Lecture 3 INF-MAT 4350 2008: LU-factorization and Positive Definite Matrices

Tom Lyche

Centre of Mathematics for Applications, Department of Informatics, University of Oslo

September 7, 2008
Plan for the day

- LU-factorization
- LDLT-factorization
- Cholesky-factorization
- Positive definite matrices
- Solution algorithms for a band matrix
LU-factorization

We say that $A = LR$ is an *LU-factorization* of $A \in \mathbb{R}^{n,n}$ if $L \in \mathbb{R}^{n,n}$ is lower triangular and $R \in \mathbb{R}^{n,n}$ is upper triangular. In addition we will assume that $L$ is unit triangular.

Example

\[
A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 3/2 \end{bmatrix}
\]
Example

Not every matrix has an LU-factorization.

- An LU-factorization of \( A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \) must satisfy the equations

\[
\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ l_1 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_3 \\ 0 & r_2 \end{bmatrix}
\]

for the unknowns \( l_1 \) in \( L \) and \( r_1, r_2, r_3 \) in \( R \).

- Get equations

\[
\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} r_1 & r_3 \\ l_1 r_1 & l_1 r_3 + r_2 \end{bmatrix}
\]

- Comparing \((1, 1)\)-elements we see that \( r_1 = 0 \),

- this makes it impossible to satisfy the condition \( 1 = l_1 r_1 \) for the \((2, 1)\) element. We conclude that \( A \) has no LU-factorization.
Suppose $\mathbf{A} \in \mathbb{C}^{n,n}$. The upper left $k \times k$ corners

$$
\mathbf{A}_k = \begin{bmatrix}
    a_{11} & \cdots & a_{1k} \\
    \vdots & & \vdots \\
    a_{k1} & \cdots & a_{kk}
\end{bmatrix}
$$

for $k = 1, \ldots, n$

of $\mathbf{A}$ are called the leading principal submatrices of $\mathbf{A}$.
Existence of LU

Theorem
Suppose the leading principal submatrices $A_k$ of $A \in \mathbb{C}^{n,n}$ are nonsingular for $k = 1, \ldots, n - 1$. Then $A$ has a unique LU-factorization.

Example

\[
\begin{bmatrix}
1 & 1 \\
0 & 0
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 1 \\
0 & 0
\end{bmatrix}
\]
Existence of LU

- **Proof by induction on** $n$
- $n = 1$: $[a_{11}] = [1][a_{11}]$. 
- Suppose that $A_{n-1}$ has a unique LU-factorization $A_{n-1} = L_{n-1}R_{n-1}$, and that $A_1, \ldots, A_{n-1}$ are nonsingular. 
- Since $A_{n-1}$ is nonsingular it follows that $L_{n-1}$ and $R_{n-1}$ are nonsingular. 
- But then

$$
A = \begin{bmatrix} A_{n-1} & b \\ c^T & a_{nn} \end{bmatrix} = \begin{bmatrix} L_{n-1} & 0 \\ c^T R_{n-1}^{-1} & 1 \end{bmatrix} \begin{bmatrix} R_{n-1} & v \\ 0 & a_{nn} - c^T R_{n-1}^{-1} v \end{bmatrix} = LR
$$

is an LU-factorization of $A$. 
- Since $L_{n-1}$ and $R_{n-1}$ are non-singular the block $(2,1)$ entry in $L$ is uniquely given and then $r_{nn}$ is also determined uniquely from the construction. Thus the LU-factorization is unique.
Using block multiplication one can show

**Lemma**

Suppose $A = LR$ is the LU-factorization of $A \in \mathbb{R}^{n,n}$. For $k = 1, \ldots, n$ let $A_k, L_k, R_k$ be the leading principal submatrices of $A, L, R$, respectively. Then $A_k = L_k R_k$ is the LU-factorization of $A_k$ for $k = 1, \ldots, n$.

**Example**

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix} = LR.$$ 

$A_1 = [1] = [1][1] = L_1 R_1$

$A_2 = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -3 \end{bmatrix} = L_2 R_2$

$R(3, 3) = 0$ and $A$ is singular.
**A converse to existence**

**Theorem**

*Suppose* $A \in \mathbb{C}^{n \times n}$ *has an LU-factorization. If* $A$ *is non-singular then the leading principal submatrices* $A_k$ *are nonsingular for* $k = 1, \ldots, n - 1$.

