

Singular Value Decomposition

Why and How

Royi Avital

June, 2011

1 Definitions and Notations

- Notations
- Definitions
- Introduction

2 Singular Value Decomposition Theorem

- SVD Theorem
- Proof of the SVD Theorem
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

1 Definitions and Notations

- Notations
- Definitions
- Introduction

- SVD Theorem
- Proof of the SVD Theorem
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

- Referring a Column of a Matrix

A^j – The j – th Column of a Matrix

1 Definitions and Notations

- Notations
- Definitions
- Introduction

2 Singular Value Decomposition Theorem

- SVD Theorem
- Proof of the SVD Theorem
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

- Unless written otherwise, The Complex Field is the default
- Conjugate Operator

$$A^*$$

- Transpose Operator

$$A^T : A^T_{ij} = A_{ji}$$

- Complex Conjugate Transpose Operator

$$A^H : A^H_{ij} = A_{ji}^*$$

- Range Space and Null Space of Operator

Let $L : X \rightarrow Y$ be an Operator (Linear or otherwise). The Range Space $\mathcal{R}(L) \subseteq Y$ is

$$\mathcal{R}(L) = \{y = Lx : x \in X\}$$

The Null Space $\mathcal{N}(L) \subseteq X$ is

$$\mathcal{N}(L) = \{x \in X : Lx = 0\}$$

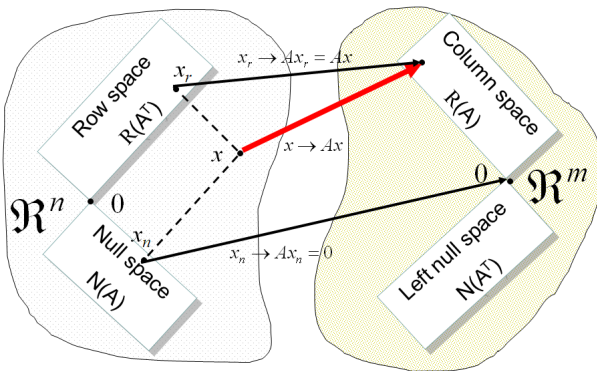
- 1 Definitions and Notations
 - Notations
 - Definitions
 - Introduction
- 2 Singular Value Decomposition Theorem
 - SVD Theorem
 - Proof of the SVD Theorem
 - SVD Properties
 - SVD Example
- 3 Applications
 - Order Reduction
 - Solving Linear Equation System
 - Total Least Squares
 - Principal Component Analysis

-
- Diagram illustrating the relationship between the domain and codomain of a linear transformation A .
- The domain is \mathbb{R}^n , which is the direct sum of the Row space $R(A^T)$ (dimension $\dim r$) and the Null space $N(A)$ (dimension $\dim n-r$).
- The codomain is \mathbb{R}^m , which is the direct sum of the Column space $R(A)$ (dimension $\dim r$) and the Left null space $N(A^T)$ (dimension $\dim m-r$).
- A green arrow labeled A points from the domain to the codomain.

- $$\mathcal{R}(A) \perp \mathcal{N}(A^H), \mathcal{R}(A^H) \perp \mathcal{N}(A)$$

$$\text{rank}(A) = \dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^H)) = \text{rank}(A^H)$$

- The action of linear operator $A \in \mathbb{C}^{m \times n}$



- The following properties hold

$$\text{rank}(A) = \text{rank}(AA^H) = \text{rank}(A^H A) = \text{rank}(A^H)$$

$$\mathcal{R}(A) = \mathcal{R}(AA^H), \mathcal{R}(A^H) = \mathcal{R}(A^HA)$$

1 Definitions and Notations

- Notations
- Definitions
- Introduction

2 Singular Value Decomposition Theorem

- SVD Theorem
- Proof of the SVD Theorem
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

Every Matrix $A \in \mathbb{C}^{m \times n}$ can be factored as $A = U\Sigma V^H$.

Where $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ are Unitary.

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

Corollary (i)

The columns of U are the Eigenvectors of AA^H (Left Eigenvectors).

$$AA^H = U\Sigma V^H (U\Sigma V^H)^H = U\Sigma V^H V \Sigma^H U^H = U\Sigma \Sigma^H U^H$$

The columns of V are the Eigenvectors of $A^H A$ (Right Eigenvectors).

$$A^H A = \left(U \Sigma V^H \right)^H U \Sigma V^H = V \Sigma^H U^H U \Sigma V^H = V \Sigma^H \Sigma V^H$$

Theorem

SVD Theorem:

Every Matrix $A \in \mathbb{C}^{m \times n}$ can be factored as $A = U \Sigma V^H$.

Where $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ are Unitary.

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

Corollary (ii)

The p singular values on the diagonal of Σ are the square roots of the non zero eigenvalues of both AA^H and $A^H A$.

The SVD is unique up to permutations of (u_i, σ_i, v_i) as long as $\sigma_i \neq \sigma_j \Leftrightarrow i \neq j$. If the "Algebraic Multiplicity" of a certain eigenvalue of $A^H A$ / AA^H is larger than 1, Then, there's a freedom of choosing the the vectors which spans the the null space $AA^H - \lambda I$ / $A^H A - \lambda I$.

1 Definitions and Notations

- Notations
- Definitions
- Introduction

2 Singular Value Decomposition Theorem

- SVD Theorem
- **Proof of the SVD Theorem**
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

Theorem

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

In order to prove the SVD Theorem 2 propositions should be used.

Proposition I

$\forall A \in \mathbb{C}^{m \times n}$ $A^H A$ or AA^H are Hermitian Matrix.

Proof.

$$\begin{aligned} C_{ij} &= A^H_i A^j = \left(A^H_i A^j \right)^{HH} = \left(A^{jH} \left(A^H_i \right)^H \right)^H \\ &= A^H_j A^i = C_{ji}^H = C_{ji}^* \end{aligned}$$



Proposition II - Spectral Decomposition

$\forall A \in \mathbb{C}^{n \times n} : A_{ij} = A_{ji}^*$ (Hermitian Matrix) Can be diagonalized using Unitary Matrix $U \in \mathbb{C}^{n \times n}$ s.t. $U^H A U = \Lambda$.

Proof.

