{U}\hat{\Sigma}V^*$$

- Grapical Illustration



Full SVD



Economy SVD

Columns of $\hat{U}^{\perp}$ span a vector space that is complementary and orthogonal to that spanned by $\hat{U}$.

Columns in U: left singular vectors of X
Columns in V: right singular vectors of X
diagonal elements of $\hat{\Sigma}$: singular values of X
ordered from largest to smallest

rank of X: number of non-zero singular values.

The singular values are crucial to perform optimal rank-r approximations of X for $r < m$.

* Computing SVD.
  For matrices $n \geq m$. QR factorization, see discussion on P6-7

one very important machine learning algorithm Principal Component Analysis (PCA) is closely related to SVD.

1.2 Matrix Approximation

- The reason SVD is so powerful and popular is that it provides an optimal low-rank approximation to a matrix X.

How does this approximation work ?

Since Σ is diagonal, we can express matrix $X = U\Sigma V^*$ as a sum of rank-one matrices.

$$X = \sum_{k=1}^{m} \sigma_k u_k v_k^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_m u_m v_m^*$$

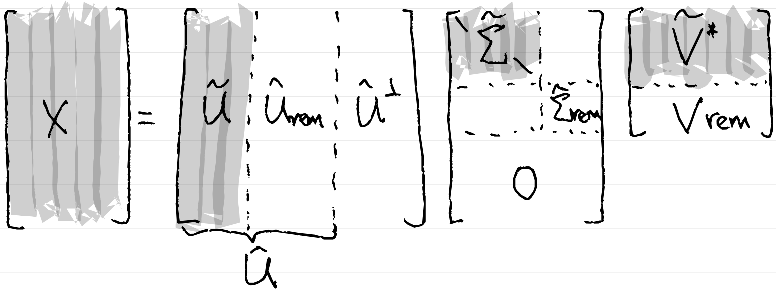
σ_k: k^{th} diagonal entry of Σ
u_k: k^{th} column of U
v_k: k^{th} column of V

Singular values σ_k are arranged in decreasing order
$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0 .$$
So every next $\sigma_k u_k v_k^*$ is less important than the previous one

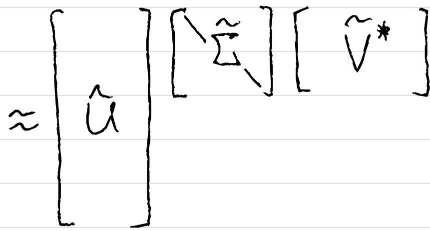
We can obtain an approximation of X by truncating at some rank r.
$$X \approx \hat{X} = \sum_{k=1}^{r} \sigma_k u_k v_k^* = \sigma_1 u_1 v_1^* + \sigma_2 u_2 v_2^* + \cdots + \sigma_r u_r v_r^*$$

Notation for truncated SVD basis
$$\tilde{X} = \tilde{U} \tilde{\Sigma} \tilde{V}^*$$
$\tilde{U}, \tilde{V}$: first r column of U, V, $\tilde{\Sigma}$: first $r \times r$ sub-block of Σ

- Graphical Illustration



Full SVD



Truncated SVD

In the least square sense, there is no better approximation for X for a given rank r.

* Optimal Approximation and Error Bounds

Theorem 1.1 (Eckart - Young):

The optimal rank-r approximation to X, in a least square sense, is given by the rank-r SVD truncation $\tilde{X}$:

$$\underset{\tilde{X},\ s.t.\ rank(\tilde{X})=r}{\operatorname{argmin}} \|X - \tilde{X}\|_F = \tilde{U}\tilde{\Sigma}\tilde{V}^*$$

Definition of Frobenius sum:

$$\|X\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}|X_{ij}|^2}$$

We can then exactly quantify the error of the rank-r SVD

$$\|X - \tilde{X}\|_F^2 = \sum_{k=r+1}^{m} \sigma_k^2$$

All rank-r matrices $\tilde{X}$ will have at least this much error.

Relative error: $$\frac{\|X - \tilde{X}\|_F^2}{\|X\|_F^2}$$

(For intuitive interpretations, see P11)

SVD also provides an optimal rank-r approximation in the matrix 2-norm, also known as the spectral norm:

$$\underset{\tilde{X},\ \text{s.t. } \operatorname{rank}(\tilde{X})=r}{\operatorname{argmin}} \|X - \tilde{X}\|_2 = \tilde{U}\tilde{\Sigma}\tilde{V}^*$$

Definition of 2-norm for matrix

$$\|X\|_2 = \max_{v \neq 0} \frac{\|Xv\|_2}{\|v\|_2}$$

2-norm error expression:

$$\|X - \tilde{X}\|_2 = \sigma_{r+1}$$

(short derivation see P11)