1. **LU decomposition (Gaussian elimination in matrix form).** If \( A \) is a square nonsingular matrix, then there exist a permutation matrix \( P \), a unit lower triangular matrix \( L \), and an upper triangular matrix \( U \) such that

\[
PA = LU.
\]

Special cases:

(a) **Cholesky decomposition.** A matrix \( A \) is symmetric positive definite if and only if there exists a unique nonsingular upper triangular matrix \( R \), with positive diagonal entries, such that

\[
A = R^T R.
\]

(b) **LDL^T factorization.** If \( A^T = A \) is nonsingular, then there exists a permutation \( P \), a unit lower triangular matrix \( L \), and a block diagonal matrix \( D \) with 1-by-1 and 2-by-2 blocks such that

\[
PAP^T = LDL^T.
\]

Applications:

- Solve \( Ax = b \).
- Compute \( \det(A) \).
- Compute \( A^{-1} \), if really necessary.

2. **QR decomposition.** Let \( A \) be \( m \)-by-\( n \) with \( m \geq n \). Suppose that \( A \) has full column rank. Then there exist a unique \( m \)-by-\( n \) orthogonal matrix \( Q \) (i.e. \( Q^T Q = I \)) and a unique \( n \)-by-\( n \) upper triangular matrix \( R \) with positive diagonal \( r_{ii} > 0 \) such that

\[
A = QR.
\]

Applications:

- Find an orthonormal basis of the subspace spanned by the columns of \( A \) (the Gram-Schmidt orthogonalization process)
- Solve the linear least squares problem \( \min_x \|Ax - b\|_2 \).

3. **Schur decomposition, eigenvalue decomposition and spectral decomposition.** Let \( A \) be of order \( n \). Then

(a) there is an \( n \times n \) unitary matrix \( U \) (i.e. \( U^H U = I \)) such that

\[
A = UTU^H,
\]

where \( T \) is upper triangular. This is called a **Schur decomposition**.

(b) The **eigenvalue decomposition**, if exists, is given by

\[
A = XAX^{-1},
\]

where \( A \) is a diagonal matrix.
(c) When $A$ is Hermitian, $A^H = A$, we have the spectral decomposition
\[ A = Q\lambda Q^H, \]
where $A$ is real and diagonal.

Applications:
- The eigenvalues of $A$ are the diagonal elements of $T$. By appropriate choice of $U$, the eigenvalues of $A$, which are the diagonal elements of $T$, may be made to appear in any order.
- Compute matrix functions $f(A) = Uf(T)U^H$.

4. **Singular Value Decomposition (SVD).** Let $A$ be an $m$-by-$n$ matrix with $m \geq n$. Then we can write
\[ A = U\Sigma V^T, \]
where $U$ is $m$-by-$m$ orthogonal matrix (i.e. $U^TU = I_m$) and $V$ is $n$-by-$n$ orthogonal matrix (i.e. $V^TV = I_n$), and $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

If $m < n$, the SVD can be defined by considering $A^T$.

The columns $u_1, u_2, \ldots, u_m$ of $U$ are called left singular vectors of $A$. The columns $v_1, v_2, \ldots, v_n$ of $V$ are called right singular vectors. The $\sigma_1, \sigma_2, \ldots, \sigma_n$ are called singular values.

Applications:
- Suppose that $A$ is $m$-by-$n$ with $m \geq n$ and has full rank, with $A = U\Sigma V^T$ being $A$’s SVD. Then the pseudo-inverse can also be written as
  \[ A^\dagger \equiv (A^TA)^{-1}A^T = V\Sigma^{-1}U^T. \]
  If $m < n$, then $A^\dagger = A^T(AA^T)^{-1}$.
- Suppose that
  \[ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} = \cdots = \sigma_n = 0, \]
  Then the rank of $A$ is $r$. The range space of $A$ is span($u_1, u_2, \ldots, u_r$). and the null space of $A$ is span($v_{r+1}, v_{r+2}, \ldots, v_n$).
- $\|A\|_2 = \sigma_1(\equiv \sigma_{\max})$
- Let $A$ be $m \times n$ with $m \geq n$. Then
  (a) eigenvalues of $A^TA$ are $\sigma_i^2$, $i = 1, 2, \ldots, n$. The corresponding eigenvectors are the right singular vectors $v_i$, $i = 1, 2, \ldots, n$.
  (b) eigenvalues of $AA^T$ are $\sigma_i^2$, $i = 1, 2, \ldots, n$ and $m - n$ zeros. The left singular vectors $u_i$, $i = 1, 2, \ldots, n$ are corresponding eigenvectors for the eigenvalues $\sigma_i^2$. One can take any $m - n$ other orthogonal vectors that are orthogonal to $u_1, u_2, \ldots, u_n$ as the eigenvectors for the eigenvalues 0.
- Principal components. The SVD of $A$ can be rewritten as
  \[ A = E_1 + E_2 + \cdots + E_p \]
  where $p = \min(m, n)$, and $E_k$ is a rank-one matrix of the form
  \[ E_k = \sigma_k u_k v_k^T, \]
  $E_k$ are referred to as component matrices, and are orthogonal to each other in the sense that
  \[ E_j E_k^T = 0, \quad j \neq k. \]
  Since $\|E_k\|_2 = \sigma_k$, the contribution each $E_k$ makes to reproduce $A$ is determined by the size of the singular value $\sigma_k$. 

2
• Optimal rank-\(k\) approximation:

\[
\min_{B : m \times n} \| A - B \|_2 = \| A - A_k \|_2 = \sigma_{k+1},
\]

where

\[
A_k = U \Sigma_k V^T, = E_1 + E_2 + \cdots + E_k,
\]

and \(\Sigma_k = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0)\) \(^1\)

• Data compression. Note that the optimal rank-\(k\) approximation \(A_k\) can be written in a compact form as

\[
A_k = U_k \hat{\Sigma}_k V_k^T,
\]

where \(U_k\) and \(V_k\) are the first \(k\) columns of \(U\) and \(V\), respectively, \(\hat{\Sigma}_k = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)\).

Therefore, \(A_k\) is represented by \(mk + k + nk = (m + n + 1)k\) elements, in contrast, \(A\) is represented by \(mn\) elements.

\[
\text{compression ratio} = \frac{(m + n + 1)k}{mn}
\]

The following plots show the original image, and three compressed ones with different compression ratios:

\(^1\)In [Golub, Hoffman and Stewart, LAA, vol.88/89, pp.317-327, 1987], it is shown how to obtain a best approximation of lower rank in which a specified set of columns of the matrix \(A\) remains fixed.