EE 263 Final Review

The key ideas from most of the course are listed here as prompts to cast your mind back to things. Interrupt to ask about lingering uncertainties along the way.
Subspaces, basis, nullspace, range, rank

- **basis**: set of vectors that are independent and span the subspace
- **dimension**: number of vectors in a basis for the subspace
- **nullspace**: set of input vectors that \( y = Ax \) maps to zero
- **range**: set of output vectors that \( y = Ax \) can land on
- **one-to-one**: \( A \) has the trivial nullspace \( \{0\} \)
- **onto**: the range of \( A \) is (all of) \( \mathbb{R}^m \)

Remember: \( \text{null}(A) \) and \( \text{range}(A) \) are spaces, not matrices!

- **rank**: \( \text{rank}(A) := \dim \text{range}(A) \)
- **full rank**: when \( \text{rank}(A) = \min(m,n) \)
- **fact**: \( \dim \text{range}(A) + \dim \text{null}(A) = n \)
Fill in this table!

<table>
<thead>
<tr>
<th>options</th>
<th>one-to-one</th>
<th>onto</th>
</tr>
</thead>
<tbody>
<tr>
<td>relates to range$(A)$ or null$(A)$?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>skinny matrix can be? (yes/no)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fat matrix can be? (yes/no)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>uniqueness or existence of solution?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>left or right inverse?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$AA^\top$ or $A^\top A$ is invertible?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>use least squares or least norm?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Orthogonality

- \( \frac{x^T y}{\| x \| \| y \|} \) relates to the angle between \( x \) and \( y \)
- **Cauchy-Schwarz**: \( | x^T y | \leq \| x \| \| y \| \)
- **orthonormal set**: all vectors both orthogonal and normalized (unit length)
- **orthogonal matrix**: square and \( U^T U = I \)
- **orthogonal complement**: \( S^\perp \), all vectors orthogonal to every vector in \( S \)
- **fact**: \( \text{range}(A)^\perp = \text{null}(A^T) \)

Questions

1. Orthonormal bases make your life better. Why?
2. Where have orthogonal matrices shown up since?
**QR factorization**

If $A$ is skinny and full rank:

- $A = QR$, where $Q^TQ = I$, $R$ upper triangular & invertible
- columns of $Q$ are orthonormal basis for $\text{range}(A)$

- For general $A$, skip dependent columns, then Gram-Schmidt produces “staircase pattern” in $R$.

**Full QR decomposition:**

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1$$

- $Q_2$ “completes” $Q_1$ to an orthonormal basis for $\mathbb{R}^n$
- $\text{range}(Q_2) = \text{null}(A^T)$
Projection and least squares

\[ x_{ls} = (A^T A)^{-1} A^T y = A^\dagger y \]

- requires \( A \) to be \textit{skinny} and \textit{full rank}
- \( Ax_{ls} \) is point in \texttt{range}(\( A \)) closest to \( y \)
- \( Ax_{ls} = \( A \)(\( A^T A \)^{-1} \( A^T \) \( y \)) is \textit{projection} of \( y \) onto \texttt{range}(\( A \))
- optimal residual \( r = y - Ax_{ls} \) is orthogonal to \texttt{range}(\( A \))
Variations of least squares

via QR factorization

\[ A^\dagger = R^{-1}Q^T \]

Least-squares fitting

- With basis functions \( f_i \) and data \((s_i, g_i)\), finding coefficients \( x_i \) so that
\[
x_1 f_1(s_i) + \cdots + x_n f_n(s_i) \approx g_i, \quad i = 1, \ldots, m
\]
can be cast as least-squares, if we want to minimize square error

Recursive least-squares

- If rows of \( A, y \) roll in sequentially, we can update \( x_{1s} \) with each new “sample”

Multi-objective least-squares

- If you have two objectives, could cast as the least-squares problem
\[
\text{minimize} \quad \|Ax - y\|^2 + \mu \|Fx - g\|^2
\]
Least-norm solution

\[ x_{ln} = A^T (A A^T)^{-1} y \]

- requires \( A \) to be \textit{fat} and \textit{full rank}
- \( x_{ln} \) is the point in the solution space closest to the origin
- \( x_{ln} + z \) for any \( z \in \text{null}(A) \) gives another solution to \( y = Ax \) with larger norm
- \( x_{ln} \) is orthogonal to \( \text{null}(A) \)

\[
\begin{align*}
\text{null}(A) &= \{ x \mid Ax = 0 \} \\
\{ x \mid Ax = y \} &= \{ x \mid A x = 0 \}
\end{align*}
\]
Regularization and general norm minimization

\[
\text{minimize } \|Ax - y\|^2 + \mu \|x\|^2
\]

is a multi-objective least-squares problem that approaches least-norm as \(\mu \to 0\).