**Proof**: Suppose $A$ is nonsingular with the LU-factorization $A = LR$.

- Since $A$ is nonsingular it follows that $L$ and $R$ are nonsingular.
- By Lemma we have $A_k = L_k R_k$.
- $L_k$ is unit lower triangular and therefore nonsingular.
- $R_k$ is nonsingular since its diagonal entries are among the nonzero diagonal entries of $R$.
- But then $A_k$ is nonsingular.

**Remark** The LU-factorization of a singular matrix need not be unique. For the zero matrix any unit lower triangular matrix can be used as $L$ in an LU-factorization.
Uniqueness

- **Corollary**

  The LU-factorization of a nonsingular matrix is unique whenever it exists.

  - **Proof.** We use the two previous theorems. The leading principal submatrices $A_k$ are nonsingular for $k = 1, \ldots, n - 1$. But then the LU-factorization is unique.
**LDLT-factorization**

- For a symmetric matrix the LU-factorization can be written in a special form.
  \[
  A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 0 & 3/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3/2 \end{bmatrix} \begin{bmatrix} 1 & -1/2 \end{bmatrix}
  \]

- In the last product the first and last matrix are transposes of each other.

- \( A = LDL^T \) LDLT factorization

- \( A = LR \) where \( R = DL^T \)

- **Definition**
  Suppose \( A \in \mathbb{R}^{n \times n} \). A factorization \( A = LDL^T \), where \( L \) is unit lower triangular and \( D \) is diagonal is called an LDLT-factorization.
Theorem

Suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$ is nonsingular. Then $\mathbf{A}$ has an LDLT-factorization if and only if $\mathbf{A} = \mathbf{A}^T$ and $\mathbf{A}_k$ is nonsingular for $k = 1, \ldots, n - 1$. The LDLT factorization is unique.
Cholesky Factorization

Used for positive definite matrices

**Definition**
A factorization $A = R^T R$ where $R$ is upper triangular with positive diagonal entries is called a *Cholesky-factorization*.

 betrifft den >LaTeX-Code:<br>
$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ has an LDLT- and Cholesky-factorization given by

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3/2 \end{bmatrix} \begin{bmatrix} 1 & -1/2 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{2} & 0 \\ -1/\sqrt{2} & \sqrt{3/2} \end{bmatrix} \begin{bmatrix} \sqrt{2} & -1/\sqrt{2} \\ 0 & \sqrt{3/2} \end{bmatrix}.$$  

**Cholesky ⇔ LDLT with positive diagonal entries**

$D = \begin{bmatrix} 2 & 0 \\ 0 & 3/2 \end{bmatrix}$, $D^{1/2} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{3/2} \end{bmatrix}$, $R = D^{1/2} L^T$.  

Suppose $A \in \mathbb{R}^{n,n}$ is a square matrix. The function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = x^T Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

is called a quadratic form.

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 + 2x_2^2 - x_1 x_2 - x_2 x_1$$
Definite matrices

We say that $A$ is

(i) **positive definite** if $A^T = A$ and $x^T A x > 0$ for all nonzero $x \in \mathbb{R}^n$.

(ii) **positive semi-definite** if $A^T = A$ and $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$.

(iii) **negative (semi-)definite** if $-A$ is positive (semi-)definite.

$\Rightarrow$ A is Positive semi-definite:

$$x^T Ax = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2 + (x_1 - x_2)^2 \geq 0$$

$\Rightarrow$ A is Positive definite: $x^T Ax = 0 \Rightarrow x = 0$. 
Some Observations

- A matrix is positive definite if it is positive semidefinite and in addition
  \[ \mathbf{x}^T \mathbf{A} \mathbf{x} = 0 \Rightarrow \mathbf{x} = \mathbf{0}. \]  
  (1)

- The zero-matrix is positive semidefinite;

- a positive definite matrix must be nonsingular.

