Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 7, No. 3, 2025

Matrix Factorization: An Introduction to Traditional Methods and a Review of Recent Application Research

Junpeng Yuan[*, #], Wenqing Zhuang[#], Jiaxin Huang[#], Runze Li[#], Jun Liu[#]

[1] D School of Mathematics and Information Sciences, Guangzhou University, Guangzhou 510006, China
[*] **Corresponding author**: Junpeng Yuan (Email: 13257516286@163.com)
[#]**These authors contributed equally.**

Abstract: Matrix decomposition, as a core mathematical tool for low-rank representation and latent feature extraction in high-dimensional data, has evolved in recent years from the realm of classical numeri-cal linear algebra to become a key component of intelligent computing engines. This paper systemat-ically reviews the mathematical mechanisms, stability conditions, and algorithmic complexity of three major classical methods: LU, Cholesky, and SVD. It identifies theoretical bottlenecks within the traditional framework concerning interpretability, computational efficiency, and complex pattern characterization. Subsequently, it focuses on three interdisciplinary scenarios—hyperspectral remote sensing, lncRNA–disease association prediction, and ground-penetrating radar clutter suppres-sion—to dissect recent advancements and failure mechanisms in non-negative matrix factorization, low-rank sparse modelling, and tensor regularization strategies. Furthermore, it proposes a synergis-tic optimization pathway integrating domain knowledge, regularization constraints, and stochastic algorithms. This demonstrates the feasibility of embedding physical equations, biological priors, and deep learning nonlinear representations into a unified loss function, enabling a paradigm shift from static data analysis to dynamic system intervention. Finally, we envision personalized computational models for millisecond-scale closed-loop neural modulation, providing an extensible methodological framework for deep applications of matrix decomposition in precision medicine and real-time engi-neering systems.

Keywords: Matrix decomposition, Low-rank approximation, Multimodal data fusion, Intelligent computing en-gine.

1. Introduction

Matrix factorisation reveals intrinsic structures within high-dimensional data and achieves low-rank representations. Its core function lies in effectively separating and extracting key latent features and patterns from complex, sparse, or noisy observational data. This enables the connection between raw data and deep-seated information. Research into matrix decomposition addresses the growing demand for data analysis across various fields by transforming complex problems into more manageable low-rank forms, thereby yielding the desired segmented data. A synthesis of the literature reveals inherent limitations of matrix decomposition in real-world deployments; accordingly, we propose targeted remedies in three dimensions—interpretability, computational efficiency, and the modeling of complex patterns. This paper aims to provide researchers with a clear overview from theoretical foundations to application frontiers, while also outlining potential future pathways for the evolution of matrix decomposition methods towards intelligent computing engines. It innovatively proposes a synergistic optimization pathway integrating domain knowledge, regularization constraints, and stochastic algorithms, finally enabling a paradigm shift from static data analysis to dynamic system intervention.

2. Introduction to Traditional Matrix Decomposition Methods

2.1. LU Decomposition

LU decomposition decomposes a square matrix into the product of a lower triangular matrix (L) and an upper triangular matrix (U), expressed as $\mathrm{A} = \mathrm{LU}$. Its core principle lies in the matrix form of Gaussian elimination, transforming the solution of a linear equation system into solving two simple triangular equation systems (forward substitution and backward substitution), thereby significantly enhancing computational efficiency.

For any square matrix, a unique LU decomposition exists (with L being the lower triangular unit matrix) if and only if all sequential principal minors (the determinants of the first k rows and first k columns) are non-zero.

The LU decomposition method embodies the matrix representation of Gaussian elimination. Through row elementary operations, the coefficient matrix is transformed into an upper triangular matrix while recording row multipliers. Each elimination step is equivalent to left multiplication by an elementary matrix. The entire elimination process can be expressed as:

$$E_{n-1} \cdots E_2 E_1 A = U \tag{1}$$

where E_i are elementary matrices. Their inverses are easily combined, and the product of these inverses precisely forms a lower triangular matrix L.

Let $L = (E_{n-1} \cdots E_2 E_1)^{-1}$, then $A = LU$. The elements of L are the multipliers used during elimination, and L is an upper triangular matrix with a unit diagonal (diagonal entries equal to 1). However, not every matrix admits a unique LU decomposition. Below we present the conditions for the existence and uniqueness of LU decomposition.

The necessary and sufficient condition for the matrix $A \in R^{n \times n}$ to possess a unique LU decomposition (where L is a lower triangular unit matrix) is that all sequential principal submatrices of A are non-singular (i.e., their sequential principal minors are non-zero).

