Vector Norms

- Properties of vector norms

- A norm is uniquely defined by its unit sphere:

- $p$-norms
Inner-Product Spaces

- **Properties of inner-product spaces**: Inner products $\langle x, y \rangle$ must satisfy
  
  \[
  \begin{align*}
  \langle x, x \rangle & \geq 0 \\
  \langle x, x \rangle = 0 & \iff x = 0 \\
  \langle x, y \rangle = \langle y, x \rangle \\
  \langle x, y + z \rangle & = \langle x, y \rangle + \langle x, z \rangle \\
  \langle \alpha x, y \rangle & = \alpha \langle x, y \rangle
  \end{align*}
  \]

- **Inner-product-based vector norms**
Matrix Norms

- Properties of matrix norms:
  
  \[ \|A\| \geq 0 \]
  
  \[ \|A\| = 0 \iff A = 0 \]
  
  \[ \|\alpha A\| = |\alpha| \cdot \|A\| \]
  
  \[ \|A + B\| \leq \|A\| + \|B\| \quad (\text{triangle inequality}) \]

- Frobenius norm:

- Operator/induced/subordinate matrix norms:
Induced Matrix Norms

- Interpreting induced matrix norms:

- General induced matrix norms:
Matrix Condition Number

**Definition:** \( \kappa(A) = \|A\| \cdot \|A^{-1}\| \) is the ratio between the shortest/longest distances from the unit-ball center to any point on the surface.

**Intuitive derivation:**

\[
\kappa(A) = \max_{\text{inputs}} \max_{\text{perturbations in input}} \frac{\text{relative perturbation in output}}{\text{relative perturbation in input}}
\]

since a matrix is a linear operator, we can decouple its action on the input \( x \) and the perturbation \( \delta x \) since \( A(x + \delta x) = Ax + A\delta x \), so

\[
\kappa(A) = \frac{|A|}{\max_{\text{perturbations in input}} \max_{\text{relative input reduction}}} \cdot \frac{1}{\|A^{-1}\|}
\]
Matrix Conditioning

- The matrix condition number $\kappa(A)$ is the ratio between the max and min distance from the surface to the center of the unit ball transformed by $\kappa(A)$:

- The matrix condition number bounds the worst-case amplification of error in a matrix-vector product:
Norms and Conditioning of Orthogonal Matrices

- Orthogonal matrices:

- Norm and condition number of orthogonal matrices:
Singular Value Decomposition

- The singular value decomposition (SVD):
Norms and Conditioning via SVD

- Norm and condition number in terms of singular values:
Visualization of Matrix Conditioning

\[ \{x : x \in \mathbb{R}^2, \|x\|_2 = 1\} \xrightarrow{A} \{Ax : x \in \mathbb{R}^2, \|x\|_2 = 1\} \]

\[ A = U \begin{bmatrix} \sigma_{\text{max}} & 0 \\ 0 & \sigma_{\text{min}} \end{bmatrix} V^T \]

\[ \kappa(A) = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} \]

\[ \|A\|_2 = \sigma_{\text{max}} \]

\[ 1/\|A^{-1}\|_2 = \sigma_{\text{min}} \]
Conditioning of Linear Systems

- Lets now return to formally deriving the conditioning of solving $Ax = b$: 
Consider perturbations to the input coefficients $\hat{A} = A + \delta A$.
Solving Basic Linear Systems

- Solve $Dx = b$ if $D$ is diagonal

- Solve $Qx = b$ if $Q$ is orthogonal

- Given SVD $A = U\Sigma V^T$, solve $Ax = b$
Solving Triangular Systems

- \( Lx = b \) if \( L \) is lower-triangular is solved by forward substitution:

\[
\begin{align*}
    l_{11}x_1 &= b_1 & x_1 &= \\
    l_{21}x_1 + l_{22}x_2 &= b_2 & \Rightarrow x_2 &= \\
    l_{31}x_1 + l_{32}x_2 + l_{33}x_3 &= b_3 & x_3 &= \\
    & \vdots & & \vdots
\end{align*}
\]

- Algorithm can also be formulated recursively by blocks:
Solving Triangular Systems

- **Existence of solution to** $Lx = b$: 

- **Uniqueness of solution:**

- **Computational complexity of forward/backward substitution:**
Properties of Triangular Matrices

- $Z = XY$ is lower triangular is $X$ and $Y$ are both lower triangular:

- $L^{-1}$ is lower triangular if it exists:
LU Factorization

- An **LU factorization** consists of a unit-diagonal lower-triangular factor $L$ and upper-triangular factor $U$ such that $A = LU$:

- Given an LU factorization of $A$, we can solve the linear system $Ax = b$: 
Gaussian Elimination Algorithm

- Algorithm for factorization is derived from equations given by $A = LU$:

- The computational complexity of LU is $O(n^3)$:
Existence of LU Factorization

- The LU factorization may not exist: Consider matrix

\[
\begin{bmatrix}
3 & 2 \\
6 & 4 \\
0 & 3
\end{bmatrix}
\]

- Permutation of rows enables us to transform the matrix so the LU factorization does exist:

\[
P A = L U
\]

always exist if \( A \) is full rank.
Gaussian Elimination with Partial Pivoting

- **Partial pivoting** permutes rows to make divisor $u_{ii}$ is maximal at each step:

- A row permutation corresponds to an application of a row permutation matrix $P_{jk} = I - (e_j - e_k)(e_j - e_k)^T$:
Partial Pivoting Example

- Let's consider again the matrix $A = \begin{bmatrix} 3 & 2 \\ 6 & 4 \\ 0 & 3 \end{bmatrix}$.

