Matrix Decomposition Proofs

Gene Yang

May 22, 2024

Contents

1 Introduction

Matrix decomposition is a way of representing one matrix as the product of several matrices with special properties. For example, factoring a matrix into two triangular matrices L and U makes it easier to solve complicated systems of equations. Spectral decomposition gives a way to look at matrices in terms of their eigenvectors and eigenvalues, but also makes it easier to raise matrices to a power. Matrix decomposition is widely used in machine learning, quantum gates, statistics, and more.

This paper covers LU decomposition, QR decomposition, spectral decomposition (with the spectral theorem), and polar decomposition.

2 LU Decomposition

Theorem 22.2.1: "Most" $n \times n$ matrices A have the form $A = LU$ for $n \times n$ lower triangular matrix L and $n \times n$ upper triangular matrix U. The matrix A is invertible precisely when the diagonal entries of L and U are nonzero. [1, p. 405]

First we prove that a $n \times n$ matrix A with an LU decomposition is only invertible when the diagonals of L and U are nonzero. Then we prove that an invertible square matrix A has an LU decomposition if and only if its leading principal minors are nonzero.

2.1 If $A = LU$, A is invertible if diagonals of L and U are nonzero

For a $n \times n$ matrix A, if an LU decomposition exists, then [1, p. 570]

$$\begin{aligned} det(A) &= det(LU) \\ &= det(L)det(U) \end{aligned}$$

A is invertible when $det(A) \neq 0$, or $det(L), det(U) \neq 0$. The determinant of a triangular matrix is the product of its diagonal entries [1, p. 561]. Therefore, if a $n \times n$ matrix A has an LU decomposition, the diagonals of L and U must be nonzero.

2.2 If $A = LU$, principal minors of A are nonzero

Using block matrix notation,

$$A = \left[\begin{array}{c|c} A' & \cdots \\ \hline \vdots & \ddots \end{array}\right] = \left[\begin{array}{c|c} L' & \mathbf{0} \\ \hline \vdots & \ddots \end{array}\right] \left[\begin{array}{c|c} U' & \cdots \\ \hline \mathbf{0} & \ddots \end{array}\right]$$

for some leading submatrix A' of A, a lower triangular L', and an upper triangular U'. Using the rules of block matrix multiplication,

$$\begin{aligned} det(A') &= det(L'U') \\ &= det(L')det(U') \end{aligned}$$

Since A is invertible, L and U have nonzero diagonals from 2.1. L' and U' also have nonzero diagonals. The determinant of a triangular matrix is the product of its diagonal entries, so

$$det(L'), det(U') \neq 0$$

$$det(A') \neq 0$$

This applies to all leading principal submatrices of A. If LU decomposition exists for an invertible $n \times n$ matrix A, then all its principal minors are nonzero.

2.3 If principal minors of A are nonzero, $A = LU$

Induction is used to prove this for the 1×1 case and generalize it to an invertible $n \times n$ matrix.

For the 1×1 case, the matrix $\begin{bmatrix} A \end{bmatrix}$ can be factored with $L = \begin{bmatrix} 1 \end{bmatrix}$, $U = \begin{bmatrix} A \end{bmatrix}$ [2, p. 17].

For the $n \times n$ case, let A_n be an invertible $n \times n$ matrix with a $(n-1) \times (n-1)$ leading principal submatrix. Assume that this submatrix A_{n-1} has an LU decomposition $A_{n-1} = L_{n-1}U_{n-1}$. Then A_n can be factored as [2, p. 17]

$$A_n = \left[\begin{array}{c|c} A_{n-1} & \mathbf{a}_1 \\ \hline \mathbf{a}_2^T & a_3 \end{array}\right] = \left[\begin{array}{c|c} L_{n-1} & \mathbf{0} \\ \hline \mathbf{l}_1^T & l_2 \end{array}\right] \left[\begin{array}{c|c} U_{n-1} & \mathbf{u}_1 \\ \hline \mathbf{0}^T & u_2 \end{array}\right]$$

$$A_{n-1} = L_{n-1}U_{n-1}$$

$$\mathbf{a}_1 = L_{n-1}\mathbf{u}_1$$

$$\mathbf{a}_2^T = \mathbf{l}_1^T U_{n-1}$$

$$a_3 = \mathbf{l}_1^T \mathbf{u}_1 + l_2 u_2$$

The block matrix determinant formula gives $det(A_n) = det(A_{n-1})det(a_3 - \mathbf{a}_2^T A_{n-1}^{-1}\mathbf{a}_1)$ [2, p. 17]. Plugging in values for $a_3 - \mathbf{a}_2^T A_{n-1}^{-1}\mathbf{a}_1$, we get

$$\begin{aligned} a_3 - \mathbf{a}_2^T A_{n-1}^{-1}\mathbf{a}_1 &= (\mathbf{l}_1^T\mathbf{u}_1 + l_2u_2) - (\mathbf{l}_1^T U_{n-1})(A_{n-1}^{-1})(L_{n-1}\mathbf{u}_1) \\ &= (\mathbf{l}_1^T\mathbf{u}_1 + l_2u_2) - (\mathbf{l}_1^T U_{n-1})(U_{n-1}^{-1}L_{n-1}^{-1})(L_{n-1}\mathbf{u}_1) \\ &= (\mathbf{l}_1^T\mathbf{u}_1 + l_2u_2) - \mathbf{l}_1^T\mathbf{u}_1 \\ &= l_2u_2 \end{aligned}$$