The Spectral Decomposition is a result of few properties of Hermitian Matrices:

- For Hermitian Matrices the Eigenvectors of distinct Eigenvalues are Orthogonal.
- Schur's Lemma

$$\forall A \in \mathbb{C}^{n \times n} \exists U \in \mathbb{C}^{n \times n} \text{ Unitary s.t. } U^H A U = T$$

Where $T \in \mathbb{C}^{n \times n}$ is Upper Triangular Matrix.

When A has n distinct Eigenvalues the Proposition II is immediate. Otherwise it can be shown that if A is Hermitian, T is Hermitian and since it is Upper Triangular, it must be Diagonal Matrix. \square

100

Where $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ are Unitary.

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

100

100

© 2001 Blackwell Science Ltd

[illegible]

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

Then the set of u_i are the Eigenvectors for AA^H .

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

For the zero Eigenvalues, The Eigenvectors must come from the Null Space of AA^H . Since the Eigenvectors with zero Eigenvalues are, By construction, Orthogonal to the Eigenvectors with non zero Eigenvalues that are in the Range of AA^H , Hence must be in the Null Space of AA^H .

Theorem

$\forall A \in \mathbb{C}^{m \times n}$ can be factored as $A = U\Sigma V^H$.

Where $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ are Unitary.

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

Proof. Continued

Examining the elements of $U^H A V$.

For $i \leq r$ the (i, j) element of $U^H A V$ is

$$u_i^H A v_j = \frac{1}{\sqrt{\lambda_j}} v_i^H A^H A v_j = \frac{\lambda_j}{\sqrt{\lambda_j}} v_i^H v_j = \sqrt{\lambda_j} \delta_{ij}$$

For $i > r$ We get

$$AA^H u_i = 0$$

Thus $A^H u_i \in \mathcal{N}(A)$ and also $A^H u_i \in \mathcal{R}(A^H)$ as a Linear Combination of the columns of A^H . Yet $\mathcal{R}(A^H) \perp \mathcal{N}(A)$ Hence $A^H u_i = 0$.

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

Since $A^H u_i = 0$ we get $u_i^H A v_j = v_j^H A^H u_i = 0$. Thus $U^H A V = \Sigma$,
Where Σ is diagonal (Main Diagonal). □

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

By Spectral Theorem:

$$U = [u_1, u_2, \dots, u_m], U \in \mathbb{C}^{m \times m}, UU^H = U^H U = I_m$$

Thus $\|A^H u_i = \sigma_i\|$ for $i = 1, 2, \dots, m$.

Theorem

$\forall A \in \mathbb{C}^{m \times n}$ can be factored as $A = U\Sigma V^H$.

Where $U \in \mathbb{C}^{m \times m}$, $V \in \mathbb{C}^{n \times n}$ are Unitary.

$\Sigma \in \mathbb{C}^{m \times n}$ has the form $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$.

Alternative Proof Continued

Examining z_j yields:

$$z_j^H z_i = u_j^H A A^H u_i = \sigma_i^2 u_j^H u_i \Rightarrow z_i = u_i \sigma_i \Rightarrow v_i = \frac{z_i}{\sigma_i} = \frac{A^H u_i}{\sigma_i}$$

Consider the following n equations for $i = 1, 2, \dots, m$:

$$Av_i = AA^H \frac{u_i}{\sigma_i} \text{ (or zero)} = \sigma_i u_i \text{ (or zero)}$$

1 Definitions and Notations

- Notations
- Definitions
- Introduction

2 Singular Value Decomposition Theorem

- SVD Theorem
- Proof of the SVD Theorem
- **SVD Properties**
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

$$\Sigma = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}$$

.....

$$\Sigma_0 = \text{diag}(\sigma_{-1}, \sigma_{-2}, \dots, \sigma_{-n})$$

TABLE 1. The 100 Most Influential Journals in the Field of Health Communication, 1980-2000

[] [] [] []

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:01 11 November 2014

The SVD can also be written as

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

The SVD can also be used to compute 2 matrix norms:

- Hilbert-Schmidt / Frobenius Norm

$$\|A\|_F^2 = \sum_{i,j} \|A_{ij}\|^2 = \sum_{i=1}^r \sigma_i^2$$

- l_2 Norm

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max(\lambda(A)) = \sigma_1$$

Which implies

$$\underset{x \neq 0}{\operatorname{argmax}} \frac{\|Ax\|}{\|x\|} = v_1, \quad \underset{x \neq 0}{\operatorname{argmax}} \frac{\|x^H A\|}{\|x\|} = u_1$$

- The range (Column Space) of a matrix is

The range of a matrix is spanned by the orthogonal set of vectors in U_1 , the first r columns of U .

- Generally, the other fundamental spaces of a matrix A can also be determined from the SVD:

$$\mathcal{R}(A) = \text{span}(U_1) = \mathcal{R}(AA^H)$$

$$\mathcal{N}(A) = \text{span}(V_2)$$

$$\mathcal{R}(A^H) = \text{span}(V_1) = \mathcal{R}(A^HA)$$

$$\mathcal{N}(A^H) = \text{span}(U_2)$$

The SVD thus provides an explicit orthogonal basis and a computable dimensionality for each of the fundamental spaces of a matrix.

Since the SVD is a decomposition of a given matrix into 2 Unitary matrices and a diagonal matrix, all matrices could be described as a rotation, scaling and another rotation.

This intuition is a result of the properties of unitary matrices which basically rotate the multiplied matrix.

This property is farther examined when dealing Linear Equations.

- 1 Definitions and Notations
 - Notations
 - Definitions
 - Introduction
- 2 Singular Value Decomposition Theorem
 - SVD Theorem
 - Proof of the SVD Theorem
 - SVD Properties
 - SVD Example
- 3 Applications
 - Order Reduction
 - Solving Linear Equation System
 - Total Least Squares
 - Principal Component Analysis

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \end{bmatrix}$$

Then

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

Where

$$U = \begin{bmatrix} -0.355 & -0.934 \\ -0.934 & 0.355 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 9.362 & 0 & 0 \\ 0 & 1.831 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.637 & -0.653 & 0.408 \\ -0.575 & -0.050 & -0.8165 \\ -0.513 & -0.754 & 0.408 \end{bmatrix}$$

Let A be a **Diagonal Matrix**

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

In this case, the U and V matrices just shuffle the columns around and change the signs to make the singular values positive.

