Linear Algebra for Wireless Communications

Lecture: 8

Singular Value Decomposition SVD

Ove Edfors
Department of Electrical and Information Technology
Lund University

A bit of repetition

A very useful matrix decomposition comes from the Spectral Theorem:

REAL CASE

A real symmetric matrix **A** can be factored into $\mathbf{Q\Lambda Q}^T$. The ortonormal eigenvectors of **A** are in the orthogonal matrix **Q** and the corresponding eigenvalues in the diagonal matrix **Λ**.

COMPLEX CASE

A Hermitian matrix **A** can be factored into $\mathbf{U\Lambda U}^H$. The ortonormal eigenvectors of **A** are in the unitary matrix **U** and the corresponding eigenvalues in the diagonal matrix **Λ**.

We have seen that we, e.g., can reduce the complexity of channel estimation in OFDM systems and decouple otherwise coupled systems linear systems.

What a pity that it only works with *symmetric* or *Hermitian* matrices, i.e., very specific types of square matrices!

Is there a generalization?

Would it be possible to replace (the restricted)

$$\mathbf{A} = \mathbf{Q\Lambda Q}^T$$

where **Q** is orthogonal and **Λ** is diagonal, with something (more general) like

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

where **U** and **V** are still orthogonal (but not necessarily the same) and **Σ** is still diagonal?

This would give us essentially the same "orthogonal transform" property as we have used when we had *symmetric* or *Hermitian* matrices.

... so, does this work?

Let's try the factorization

Let's assume that a factorization

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

extists. What would that imply?

First look at $\mathbf{AA}^T$:

$$\underbrace{\mathbf{AA}^T}_{\text{symmetric}} = \left(\mathbf{U\Sigma V}^T\right)\mathbf{V\Sigma}^T\mathbf{U}^T = \mathbf{U\Sigma\Sigma}^T\mathbf{U}^T$$

$$= \mathbf{Q}_1\mathbf{\Lambda}_1\mathbf{Q}_1^T \text{ (spectral theorem!)}$$

... then at $\mathbf{A}^T\mathbf{A}$:

$$\underbrace{\mathbf{A}^T\mathbf{A}}_{\text{symmetric}} = \mathbf{V\Sigma}^T\mathbf{U}^T\left(\mathbf{U\Sigma V}^T\right) = \mathbf{V\Sigma}^T\mathbf{\Sigma V}^T$$

$$= \mathbf{Q}_2\mathbf{\Lambda}_2\mathbf{Q}_2^T \text{(spectral theorem!)}$$

This seem to work only if $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$ have some very specific relations between their eigenvalues.

Eigenvalues of $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$

Let's find out the properties of the eigenvalues of $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$.

Do they have the same eigenvalues?

If $\mathbf{x}$ is an eigenvector of $\mathbf{AA}^T$ and λ the corresponding eigenvalue

$$\mathbf{AA}^T\mathbf{x} = \lambda\mathbf{x}$$

then, multiplying by $\mathbf{A}^T$ from the left,

$$\mathbf{A}^T\underline{\mathbf{AA}^T\mathbf{x}} = \lambda\underline{\mathbf{A}^T\mathbf{x}}$$

shows us that $\mathbf{A}^T\mathbf{x}$ is and eigenvector of $\mathbf{A}^T\mathbf{A}$ and the same λ is its eigenvalue. **YES!**

Any other useful properties?

Since $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$ are symmetric, the eigenvalues are real.

Since the quadratic form $\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = ||\mathbf{Ax}||^2 \geq 0$, all eigenvalues of $\mathbf{A}^T\mathbf{A}$ (and $\mathbf{AA}^T$) are non-negative.

Note that there are no restrictions on **A**. It doesn't even have to be square.

The singular value decomposition

The above leads us to the following theorem:

Singular Value Decomposition

Any M x N matrix **A** can be factored into

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

The columns of **U** (M x M) are eigenvectors of $\mathbf{AA}^T$, and the columns of **V** (N x N) are eigenvectors of $\mathbf{A}^T\mathbf{A}$. The r singular values on the diagonal of **Σ** (M x N) are the square roots of the nonzero eigenvalues of both $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$.