- Indeed, if \( \mathbf{A} \mathbf{x} = \mathbf{0} \) for some \( \mathbf{x} \in \mathbb{R}^n \) then \( \mathbf{x}^T \mathbf{A} \mathbf{x} = 0 \) which by (1) implies that \( \mathbf{x} = \mathbf{0} \).
Example $B^T B$

- Let $A = B^T B$, where $B \in \mathbb{R}^{m,n}$ and $m, n$ are positive integers.
- $B^T B$ is positive semi-definite
  1. $B^T B$ is symmetric
  2. $x^T A x = x^T B^T B x = (Bx)^T (Bx) = \|Bx\|_2^2 \geq 0$.
- $B^T B$ is positive definite if and only if $B$ has linearly independent columns.
- Positive definite implies $m \geq n$
The Hessian Matrix

- Suppose $F(t) = F(t_1, \ldots, t_n)$ is a real valued function of $n$ variables which has continuous 1. and 2. order partial derivatives for $t$ in some domain $\Omega$. For each $t \in \Omega$ the gradient and Hessian of $F$ are given by

$$\nabla F(t) = \begin{bmatrix} \frac{\partial F(t)}{\partial t_1} \\ \vdots \\ \frac{\partial F(t)}{\partial t_n} \end{bmatrix} \in \mathbb{R}^n, \quad H(t) = \begin{bmatrix} \frac{\partial^2 F(t)}{\partial t_1 \partial t_1} & \cdots & \frac{\partial^2 F(t)}{\partial t_1 \partial t_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F(t)}{\partial t_n \partial t_1} & \cdots & \frac{\partial^2 F(t)}{\partial t_n \partial t_n} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

- It is shown in advanced calculus texts that under suitable conditions on the domain $\Omega$ the matrix $H(t)$ is symmetric for each $t \in \Omega$.

- If $\nabla F(t^*) = 0$ and $H(t^*)$ is positive definite then $t^*$ is a local minimum for $F$.

- This can be shown using second-order Taylor approximation of $F$.

- $t^*$ is a local maximum if $\nabla F(t^*) = 0$ and $H(t^*)$ is negative definite.
When is a Matrix Positive Definite?

- Not all symmetric matrices are positive definite, and sometimes we can tell just by glancing at the matrix that it cannot be positive definite. Examples:

\[
A_1 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}.
\]

- **A_1, A_3**: Nonnegative diagonal entry: \( a_{ii} \leq 0 \) for some \( i \).

\[
e_i^T A e_i = a_{ii} \leq 0
\]

- **A_2**: if the absolute value of the largest entry of \( A \) is not (only) on the diagonal then \( A \) is not positive definite.

- To show this suppose \( a_{ij} \geq a_{ii} \) and \( a_{ij} \geq a_{jj} \) for some \( i \neq j \). Since \( A \) is symmetric we obtain

\[
(e_i - e_j)^T A (e_i - e_j) = a_{ii} + a_{jj} - 2a_{ij} \leq 0
\]

which implies that \( x^T A x \leq 0 \) for some \( x \neq 0 \).
Lemma
The leading principal submatrices of a positive definite matrix are positive definite and hence nonsingular.

Proof Consider a leading principal submatrix $A_k$ of the positive definite matrix $A \in \mathbb{R}^{n,n}$.

- $A_k$ is symmetric.
- Let $x \in \mathbb{R}^k$ be nonzero, set $y = \begin{bmatrix} x^T & 0 \end{bmatrix} \in \mathbb{R}^n$, and partition $A$ conformally with $y$ as $A = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix}$, where $D_k \in \mathbb{R}^{n-k,n-k}$.

Then

$$0 < y^T Ay = [x^T \ 0^T] \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = x^T A_k x.$$
Lemma

A matrix is positive definite if and only if it has a Cholesky factorization.

Proof: Suppose $A$ is positive definite.

1. $A_k$ is nonsingular for $k = 1, \ldots, n - 1$
2. $A = LDL^T$, $d_{ii} > 0$?
3. $d_{ii} = e_i^T D e_i = e_i^T L^{-1} A L^{-T} e_i = x_i^T A x_i > 0$, since $L$ is nonsingular and so $x_i = L^{-T} e_i$ is nonzero.