The proof proceeds as follows:

Necessity: For any$k(1 \le k \le n-1)$, partition A into k rows and k columns:

$$A = \begin{pmatrix} A_k & * \\ * & * \end{pmatrix}, \quad L = \begin{pmatrix} L_k & 0 \\ * & 1 \end{pmatrix}, \quad U = \begin{pmatrix} U_k & * \\ 0 & * \end{pmatrix} \quad (2)$$

where A_k, L_k, U_k are the k×k sequential principal submatrices of A, L, and U respectively. By block matrix multiplication: $A_k = L_k U_k$. Since L is a lower triangular unit matrix, L_k is also a lower triangular unit matrix, hence $det(L_k) = 1$. As U is an upper triangular matrix, U_k is also an upper triangular matrix, whose determinant is the product of its diagonal elements:$det(U_k) = \prod_{i=1}^{k} u_{ii}$.

Thus we obtain: $\det(A_k) = \det(L_k)\det(U_k) = 1 \cdot \prod_{i=1}^{k} u_{ii} \neq 0$. Since LU decomposition is unique, the diagonal entries of U, u_{ii}, are all non-zero (otherwise U would be non-invertible, rendering the decomposition non-unique or non-existent). Therefore, $\det(A_k) \neq 0$. Given the arbitrariness of k, all sequential principal minors are non-zero.

Sufficiency: Constructing the decomposition by induction

When n=1, let $A = [a_{11}]$. By the conditions, we have $a_{11} \neq 0$. Taking $L = [1]$ and $U = [a_{11}]$, we obtain $A = LU$, and the decomposition is unique.

Suppose that for any n-1 matrix, if all its sequential principal minors are non-zero, then a unique LU decomposition exists.

Partition A into blocks:

$$A = \begin{pmatrix} A_{n-1} & b \\ c^{\top} & d \end{pmatrix} \quad (3)$$

where A_{n-1} is the (n-1)th sequential principal submatrix of A, given by the conditiondet $(A_{n-1}) \neq 0$.

By the inductive hypothesis, A_{n-1} possesses a unique LU decomposition:

$$A_{n-1} = L_{n-1}U_{n-1} \quad (4)$$

where L_{n-1} is the unit lower triangular matrix and U_{n-1} is the upper triangular matrix.

Construct candidate forms for L and U:

$$L = \begin{pmatrix} L_{n-1} & 0 \\ l^{\top} & 1 \end{pmatrix}, U = \begin{pmatrix} U_{n-1} & u \\ 0^{\top} & v \end{pmatrix} \quad (5)$$

where $l, u \in \mathbb{R}^{n-1}$ and $v \in \mathbb{R}$ are to be determined. Expanding $A = LU$:

$$\begin{pmatrix} A_{n-1} & b \\ c^{\top} & d \end{pmatrix} = \begin{pmatrix} L_{n-1}U_{n-1} & L_{n-1}u \\ l^{\top}U_{n-1} & l^{\top}u + v \end{pmatrix} \quad (6)$$

Comparing both sides yields the following system of equations:

$$A_{n-1} = L_{n-1}U_{n-1} \quad (7)$$

$$b = L_{n-1}u \quad (8)$$

$$c^{\top} = l^{\top}U_{n-1} \quad (9)$$

$$d = l^{\top}u + v \quad (10)$$

From (8), L_{n-1} is invertible(as a lower triangular matrix), hence $u = L_{n-1}^{-1}b$.

From (9), we obtain U_{n-1} is invertible(since $\det(U_{n-1}) = \det(A_{n-1}) \neq 0$), hence $l^{\top} = c^{\top}U_{n-1}^{-1}$, i.e.$l = U_{n-1}^{-\top}c$.

From (10), it follows readily that $v = d - l^{\top}u$. The above construction yields L and U satisfying $A = LU$, where L is lower triangular and U is upper triangular. Hence the decomposition exists. Having established the existence of the decomposition, we must now prove its uniqueness: Suppose A has two LU decompositions $A = L_1U_1 = L_2U_2$, where L_1, L_2 is the lower unit triangular matrix and U_1, U_2 is the upper unit triangular matrix. Then $L_2^{-1}L_1 = U_2U_1^{-1}$. The left-hand side $L_2^{-1}L_1$ is the product of lower unit triangular matrices, which remains a lower unit triangular matrix. The right-hand side $U_2U_1^{-1}$ is the product of upper unit triangular matrices, which remains an upper unit triangular matrix. Therefore, $L_2^{-1}L_1 = U_2U_1^{-1}$ must be both a lower unit triangular matrix and an upper unit triangular matrix, and can only be the unit matrix, i.e.:

$$L_2^{-1}L_1 = I \implies L_1 = L_2, U_2U_1^{-1} = I \implies U_1 = U_2 \quad (11)$$

Hence the decomposition is unique.

