

Maths for Signals and Systems

Session 4 [Week 6]

Pier Luigi Dragotti

EEE Department

Imperial College London

p.dragotti@imperial.ac.uk

Outline

- The goal of this group of lectures is:
 - to introduce the SVD
 - to relate the SVD to the pseudo-inverse
 - to show some applications of the SVD

Singular Value Decomposition (SVD)

- The so called **Singular Value Decomposition (SVD)** is one of the main tools in Linear Algebra.
- Consider a matrix A of dimension $m \times n$ and rank r .
- We would like to diagonalize A . What we know so far is $A = S\Lambda S^{-1}$. This diagonalization has the following weaknesses:
 - A has to be square.
 - There are not always enough eigenvectors (defective matrices).
 - For example consider the matrix $\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$, $a \neq 0$. It only has the eigenvector $\begin{bmatrix} x & 0 \end{bmatrix}^T$.
- **Goal: Let's look for a type of decomposition which can be applied to any matrix.**

Singular Value Decomposition (SVD) cont.

- **Theorem** (without proof): Every matrix $A \in \mathbb{C}^{m \times n}$ can be factored as $A = U\Sigma V^H$ where
 - U is a unitary matrix ($U^H U = I$) with columns u_i and dimension $m \times m$.
 - Σ is an $m \times n$ rectangular matrix with non-negative real entries only along the main diagonal. The main diagonal is defined by the elements σ_i , $i = 1, \dots, p$ and $p = \min(m, n)$;
 - V is a unitary matrix ($V^H V = I$) with columns v_i and dimension $n \times n$.
- U is, in general, different to V .
- The above type of decomposition is called **Singular Value Decomposition**.
- The diagonal elements of Σ are called **Singular Values** of matrix A . They are chosen to be **positive** and are usually ordered.
- When A is a square invertible matrix then $A = S\Lambda S^{-1}$.
- When A is a symmetric matrix, the eigenvectors of S are orthonormal, so $A = Q\Lambda Q^T$.
- Therefore, for symmetric matrices SVD is effectively an eigenvector decomposition $U = Q = V$ and $\Lambda = \Sigma$ (if eigenvalues are all positive).

Singular Value Decomposition (SVD) cont.

- Remember that U and V are unitary square matrices and therefore,

$$U^H U = U U^H = V^H V = V V^H = I$$

- Σ is instead rectangular so if $m = 2$ and $n = 3$, $\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}$. If $m = 3$ and

$$n = 2, \Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix}$$

- We note that:

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

- In the above expression $\Sigma^T \Sigma$ is a diagonal matrix of dimension $(n \times m) \times (m \times n) = n \times n$ (square matrix).
- Therefore, the above expression shows that σ_i^2 are the eigenvalues of $A^H A$ and V the eigenvectors

Singular Value Decomposition (SVD) cont.

- Similarly,

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

- In the above expression $\mathbf{\Sigma}\mathbf{\Sigma}^T$ is a matrix of dimension $(m \times n) \times (n \times m) = m \times m$.
- Therefore, the above expression shows that σ_i^2 are the eigenvalues of $\mathbf{A}\mathbf{A}^H$ and \mathbf{U} are the eigenvectors
- Based on the properties stated in previous slides, the number and values of non-zero elements of matrices $\mathbf{\Sigma}\mathbf{\Sigma}^T$ and $\mathbf{\Sigma}^T\mathbf{\Sigma}$ are identical. Note that these two matrices have different dimensions if $m \neq n$. In that case one of them (the bigger one) has at least one zero element in its diagonal
- From this discussion, we conclude that we can determine all the factors of SVD by the eigenvector decompositions of matrices $\mathbf{A}^H\mathbf{A}$ and $\mathbf{A}\mathbf{A}^H$.

Useful properties

- Matrices A , $A^H A$ and AA^H have the same rank.
- Let A be an $m \times n$ matrix with rank r . The matrix A has r **singular values**. Both $A^H A$ and AA^H have r non-zero eigenvalues which are the squares of the singular values of A . Furthermore:
 - $A^H A$ is of dimension $n \times n$. It has r eigenvectors $[v_1 \dots v_r]$ associated with its r non-zero eigenvalues and $n - r$ eigenvectors associated with its $n - r$ zero eigenvalues.
 - AA^H is of dimension $m \times m$. It has r eigenvectors $[u_1 \dots u_r]$ associated with its r non-zero eigenvalues and $m - r$ eigenvectors associated with its $m - r$ zero eigenvalues.

Singular Value Decomposition (SVD) cont.

