Discussion Session 6

October 23, 2020

- Least-norm solutions
- Gauss-Newton method
- Eigenvectors and diagonalization
- Symmetric matrices, quadratic forms and ellipsoids
- Singular value decomposition
Least-norm solutions

\[ \{ x \mid Ax = y \} \]

\[ \text{null}(A) = \{ x \mid Ax = 0 \} \]

To solve the optimization problem

\[
\begin{align*}
\text{minimize} & \quad \| x \| \\
\text{subject to} & \quad Ax = y
\end{align*}
\]

take \( x_{ln} = A^T(AA^T)^{-1}y \).

\( x_{ln} \) 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 \( \text{null}(A) \)
Gauss-Newton method

Non-linear least squares:

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

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

If \( A \in \mathbb{C}^{n \times n} \), we call a scalar \( \lambda \in \mathbb{C} \) an \textit{eigenvalue of }\( A \) if there is a vector \( v \neq 0 \) satisfying \( Av = \lambda v \), and we call every such \( v \) an \textit{eigenvector corresponding to }\( \lambda \).

If there exists an invertible matrix \( T \) such that \( T^{-1}AT \) is diagonal, we say that \( A \) is \textit{diagonalizable} and that \( T \) diagonalizes \( A \).

\( A \in \mathbb{C}^{n \times n} \) is diagonalizable if and only if it has \( n \) linearly independent eigenvectors \( v_1, \ldots, v_n \). Then if \( T = [v_1 \ \cdots \ v_n] \) and \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \), \( T^{-1}AT = \Lambda \).

If \( A \) has distinct eigenvalues, then it has linearly independent eigenvectors.

\textit{Take a minute to think about what image is in your head when you think about eigenvectors.}
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 $Q$
- *How does that relate to diagonalization?*

- Quadratic form: $f(x) = x^T Ax$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- Always possible to choose symmetric $A$ without affecting $f$ (*why?*)
Positive definiteness and matrix order

- We say $A$ is **positive semidefinite**, or $A \succeq 0$, if $x^T Ax \geq 0$ for all $x$
- $A \succeq 0$ if and only if all its eigenvalues are nonnegative
- Analogously for positive definite, and negative (semi)definite
- Not all matrices fall into one of these categories (*why?*)

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

The norm of a matrix $A$ is defined as

$$\| A \| = \max_{x \neq 0} \frac{\| Ax \|}{\| x \|}$$

and is equal to $\| A \| = \sqrt{\lambda_{\text{max}}(A^T A)}$.

We can think of the eigenvectors of $A^T A$ as the ‘input directions’ yielding ‘gains’ associated with corresponding eigenvalues for the system $y = Ax$.

- What do we mean by ‘gain’?
- Which input direction yields the smallest gain?
Ellipsoids

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

\[
\mathcal{E} = \{ \, x \mid x^T Ax \leq 1 \, \}
\]

with \( A > 0 \).

- eigenvectors determine directions of semiaxes
- eigenvalues determine lengths of semiaxes

Take a minute to think through some special cases: the unit ball, a ball of radius \( r \), an extremely “flat” ellipsoid, and extremely “skinny” ellipsoid...
Singular value decomposition

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

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

- $U \in \mathbb{R}^{m \times r}$ has orthonormal columns,
- $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$, where $\sigma_1 \geq \cdots \geq \sigma_r > 0$
- $V \in \mathbb{R}^{n \times r}$ has orthonormal columns

Can extend to full SVD by extending $U, V$ to be orthogonal (square), and padding $\Sigma$ with zeros
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!): \( A^\dagger = V \Sigma^{-1} U^T \).
Question

Given $A \in \mathbb{R}^{m \times n}$, how can we find vectors $x$ and $y$ that maximize $y^T A x$, subject to $\|y\| = 1$, $\|x\| = 1$?

Things to think about

- What is the effect of the constraints? (What if you didn’t have them?)
- Describe what happens if you’re slightly off the optimal $x$ or $y$. 