The LU decomposition method decomposes an invertible matrix into the product of an lower triangular matrix L and an upper triangular matrix U. Its classical theory requires all sequential principal minors to be non-zero to ensure the uniqueness of the decomposition. However, in practical numerical computations, this condition is rather stringent. Typically, partial pivot selection techniques introduce row swaps to yield the more general PLU decomposition (PA=LU), which applies to all invertible matrices and effectively enhances numerical stability. For singular or ill-conditioned matrices, LU decomposition may fail or become unstable. In such cases, alternative methods such as QR decomposition or singular value decomposition are recommended. Consequently, LU decomposition serves as a core tool for solving linear equations and matrix inversion problems, though its practical applicability requires integration with row echelon equivalence techniques.

2.2. Cholesky Decomposition

Cholesky decomposition is a specialized triangularization method for positive definite symmetric matrices. It decomposes the matrix A into the product of a lower triangular matrix L and its transpose, namely:

$$A = LL^T \quad (12)$$

where$A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix (satisfying $A = A^T$ and possessing all non-zero eigenvalues); L is a lower triangular matrix with positive diagonal elements.

Since A is symmetric and positive definite, all its sequential principal submatrices are positive definite. This guarantees the existence of the Cholesky decomposition, and when the diagonal elements of L are specified to be positive, the decomposition is unique.

Let $L = (l_{ij})$, where $l_{ij} = 0\ (i < j)$. Expanding the elements according to $A = LL^T$:

$$a_{ij} = \sum_{k=1}^{\min(i,j)} l_{ik}\, l_{jk}, 1 \leq i, j \leq n \quad (13)$$

Diagonal elements ($i = j$):

$$a_{ii} = \sum_{k=1}^{i} l_{ik}^2 \Rightarrow l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2} \quad (14)$$

Since A is positive definite, the expression under the square root is always positive. Off-diagonal elements ($i > j$):

$$a_{ij} = \sum_{k=1}^{j} l_{ik}\, l_{jk} \Rightarrow l_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} l_{ik}\, l_{jk}}{l_{jj}}, i > j \quad (15)$$

During the computation, calculate column by column starting from the first column:L. First, set L to the zero matrix. For each column $j = 1,2,\dots,n$, compute the diagonal elements:

$$l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2} \quad (16)$$

Compute the elements below this column ($i = j + 1, \dots, n$):

$$l_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} l_{ik}\, l_{jk}}{l_{jj}} \quad (17)$$

The Cholesky decomposition is applicable for the triangular decomposition of symmetric positive definite matrices, with the core condition being that the matrix must be symmetric and possess only positive eigenvalues, i.e., positive definiteness. This decomposition represents the matrix as the product of a lower triangular matrix and its transpose, offering computational efficiency (requiring approximately half the computational effort of LU decomposition), numerical stability (typically without the need for principal element selection), and storage economy (requiring storage of only half the elements). It finds extensive application in solving linear equations, Kalman filtering, covariance matrix processing, and Monte Carlo simulations. Provided the conditions of symmetry and positive definiteness are satisfied, computational performance can be further enhanced through block-wise parallelization and parallelization designs (such as FPGA or GPU implementations).

2.3. Singular Value Decomposition (SVD) Method

For any matrix$A \in \mathbb{C}^{m\times n}$(or$\mathbb{R}^{m\times n}$), there exist orthogonal (or unitary) matrices$U \in \mathbb{C}^{m\times m}$ and$V \in \mathbb{C}^{n\times n}$, along with a non-negative diagonal matrix$\Sigma \in \mathbb{R}^{m\times n}$, such that:

$$A = U\Sigma V^* \tag{18}$$

where $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$ and $p = \min(m, n)$; $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$is termed the singular value; the column vectors $\{u_i\}$ of U form the left singular vectors, constituting an orthogonal basis for the column space A; the column vectors$\{v_i\}$ of V form the right singular vectors, constituting an orthogonal basis for the row space$A$. Existence theorem: Any matrix possesses an SVD decomposition, and when singular values are arranged in descending order, the decomposition is unique. Below, we shall further elucidate the mathematical principles underpinning the SVD decomposition method.