- We can write $\mathbf{V} = [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_r \quad \mathbf{v}_{r+1} \quad \dots \quad \mathbf{v}_n]$ and $\mathbf{U} = [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_r \quad \mathbf{u}_{r+1} \quad \dots \quad \mathbf{u}_m]$.
- Matrices \mathbf{U} and \mathbf{V} have already been defined previously.
- Note that in the above matrices, we put first in the columns the eigenvectors of $\mathbf{A}^H \mathbf{A}$ and $\mathbf{A} \mathbf{A}^H$ which correspond to non-zero eigenvalues.
- To take the above even further, we order the eigenvectors according to the magnitude of the associated eigenvalue.
- The eigenvector that corresponds to the maximum eigenvalue is placed in the first column and so on.

Singular Value Decomposition (SVD) cont.

- As already shown, from $A = U\Sigma V^H$ we obtain that $AV = U\Sigma$ or

$$A[v_1 \quad \dots \quad v_r \quad v_{r+1} \quad \dots \quad v_n] = [u_1 \quad \dots \quad u_r \quad u_{r+1} \quad \dots \quad u_m]\Sigma$$
- Therefore, we can break $AV = U\Sigma$ into a set of relationships of the form $Av_i = \sigma_i u_i$. Note that σ_i is a scalar and v_i and u_i vectors.
- For $i \leq r$ the relationship $AV = U\Sigma$ tells us that:
 - The vectors u_1, u_2, \dots, u_r are in the column space of A . This observation comes directly from $u_i = \frac{1}{\sigma_i} Av_i$, $\sigma_i \neq 0$, i.e., u_i s are linear combinations of columns of A . Furthermore, the u_i s associated with $\sigma_i \neq 0$ are orthonormal. Thus, they form a basis of the column space.
 - The vectors v_1, v_2, \dots, v_r are in the row space of A . This is because from $AV = U\Sigma$ we have $U^H AVV^H = U^H U \Sigma V^H \Rightarrow U^H A = \Sigma V^H \Rightarrow v_i^H = \frac{1}{\sigma_i} u_i^H A$, $\sigma_i \neq 0$.
Furthermore, since the v_i 's associated with $\sigma_i \neq 0$ are orthonormal, they form a basis of the row space.

Singular Value Decomposition (SVD) cont.

- Based on the facts that:
 - $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$,
 - \mathbf{v}_i form an orthonormal basis of the row space of A ,
 - \mathbf{u}_i form an orthonormal basis of the column space of A , we conclude that:
- The $n - r$ additional \mathbf{v} 's which correspond to the zero eigenvalues of matrix $A^H A$ are taken from the null space of A .
- The $m - r$ additional \mathbf{u} 's which correspond to the zero eigenvalues of matrix AA^H are taken from the null space of A^H

Singular Value Decomposition (SVD) cont.

To summarize:

- The four subspaces are related as follows:
 - $\mathbb{C}^m = \mathfrak{R}(A) \oplus \mathcal{N}(A^H)$, $\mathfrak{R}^\perp(A) = \mathcal{N}(A^H)$, $\mathcal{N}^\perp(A^H) = \mathfrak{R}(A)$,
 - $\mathbb{C}^n = \mathfrak{R}(A^H) \oplus \mathcal{N}(A)$, $\mathfrak{R}^\perp(A^H) = \mathcal{N}(A)$, $\mathcal{N}^\perp(A) = \mathfrak{R}(A^H)$,
- We split $[v_1 \ \dots \ v_r \ v_{r+1} \ \dots \ v_n] = [V_1 \ V_2]$ and $[u_1 \ \dots \ u_r \ u_{r+1} \ \dots \ u_m] = [U_1 \ U_2]$, it then follows that
 - $\mathfrak{R}(A) = \text{span}\{U_1\}$,
 - $\mathcal{N}(A) = \text{span}\{U_2\}$,
 - $\mathfrak{R}(A^H) = \text{span}\{V_1\}$,
 - $\mathcal{N}(A^H) = \text{span}\{V_2\}$

Truncated or Reduced Singular Value Decomposition

- In the expression for SVD we can reformulate the dimensions of all matrices involved by ignoring the eigenvectors which correspond to zero eigenvalues.
- In that case we have:

$$A \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_r \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix} \Rightarrow A = u_1 \sigma_1 v_1^H + \dots + u_r \sigma_r v_r^H$$

where:

- The dimension of A is $m \times n$.
- The dimension of $\begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_r \end{bmatrix}$ is $n \times r$.
- The dimension of $\begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r \end{bmatrix}$ is $m \times r$.
- The dimension of Σ is $r \times r$.
- The above formulation is called **Truncated or Reduced Singular Value Decomposition**.
- So the SVD can be written as a sum for r rank one matrices:

