SYMMETRIC POSITIVE DEFINITE LINEAR SYSTEMS OF EQUATIONS

- Symmetric positive definite matrices.
- The $LDL^T$ decomposition; The Cholesky factorization
A real matrix is said to be positive definite if

$$(Au, u) > 0 \text{ for all } u \neq 0, u \in \mathbb{R}^n$$

Let $A$ be a real positive definite matrix. Then there is a scalar $\alpha > 0$ such that

$$(Au, u) \geq \alpha \|u\|^2.$$ 

Consider now the case of Symmetric Positive Definite (SPD) matrices.

Consequence 1: $A$ is nonsingular

Consequence 2: the eigenvalues of $A$ are (real) positive
A few properties of SPD matrices

- Diagonal entries of $A$ are positive
- Recall: the $k$-th principal submatrix $A_k$ is the $k \times k$ submatrix of $A$ with entries $a_{ij}$, $1 \leq i, j \leq k$ (Matlab: $A(1:k, 1:k)$).

D1 Each $A_k$ is SPD

D2 Consequence: $\det(A_k) > 0$ for $k = 1, \cdots, n$.

D3 If $A$ is SPD then for any $n \times k$ matrix $X$ of rank $k$, the matrix $X^TAX$ is SPD.

- The mapping $\langle x, y \rangle \rightarrow (x, y)_A \equiv (Ax, y)$ defines a proper inner product on $\mathbb{R}^n$. The associated norm, denoted by $\| \cdot \|_A$, is called the energy norm, or simply the $A$-norm:

$$\| x \|_A = (Ax, x)^{1/2} = \sqrt{x^T A x}$$
Related measure in Machine Learning, Vision, Statistics: the Mahalanobis distance between two vectors:

\[ d_A(x, y) = \|x - y\|_A = \sqrt{(x - y)^T A (x - y)} \]

Appropriate distance (measured in \# standard deviations) if \( x \) is a sample generated by a Gaussian distribution with covariance matrix \( A \) and center \( y \).
More terminology

- A matrix is **Positive Semi-Definite** if:
  \[(Au, u) \geq 0 \text{ for all } u \in \mathbb{R}^n\]

- Eigenvalues of symmetric positive semi-definite matrices are real nonnegative, i.e., ...

- ... \(A\) can be singular [If not, \(A\) is SPD]

- A matrix is said to be **Negative Definite** if \(-A\) is positive definite. Similar definition for Negative Semi-Definite

- A matrix that is neither positive semi-definite nor negative semi-definite is **indefinite**

**Show** that if \(A^T = A\) and \((Ax, x) = 0 \forall x\) then \(A = 0\)

**Show:** \(A \neq 0\) is indefinite iff \(\exists x, y: (Ax, x)(Ay, y) < 0\)
The LDL<sup>T</sup> and Cholesky factorizations

The LU factorization of an SPD matrix $A$ exists

- Let $A = LU$ and $D = diag(U)$ and set $M \equiv (D^{-1}U)^T$.

Then

$$A = LU = LD(D^{-1}U) = LDM^T$$

- Both $L$ and $M$ are unit lower triangular
- Consider $L^{-1}AL^{-T} = D^{M^T}L^{-T}$
- Matrix on the right is upper triangular. But it is also symmetric. Therefore $M^TL^{-T} = I$ and so $M = L$
- The diagonal entries of $D$ are positive [Proof: consider $L^{-1}AL^{-T} = D$]. In the end:

$$A = LDL^T = GG^T$$

where $G = LD^{1/2}$
Cholesky factorization is a specialization of the LU factorization for the SPD case. Several variants exist.

First algorithm: row-oriented LDLT

Adapted from Gaussian Elimination [Work only on upper triang. part]

1. For $k = 1 : n - 1$ Do:
2. For $i = k + 1 : n$ Do:
3. $piv := a(k, i)/a(k, k)$
4. $a(i, i : n) := a(i, i : n) - piv * a(k, i : n)$
5. End
6. End

This will give the U matrix of the LU factorization. Therefore $D = \text{diag}(U), L^T = D^{-1}U$. 
Row-Cholesky (outer product form)

Scale the rows as the algorithm proceeds. Line 4 becomes

\[ a(i,:) := a(i,:) - \left[ \frac{a(k,i)}{\sqrt{a(k,k)}} \right] \ast \left[ \frac{a(k,:)}{\sqrt{a(k,k)}} \right] \]

**ALGORITHM : 1. Outer product Cholesky**

1. For \( k = 1 : n \) Do:
2. \( A(k, k : n) = \frac{A(k, k : n)}{\sqrt{A(k,k)}} \);
3. For \( i := k + 1 : n \) Do:
4. \( A(i, i : n) = A(i, i : n) - A(k,i) \ast A(k, i : n) \);
5. End
6. End

Result: Upper triangular matrix \( U \) such \( A = U^T U \).
Example:

\[
A = \begin{pmatrix}
1 & -1 & 2 \\
-1 & 5 & 0 \\
2 & 0 & 9
\end{pmatrix}
\]

Is \(A\) symmetric positive definite?

What is the \(LDL^T\) factorization of \(A\)?

What is the Cholesky factorization of \(A\)?
**Column Cholesky.** Let \( A = GG^T \) with \( G = \) lower triangular. Then equate \( j \)-th columns:

\[
a(i, j) = \sum_{k=1}^{j} g(j, k)g^T(k, i) \rightarrow \]

\[
A(:, j) = \sum_{k=1}^{j} G(j, k)G(:, k) = G(j, j)G(:, j) + \sum_{k=1}^{j-1} G(j, k)G(:, k) \rightarrow G(j, j)G(:, j) = A(:, j) - \sum_{k=1}^{j-1} G(j, k)G(:, k)
\]
Assume that first \( j - 1 \) columns of \( G \) already known.

Compute unscaled column-vector:

\[
v = A(:, j) - \sum_{k=1}^{j-1} G(j, k)G(:, k)
\]

Notice that \( v(j) \equiv G(j, j)^2 \).

Compute \( \sqrt{v(j)} \) and scale \( v \) to get \( j \)-th column of \( G \).
ALGORITHM : 2 \hspace{1em} \textit{Column Cholesky}

1. For $j = 1 : n$ do
2. \hspace{0.5em} For $k = 1 : j - 1$ do
3. \hspace{1.5em} $A(j : n, j) = A(j : n, j) - A(j, k) \times A(j : n, k)$
4. \hspace{1.5em} EndDo
5. \hspace{0.5em} If $A(j, j) \leq 0$ ExitError(“Matrix not SPD”) \hspace{0.5em}  
6. \hspace{1.5em} $A(j, j) = \sqrt{A(j, j)}$
7. \hspace{1.5em} $A(j + 1 : n, j) = A(j + 1 : n, j)/A(j, j)$
8. \hspace{1.5em} EndDo

Try algorithm on:

\[
A = \begin{pmatrix}
1 & -1 & 2 \\
-1 & 5 & 0 \\
2 & 0 & 9
\end{pmatrix}
\]