This is a very powerful decomposition ... it works on everything and gives us the practical (orthogonal)(diagonal)(orthogonal) form that simplifies many problems!

The singular value decomposition

About the structure:

Are often sorted $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$

$$\underset{M \times N}{\mathbf{A}} = \underset{M \times M}{\mathbf{U}}\underset{M \times N}{\mathbf{\Sigma}}\underset{N \times N}{\mathbf{V}^T} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_m \\ | & & | \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & & \end{bmatrix}\begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}$$

$\mathbf{u}_1$ to $\mathbf{u}_r$ spans $C(\mathbf{A})$
$\mathbf{u}_{r+1}$ to $\mathbf{u}_M$ spans $N(\mathbf{A}^T)$

Implies that r = rank(A)

Since the singular values are the square-roots of the non-zero eigenvalues of both $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$, there can be at most min(m,n) of them, i.e., $r \leq \min(m,n)$.

$\mathbf{v}_1$ to $\mathbf{v}_r$ spans $C(\mathbf{A}^T)$
$\mathbf{v}_{r+1}$ to $\mathbf{v}_N$ spans $N(\mathbf{A})$

In certain applications, e.g. linear estimation with strong correlation, we have the property that $r \ll \min(m,n)$ and we can use this to simplify calculations.

SVD of symmetric matrices?

What about

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

when $\mathbf{A} = \mathbf{A}^T$?

We know that there is a spectral factorization $\mathbf{A} = \mathbf{Q\Lambda Q}^T$ of such a symmetric matrix, and we have

$$\mathbf{AA}^T = \mathbf{Q\Lambda Q}^T\mathbf{Q\Lambda}^T\mathbf{Q}^T = \mathbf{Q\Lambda}^2\mathbf{Q}^T$$

Real eigenvalues

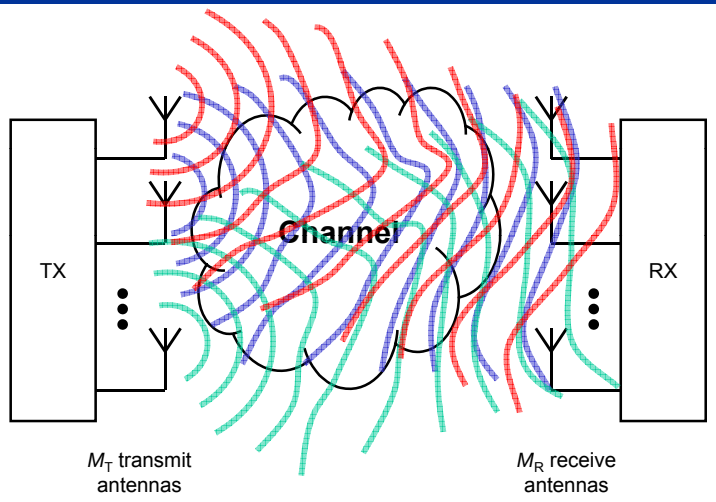
Square root of these diagonal elements are singular values

Conclusion: The singular values of a symmetric matrix are the absolute values of its nonzero eigenvalues.

Can you find a complete SVD from the spectral factorization?

Capacity of MIMO-systems

A MIMO system



A general (narrow-band) model

The "general" case with M_T TX antennas and M_R RX antennas:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{M_R} \end{bmatrix} = \underbrace{\begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,M_T} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,M_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_R,1} & h_{M_R,2} & \cdots & h_{M_R,M_T} \end{bmatrix}}_{\text{Channel matrix } (M_R \times M_T)} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{M_T} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{M_T} \end{bmatrix} = \mathbf{Hx} + \mathbf{n}$$

Received vector (M_R x 1)

Transmitted vector (M_T x 1)

Receiver noise vector (M_R x 1)

It is not trivial to figure out the capacity of this MIMO system, but the rectangular channel matrix calls for trying a singular value decomposition to obtain a somewhat simpler system to analyze.