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\| \\
\text{subject to} & \quad Cx = d
\end{align*}
\]

is the most general form of norm minimization with affine equality constraint, and includes least-squares and least-norm as special cases.
**Gauss-Newton method**

Non-linear least squares:

- Find $x \in \mathbb{R}^n$ to minimize $\|r(x)\|^2$, where $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

**Gauss-Newton method:**

given starting guess for $x$
repeat
  linearize $r$ near current guess
  new guess is linear LS solution, using linearized $r$
until convergence

- “linearize” means take the first Taylor polynomial
  (more details in lecture slides)
Eigenvectors and diagonalization

- **eigenvalue** $\lambda \in \mathbb{C}$, **eigenvector** $v \in \mathbb{C}^n$: satisfies $Av = \lambda v$
  (we call $\lambda$ and $v$ corresponding to each other)

- **conjugate symmetry**: if $A$ real and $Av = \lambda v$, then $A\bar{v} = \bar{\lambda}\bar{v}$

- **diagonalizable**: there exists invertible $T$ such that $T^{-1}AT = \Lambda$ is diagonal
  (we say that $T$ diagonalizes $A$)

facts:

- $A$ is diagonalizable if and only if it has $n$ linearly independent eigenvectors, and those eigenvectors and eigenvalues are in $T$ and $\Lambda$

- if $A$ has distinct eigenvalues, then it has linearly independent eigenvectors

questions:

- Interpretation of eigenvectors in mapping $y = Ax$?

- Interpretation of eigenvectors in LDS $\dot{x} = Ax$, $x(t+1) = Ax$?
Symmetric matrices and quadratic forms

- Symmetric matrices have real eigenvalues and orthonormal eigenvectors
- Symmetric $A$ can be written as $A = Q\Lambda Q^T$ for orthogonal $Q$

- **quadratic form**: $f(x) = x^TAx$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- Always possible to choose symmetric $A$ without affecting $f$ (why?)

- **positive semidefinite**, or $A \geq 0$: when $x^TAx \geq 0$ for all $x$
- $A \geq 0$ if and only if all its eigenvalues are nonnegative

- We say that $A \geq B$ if $A - B \geq 0$ (and analogously for other relationships)
- Matrix inequality is only a **partial order** (why?)
Ellipsoids

An ellipsoid in $\mathbb{R}^n$ is a set

$$\mathcal{E} = \{ \, x \mid x^T A x \leq 1 \, \}$$

with $A > 0$.

eigenvectors/eigenvalues determine directions/lengths of semiaxes $s_i = \lambda_i^{-1/2} q_i$
Gain and matrix norm

- **gain**: \( \frac{\|Ax\|}{\|x\|} \), varies with direction of input \( x \)

- **norm** of a matrix: maximum gain

\[
\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sqrt{\lambda_{\text{max}}(A^T A)}
\]

- eigenvectors of \( A^T A \) are like ‘input directions yielding ‘gains’ (eigenvalues)

- **Frobenius norm**: \( \|A\|_F = \left( \sum_i \sum_j |a_{ij}|^2 \right)^{\frac{1}{2}} \)

questions:

1. Which input direction yields the smallest gain?
2. How is this idea different from the eigenvalues of \( A \)?
Singular value decomposition

For $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = r$, the thin SVD is

$$A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

$\Rightarrow U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r}$ have orthonormal columns,

$\Rightarrow \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$, where $\sigma_1 \geq \cdots \geq \sigma_r > 0$

full SVD: extend $U, V$ to be orthogonal (square) and pad $\Sigma$ with zeros

Question: What (practical) limitation of rank does the SVD help resolve?
Singular value decomposition

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

- \( u_1, \ldots, u_r \) (columns of \( U \)) are a basis for \( \text{range}(A) \) (span of columns of \( A \))
- \( v_1, \ldots, v_r \) (rows of \( V^T \)) are a basis for \( \text{range}(A^T) \) (span of rows of \( A \))
- \( v_i \) represent ‘input directions’, \( \sigma_i \) ‘gains’, \( u_i \) ‘output directions’, \( Av_i = \sigma_i u_i \)

Moore-Penrose inverse or pseudo-inverse of \( A \) (finally!): \( A^\dagger = V \Sigma^{-1} U^T \).
Applications of SVD

