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 \( Q\Lambda Q^T \). The orthonormal eigenvectors of \( A \) are in the orthogonal matrix \( Q \) and the corresponding eigenvalues in the diagonal matrix \( \Lambda \).

COMPLEX CASE
A Hermitian matrix \( A \) can be factored into \( U\Lambda U^H \). The orthonormal eigenvectors of \( A \) are in the unitary matrix \( U \) and the corresponding eigenvalues in the diagonal matrix \( \Lambda \).

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)

\[
A = Q\Lambda Q^T
\]

where \( Q \) is orthogonal and \( \Lambda \) is diagonal, with something (more general) like

\[
A = U\Sigma V^T
\]

where \( U \) and \( V \) are still orthogonal (but not necessarily the same) and \( \Sigma \) 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

\[
A = U\Sigma V^T
\]

exists. What would that imply?

First look at \( AA^T \):

\[
AA^T = (U\Sigma V^T)\Sigma V^T U^T = U\Sigma \Sigma^T U^T = Q_1\Lambda_1 Q_1^T \quad \text{(spectral theorem!)}
\]

... then at \( A^TA \):

\[
A^TA = V\Sigma^T U^T (U\Sigma V^T) = V\Sigma^T \Sigma V^T = Q_2\Lambda_2 Q_2^T \quad \text{(spectral theorem!)}
\]

This seem to work only if \( AA^T \) and \( A^TA \) have some very specific relations between their eigenvalues.
Eigenvectors of $AA^T$ and $A^TA$

Let’s find out the properties of the eigenvalues of $AA^T$ and $A^TA$.

Do they have the same eigenvalues?

If $x$ is an eigenvector of $AA^T$ and $\lambda$ the corresponding eigenvalue

$$AA^T x = \lambda x$$

then, multiplying by $A^T$ from the left,

$$A^T AA^T x = \lambda A^T x$$

shows us that $A^T x$ is an eigenvector of $A^TA$ and the same $\lambda$ is its eigenvalue. **YES!**

Any other useful properties?

Since $AA^T$ and $A^TA$ are symmetric, the eigenvalues are **real**.

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

---

The singular value decomposition

The above leads us to the following theorem:

**Singular Value Decomposition**

Any $M \times N$ matrix $A$ can be factored into

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

The columns of $U$ ($M \times M$) are eigenvectors of $AA^T$, and the columns of $V$ ($N \times N$) are eigenvectors of $A^TA$. The $r$ singular values on the diagonal of $\Sigma$ ($M \times N$) are the square roots of the nonzero eigenvalues of both $AA^T$ and $A^TA$.

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:

$$A = U \Sigma V^T = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix}$$

$u_1$, to $u_r$, spans $C(A)$

$u_{r+1}$ to $u_m$ spans $N(A^\perp)$

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

$\sigma_1$ to $\sigma_r$ are the square-roots of the non-zero eigenvalues of both $AA^T$ and $A^TA$, there can be at most $\min(m,n)$ of them, i.e., $r \leq \min(m,n)$.

$v_1$ to $v_r$ spans $C(A^\perp)$

$v_{r+1}$ to $v_n$ spans $N(A)$

Implies that $r = \text{rank}(A)$

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

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

when $A = A^T$?

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

$$AA^T = Q \Lambda Q^T Q \Lambda Q^T = QA^2 Q^T$$

**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?
A general (narrow-band) model

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

$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{M_R} \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,M_T} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,M_T} \\ \vdots & \vdots & & \vdots \\ h_{M_R,1} & h_{M_R,2} & \cdots & h_{M_R,M_T} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{M_T} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{M_T} \end{bmatrix} = Hx + n$

Channel matrix ($M_R \times M_T$) Transmitted vector ($M_T \times 1$) Receiver noise vector ($M_R \times 1$)

Received vector ($M_R \times 1$)

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.

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

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

where $U$ ($M_R \times M_R$) and $V$ ($M_T \times M_T$) are unitary matrices and $\Sigma$ ($M_R \times M_T$) is a matrix containing the singular values on its diagonal.

Multiply by $U^H$ from left:

$U^H y = \Sigma V^H x + U^H n \quad \rightarrow \quad \hat{y} = \Sigma \hat{x} + \hat{n}$

Only "rotations" of $y$, $x$ and $n$. All-zero, except diagonal.
What have we obtained?
Parallel independent channels:
\[
\mathbf{y} = \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_r \\ 0 \\ \vdots \end{bmatrix} \tilde{x} + \mathbf{n}
\]
Number of singular values \(r = \text{rank}(\mathbf{H})\).