The mathematical principles of singular value decomposition(SVD) are grounded in spectral theory within linear algebra. It decomposes any matrix into the product of three structurally simple matrices:

$$A = U\Sigma V^* \tag{19}$$

Its core concept involves constructing the decomposition through the characteristic analysis of the symmetric matricesA^*A and AA^*: The singular values σ_i are the non-negative square roots of the eigenvalues of A^*A, the left singular vectors u_i are the unit eigenvectors of AA^*, and the right singular vectors v_i are the unit eigenvectors of A^*A. These singular values originate from the eigenvalues of the symmetric, positive semidefinite matrix A^*A:

$$\sigma_i = \sqrt{\lambda_i(A^*A)}, i = 1, \dots, p \tag{20}$$

Moreover, the left singular vector u_i is an eigenvector of AA^*; the right singular vector v_i is an eigenvector of A^*A.

This decomposes any linear transformation into a geometric process of rotation (or reflection) – scaling – rotation. The Eckart-Young theorem further endows it with explicit approximation meaning: Let the rank of A be$r$. Then its optimal rank approximation $k(k < r)$ is given by:

$$A_k = \sum_{i=1}^{k} \sigma_i\, u_i v_i^T \tag{21}$$

with the error satisfying:

$$\| A - A_k \|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}, \| A - A_k \|_2 = \sigma_{k+1} \tag{22}$$

$\mathrm{rank}(A)$ is the number of non-zero singular values. The null space is $\mathcal{N}(A) = \mathrm{span}\{v_{r+1}, \dots, v_n\}$. $\| A \|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$;$\| A \|_2 = \sigma_1$.

$$A^\dagger = V\Sigma^\dagger U^*, \Sigma^\dagger = \mathrm{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1}, 0, \dots, 0) \tag{23}$$

The matrix reconstructed by retaining the first$k$singular values, $A_k = \sum_{i=1}^{k} \sigma_i\, u_i v_i^T$, constitutes an optimal rank approximation k under both the Frobenius norm and spectral norm, with the error determined by the sum of squares of the discarded singular values.

Singular value decomposition (SVD) exhibits remarkable universality, being applicable to any real or complex matrix regardless of whether it is square, full-rank, or invertible. It imposes no symmetry or positive-definiteness requirements and handles ill-conditioned or rank-deficient scenarios stably. Through the analysis of the matrix AA and its eigenvectors, SVD extracts singular values and an orthogonal basis representing the data structure. Consequently, it finds core applications in data compression, dimensionality reduction (e.g., PCA), noise filtering, recommendation systems, image processing, and least-squares solutions. Its decomposition results exhibit excellent numerical stability and clear geometric interpretation, making it particularly suitable for feature extraction and model interpretability construction in high-dimensional data analysis, signal processing, and machine learning.

3. Recent Research on Applications of Matrix Decomposition Methods

Literature reviews indicate that matrix decomposition methods demonstrate invaluable utility across multiple scientific domains by effectively separating and extracting key latent information from high-dimensional, sparse, or noisy complex observational data. In recent years, notable progress has been achieved in hyperspectral remote sensing image processing[1]. Among these, non-negative matrix factorization has emerged as a tool for feature extraction in remote sensing imagery, owing to its ability to represent partial components within a holistic framework. This approach decomposes image data into a low-rank basis matrix with explicit physical significance and its corresponding coefficient matrix, enabling precise reconstruction of surface information within a low-rank representation framework. This technical pathway offers solutions for enhancing the interpretability of remote sensing imagery, demonstrating strong potential for practical applications. Presently, the research frontier of matrix decomposition is expanding from traditional domains into emerging interdisciplinary directions. Within bioinformatics and geophysical detection, focus has shifted towards novel challenges such as predicting disease associations with long non-coding RNAs (LncRNAs) [2] and suppressing ground-penetrating radar (GPR) clutter[3]. Concurrently, low-rank sparse modelling and high-order/global feature learning have become key development priorities. Within this trend, research activity has intensified on advanced methods integrating sophisticated regularization strategies[4] and tensor theory[5], such as SDPRMF, HP-TRMF, and Schatten p-TRPCA. These approaches have yielded notable outcomes, particularly in disease mechanism exploration and precise underground target identification. Nevertheless, existing methods still face significant challenges. In biological association prediction, models often become trapped in local optima or overlook global nonlinear relationships, thereby limiting their ability to predict novel associations. In signal processing, traditional algorithms exhibit high computational complexity and insufficient adaptability to complex field scenarios, severely hindering real-time processing and engineering deployment efficiency. These common issues constitute the primary bottlenecks in the practical application of matrix factorization.