Conversely, suppose $A$ has a Cholesky factorization $A = R^T R$

1. $R$ is nonsingular and $A$ is positive definite by $B^T B$ example.
Lemma

A matrix is positive definite if and only if it is symmetric and has positive eigenvalues.

Proof

⇒:

- $A$ is symmetric and $\lambda = \frac{x^T A x}{x^T x} > 0$

⇐:

- Use the spectral theorem
Necessary and Sufficient Conditions

Theorem

The following is equivalent for a symmetric matrix $A \in \mathbb{R}^{n,n}$

1. $A$ is positive definite.
2. $A$ has only positive eigenvalues.
3. Sylvester’s criterium

$$\begin{vmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \ldots & a_{kk} \end{vmatrix} > 0 \text{ for } k = 1, \ldots, n$$

4. $A = B^T B$ for a nonsingular $B \in \mathbb{R}^{n,n}$
Proof 1 ⇒ 3:

- **Proof** By Lemma 10 we know that $1 ⇔ 2$. We can show that $1 ⇒ 3 ⇒ 4 ⇒ 1$.
- Show only $1 ⇔ 3$
- Each $A_k$ is positive definite, and hence has positive eigenvalues.
- The determinant of a matrix equals the product of its eigenvalues.
- $\det(A_k) > 0$ for $k = 1, \ldots, n$. 
trace, determinant and eigenvalues

▶ \( \text{trace}(\mathbf{A}) := a_{11} + \cdots + a_{nn} = \lambda_1 + \cdots + \lambda_n, \)
▶ \( \text{det}(\mathbf{A}) = \lambda_1 \cdots \lambda_n \)
▶ Proof \( n = 2 \)

\[
p_{\mathbf{A}}(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{21}a_{12}
\]
\[
= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{21}a_{12}
\]
\[
= \lambda^2 - \text{trace}(\mathbf{A})\lambda + \text{det}(\mathbf{A}).
\]

▶ On the other hand
\[
p_{\mathbf{A}}(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2.
\]
▶ \( \text{trace}(\mathbf{A}) = \lambda_1 + \lambda_2, \quad \text{det}(\mathbf{A}) = \lambda_1\lambda_2 \)
Positive definite example

\[ C := \text{tridiag}(a, b, a) = \begin{bmatrix} b & a & 0 & \ldots & 0 \\ a & b & a & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & a & b & a \\ 0 & \ldots & a & b & \end{bmatrix} \in \mathbb{R}^{m,m}. \]

\[ (2) \]

\[ \text{\begin{itemize}
  \item $C$ is symmetric $C^T = C$.
  \item $C = T = \text{tridiag}(-1, 2, 1)$ when $a = -1$ and $b = 2$.
  \item $C = N_1 = \text{tridiag}(1, 4, 1)$ the cubic spline matrix when $a = 1$ and $b = 4$.
\end{itemize}} \]
\[
C = \text{tridiag}(a, b, a) \text{ is positive definite if } b > 0 \text{ and } b \geq 2|a|
\]

- Suppose \( b > 0 \) and \( b \geq 2|a| \)
- Since \( C \) is symmetric it is enough to show that the smallest eigenvalue \( \lambda_{\min} \) is positive.
- The eigenvalues are \( \lambda_j = b + 2a \cos(j\pi h) \) for \( j = 1, \ldots, m \), where \( h = 1/(m+1) \).
- For \( C \) to be positive definite it is necessary that the diagonal entry \( b > 0 \).
- If \( b > 0 \) then \( \lambda_{\min} = b - 2|a| \cos(\pi h) > b - 2|a| \geq 0 \)
- Thus \( C \) is positive definite
- \( T \) and \( N_1 \) are both positive definite.
Solving Positive Definite Systems by Cholesky

- Given $Ax = b$ where $A$ is positive definite
- $A = R^T R$ (Cholesky)
- $R^T y = b$ (forward substitution)
- $Rx = y$ (backward substitution)
Cholesky Factorization Algorithm

\[
(A)_{ij} = (R^T R)_{ij}
\]

For \( j \geq i \):
\[
a_{ij} = \sum_{k=1}^{n} r_{ik}^T r_{kj} = \sum_{k=1}^{n} r_{ki} r_{kj} = \sum_{k=1}^{i} r_{ki} r_{kj}.
\]