Let A be a **Square Symmetric Matrix**

$$A = \begin{bmatrix} 5 & 6 & 2 \\ 6 & 1 & 4 \\ 2 & 4 & 7 \end{bmatrix} = U \Sigma V^H$$

Where

$$U = V = \begin{bmatrix} 0.592 & -0.616 & 0.518 \\ 0.526 & -0.191 & 0.828 \\ 0.610 & 0.763 & -0.211 \end{bmatrix}, \Sigma = \begin{bmatrix} 12.391 & 0 & 0 \\ 0 & 4.383 & 0 \\ 0 & 0 & 3.774 \end{bmatrix}$$

In this case, the SVD is the regular Eigen Decomposition.

1 Definitions and Notations

- Notations
- Definitions
- Introduction

2 Singular Value Decomposition Theorem

- SVD Theorem
- Proof of the SVD Theorem
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

The SVD of a matrix can be used to determine how near (In the sense of l_2 -norm) the matrix is to a matrix of a lower rank. It can also be used to find the nearest matrix of a given lower rank.

Theorem

Let A be an $m \times n$ matrix with $\text{rank}(A) = r$ and let $A = U\Sigma V^H$. Let $k < r$ and let

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^H = U \Sigma_k V^H$$

where

$$\Sigma_k = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$$

Then $\|A - A_k\|_2 = \sigma_{k+1}$, and A_k is the nearest matrix of rank k to A (In the sense of l_2 -norm / Frobenius norm):

$$\min_{\text{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2$$

Proof.

Since $A - A_k = U \text{diag}(0, 0, \dots, 0, \sigma_{k+1}, \dots, \sigma_r, 0, \dots, 0) V^H$ it follows that $\|A - A_k\|_2 = \sigma_{k+1}$.

The second part of the proof is a "Proof by Inequality". By Definition of the matrix norm, for any unit vector z the following holds:

$$\|A - B\|_2^2 \geq \|(A - B)z\|_2^2$$

Let B be a rank - k matrix of size $m \times n$. Then there exist vectors $\{x_1, x_2, \dots, x_{n-k}\}$ that span $\mathcal{N}(B)$ where $x_i \in \mathbb{C}^n$.

Consider the vectors from the matrix V of the SVD, $\{v_1, v_2, \dots, v_{k+1}\}$ where $v_i \in \mathbb{C}^n$.

Proof. Continued.

The intersection, $\text{span}(x_1, \dots, x_{n-k}) \cap \text{span}(v_1, \dots, v_{k+1}) \subseteq \mathbb{R}^n$, cannot be zero since there are total of $n+1$ vectors. Let z be a vector from this intersection, normalized s.t. $\|z\|_2 = 1$. Then:

$$\|A - B\|_2^2 \geq \|(A - B)z\|_2^2 = \|Az\|_2^2$$

Since $z \in \text{span}(v_1, v_2, \dots, v_{k+1})$, $Az = \sum_{i=1}^{k+1} \sigma_i (v_i^H z) u_i$ Now

$$\|A - B\|_2^2 \geq \|A_Z\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 \left(v_i^H Z \right)^2 \geq \sigma_{k+1}^2$$

Lower bound is achieved by $B = \sum_{i=1}^k \sigma_i u_i v_i^H$, with $z = v_{k+1}$.

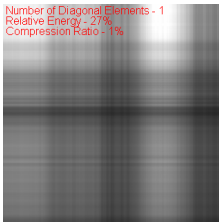


Applications of Order Reduction:

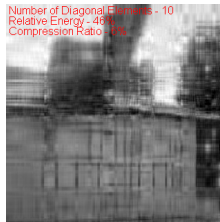
- Image Compression



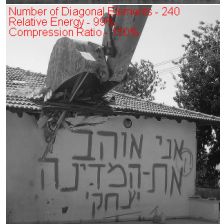
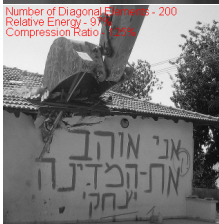
Number of Diagonal Elements - 100
Relative Energy - 84%
Compression Ratio - 65%



Number of Diagonal Elements - 200
Relative Energy - 97%
Compression Ratio - 50%



Number of Diagonal Elements - 240
Relative Energy - 99%
Compression Ratio - 40%



Applications of Order Reduction:

- Noise Reduction

Basic assumption - Noise is mainly pronounced in the small singular values.



Noiseless Matrix



Noisy Matrix - Std 1

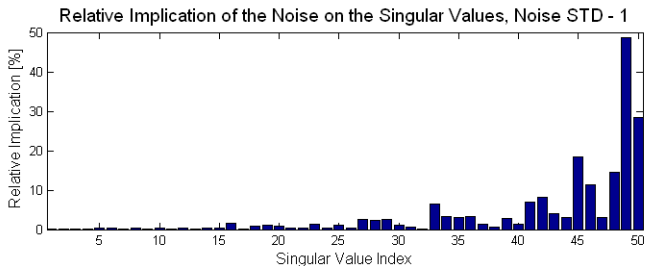
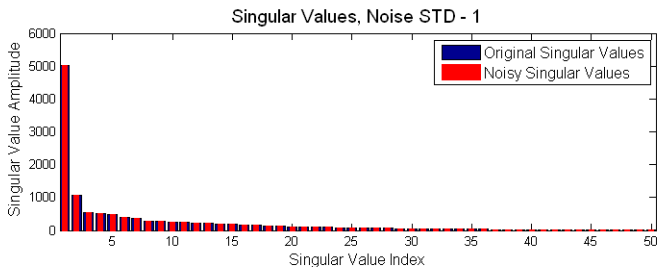


Noisy Matrix - Std 6

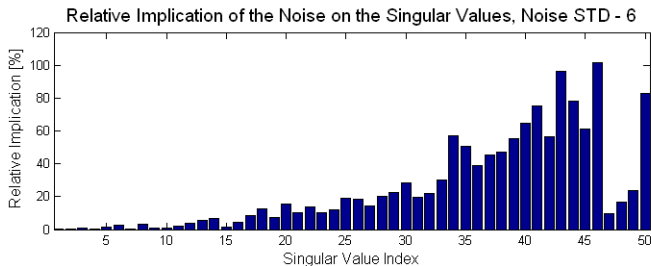
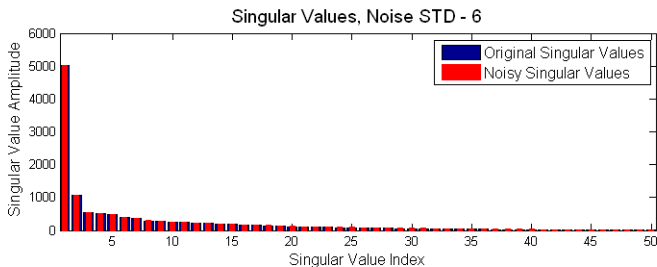


