Special Matrices
(Strict) Diagonal Dominance

• The magnitude of each diagonal element is (either):
  • strictly larger than the sum of the magnitudes of all the other elements in its **row**
  • strictly larger than the sum of the magnitudes of all the other elements in its **column**

• One may row/column scale and permute rows/columns to achieve diagonally dominance (since it is just a rewriting of the equations)
  • Recall: choosing the form of the equations wisely is important

• E.g. consider $\begin{pmatrix} 3 & -2 \\ 5 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 9 \\ 4 \end{pmatrix}$

• Switch rows $\begin{pmatrix} 5 & 1 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 9 \end{pmatrix}$ and column scale $\begin{pmatrix} 5 & -2 \\ 4 & -0.5c_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 9 \end{pmatrix}$
(Strict) Diagonal Dominance

• Strictly diagonally dominant (square) matrices are guaranteed to be non-singular.
• Since $\det(A) = \det(A^T)$, either row or column diagonal dominance is enough.

• Column diagonal dominance guarantees pivoting is not required during $LU$ factorization.
• However, pivoting still improves robustness.
• E.g. consider $\begin{pmatrix} 4 & 3 \\ -2 & 50 \end{pmatrix}$ where 50 is more desirable than 4 for $a_{11}$. 
Inner Product

• Consider the space of all vectors with length $m$

• The dot/inner product of two vectors is $u \cdot v = \Sigma_i u_i v_i$

• The magnitude of a vector is $\|v\|_2 = \sqrt{v \cdot v}$ ($\geq 0$)

• Alternative notations: $<u, v> = u \cdot v = u^T v$

• Weighted inner product defined via an $n \times n$ matrix $A$

• $<u, v>_A = u \cdot Av = u^T Av$

• Since $<v, u>_A = v^T Au = u^T A^T v$, weighted inner products commute when $A$ is symmetric

• The standard dot product uses identity matrix weighting: $<u, v> = <u, v>_I$
Definiteness

• Assume $A$ is symmetric so that $< u, v >_A = < v, u >_A$

• $A$ is **positive definite** if and only if $< v, v >_A = v^T A v > 0$ for $\forall v \neq 0$
• $A$ is **positive semi-definite** if and only if $< v, v >_A = v^T A v \geq 0$ for $\forall v \neq 0$
• We abbreviate with SPD and SP(S)D

• $A$ is **negative definite** if and only if $< v, v >_A = v^T A v < 0$ for $\forall v \neq 0$
• $A$ is **negative semi-definite** if and only if $< v, v >_A = v^T A v \leq 0$ for $\forall v \neq 0$
• If $A$ is negative (semi) definite, then $-A$ is positive (semi) definite (and vice versa)
• Thus, can convert such problems to SPD/SP(S)D instead

• $A$ is considered **indefinite** when it is neither positive/negative semi-definite
Eigenvalues

• SPD matrices have all eigenvalues $> 0$
• SP(S)D matrices have all eigenvalues $\geq 0$

• Symmetric negative definite matrices have all eigenvalues $< 0$
• Symmetric negative semi-definite matrices have all eigenvalues $\leq 0$

• Indefinite matrices have both positive and negative eigenvalues
SVD Construction (Important Detail)

• Let \( A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \) so that \( A^T A = A A^T = I \), and thus \( U = V = \Sigma = I \)

• But \( A \neq U \Sigma V^T = I \) What’s wrong?

• Given a column vector \( v_k \) of \( V \), \( Av_k = U \Sigma V^T v_k = U \Sigma \hat{e}_k = U \sigma_k \hat{e}_k = \sigma_k u_k \) where \( u_k \) is the corresponding column of \( U \)

• \( Av_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = u_1 \) but \( Av_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 1 \end{pmatrix} = u_2 \)

• Although eigenvectors may be scaled by an arbitrary constant, the constraint that \( U \) and \( V \) be orthonormal forces their columns to be unit length

• However, there are still two choices for the direction of each column

• Multiplying \( u_2 \) by \(-1\) makes \( U = A \) and thus \( A = U \Sigma V^T \) as desired
SVD Construction (Important Detail)

• An orthogonal matrix has determinant equal to $\pm 1$, where $-1$ indicates a reflection of the coordinate system

• If $\det V = -1$, flip the direction of any column to make $\det V = 1$ (so $V$ does not contain a reflection)

• Then, for each $v_k$, compare $Av_k$ to $\sigma_k u_k$ and flip the direction of $u_k$ when necessary in order to make $Av_k = \sigma_k u_k$

• $\det U = \pm 1$ and may contain a reflection

• When $\det U = -1$, one can flip the sign of the smallest singular value in $\Sigma$ to be negative, whilst also flipping the direction of the corresponding column in $U$ so that $\det U = 1$

• This embeds the reflection into $\Sigma$ and is called the polar-SVD (Irving et al. 2004)
Symmetric Matrices (SVD)