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

Pseudoinverses and the SVD

- We now go back to the solution to $Ax = b$ when A is neither full column or row rank.
- Claim:** The Minimum Norm Least Squares solution to $Ax = b$ is given by $x = A^+b$ where A^+ is the pseudoinverse of $A = U\Sigma V^H$ and is given by $A^+ = V\Sigma^+U^H$ where

$$\Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix}$$

Pseudoinverses and the SVD

- **Proof:** Consider first the following case:

- $$\begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

- The solution to the normal equation is
$$\begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \\ \hat{x}_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ 0 \\ 0 \end{pmatrix}$$

- The null space of A is $[0, 0, 1, 0]^T$ and $[0, 0, 0, 1]^T$. The general solution is then

- $$\hat{\mathbf{x}} = \begin{pmatrix} b_1/\sigma_1 \\ b_2/\sigma_2 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \mathbf{x}_r + \mathbf{x}_n \quad \text{with } s, t \in \mathbb{R}$$

Pseudoinverses and the SVD

- Because of the orthogonality of \mathbf{x}_r and \mathbf{x}_n the norm is $\|\hat{\mathbf{x}}\|^2 = \|\mathbf{x}_r\|^2 + \|\mathbf{x}_n\|^2$
- The minimum norm solution is then obtained by taking $\mathbf{x}_n = 0$ and is given by

$$\hat{\mathbf{x}} = \begin{pmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

- In general when \mathbf{A} is diagonal with r nonzero entries the pseudoinverse is:

$$\mathbf{A}^+ = \mathbf{\Sigma}^+ = \begin{pmatrix} \frac{1}{\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \end{pmatrix}$$

Pseudoinverses and the SVD

- In the general case we have:
- $\min \|Ax - b\| = \min \|U\Sigma V^H x - b\| = \min \|\Sigma V^H x - U^H b\|$ where the last equality follows from the fact that U is unitary.
- We denote with $v = V^H x$ and $\hat{b} = U^H b$ then the least-squares problem reduces to $\min \|\Sigma v - \hat{b}\|$ for which we know the solution is $v = \Sigma^+ \hat{b}$.
- By working backwards we then find the MNLS solution: $V^H \hat{x} = \Sigma^+ U^H b \Rightarrow$

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

Ax=b Summary

Scenario	Number of exact Solutions	Type of Solution	Solution
A 'tall' and full column rank	0 or 1	Least square: $\min_{\mathbf{x}} \ A\mathbf{x} - \mathbf{b}\ $	$\mathbf{x}_{ls} = (A^H A)^{-1} A^H \mathbf{b}$
A 'fat' and full row rank	Infinitely many solutions	Minimum-norm solution: $\min \ \mathbf{x}\ $ subject to $A\mathbf{x} = \mathbf{b}$	$\mathbf{x}_{MN} = A^H (A A^H)^{-1} \mathbf{b}$
A neither full row nor full column rank	0 or infinitely many solutions	Minimum-norm least square solution	$\hat{\mathbf{x}} = V \Sigma^+ U^H \mathbf{b}$ with $A = U \Sigma V^H$
A square and full rank (trivial)	1	Unique exact solution	$\mathbf{x} = A^{-1} \mathbf{b}$

Other Applications of the SVD : Total Least squares problems

- Let's look again at the minimization of $\|Ax - b\|$ when A is full column rank.
- The LS solution essentially finds a \hat{b} such that $\|b - \hat{b}\|$ is minimized.
- This is like assuming that A is noiseless and that the b we observe has been corrupted by noise. The x we find is the solution to $Ax = b + r$ where r is a perturbation we want to minimize
- In some cases both A and b are corrupted by noise
- We then want to find a solution to the perturbed equation: $(A + E)x = b + r$
- It turns out that this solution is found by using the SVD

Other Applications of the SVD : Total Least squares problems

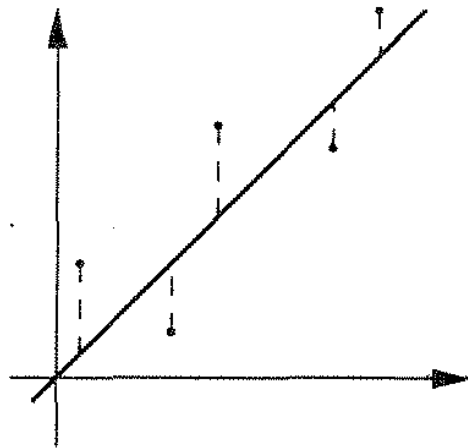
- When the solution exists is given by
 1. Compute the SVD $[A \ b] = U\Sigma V^H$
 2. Then $\hat{x} = -v_{n+1}(1:n)/v_{n+1}(n+1)$