Noisy Matrix - Std 11

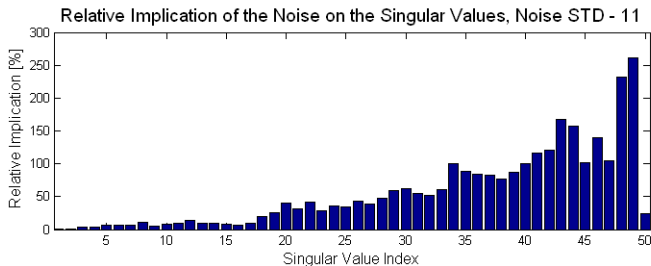
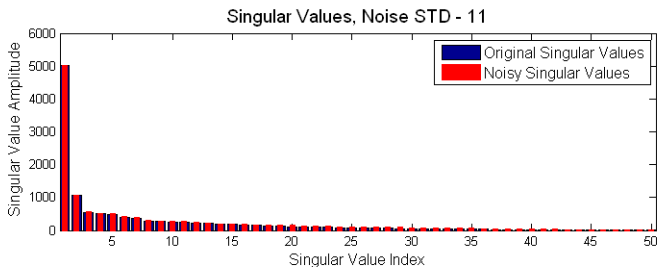
Analyzing the effect of noise on the Singular Values

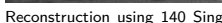


Analyzing the effect of noise on the Singular Values



Analyzing the effect of noise on the Singular Values





Values

- Notations

- Definitions
- Introduction

- SVD Theorem

- Proof of the SVD Theorem
- SVD Properties
- SVD Example

- Order Reduction

- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

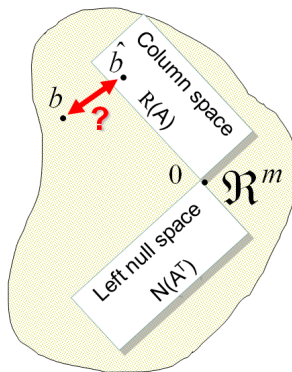
Consider the solution of the equation $Ax = b$.

- If $b \in \mathcal{R}(A)$ there is at least one solution:
 - If $\dim(\mathcal{N}(A)) = 0$ there is only one unique solution, $x_r \in \mathcal{R}(A^H)$, s.t. $Ax_r = b$.
 - If $\dim(\mathcal{N}(A)) \geq 1$, The columns of A are not independent, there are infinite solutions. Any vector of the form $\hat{x} = x_r + x_n$ where $x_r \in \mathcal{R}(A^H)$ is the solution from the previous case and $x_n \in \mathcal{N}(A)$ is a solution s.t. $A(x_r + x_n) = b$.
Which solution should be chosen?
Usually the solution with the minimum norm, x_r .
- If $b \notin \mathcal{R}(A)$ there is no solution.
Usually, the following vector is searched, \hat{x} s.t. $\|A\hat{x} - b\|_2$ is brought to minimum.

Assuming $\hat{x} = \min_x \|Ax - b\|_2$.

By definition $\hat{b} = A\hat{x} \in \mathcal{R}(A)$.

Meaning, the search is for \hat{b} s.t. $\|\hat{b} - b\|_2$ is minimized.



According to the Projection Theorem, only one vector \hat{b} exists s.t. $\|\hat{b} - b\|_2$ is minimized. This vector is the projection of b on $\mathcal{R}(A)$.

$$\hat{b} = Proj_{\mathcal{R}(A)}(b) = A^H b$$

Moreover,

$$\hat{x} = \min_x \|Ax - b\|_2 \Leftrightarrow (A^H A) \hat{x} = A^H b$$

Intuitively, the procedure is as following:

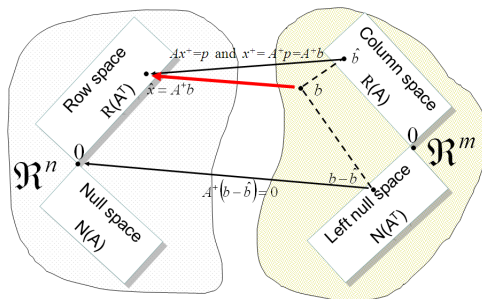
- Project b onto the Column Space $\mathcal{R}(A)$, namely, $\hat{b} = Proj_{\mathcal{R}(A)}(b) = A^H b$.
- Project \hat{x} onto the Row Space $\mathcal{R}(A^H)$, namely, $Proj_{\mathcal{R}(A^H)}(\hat{x}) = Ax$.
- Project the previous result Ax onto the Column Space $\mathcal{R}(A)$, namely, $Proj_{\mathcal{R}(A)}(A\hat{x}) = A^H(A\hat{x})$.

The equation $(A^H A) \hat{x} = A^H b$ is called the Normal Equations. If the columns of A are independent then $A^H A$ is invertible and \hat{x} could be calculated as the following:

$$\hat{x} = (A^H A)^{-1} A^H b$$

This is the Least Squares solution using the Pseudo Inverse of A :

$$A^+ = (A^H A)^{-1} A^H$$



Yet, if the columns of A are linearly dependent the Pseudo Inverse of A can't be calculated directly.

If A has dependent columns, then the null space of A is not trivial and there is no unique solution.

The problem becomes selecting one solution out of the infinite number of possible solutions.

As mentioned, commonly accepted approach is to select the solution with the smallest norm (Length).

This problem could be solved using the SVD and definition of the generalized Pseudo Inverse of a matrix.

The Pseudo Inverse of a matrix $A = U\Sigma V^H$, denoted A^\dagger is given by

$$A^\dagger = V \Sigma^\dagger U^H$$

Where Σ^+ is obtained by transposing Σ and inverting all non zero entries.

Proposition III

Let $A = U\Sigma V^H$ and $x^\dagger = A^\dagger b = V\Sigma^\dagger U^H b$. Then $A^H A x^\dagger = A^H b$.