\[
\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ 0 & 3 \end{bmatrix}
\]

\[
\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}
\]

\[
\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ 0 & 3 \end{bmatrix}
\]

\[
\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 3 & 2 \\ 0 & 3 \end{bmatrix}
\]
Complete Pivoting

\[ \eta_2 A P_2 x = b \]
\[ \eta_2 A = L U \quad \text{or} \quad \eta_2 A \hat{P}_2 = L U \]

- **Complete pivoting** permutes rows and columns to make divisor \( u_{ii} \) maximal at each step:

\[
P_2 A P_2^T = L U \quad |e_{ij}| \leq 1
\]

\[
P_2 A \hat{P}_2 = \begin{bmatrix} d_{11} & d_{12} \\ \hat{L}_{21} & \hat{L}_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}
\]

- Complete pivoting is noticeably more expensive than partial pivoting:

\[
p - p \quad O(n^3) \quad \text{comparisons for each column}
\]
\[
c - p \quad \text{complete pivoting, need } O(n^3) \quad \text{comparisons per column}
\]
Round-off Error in LU

- Let's consider factorization of \( \begin{bmatrix} \epsilon & 1 \\ 1 & 1 \end{bmatrix} \) where \( \epsilon < \epsilon_{\text{mach}} \):

\[
L = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix}, \quad U = \begin{bmatrix} \epsilon & 1 \\ 1 & 1 - \epsilon \end{bmatrix}, \quad f\left(\frac{x+y}{y} \right) = \frac{x}{y} \text{ if } y < \epsilon_{\text{mach}}x
\]

- Permuting the rows of \( A \) in partial pivoting gives \( PA = \begin{bmatrix} 1 & 1 \\ \epsilon & 1 \end{bmatrix} \):

\[
\begin{bmatrix} L \\ U \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \epsilon & 1 - \epsilon \end{bmatrix}, \quad f\left(\frac{x+y}{y} \right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ since } f\left(1+\epsilon\right) = 1
\]

\[
L \cdot f\left(U\right) = \begin{bmatrix} 1 & 1 \\ \epsilon & 1 \end{bmatrix} \quad A - L \cdot f\left(U\right) = \begin{bmatrix} 0 & 0 \\ 0 & \epsilon \end{bmatrix}
\]
Error Analysis of LU

- The main source of round-off error in LU is in the computation of the Schur complement:

\[
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
= \begin{bmatrix}
L_{11} & L_{12} \\
L_{21} & L_{22}
\end{bmatrix}
\begin{bmatrix}
u_{11} & u_{12} \\
u_{21} & u_{22}
\end{bmatrix}
\]

\[
S_{22} = (A_{22} - L_{21}U_{12})
\]

- When computed in floating point, absolute backward error $\delta A$ in LU (so $\hat{L}\hat{U} = A + \delta A$) is

\[
|\delta a_{ij}| \leq \varepsilon_{\text{mach}} (|\hat{L}| \cdot |\hat{U}|)_{ij}
\]

\[
L \preceq \begin{bmatrix}
l_1 & 0 \\
\vdots & \ddots \\
l_n & 0
\end{bmatrix}
\]

\[
U \succeq \begin{bmatrix}
u_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
u_n & \cdots & 0
\end{bmatrix}
\]

\[
\begin{align*}
a_{ij} &\preceq \langle e_i, u_j \rangle \\
\|a_{ij} - \langle e_i, u_j \rangle\|_2 &\preceq \varepsilon \|e_i\|_2 \|u_j\|_2
\end{align*}
\]
Helpful Matrix Properties

- **Matrix is diagonally dominant**, so \( \sum_{i \neq j} |a_{ij}| \leq |a_{ii}| \):

- **Matrix is symmetric positive definite (SPD)**, so \( \forall x \neq 0, x^T A x > 0 \):
  
  \[ A = L L^T \]
  
  Cholesky

- **Matrix is symmetric but indefinite**:
  
  \[ A = A^T \]

  \[ \text{PAP}^T = L D L^T \]

  \[ \text{unit-diagonal} \]

- **Matrix is banded**, \( a_{ij} = 0 \) if \( |i - j| > b \):

  \[ \begin{bmatrix}
  a_{11} & a_{12} & 0 & \cdots \\
  a_{21} & a_{22} & a_{23} & \cdots \\
  0 & a_{32} & a_{33} & \cdots \\
  \cdots & \cdots & \cdots & \cdots
  \end{bmatrix} =
  \begin{bmatrix}
  1 & 1 & 0 & \cdots \\
  1 & 1 & 1 & \cdots \\
  0 & 1 & 1 & \cdots \\
  \cdots & \cdots & \cdots & \cdots
  \end{bmatrix}
  \begin{bmatrix}
  c_1 \\
  c_2 \\
  0 \\
  \cdots
  \end{bmatrix}
  \]
  
  \( O(n^2) \)
Solving Many Linear Systems

- Suppose we have computed $A = LU$ and want to solve $AX = B$ where $B$ is $n \times k$ with $k < n$:

- Suppose we have computed $A = LU$ and now want to solve a perturbed system $(A - uv^T)x = b$:
  Can use the **Sherman-Morrison-Woodbury** formula

\[
(A - uv^T)^{-1} = (A^{-1})^{-1} + \frac{A^{-1}uv^TA^{-1}}{1 - v^TA^{-1}u}
\]

**Demo:** Sherman-Morrison-Woodbury Formula
\[ A^T A x = A^T b \]

Cholesky \((A^T A)\)

\[ x^T A^T A x = y^T y \geq 0 \]