Capacity – No fading & AWGN [1]

Singular value decomposition of the (fixed) channel **H**:

$$\mathbf{y} = \mathbf{Hx} + \mathbf{n} = \mathbf{U\Sigma V}^H\mathbf{x} + \mathbf{n}$$

where **U** (M_R x M_R) and **V** (M_T x M_T) are unitary matrices and **Σ** (M_R x M_T) is a matrix containing the singular values on its diagonal.

Multiply by $\mathbf{U}^H$ from left:

$$\underbrace{\mathbf{U}^H\mathbf{y}}_{\tilde{\mathbf{y}}} = \mathbf{\Sigma}\underbrace{\mathbf{V}^H\mathbf{x}}_{\tilde{\mathbf{x}}} + \underbrace{\mathbf{U}^H\mathbf{n}}_{\tilde{\mathbf{n}}} \quad \Longrightarrow \quad \tilde{\mathbf{y}} = \mathbf{\Sigma}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}$$

Only "rotations" of **y**, **x** and **n**.

All-zero, exept diagonal.

Capacity – No fading & AWGN [2]

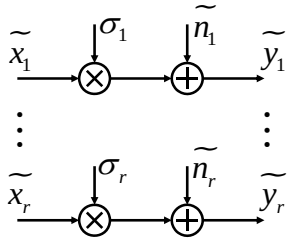
What have we obtained?

Parallel independent channels:

$$\tilde{\mathbf{y}} = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} \tilde{\mathbf{x}} + \tilde{\mathbf{n}}$$

Number of singular values r = rank(**H**).

Shannon's "standard case":



(+ channels with $\sigma_k = 0$)

Capacity – No fading & AWGN [3]

Shannon: The total capacity of parallel independent channels is the sum of their individual capacities.

$$C_k = \log_2\left(1 + SNR_k\right)$$

$$C = \sum C_k = \sum \log_2\left(1 + SNR_k\right)$$

Equal power distribution (channel not known at TX):

Constant dep. on e.g. TX power and noise.

$$C = \sum C_k = \sum_{k=1}^{r} \log_2\left(1 + \alpha\sigma_k^2\right) = \boxed{\log_2 \prod_{k=1}^{r}\left(1 + \alpha\sigma_k^2\right)}$$

This doesn't look like what we usually see about MIMO capacity.

Capacity – No fading & AWGN [4]

MIMO capacity is often seen on the form:

$$C = \log_2 \det\left(\mathbf{I}_{M_R} + \alpha\mathbf{H}\mathbf{H}^H\right)$$

... so, let's see if they are the same!

What we derived!

$$\boxed{C = \log_2 \prod_{k=1}^{r}\left(1 + \alpha\sigma_k^{\,2}\right)} = \log_2 \det\left(\begin{bmatrix} 1+\alpha\sigma_1^{\,2} & & & & \\ & \ddots & & & \\ & & 1+\alpha\sigma_r^{\,2} & & \\ & & & 1 & \\ & & & & \ddots \end{bmatrix}\right) =$$

$$= \log_2 \det\left(\mathbf{I}_{M_R} + \alpha\mathbf{\Sigma}\mathbf{\Sigma}^H\right) = \log_2 \det \mathbf{U}\left(\mathbf{I}_{M_R} + \alpha\mathbf{\Sigma}\mathbf{\Sigma}^H\right)\mathbf{U}^H$$

YES!

$$= \log_2 \det\left(\underbrace{\mathbf{U}\mathbf{U}^H}_{\mathbf{I}_{M_R}} + \alpha\underbrace{\mathbf{U}\mathbf{\Sigma}\mathbf{V}^H}_{\mathbf{H}}\underbrace{\mathbf{V}\mathbf{\Sigma}^H\mathbf{U}^H}_{\mathbf{H}^H}\right) \boxed{= \log_2 \det\left(\mathbf{I}_{M_R} + \alpha\mathbf{H}\mathbf{H}^H\right)}$$

Capacity – No fading & AWGN [5]

Singular values of **H** alone determine the MIMO channel capacity (at a given SNR)!