Namely, using the solution given by the Pseudo Inverse matrix calculated using the SVD holds the Normal Equations. **This definition of Pseudo Inverse exists for any matrix.**

Proposition III

Let $A = U\Sigma V^H$ and $x^\dagger = A^\dagger b = V\Sigma^\dagger U^H b$. Then $A^H A x^\dagger = A^H b$.

proof

It's sufficient to show that $A^H (A x^\dagger - b) = 0$.

$$\begin{aligned} A x^\dagger - b &= (U \Sigma V^H) V \Sigma^\dagger U^H b - b \\ &= (U \Sigma \Sigma^\dagger U^H - I) b \\ &= (U (\Sigma \Sigma^\dagger - I) U^H) b \end{aligned}$$

Proof. Continued.

Thus,

$$\begin{aligned} A^H (Ax^\dagger - b) &= V \Sigma^H U^H U (\Sigma \Sigma^\dagger - I) U^H b \\ &= V \Sigma^H (\Sigma \Sigma^\dagger - I) U^H b \end{aligned}$$

One should observe that

$$\Sigma^H = \begin{bmatrix} \Sigma_{r \times r}^H & 0_{r \times (m-r)} \\ 0_{(n-r) \times r} & 0_{(m-r) \times (m-r)} \end{bmatrix}$$

Where Σ_r is r by r submatrix of non zero diagonal entries in Σ and

$$\Sigma \Sigma^\dagger - I = \begin{bmatrix} 0_{r \times r} & 0_{r \times (m-r)} \\ 0_{(n-r) \times r} & -I_{(m-r) \times (m-r)} \end{bmatrix}$$

Hence the multiplication yields the zero matrix.



Proposition IV

The vector $\hat{x} = A^+b$ is the shortest Least Squares solution to $Ax = b$, namely,

$$\|\hat{x}\|_2 = \min \{ \|x\|_2 : \|Ax - b\|_2 \text{ is minimal} \}$$

proof

Using the fact both U and V are Unitary

$$\begin{aligned} \min_{\|x\|_2} \left\{ \min_x \|Ax - b\|_2 \right\} &= \min_{\|x\|_2} \left\{ \min_x \|U\Sigma V^H x - b\|_2 \right\} \\ &= \min_{\|V^H x\|_2} \left\{ \min_x \|\Sigma V^H x - U^H b\|_2 \right\} \\ &= \min_{\|y\|_2} \left\{ \min_x \|\Sigma y - U^H b\|_2 \right\} \end{aligned}$$

Proof.

Observing at $\min_{\|y\|_2} \left\{ \min_x \|\Sigma y - U^H b\|_2 \right\}$. Since Σ is diagonal (Main diagonal to the least) there's only one Least Squares solution, $\hat{y} = \Sigma^+ U^H b$.

Thus,

$$\hat{x} = V \hat{y} = V \Sigma^+ U^H b$$

will attain the minimum norm. □

As written previously, any solution which holds the Normal Equations is the Least Squares solution.

$$\hat{x} = \min_x \|Ax - b\|_2 \Leftrightarrow (A^H A) \hat{x} = A^H b$$

Yet, one should observe $\hat{x} \in \mathcal{R}(A^H)$, namely, the solution lies in the Row Space of A . Hence, its norm is minimal among all solutions. In short, the Pseudo Inverse simultaneously minimizes the norm of the error as well as the norm of the solution.

Example I

Examining the following Linear System:

$$Ax = b$$

Where,

$$A = \begin{bmatrix} 8 & 10 & 3 & 30 \\ 9 & 6 & 6 & 18 \\ 1 & 1 & 10 & 3 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 6 \end{bmatrix}, \quad b = \begin{bmatrix} 217 \\ 147 \\ 51 \end{bmatrix}$$

Obviously, A^{-1} can't be calculated. Moreover, since $\text{rank}(A) = 3$ neither $(A^H A)^{-1}$ exists. Yet the Pseudo Inverse using the SVD does exist.

Using the SVD approach $A = U\Sigma V^H$.

Hence, $A^\dagger = V\Sigma^\dagger U^H$.

Using MATLAB to calculate the SVD yields:

$$\Sigma = \begin{bmatrix} 39.378 & 0 & 0 & 0 \\ 0 & 10.002 & 0 & 0 \\ 0 & 0 & 3.203 & 0 \end{bmatrix} \rightarrow \Sigma^\dagger = \begin{bmatrix} 0.025 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.312 \\ 0 & 0 & 0 \end{bmatrix}$$

Calculating \hat{x} yields:

$$\hat{x} = V\Sigma^\dagger U^H b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 6 \end{bmatrix} = x$$

The SVD canceled the 4th column which is dependent on the 2nd column of A .

Since $b \in \mathcal{R}(A)$ the exact solution could be calculated.

Example II

In this case

$$A = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 6 \end{bmatrix}, \quad b = \begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix}$$

Obviously, $b \notin \mathcal{R}(A)$. Neither A^{-1} nor $(A^H A)^{-1}$ exist. Using the SVD Pseudo Inverse:

$$\hat{x} = V \Sigma^+ U^H b = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

Examining the solution using the SVD. First, Since $\text{rank}(A) = 2$ its Column Space is spanned by the first 2 columns of U .

Calculating the projection of b onto the Column Space of A is

$$\text{given by } \hat{b} = \text{Proj}_{\mathcal{R}(A)}(b) = \sum_{i=1}^2 U^i H b U^i = \begin{bmatrix} 5 \\ 4 \\ 0 \end{bmatrix}.$$

Now given the updated Linear System $A\hat{x} = \hat{b}$ which has infinite number of solutions.

$$\text{One could calculate that } \mathcal{N}(A) = \text{span} \left(\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

$$\text{Hence } \hat{x} = \begin{bmatrix} \frac{\hat{b}_1}{A_{1,1}} \\ \frac{\hat{b}_2}{A_{2,2}} \\ 0 \\ 0 \end{bmatrix} + \left(s \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right) = \hat{x}_r + \hat{x}_n \text{ where } s, t \in \mathbb{R}.$$

The target is the solution with the minimum norm.

Since $\hat{x}_r \perp \hat{x}_n$ the norm of this solution is

$$\|\hat{x}\|^2 = \|\hat{x}_r\|^2 + \|\hat{x}_n\|^2$$

The minimum norm solution is obtained by taking $x_n = 0$.

This results in the Pseudo Inverse solution as above.