\( j = i \)
\[
a_{ii} = \sum_{k=1}^{i} r_{ki}^2 = \sum_{k=1}^{i-1} r_{ki}^2 + r_{ii}^2,
\]

\[
r_{ii} = (a_{ii} - \sum_{k=i_m}^{i-1} r_{ki}^2)^{1/2}, \text{ where } i_m = 1 \text{ for a full matrix.}
\]

\[
a_{ij} = \sum_{k=1}^{i-1} r_{ki} r_{kj} + r_{ii} r_{ij}, \text{ for } j > i
\]

\[
r_{ij} = (a_{ij} - \sum_{k=i_m}^{i-1} r_{ki} r_{kj})/r_{ii} \quad j = i + 1, \ldots, i_p, \text{ where } i_m = 1 \text{ and } i_p = n \text{ for a full matrix.}
\]

\[
y_i = (b_i - \sum_{k=i_m}^{i-1} r_{ki} y_k)/r_{ii}, \text{ for } i = 1, \ldots, n,
\]

\[
x_i = (y_i - \sum_{k=i+1}^{i_p} r_{ik} x_k)/r_{ii}, \text{ for } i = n, n-1, \ldots, 1,
\]
Band Matrix

- **A** has bandwidth *d* if \( a_{ij} = 0 \) for \( |i - j| > d \).
- A tridiagonal matrix has bandwidth *d* = 1.

**Lemma**

Suppose **A** is positive definite with Cholesky-factorization \( **A = R^T R** \). If **A** has bandwidth *d*, then **R** has bandwidth *d*.

- For fixed *i* use previous formulas with

\[
i_m := \max(i - d, 1), \text{ and } i_p := \min(i + d, n).
\]  

(3)
vectorized versions

\[ s_j^T := [r_{i_m,j}, \ldots, r_{i-1,j}] \]

\[ r_{ii} = (a_{ii} - s_i^T s_i)^{1/2}, \]

\[ [r_{i,i+1}, \ldots, r_{i,i_p}] = ([a_{i,i+1}, \ldots, a_{i,i_p}] - s_i^T [s_{i+1}, \ldots, s_{i_p}]) / r_{ii}, \]

\[ y_i = (b_i - s_i^T \begin{bmatrix} y_{i_m} \\ \vdots \\ y_{i-1} \end{bmatrix}) / r_{ii}, \]

\[ x_i = (y_i - [r_{i+1,i}, \ldots, r_{i_p,i}] \begin{bmatrix} x_{i+1} \\ \vdots \\ x_{i_p} \end{bmatrix}) / r_{ii} \]
\[ r_{ii} = \left( a_{ii} - \sum_{k=1}^{i-1} r_{ki}^2 \right)^{1/2}, \]
\[ r_{ij} = \frac{a_{ij} - \sum_{k=i_m}^{i-1} r_{ki} r_{kj}}{r_{ii}} \quad j = i + 1, \ldots, n, \]
\[ \text{The number of flops needed for the full Cholesky-factorization} \]
\[ \sum_{i=1}^{n} (2i-2 + (2i-1)(n-i)) \approx \sum_{i=0}^{n} 2i(n-i) \approx \int_{0}^{n} 2x(n-x)dx = \frac{n^3}{3}. \]
\[ \text{In addition we need to take } n \text{ square roots.} \]
\[ \text{This is half the number of flops needed on Gaussian elimination of an arbitrary matrix.} \]
\[ \text{We obtain this reduction since the Cholesky factorization takes advantage of the symmetry of } A. \]
# flops $d$-banded Cholesky

- $O(2nd^2)$ for the banded Cholesky
- $O(4nd)$ for backward and forward substitution.
- When $d$ is small compared to $n$ we obtain an $O(n)$ algorithm
- these numbers are considerably smaller than the $O(n^3/3)$ and $O(2n^2)$ counts for the factorization of a full matrix.
Conclusions

- LU-factorization does not always exist
- LDLT-factorization for symmetric matrices
- Cholesky-factorization for positive definite matrices
- Positive definite matrices
- $O(nd^2)$ algorithm for a positive definite band matrix