• Since $A^T A = A A^T = A^2$, both the columns of $U$ and the columns of $V$ are eigenvectors of $A^2$

• They are identical (but potentially opposite) directions: $u_k = \pm v_k$

• Thus, $A v_k = \sigma_k u_k$ implies $A v_k = \pm \sigma_k v_k$

• That is, the $v_k$ (and $u_k$) are eigenvectors of $A$ with eigenvalues $\pm \sigma_k$

• Similar to the polar SVD, pull negative signs out of the columns of $U$ into the $\sigma_k$ to obtain $U = V$ and $A = V \Lambda V^T$ as a modified SVD

• $A = V \Lambda V^T$ implies $AV = V \Lambda$ which is the matrix form of the eigensystem of $A$

• Thus, $\Lambda$ contains the positive and negative eigenvalues of $A$
SPD Matrices

- When $A$ is SP(S)D, $\Lambda = \Sigma$ and the **standard SVD** is $A = V\Sigma V^T$ (i.e. $U = V$)

- The singular values are the (all positive) eigenvalues of $A$ (since $AV = V\Sigma$)

- Constructing $V$ with $\det V = 1$, all $\sigma_k > 0$ implies that there are no reflections

- Since all $\sigma_k > 0$, the matrix has full rank and is invertible

- SP(S)D (and not SPD) has at least one $\sigma_k = 0$ and a null space

- Often, one can use modified SPD techniques for SP(S)D matrices

- Unfortunately, indefinite matrices are significantly more challenging
Making/Breaking Symmetry

• Row/column scaling breaks symmetry:
  • Row scaling \( \begin{pmatrix} 5 \\ 3 \\ -4 \end{pmatrix} \) by \(-2\) gives a non-symmetric \( \begin{pmatrix} 5 \\ -6 \\ 3 \\ 8 \end{pmatrix} \)
  • Additional column scaling by \(-2\) gives \( \begin{pmatrix} 5 \\ -6 \\ -6 \\ -16 \end{pmatrix} \)

• Scaling the same row/column together in the same way preserves symmetry

• Important: a nonsymmetric matrix might be inherently symmetric when properly rescaled/rearranged
Rules Galore

• There are many rules/theorems regarding special matrices (especially for SPD)
• Important to be aware of reference material (and to look things up)

• Examples:
  • SPD matrices don’t require pivoting during $LU$ factorization
  • A symmetric (strictly) diagonally dominant matrix with positive diagonal entries is positive definite
  • Jacobi and Gauss-Seidel iteration converge when a matrix is strictly (or irreducibly) diagonally dominant
  • Etc.
Cholesky Factorization

- SPD matrices have $LU$ factorization of $LL^T$ and don’t require elimination to find it.

Consider

\[
\begin{pmatrix}
  a_{11} & a_{21} \\
  a_{21} & a_{22}
\end{pmatrix}
= \begin{pmatrix}
  l_{11} & 0 \\
  l_{21} & l_{22}
\end{pmatrix}
\begin{pmatrix}
  l_{11} & l_{21} \\
  0 & l_{22}
\end{pmatrix}
= \begin{pmatrix}
  l_{11}^2 & l_{11}l_{21} \\
  l_{11}l_{21} & l_{21}^2 + l_{22}^2
\end{pmatrix}
\]

Then $l_{11} = \sqrt{a_{11}}$ and $l_{21} = \frac{a_{21}}{l_{11}}$ and $l_{22} = \sqrt{a_{22} - l_{21}^2}$

\[
\text{for}(j=1,n)\{ \\
  \text{for}(k=1,j-1) \text{ for}(i=j,n) \ a_{ij} \leftarrow a_{ik}a_{jk} ; \\
  a_{jj} = \sqrt{a_{jj}} ; \text{ for}(k=j+1,n) \ a_{kj} /\!\!/ = a_{jj} ; \\
\}
\]

\text{For each column $j$ of the matrix}
\text{Loop over all previous columns $k$, and subtract a multiple of column $k$ from the current column $j$}
\text{Take the square root of the diagonal entry, and scale column $j$ by that value}

- This algorithm factors the matrix “in place” replacing $A$ with $L$
Incomplete Cholesky Preconditioner

- Cholesky factorization can be used to construct a preconditioner for a sparse matrix.
- The full Cholesky factorization would fill in too many non-zero entries.
- So, **incomplete** Cholesky preconditioning uses Cholesky factorization with the caveat that only the nonzero entries are modified (all zeros remain zeros).
Symmetric Approximation

• For non-symmetric $A$, a symmetric $\hat{A} = \frac{1}{2} (A + A^T)$ averages off-diagonal components

• Solving the symmetric $\hat{A}c = b$ instead of the non-symmetric $Ac = b$ gives a faster/easier (though error prone) approximation to a problem that might not require too much accuracy

• Alternatively, the inverse of $\hat{A}$ (or the notion thereof) may be used to devise a preconditioner for $Ac = b$