CONCLUSION:

$$C = \log_2 \prod_{k=1}^{r}\left(1 + \alpha\sigma_k^2\right) = \log_2 \det\left(\mathbf{I}_{M_R} + \alpha\mathbf{H}\mathbf{H}^H\right) \quad \text{[bit/sec/Hz]}$$

Normalization: ρ - SNR at each receiver branch

$$C = \log_2 \det\left(\mathbf{I}_{M_R} + \frac{\rho}{M_T}\mathbf{H}\mathbf{H}^H\right)$$

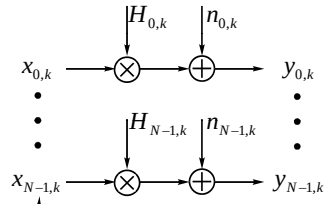
This relation is also derived (in a differnt way) in *e.g*
G.J. Foschini and M.J. Gans. **On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas.** Wireless Personal Communications, no 6, pp. 311-335, 1998.

Channel estimation in OFDM with pilot-symbol assisted modulation

OFDM System model

From Lec. 7

Simplified model under ideal conditions (slow enough fading and sufficient CP)



OFDM systems have *N* between 64 (WLAN) and 8192 (DigTV).

Total filter in the signal path:

$$h_{signal}(t) = h_{TX}(t) * h(t) * h_{RX}(t)$$

$$H_{signal}(f) = H_{TX}(f) * H(f) * H_{RX}(f)$$

Given that subcarrier *n* is transmitted at frequency f_n the attenuations become:

$$H_{n,k} = H_{signal}(f_n)$$

OFDM System model [cont.]

From Lec. 7

We have ended up with an OFDM matrix model:

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n}$$

where **y** is the received vector, **X** a diagonal matrix with the transmitted constellation points on its diagonal, **h** a vector of channel attenuations, and a vector **n** of receiver noise.

For the purpose of channel estimation, assume that all ones are transmitted, i.e., that **X** = **I**. We now have a simplified model:

Same numer of measurements as parameters to estimate!

$$\mathbf{y} = \mathbf{h} + \mathbf{n}$$

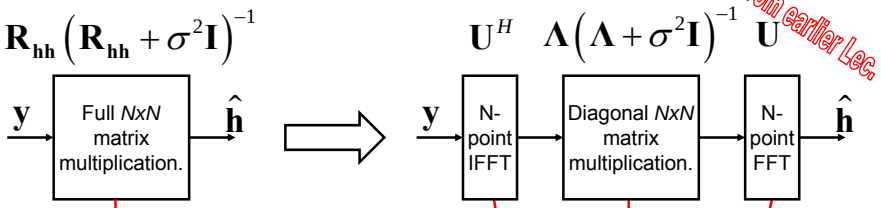
All matrices in calculations are square, of equal size (and symmetric).

Further assume that the channel is zero-mean and has autocorrelation $\mathbf{R}_{\mathbf{hh}}$ while the noise is i.i.d zero mean complex Gaussian with autocorrelation $\mathbf{R}_{\mathbf{nn}} = \sigma^2\mathbf{I}$. We also assume that **h** and **n** are independent.

OFDM Channel estimation

What have we obtained?

From earlier Lec.