A_n is invertible, so L_n and U_n have nonzero diagonals by 2.1 and $l_2u_2 \neq 0$. By 2.2, $det(A_{n-1})$ is nonzero because the assumption was that it has an LU decomposition. Therefore,

$$det(A_n) = det(A_{n-1})det(a_3 - \mathbf{a}_2^T A_{n-1}^{-1}\mathbf{a}_1)$$

$$det(A_n) \neq 0$$

If an invertible $n \times n$ matrix A has nonzero principal minors, then it has an LU decomposition.

2.4 Result

1. If an LU decomposition exists for A, its principal minors are nonzero. (2.2)

2. If the principal minors of A are nonzero, an LU decomposition exists for A. (2.3)

Invertible $n \times n$ matrices have LU decomposition if and only if their principal minors are nonzero.

3 QR Decomposition

Theorem 22.2.1: An invertible $n \times n$ matrix A can be written as $A = QR$ where Q is an $n \times n$ orthogonal matrix and R is an $n \times n$ upper triangular matrix.

Rewrite A in terms of its column vectors, so that

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ | & | & & | \end{bmatrix}$$

Carrying out the Gram-Schmidt process on the columns of A,

$$\begin{gathered} \mathbf{q}_1 = \mathbf{a}_1/\|\mathbf{a}_1\| \\ \mathbf{q}_2 = (\mathbf{a}_2 - \frac{\mathbf{q}_1 \cdot \mathbf{a}_2}{\mathbf{q}_1 \cdot \mathbf{q}_1}\mathbf{q}_1)/\|\mathbf{a}_2 - \frac{\mathbf{q}_1 \cdot \mathbf{a}_2}{\mathbf{q}_1 \cdot \mathbf{q}_1}\mathbf{q}_1\| \\ \vdots \\ \mathbf{q}_n = (\mathbf{a}_n - \sum_{i=1}^{n-1} \frac{\mathbf{q}_i \cdot \mathbf{a}_n}{\mathbf{q}_i \cdot \mathbf{q}_i}\mathbf{q}_i)/\|\mathbf{a}_n - \sum_{i=1}^{n-1} \frac{\mathbf{q}_i \cdot \mathbf{a}_n}{\mathbf{q}_i \cdot \mathbf{q}_i}\mathbf{q}_i\| \end{gathered}$$

Solving for $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$ gives

$$\mathbf{a}_1 = \|\mathbf{a}_1\|\mathbf{q}_1$$
$$\mathbf{a}_2 = \|\mathbf{a}_2 - \frac{\mathbf{q}_1 \cdot \mathbf{a}_2}{\mathbf{q}_1 \cdot \mathbf{q}_1}\mathbf{q}_1\|\mathbf{q}_2 + \frac{\mathbf{q}_1 \cdot \mathbf{a}_2}{\mathbf{q}_1 \cdot \mathbf{q}_1}\mathbf{q}_1$$
$$\vdots$$
$$\mathbf{a}_n = \|\mathbf{a}_n - \sum_{i=1}^{n-1} \operatorname{proj}_{\mathbf{q}_i} \mathbf{a}_n\|\mathbf{q}_n + \sum_{i=1}^{n-1} \operatorname{proj}_{\mathbf{q}_i} \mathbf{a}_n$$

A can now be factored as QR, with [3, p. 3]

$$\begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \dots & \mathbf{q}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \|\mathbf{a}_1\| & \frac{\mathbf{q}_1 \cdot \mathbf{a}_2}{\mathbf{q}_1 \cdot \mathbf{q}_1} & \dots & \frac{\mathbf{q}_1 \cdot \mathbf{a}_n}{\mathbf{q}_1 \cdot \mathbf{q}_1} \\ 0 & \|\mathbf{a}_2 - \frac{\mathbf{q}_1 \cdot \mathbf{a}_2}{\mathbf{q}_1 \cdot \mathbf{q}_1}\mathbf{q}_1\| & \dots & \frac{\mathbf{q}_2 \cdot \mathbf{a}_n}{\mathbf{q}_2 \cdot \mathbf{q}_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \|\mathbf{a}_n - \sum_{i=1}^{n-1} \operatorname{proj}_{\mathbf{q}_i} \mathbf{a}_n\| \end{bmatrix}$$

Any invertible $n \times n$ matrix A can be factored as $A = QR$. Q is an orthogonal matrix, since its columns were found using the Gram-Schmidt process. R is an upper triangular matrix.