Numerically Sensitive Problems

Systems of equations that are poorly conditioned are sensitive to small change in values.

Since, practically speaking, there are always inaccuracies in measured data, the solution to these equations may be almost meaningless.

The SVD can help with the solution of ill-conditioned equations by identifying the direction of sensitivity and discarding that portion of the problem. The procedure will be illustrated by the following example.

Example III

Examining the following system of equations $Ax = b$

$$\begin{bmatrix} 1 + 3\epsilon & 1 - 3\epsilon \\ 3 - \epsilon & 3 + \epsilon \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

The SVD of A is

$$A = \frac{1}{\sqrt{20}} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 2\sqrt{5} & \\ & 2\epsilon\sqrt{5} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

From which the exact inverse of A is

$$A^{-1} = \sqrt{20} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{2\sqrt{5}} & \\ & \frac{1}{2\epsilon\sqrt{5}} \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 1 + \frac{3}{\epsilon} & 3 - \frac{1}{\epsilon} \\ 1 - \frac{3}{\epsilon} & 3 + \frac{1}{\epsilon} \end{bmatrix}$$

Easily, one can convince himself that for small ϵ the matrix A^{-1} has large entries which makes $x = A^{-1}b$ unstable.

Observe that the entry $\frac{1}{2\epsilon\sqrt{5}}$ multiplies the column $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. This is the sensitive direction. As b changes slightly, the solution changes in a direction mostly along the sensitive direction.

If ϵ is small, $\sigma_2 = 2\epsilon\sqrt{5}$ may be set to zero to approximate A .

$$A \approx \frac{1}{\sqrt{20}} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} 2\sqrt{5} & \\ & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The Pseudo Inverse is

$$A^{\dagger} = \sqrt{20} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{2\sqrt{5}} & \\ & 0 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & -1 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix}$$

In this case the multiplier of the sensitive direction vector is zero, no motion in the sensitive direction occurs. Any Least Squares solution to the equation $Ax = b$ is of the form $\hat{x} = A^\dagger b$ so that $\hat{x} = c \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for $c \in \mathbb{R}$, meaning perpendicular to the sensitive direction.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ | ≡ ↺ 🔍 ↻

$$Ax_0 = b_0 \Rightarrow x_0 = A^{-1}b_0$$

Let $b = b_0 + \delta b$ where δb is the error or noise, etc.

Therefore

$$Ax = b_0 + \delta b \Rightarrow x = A^{-1}b_0 + A^{-1}\delta b = x_0 + \delta x$$

Investigating how small or large is this error in the answer for a given amount of error. Note that

$$\delta x = A^{-1} \delta b \Rightarrow \|\delta x\| \leq \|A^{-1}\| \|\delta b\|$$

Or since $\|A^{-1}\| = \sigma_{\max}(A^{-1}) = \frac{1}{\sigma_{\min}(A)}$ the following holds

$$\|\delta x\| \leq \frac{\|\delta b\|}{\sigma_{min}(A)}$$

However recalling that $x_0 = A^{-1}b_0$ and therefore

$$\|x_0\| \geq \sigma_{\min}(A^{-1}) \|b_0\| = \frac{\|b_0\|}{\sigma_{\max}(A)}$$

Combining the equations yields

$$\frac{\|\delta x\|}{\|x_0\|} \leq \frac{\|\delta b\|}{\|b_0\|} \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

The last fraction, $\frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$, is called 'The Condition Number of A'. This number is indicative of the magnification of error in linear equation of interest. In most problems, a matrix with very large condition number is called ill conditioned and will result in severe numerical difficulties.

The solution to those numerical difficulties using the SVD is basically rank reduction:

- 1 Compute the SVD of A .
- 2 Examine the singular values of A and zero out any that are "small" to obtain a new approximate Σ matrix.
- 3 Compute the solution by $\hat{x} = V\Sigma^+U^Hb$.

Determining which singular values are "small" is problem dependent and requires some judgment.

1 Definitions and Notations

- Notations
- Definitions
- Introduction

- SVD Theorem
- Proof of the SVD Theorem
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- **Total Least Squares**
- Principal Component Analysis

In the classic Least Squares problems, the solution minimizing $\|Ax - b\|_2$ is sought after. The hidden assumption is that matrix A is correct, any error in the problem is in b .

The Least Squares problem finds a vector \hat{x} s.t.

$$\|A\hat{x} - b\|_2 = \min$$

which is accomplished by finding some perturbation r of the right hand side of minimum norm

$$Ax = b + r$$

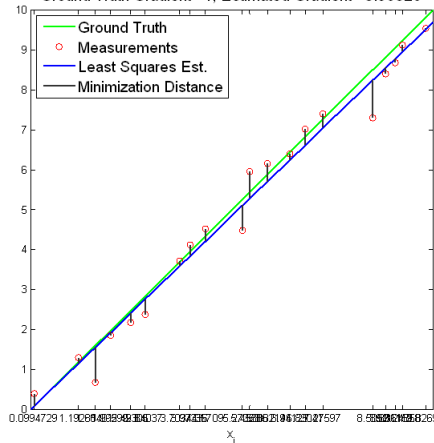
s.t. $(b + r) \in \mathcal{R}(A)$. In the Total Least Squares problem, both the right and left side of the equation assumed to have errors. The solution of the perturbed equation

$$(A + E)x = b + r$$

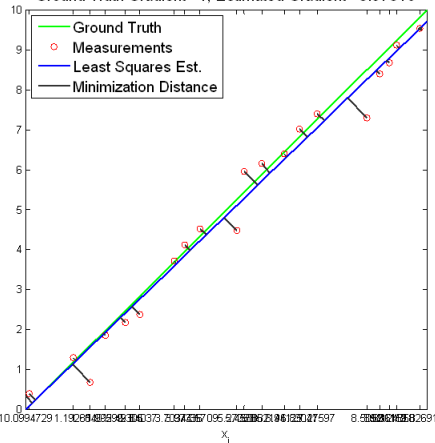
is sought s.t. $(b + r) \in \mathcal{R}(A + E)$ and the norm of the perturbations is minimized.

Intuitively, The right hand side is "bent" toward the left hand side while the left hand side is "bent" toward the right hand side.