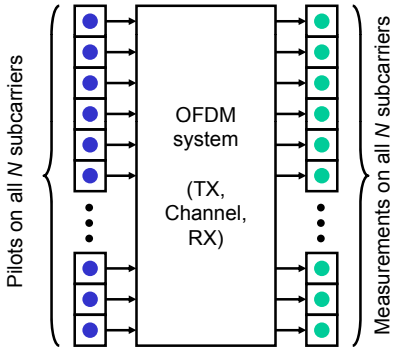
COMPLEXITY (operations)

$$N^2$$

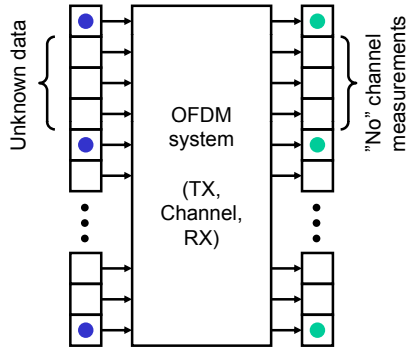
$$N\log_2 N + N + N\log_2 N$$
$$= 2N\log_2 N + N$$

OFDM
Channel estimation (scattered pilots)

From earlier lecture



Pilots scattered in unknown data



OFDM
Channel estimation (scattered pilots)

During Lecture 7 we modelled the case with "all pilots" as a completely known diagonal data matrix **X** = **I**.

We now have

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n} = \begin{bmatrix} 1 & & & & & & \\ & x_2 & & & & & \\ & & x_3 & & & & \\ & & & 1 & & & \\ & & & & x_4 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \mathbf{h} + \mathbf{n}$$

If the unknown data (x_ks) are zero mean and independent, we cannot obtain any information about the channel (**h**) from measurements (**y**) on those subcarriers. We have to rely on the subcarriers where we have known data (pilots).

OFDM
Channel estimation (scattered pilots)

To simplify the situation, assume a new (rediced size) model where we only collect measurements from the *P* pilot subcarriers.

$$\mathbf{y} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \mathbf{h} + \mathbf{n}$$

Measurements on the *P* pilot subcarriers (*P* x 1)

Pilot matrix [data rows removed] (*P* x *N*)

Full channel vector (*N* x 1)

Receiver noise vector (*P* x 1)

We have a rectangular matrix and the nice calculations from Lecture 7 have to be done in a different way!

Smells like an SVD!

OFDM
LMMSE Scattered Pilot Estimation

The linear estimator we are looking for is

$$\hat{\mathbf{h}}_{LMMSE} = \mathbf{A}\mathbf{y}$$

where **A** is an *N* x *P* matrix on the form

$$\mathbf{A} = \mathbf{R}_{hy}\mathbf{R}_{yy}^{-1}$$

using the cross correlation $\mathbf{R}_{hy}$ between the *N* channel coefficients **h** and the *P* measurements **y** and the autororrelation $\mathbf{R}_{yy}$ of the *P* measurements.

Number of operations required to perform one estimation is: *PN*

If we want to reduce the complexity of this estimator, we can use the theory of low-rank estimators. [Details can be found in books about estimation theory.]

OFDM
Channel est. low-rank approximation

By performing an SVD on the following matrix:

$$\mathbf{R}_{hy}\mathbf{R}_{yy}^{-1/2} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$$

... the "best rank-*q* estimator" is given by a rank-*q* approximation of the **A** matrix:

$$\mathbf{A} \approx \mathbf{A}_q = \mathbf{U}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_q \\ & & \end{bmatrix}\mathbf{V}^H\mathbf{R}_{yy}^{-1/2}$$

OFDM
Channel est. low-rank approximation

$$\hat{\mathbf{h}}_{LMMSE-q} = \mathbf{U}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_q \\ & & \end{bmatrix}\mathbf{V}^H\mathbf{R}_{yy}^{-1/2}\mathbf{y}$$

[... keep only the *q* used columns of **U** and **V** ...]

$$= \underbrace{\underset{N \times q}{\mathbf{U}_q}\underset{q \times q}{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_q \end{bmatrix}}\underset{q \times P}{\mathbf{V}_q^H}\underset{P \times P}{\mathbf{R}_{yy}^{-1/2}}}_{\text{Combine to one matrix, } q \times P}\mathbf{y}$$

If the channel has strong correlation the rank *q* can be made small. Compare to the *NP* required without SVD.

Number of operations required to perform one estimation: $qP + qN = q(P+N)$