4 Spectral Decomposition

First the Spectral Theorem is proven, then the existence of matrix decomposition with the Spectral Theorem is proven.

4.1 Spectral Theorem

Theorem 24.1.4: Let A be a *symmetric* $n \times n$ matrix. There is an orthgonal basis $\mathbf{w}_1$, $\mathbf{w}_2$, ... , $\mathbf{w}_n$ consisting of eigenvectors for A. The corresponding eigenvalues are all of the eigenvalues for A: if $\mathbf{w}_j$ has eigenvalue λ_j then any eigenvalue of A equals some λ_j. [1, p. 446]

4.1.1 An eigenvector exists for a $n \times n$ symmetrical matrix A

We can prove that an eigenvector of A exists at a maximum of the Rayleigh quotient $\frac{q_A(\mathbf{v})}{\|\mathbf{v}\|^2}$ [1, p. 543]. Lagrange multipliers can be used to maximize $f(\mathbf{v}) = q_A(\mathbf{v})$ with constraint $g(\mathbf{v}) = \|\mathbf{v}\|^2 = 1$.

$$(\nabla f)(\mathbf{v}) = \nabla(\mathbf{v}^T A\mathbf{v}) = 2A\mathbf{v}$$

$$(\nabla g)(\mathbf{v}) = \nabla(\mathbf{v} \cdot \mathbf{v}) = 2\mathbf{v}$$

$2\mathbf{v} \neq 0$, because $\mathbf{v}$ is an eigenvector of A, and eigenvectors are nonzero. Then $2A\mathbf{v} = \lambda 2\mathbf{v}$, or $A\mathbf{v} = \lambda\mathbf{v}$, which is the definition of an eigenvector [1, p. 545]. Therefore, an eigenvector exists for any $n \times n$ symmetrical matrix A.

4.1.2 n eigenvectors exist for a $n \times n$ symmetrical matrix A

From the last section, a symmetric $n \times n$ matrix A has at least 1 eigenvector. Let $\mathbf{v} \in \mathbf{R}^n$ be an eigenvector of A. An orthonormal basis $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{n-1}$ for $\mathbf{v}^\perp$ can be made using the Gram-Schmidt process [1, p. 542]. Vectors in $\mathbf{v}^\perp$ are represented as $\mathbf{h} = \sum_{i=1}^{n-1} c_i\mathbf{h}_i$ for some constants c_i.

A $(n-1) \times (n-1)$ matrix B that uses the vectors $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{n-1}$ as its columns can be constructed. Then the column space of B is $\mathbf{v}^\perp$.

B is symmetric if $(B\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (B\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in \mathbf{R}^{n-1}$ [1, p. 541]. This is a special case of $(A\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (A\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in \mathbf{R}^{n-1}$ [1, p. 542]. $(A\mathbf{v}) \cdot \mathbf{w}$ is equal to $\mathbf{v} \cdot (A\mathbf{w})$ because A is symmetric, so B is also symmetric and it has at least 1 eigenvector.

Similarly, $B\mathbf{x} \in \mathbf{R}^{n-1}$ is a special case of $A\mathbf{h} \in \mathbf{R}^{n-1}$, where $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \end{bmatrix}$ and $\mathbf{h} = \sum_{i=1}^{n-1} x_i \mathbf{h}_i$ [1, p. 542]. Based on this, $B\mathbf{x} = \lambda\mathbf{x}$ is equivalent to $A\mathbf{h} = \lambda\mathbf{h}$ for some λ [1, p. 542]. The eigenvectors of B are also the eigenvectors of A. B's eigenvectors are orthogonal to A's eigenvector $\mathbf{v}$, because $\mathbf{x} \in \mathbf{v}^{\perp}$.

The process of finding eigenvalues can be continued from B to a $(n-2) \times (n-2)$ matrix, $(n-3) \times (n-3)$ matrix, and so on. In total, there are n eigenvectors of A, and since they're orthogonal to each other, they form a basis.