Least Squares Minimizes Vertical Distance to Line
Ground Truth Gradient - 1. Estimated Gradient - 0.96826



Total Least Squares Minimizes Distance to Line
Ground Truth Gradient - 1. Estimated Gradient - 0.97016



Let A be an $m \times n$ matrix. To find the solution to the TLS problem one may observe the homogeneous form

$$[A + E | b + r] \begin{bmatrix} x \\ -1 \end{bmatrix} = 0 \rightarrow ([A | b] + [E | r]) \begin{bmatrix} x \\ -1 \end{bmatrix} = 0$$

Let $C = [A | b] \in \mathbb{C}^{m \times (n+1)}$ and let $\Delta = [E | r]$ be the perturbation of the data. In order for the homogeneous form to have solution the vector $\begin{bmatrix} x \\ -1 \end{bmatrix}$ must lie in the Null Space of $C + \Delta$ and in order for the solution not to be trivial, the perturbation Δ must be such that $C + \Delta$ is rank deficient.

Analyzing the TLS problem using the SVD. We bring $(A + E)x = (b + r)$ into the form

$$[A + E | b + r] \begin{bmatrix} x \\ -1 \end{bmatrix} = 0$$

Let $[A + E | b + r] = U\Sigma V^H$ be the SVD of the above form. If $\sigma_{n+1} \neq 0$ then $\text{rank}([A + E | b + r]) = n + 1$ which means the $\mathcal{R}([A + E | b + r]) = \mathbb{R}^{n+1}$, hence there's no nonzero vector in the orthogonal complement of the Row Space hence the set of equations is incompatible. To obtain solution the rank of $[A + E | b + r]$ must be reduced to n .

As shown before the best approximation of rank n in both Frobenius and l_2 norm is given by the SVD

$$[\hat{A} | \hat{b}] = U\hat{\Sigma}V^H, \quad \hat{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n, 0)$$

The minimal TLS correction is given by

$$\sigma_{n+1} = \min_{rank(\begin{bmatrix} \hat{A} | \hat{b} \end{bmatrix})=n} \|\begin{bmatrix} A | b \end{bmatrix} - \begin{bmatrix} \hat{A} | \hat{b} \end{bmatrix}\|_F$$

Attained for

$$[E|r] = \sigma_{n+1} u_{n+1} v_{n+1}^H$$

Note that the TLS correction matrix has rank one.

It is clear that the approximate set $[\hat{A}|\hat{b}] \begin{bmatrix} x \\ -1 \end{bmatrix} = 0$ is compatible and the solution is given by the only vector, v_{n+1} , that belongs to $\mathcal{N}([\hat{A}|\hat{b}])$.

The TLS solution is obtained by scaling v_{n+1} until its last component equals to -1 , or

$$\begin{bmatrix} x \\ -1 \end{bmatrix} = \frac{-1}{V_{n+1,n+1}} v_{n+1}$$

For simplicity it assumed that $V_{n+1,n+1} \neq 0$ and $\sigma_n > \sigma_{n+1}$ hence the solution exists and it is unique. Otherwise, the solution might not exists or isn't unique (Any superposition of few columns of V). For complete analysis of the existence and uniqueness of the solution see [1].

Basic algorithm of the TLS would be:

Given $Ax \approx b$, where $A \in \mathbb{C}^{m \times n}$, $b \in \mathbb{C}^m$ the TLS solution could be obtained by

- Compute the SVD of $[A|b] = U\Sigma V^H$.
- If $V_{n+1,n+1} \neq 0$ the TLS solution would be

$$x_{TLS} = \frac{-1}{V_{p+1,p+1}} v_{n+1} (1:n)$$

Let $C = U\Sigma V^H$ From the definition of the $\| \cdot \|_F$ Norm of a matrix

$$\frac{\|Cv\|_2}{\|v\|_2} \geq \sigma_{n+1}$$

Where $\|v\|_2 \neq 0$. Equality holds if and only if $v \in S_c$ where $S_c = \text{span}\{v_i\}$ and v_i are the columns of V which satisfy $u_i^H C v_i = \sigma_{n+1}$.

The TLS problem amounts to finding vector x s.t.

$$\frac{\left\| [A|b] \begin{bmatrix} x \\ -1 \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} x \\ -1 \end{bmatrix} \right\|_2} = \sigma_{n+1}$$

By squaring everywhere

$$\min_x \frac{\left\| \begin{bmatrix} A|b \end{bmatrix} \begin{bmatrix} x \\ -1 \end{bmatrix} \right\|_2^2}{\left\| \begin{bmatrix} x \\ -1 \end{bmatrix} \right\|_2^2} = \min_x \sum_{i=1}^m \frac{|A_i^H x - b_i|^2}{x^H x + 1}$$

The quantity $\frac{|A_i^H x - b_i|^2}{x^H x + 1}$ is the square of the distance from the point $\begin{bmatrix} A_i^H \\ b \end{bmatrix} \in \mathbb{C}^{n+1}$ to the nearest point on the hyperplane P defined by

$$P = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} \mid a \in \mathbb{C}^n, b \in \mathbb{C}, b = x^H a \right\}$$

So the TLS problem amounts to finding the closest hyperplane to the set of points $\left\{ \begin{bmatrix} A_1^H \\ b_1 \end{bmatrix}, \begin{bmatrix} A_2^H \\ b_2 \end{bmatrix}, \dots, \begin{bmatrix} A_m^H \\ b_m \end{bmatrix} \right\}$.

The minimum distance property can be shown as following. Let P be the plane orthogonal to the normal vector $n \in \mathbb{R}^{n+1}$ s.t.

$$P = \left\{ r \in \mathbb{C}^{n+1} : r^H n = 0 \right\}$$

and let n have the following form $n = \begin{bmatrix} x \\ -1 \end{bmatrix}$. Let $p = \begin{bmatrix} A_m^H \\ b_m \end{bmatrix}$ be a point in \mathbb{C}^{n+1} . Finding a point, $q \in \mathbb{C}^{n+1}$ which belongs to the plane P and is closest to the point p is a constrained optimization problem, minimize $\|p - q\|$ subject to $n^H q = 0$. The minimization function

$$\begin{aligned} J(q) &= \|p - q\|^2 + 2\lambda n^H q = p^H p - 2p^H q + 2\lambda n^H q + q^H q \\ &= (q - p + \lambda n)^2 (q - p + \lambda n) + 2\lambda p^H n - \lambda^2 n^H n \end{aligned}$$

This is clearly minimized when $q = p - \lambda n$.