Relative error analysis

- **condition number**: \( \kappa(A) = \|A\|\|A^{-1}\| = \sigma_{\text{max}}(A)/\sigma_{\text{min}}(A) \)

- measure of “how singular”, or how input *relative error* \( \|\delta x\|/\|x\| \) affects \( \|\delta y\|/\|y\| \)

- poorly conditioned (large \( \kappa \)) is like “pretty much singular”

Low-rank approximation

- Take the SVD and remove the smallest singular values (call this \( \hat{A} \))

- This minimizes \( \|A - \hat{A}\| \), which is then \( \lambda_{p+1} \)

- interpretation: \( \sigma_i = \min \{ \|A - B\| \mid \text{rank}(B) \leq i - 1 \} \)
  (distance to nearest rank \( i - 1 \) matrix)
Autonomous linear dynamical systems

- $\dot{x} = Ax$ (continuous), $x(t + 1) = Ax$ (discrete)

- Special structures of $A$ (e.g. triangular) can simplify things

- Lots of (real-life) systems are approximately linear

- Lots aren't. Say $\dot{x} = f(x)$. Sometimes *linearizing near an equilibrium point* $x_e$ (satisfying $f(x_e) = 0$) helps: define $\delta x(t) = x(t) - x_e$, then take

  $$\delta \dot{x}(t) \approx Df(x_e)\delta x(t)$$
Matrix exponential

\[ e^A = I + \sum_{n=1}^{\infty} \frac{A^n}{n!} = \lim_{k \to \infty} \left( I + \frac{A}{k} \right)^k \]

- \( x(t) = e^{tA}x(0) \) is a solution of \( \dot{x} = Ax \)

- In a continuous LDS, \( e^{tA} \) steps \( t \) seconds forward in time

question:

1. what is \( e^A \) if \( A \) is diagonal?
Eigenvectors in dynamical systems

- if \(x(0) = v\) is an eigenvector, \(x(t) = e^{\lambda t}v\) is a **mode** of \(\dot{x} = Ax\)

- start on an eigenvector, stay on the eigenvector (i.e., that line is *invariant*)

- if \(\lambda, v\) are complex, then the plane \(\text{span}\{\Re(v), \Im(v)\}\) is invariant (trajectory rotates and grows/decays in that plane)

- \(\Re \lambda_j\) gives (exp.) growth/decay; \(\Im \lambda_j\) gives (sin.) oscillation frequency

**stable**: \(x(t) \to 0\) for any \(x(0)\)

- in continuous LDS, **stable**: \(e^{tA} \to 0\), or \(\Re \lambda_i < 0\) for all \(i = 1, \ldots, n\)

Modal form

- if \(A\) is diagonalizable, can recast (change coordinates) LDS into modes:
  \[x = T\tilde{x},\ \text{then}\ \dot{\tilde{x}} = T^{-1}AT\tilde{x}, \ i.e. \ \dot{\tilde{x}} = \Lambda\tilde{x}.

- if some eigenvalues are complex, can use real modal form (with \(2 \times 2\) blocks)
Input and outputs

Continuous time

\[ \dot{x} = Ax + Bu, \quad y = Cx + Du \]

solved by

\[ x(t) = e^{tA}x(0) + \int_0^t e^{(t-\tau)A}Bu(\tau) \, d\tau \]

Discrete time

\[ x(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t) \]

solved by

\[ x(t) = A^tx(0) + \sum_{\tau=0}^{t-1} A^{(t-1-\tau)}Bu(\tau) \]

*state* \( x(t) \) summarizes effect of past inputs on future output

Ideas: impulse response, step response, DC gain, causality
Controllability and state transfer

- **reachable** in $t$: you can get to $x(t)$ from $x(0)$ (in $t$ seconds/epochs)

- $\mathcal{R}_t$ is all points reachable in time $t$; $\mathcal{R}$ all points reachable ever

- in discrete time: $\mathcal{R}_t = \text{range}(C_t)$ where $C_t = [B \ AB \ \cdots \ \ A^{t-1}B]$

- in discrete time: all reachable states given by $C = C_n$

- **reachable system**: when $\mathcal{R} = \mathbb{R}^n$

- **minimum energy input**: minimize $\sum_{\tau=0}^{t-1} \|u(\tau)\|^2$ to reach $x(t) = x_{\text{des}}$, can be cast as a least norm problem

- in continuous time: $\mathcal{R} = \text{range}(C)$, where $C = [B \ AB \ \cdots \ \ A^{n-1}B]$