- Sketch of the proof:
 - We write $Ax = b$ in homogeneous form: $Cy=0$ with $C = [A \ b]$ and $y = \begin{pmatrix} x \\ -1 \end{pmatrix}$
 - We try to find $\min \|Cy\|^2$ such that $\|y\|^2 \neq 0$
 - We use Lagrangian multipliers to solve this constrained minimization:

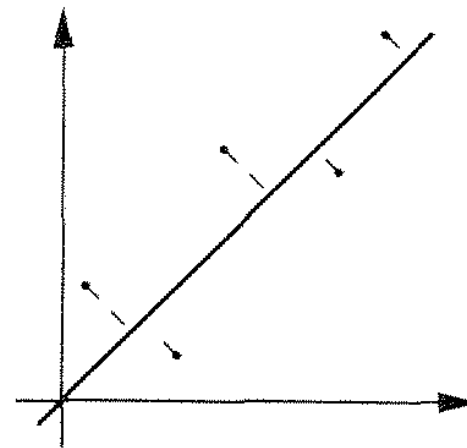
$$J = y^H C^H C y - \lambda y^H y$$
 - Taking the gradient with respect to y and equating to zero yields $C^H C y = \lambda y$
 - Solution is in the direction of the eigenvector corresponding to smallest eigenvalue of $C^H C$ this is equivalent to the column of V corresponding to smallest singular value: $y = v_{n+1}$
 - In order to find x we rescale v_{n+1} properly and we arrive at the solution

$$\hat{x} = -v_{n+1}(1:n)/v_{n+1}(n+1)$$

Least-squares vs total least-squares



(a) LS: Minimize vertical distance to line



(b) TLS: Minimize total distance to line

Figure 7.2: Comparison of least-squares and total least-squares fit

- **Note:** Figure taken from T. Moon and W. Stirling, “Mathematical Methods and Algorithms for Signal Processing”, page 382.

Other Applications of the SVD : Low-Rank Approximation

- You are given a matrix \mathbf{D} and you want to approximate it with a rank r matrix $\hat{\mathbf{D}}$
- This may happen because, for example, you observe a noisy \mathbf{D} and you know that the original one was rank deficient, you want a solution such that $\|\mathbf{D} - \hat{\mathbf{D}}\|$ is minimized.
- Recall that the Frobenius norm of a matrix \mathbf{A} is:

$$\|\mathbf{A}\| = \sqrt{\sum |a_{i,j}|^2} = \left(\text{trace}(\mathbf{A}^H \mathbf{A}) \right)^{\frac{1}{2}} = (\sum_i \sigma_i^2)^{\frac{1}{2}}$$

- Therefore, let $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, then $\hat{\mathbf{D}} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^H$

Other Applications of the SVD : PCA

- You have a data matrix X with n columns and m rows.
- The columns are the number of samples and for each sample we measure m variables (e.g., height, weight, etc.)
- We remove the mean value along each row
- The principal components of the data are selected so that the i –th principal component is the linear combination of the data that accounts for the i –th largest portion of the variance of the data

Other Applications of the SVD : PCA

- The first principal component \mathbf{a}_1 is a linear combination $\mathbf{y}_1^T = \mathbf{a}_1^T \mathbf{X}$ such that the sample variance of $y_{1,i}$ is maximized subject to $\|\mathbf{a}_1\| = 1$.
- $\sigma_{y_1}^2 = \frac{1}{n-1} \sum_{i=1}^n y_{1,i}^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{a}_1^T \mathbf{x}_i)^2 = \mathbf{a}_1^T \mathbf{S} \mathbf{a}_1$ where $\mathbf{S} = \frac{1}{n-1} \mathbf{X} \mathbf{X}^T$
- Maximizing $\mathbf{a}_1^T \mathbf{S} \mathbf{a}_1$ subject to $\|\mathbf{a}_1\| = 1$ is a problem we have met before and the solution is the eigenvector corresponding to the largest eigenvalue of \mathbf{S} which is equivalent to the left singular vector of \mathbf{X} related to its largest singular value
- The second principal component is chosen so that $\mathbf{y}_2^T = \mathbf{a}_2^T \mathbf{X}$ is uncorrelated with \mathbf{y}_1 which leads to the constraint $\mathbf{a}_2^T \mathbf{a}_1 = 0$.
- This is the singular vector related to the second largest singular value of \mathbf{X}

PCA: A visual illustration

- Assume you have a data matrix X related to the age and height of 25 teenagers so $n=25$ $m=2$.
- The two variables are clearly correlated but you are after the principal axis
- Remove the mean from X to obtain X_0
- Compute the left singular vector $\mathbf{u}_1 = [u_{11} \ u_{21}]^T$ of X_0 related to the largest singular value. Correlation is in this case positive (i.e., older kids are normally taller)
- The second principal component is also visualized