4.2 Matrix Decomposition with the Spectral Theorem

Theorem 24.4.1: Let A be a symmetric $n \times n$ matrix with orthogonal eigenvectors $\mathbf{w}_1$, $\mathbf{w}_2$, ... , $\mathbf{w}_n$, having corresponding eigenvalues λ_1, λ_2, ... , λ_n. Let W be the $n \times n$ matrix whose columns are the respective unit eigenvectors

$$\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}, \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|}, \ldots, \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|}$$

Then W is an orthogonal matrix and $A = WDW^T = WDW^{-1}$ for the diagonal matrix

$$D = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

whose entries are the corresponding eigenvalues. [1, p. 454]

For the matrix W and D,

$$AW = A\begin{bmatrix} | & | & & | \\ \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} & \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} & \dots & \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A(\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}) & A(\frac{\mathbf{w}_2}{\|\mathbf{w}_2\|}) & \dots & A(\frac{\mathbf{w}_n}{\|\mathbf{w}_n\|}) \\ | & | & & | \end{bmatrix}$$

$$WD = \begin{bmatrix} | & | & & | \\ \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} & \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} & \dots & \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|} \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1(\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}) & \lambda_2(\frac{\mathbf{w}_2}{\|\mathbf{w}_2\|}) & \dots & \lambda_n(\frac{\mathbf{w}_n}{\|\mathbf{w}_n\|}) \\ | & | & & | \end{bmatrix}$$

$AW = WD$, since for all i, $A\mathbf{w}_i = \lambda_i \mathbf{w}_i$ by definition of eigenvectors and eigenvalues, so $A(\frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}) = \lambda_i(\frac{\mathbf{w}_i}{\|\mathbf{w}_i\|})$.

From the Spectral Theorem, eigenvectors $\mathbf{w}_1$, $\mathbf{w}_2$, ... , $\mathbf{w}_n$ of A are orthogonal to each other. Therefore, W must be an orthogonal matrix, since its columns $\frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}$, $\frac{\mathbf{w}_2}{\|\mathbf{w}_2\|}$, ... ,$\frac{\mathbf{w}_n}{\|\mathbf{w}_n\|}$ are an orthonormal basis [1, p. 359]. The inverse and transpose of orthogonal matrices are equal.

$$AW = WD$$

$$A = WDW^{-1} \text{ or } A = WDW^T$$

A symmetric $n \times n$ matrix has orthogonal eigenvectors from the Spectral Theorem, and can be factored as $A = WDW^{-1}$ or $A = WDW^T$. W is the orthogonal matrix with unit eigenvectors as its columns, and D is the diagonal matrix with eigenvalues in its diagonal.

5 Polar Decomposition

Theorem B.4.2: If A is an invertible $n \times n$ matrix, then we can uniquely write $A = QS$ where Q is an orthogonal $n \times n$ matrix and S is a positive-definite symmetric $n \times n$ matrix. [1, p. 550]

Singular Value Decomposition says that a $n \times n$ matrix can be factored as $A = QDQ'^T$. D is a diagonal matrix with a positive diagonal. Q and Q'^T are orthogonal matrices [1, p. 510].

$$A = Q_0 D_0 Q_0'^T$$
$$A = Q_0 (Q_0'^{-1} Q_0') D_0 Q_0'^T$$
$$A = (Q_0 Q_0'^{-1})(Q_0' D_0 Q_0'^T)$$

Q_0' is an orthogonal matrix, so its inverse $Q_0'^{-1}$ is also orthogonal [1, p. 551]. A product of orthogonal matrices is orthogonal, so $(Q_0 Q_0'^{-1})$ is also orthogonal [1, p. 359].

Writing out $Q_0' D_0 Q_0'^T$,

$$\begin{aligned} Q_0' D_0 Q_0'^T &= \begin{bmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \dots & q_{nn} \end{bmatrix} \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{nn} \end{bmatrix} \begin{bmatrix} q_{11} & q_{21} & \dots & q_{n1} \\ q_{12} & q_{22} & \dots & q_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ q_{1n} & q_{2n} & \dots & q_{nn} \end{bmatrix} \\ &= \begin{bmatrix} q_{11}^2 d_{11} & 0 & \dots & 0 \\ 0 & q_{22}^2 d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & q_{nn}^2 d_{nn} \end{bmatrix} \end{aligned}$$

The diagonal entries d_{11}, d_{22}, ... , d_{nn} of D_0 are positive, so $Q_0' D_0 Q_0'^T$ is a diagonal matrix with positive diagonal entries, or a positive-definite symmetric matrix.

An invertible $n \times n$ matrix A can be factored as $A = QS$, where Q is an orthogonal $n \times n$ matrix and S is a positive-definite symmetric $n \times n$ matrix.

References

[1] Stanford University Math Department. *Linear Algebra, Multivariable Calculus, and Modern Applications*. 2023.

[2] Jun Lu. *Matrix Decomposition and Applications*. URL: `https://arxiv.org/pdf/2201.00145`. (accessed: 5.18.2024).

[3] Padraic Bartlett. *Applications of Orthogonality: QR Decompositions*. URL: `https://web.math.ucsb.edu/~padraic/ucsb_2013_14/math108b_w2014/math108b_w2014_lecture4.pdf`. (accessed: 5.19.2024).