To overcome these limitations and propel matrix factorization from an effective data analysis tool to a precise intelligent computing engine, deep technological integration is required. The specific pathway manifests across three dimensions: In terms of interpretability, formalizing domain knowledge (such as physical laws or biological constraints) into regularization terms within the loss function[6]—such as physical information constraints and structured sparsity constraints—can guide models to learn feature representations with explicit real-world significance; In computational efficiency, integrating randomized algorithms with online learning frameworks and hardware architectures tailored for specific computations holds promise for real-time processing of massive datasets and significant energy efficiency gains; For complex pattern modelling, constructing a unified framework capable of capturing dynamic system evolution and cross-modal correlations requires organically integrating: the powerful nonlinear representation capabilities of deep learning[7] with the structural description advantages of tensor theory for high-dimensional correlated data. The synergistic advancement of these three dimensions will propel matrix factorization technology forward, establishing it as a core component for next-generation, multifunctional scientific computing models. Once breakthroughs are achieved in these critical technical bottlenecks, matrix decomposition will evolve from a static data analysis tool into an intelligent engine capable of dynamically simulating and even intervening in complex systems. This leap in capability will extend to other potential applications. For instance, in neuroscience and biomedical engineering, dynamic tensor decomposition techniques could be employed to construct computational models of individual brain functions based on multimodal neural data streams, enabling layer-by-layer analysis of cognitive and emotional states. Integrated with specialized computing chips, this system is projected to identify millisecond-level activity in specific pathological neural circuits. Through closed-loop neuromodulation devices, it could deliver personalized precision interventions, thereby advancing diagnostic and therapeutic approaches for neuropsychiatric disorders. This marks a shift in matrix decomposition technology's application—from serving cognitive understanding of the objective world to directly participating in the monitoring and regulation of complex biological processes.

4. Conclusion

This paper systematically reviews the core theories and cutting-edge advancements in matrix decomposition. It begins with an in-depth analysis of the mathematical principles, algorithmic characteristics, and applicability of LU, QR, Cholesky, singular value decomposition (SVD), and full-rank decomposition, clearly outlining the respective advantages and limitations of each method. Subsequently, the paper focuses on interdisciplinary fields such as hyperspectral remote sensing, bioinformatics, and geophysical exploration. It reviews breakthroughs and value in emerging applications represented by techniques like non-negative matrix factorization and low-rank sparse modelling, while identifying current challenges in model interpretability, computational efficiency, and complex pattern modelling alongside corresponding solutions. Finally, the paper outlines a developmental trajectory for advancing matrix decomposition techniques into intelligent computational engines through the integration of regularization constraints, tensor theory, and deep learning. This provides a clear reference application process for academic research and technological implementation across relevant fields.

References

[1] Yang Yutong, Yi Qi, Sun Zhengbao, et al. Research progress on matrix factorization methods. J. Yunnan Univ. Nat. Sci. Ed., 2025, 47(4): 635–647..

[2] Li Dayin. Matrix decomposition-based lncRNA-disease prediction method. Master’s thesis. Guangdong University of Technology, 2025.

[3] Zhou Hao. Research on GPR noise suppression method based on low-rank sparse matrix factorisation. Master’ s thesis. Inner Mongolia University of Science and Technology, 2025.

[4] Chen Yaru, Wang Xiaojie, Li Qing, et al. Generative adversarial networks with information distillation based on supervised contrastive regularisation terms. J. Tianjin Univ. Sci. Tech., 2025, 40(2): 61–70.

[5] Zhang Qi. Research on Multi-Parameter Joint Estimation Methods for MIMO Radar Targets Based on Tensor Theory [D]. Jilin University, 2025. DOI:10.27162/d.cnki.gjlin.2025.000267.

[6] Zhang Bo, Liu Genyou, Zhao Wenhao, et al. Application of adaptive robust loss functions in multi-sensor fusion positioning. GNSS World of China, 2025, 50(4): 68–77.

[7] Fei Yiming. Trajectory tracking control for uncertain nonlinear systems based on neural network learning control. Master’s thesis. Harbin Institute of Technology, 2023.