Determining λ by the constrain

$$n^H q = n^H p - \lambda n^H n = 0 \rightarrow \lambda = \frac{n^H p}{n^H n}$$

Inserting results into the "Minimization Function" yields

$$\begin{aligned} J(q) &= 2\lambda p^H n - \lambda^2 n^H n \\ &= \frac{2n^H p p p^H n}{n^H n} - \frac{n^H n p^H p n^H n}{n^H n n^H n} \\ &= \frac{(n^H p)^2}{n^H n} = \frac{(x^H A_m - b_m)^2}{x^H x + 1} \end{aligned}$$

Alternative solution using the "Projection Theorem". The distance from the point p to the plane P can be found by finding the length of the projection of p onto n , which yields

$$d_{min}^2(p, P) = \frac{\langle p, n \rangle^2}{\|n\|^2} = \frac{[x^H, -1] \begin{bmatrix} A_m^H \\ b_m \end{bmatrix}}{x^H x + 1}$$

1 Definitions and Notations

- Notations
- Definitions
- Introduction

2 Singular Value Decomposition Theorem

- SVD Theorem
- Proof of the SVD Theorem
- SVD Properties
- SVD Example

3 Applications

- Order Reduction
- Solving Linear Equation System
- Total Least Squares
- Principal Component Analysis

Principal Component Analysis (PCA) is the mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called Principal Components.

The number of Principal Components is less than or equal to the number of original variables.

This transformation is defined in such way that the first component has a variance as high as possible (That's, accounts for as much of the variability in data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (Uncorrelated with) the preceding components.

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (Called the first Principal Component), the second greatest variance on the second coordinate and so on.

Assuming given a collection of data of columns vectors

$$a_1, a_2, \dots, a_m \in \mathbb{R}^n.$$

The projection of the data onto a subspace $U \in \mathbb{R}^r, r \leq m$ which is spanned by the orthogonal basis u_1, u_2, \dots, u_r is given by

$$a_i = f_{i1}u_1 + f_{i2}u_2 + \dots + f_{ir}u_r, i = 1 : m$$

for some coefficients f_{ij} .

Note that $f_{ij} = a_i^H u_j$, the projection of a_i along the direction of u_j .

By the Projection Theorem This projection is the closest in the l_2 - norm sense to the data given by a_i .

The search is after the orthogonal basis u_1, u_2, \dots, u_r .

Formulation the constraint of maximization of the variance along the direction of u_1 yields

$$\max_{\|w\|=1} \sum_{i=1}^m |a_i^H w|^2 = \|A^H w\|^2 = (A^H w)^H (A^H w) = w^H A A^H w$$

Using the SVD of $A = U \Sigma V^H$, Then $A A^H = U \Sigma \Sigma^H U^H$.

Observing,

$$\frac{w^H A A^H w}{w^H w} = \frac{(U^H w)^H \Sigma \Sigma^H (U^H w)}{(U^H w)^H (U^H w)}$$

Noticing that there are only r non zero entries in Σ by the properties the SVD. Defining $x = U^H w$ yields

$$\frac{w^H A A^H w}{w^H w} = \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \dots + \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \dots + x_m^2}$$

Now we have

$$\max_{w \neq 0} \frac{w^H A A^H w}{w^H w} = \max_{w \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \dots + \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \dots + x_m^2}$$

Assuming $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$.

Then

$$\max_{w \neq 0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \dots + \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \dots + x_m^2} = \sigma_1^2 = \lambda_1$$

Which the largest eigenvalue of A . The vector x which makes the maximum is $x_1 = 1$ and $x_i = 0$ for $i = 2 : m$. Which corresponds to $w = Ux = u_1$.

The first Principal Component is indeed achieved by the first eigenvector u_1 of AA^H .

Calculating the second Principal Component under the constraint being orthogonal to the first and maximizing the projection

$$\max_{\|w\|=1, w^H u_1=0} \sum_{i=1}^m |a_i^H w|^2 = \max_{w \neq 0, w^H u_1=0} \frac{w^H (AA^H) w}{w^H w}$$

Using the definition from above yields

$$\begin{aligned} \max_{x \neq 0, x^H U^H u_1=0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \dots + \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \dots + x_m^2} = \\ \max_{x \neq 0, x_1=0} \frac{\sigma_1^2 x_1^2 + \sigma_2^2 x_2^2 + \dots + \sigma_r^2 x_r^2}{x_1^2 + x_2^2 + \dots + x_m^2} = \sigma_2^2 = \lambda_2 \end{aligned}$$

Which is the second largest eigenvalue of AA^H .

The vector x which makes the maximum is $x_2 = 1$ and $x_i = 0$ for $i = 1, 3 : m$.

This corresponds to $w = Ux = u_2$, The second Eigenvector, u_2 , of AA^H .

Continuing this pattern, u_i is the i th Principal Component.

The set of orthogonal vectors which spans the subspace the data is projected to and maximizes the variance of the data is the first r vectors which consists the orthogonal matrix from the SVD, U .

Observing the SVD yields the the result immediately

$$A = U\Sigma V^H \rightarrow Y = U^H A = \Sigma V^H$$

Observing the scatter matrix of Y

$$C_Y = YY^H = U^H A (U^H A)^H = U^H A A^H U = U^H C_X U$$

Since the matrix U is the Eigenvectors matrix of $C_X = XX^H$ by the Diagonalization Theorem C_Y is diagonal. Another look yields

$$YY^H = \Sigma V^H (\Sigma V^H)^H = \Sigma V^H V \Sigma^H = \Sigma \Sigma^H = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2)$$

Namely, the Scatter matrix, hence the Covariance Matrix of Y is diagonal. Moreover, The constraint on the variance holds.

- The SVD is a decomposition **which can be applied on any matrix**.
- The SVD **exposes fundamental properties** of a linear operator such as the fundamental spaces, Frobenius Norm and l_2 Norm.
- The SVD **can be utilized in many applications** such as solving linear systems (Least Squares, Total Least Squares) and order reduction (Compression, Noise Reduction, Principal Component Analysis).
- To Be Continued
 - Regulating Linear Equations System.



A. Author.

Handbook of Everything.

Some Press, 1990.



S. Someone.

On this and that.

Journal on This and That. 2(1):50–100, 2000.



R. Avital.

On this and that.

Journal on This and That. 2(1):50–100, 2000.