Shannon's "standard case":
\[
\begin{array}{c}
\tilde{x}_1 \\
\vdots \\
\tilde{x}_r \\
\end{array} + 
\begin{array}{c}
\sigma_1 \\
\vdots \\
\sigma_r \\
\end{array}
\mathbf{n}
\Rightarrow 
\begin{array}{c}
\tilde{y}_1 \\
\vdots \\
\tilde{y}_r \\
\end{array}
\]
(+ channels with \(\sigma_k = 0\))

Shannon: The total capacity of parallel independent channels is the sum of their individual capacities.
\[
C_k = \log_2 \left(1 + \text{SNR}_k\right)
\]
\[
C = \sum C_k = \sum \log_2 \left(1 + \text{SNR}_k\right)
\]

Equal power distribution (channel not known at TX):
\[
C = \sum C_k = \sum \log_2 \left(1 + \alpha \sigma_k^2\right) = \log_2 \prod_{k=1}^{r} \left(1 + \alpha \sigma_k^2\right)
\]

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

MIMO capacity is often seen on the form:
\[
C = \log_2 \det \left( \mathbf{I}_{M_T} + \alpha \mathbf{H} \mathbf{H}^H \right)
\]

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

What we derived!
\[
C = \log_2 \prod_{k=1}^{r} \left(1 + \alpha \sigma_k^2\right) = \log_2 \det \left( \begin{bmatrix} 1 + \alpha \sigma_1^2 \\ & \ddots \\ & & 1 + \alpha \sigma_r^2 \\ & & & 1 \end{bmatrix} \right) = \log_2 \det \left( \mathbf{I}_{M_T} + \alpha \mathbf{H} \mathbf{H}^H \right)
\]

\[
= \log_2 \det \left( \mathbf{I}_{M_T} + \alpha \mathbf{\Sigma} \mathbf{\Sigma}^H \right) = \log_2 \det \left( \mathbf{I}_{M_T} + \alpha \mathbf{\Sigma} \mathbf{\Sigma}^H \mathbf{U} \mathbf{U}^H \right) = \log_2 \det \left( \mathbf{I}_{M_T} + \alpha \mathbf{H} \mathbf{H}^H \right)
\]

Normalization: \(\rho - \text{SNR} \) at each receiver branch
\[
C = \log_2 \det \left( \mathbf{I}_{M_T} + \frac{\rho}{M_T} \mathbf{H} \mathbf{H}^H \right)
\]

This relation is also derived (in a different 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 [cont.]

We have ended up with an OFDM matrix model:

\[ y = Xh + n \]

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:

\[ y = h + n \]

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

Same number of measurements as parameters to estimate!

OFDM Channel estimation

What have we obtained?

\[
\begin{align*}
R_{hh} & \left( R_{hh} + \sigma^2 I \right)^{-1} \\
U^H & \Lambda \left( \Lambda + \sigma^2 I \right)^{-1} U \\
\end{align*}
\]

\[
\begin{align*}
\text{N-point} & \quad \text{N-point} \\
\text{FFT} & \quad \text{IFFT} \\
\end{align*}
\]

\[
\begin{align*}
\text{FullNxN} & \quad \text{Diagonal NxN} \\
\text{matrix} & \quad \text{matrix} \\
\text{multiplication.} & \quad \text{multiplication.} \\
\end{align*}
\]

\[
\begin{align*}
\text{COMPLEXITY} & \quad \text{(operations)} \\
\text{N^2} & \quad \text{N log}_2 N + N + N log_2 N \\
& \quad = 2N log_2 N + N
\end{align*}
\]
OFDM Channel estimation (scattered pilots)

From earlier lecture, we modelled the case with "all pilots" as a completely known diagonal data matrix $X = I$.

We now have

$$ y = Xh + n $$

If the unknown data $(x_k)$ 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 LMMSE Scattered Pilot Estimation

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

The linear estimator we are looking for is

$$ \hat{h}_{LMMSE} = Ay $$

where $A$ is an $N \times P$ matrix on the form

$$ A = R_{yy}^{-1} R_{y} $$

using the cross correlation $R_{yy}$ between the $N$ channel coefficients $h$ and the $P$ measurements $y$ and the autorrelation $R_{yy}$ of the $P$ measurements.

Number of operations required to perform one estimation is: $PN$
By performing an SVD on the following matrix:

\[ \frac{1}{2} H = U \Sigma V^H \]

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

\[ A \approx A_q = U_q \begin{bmatrix} \sigma_1 & \cdots & \sigma_q \end{bmatrix} V_q^H R_{yy}^{-1/2} \]

If the channel has strong correlation the \(k\) can be made small. Compare to the \(NP\) required without SVD.

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

\[ \hat{h}_{LMMSE-q} = U_q \begin{bmatrix} \sigma_1 & \cdots & \sigma_q \end{bmatrix} V_q^H R_{yy}^{-1/2} y \]