International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 6.4 Website: https://ijarmt.com ISSN No.: 3048-9458

A Review Study of Eigenvalue–Eigenvector Methods in Covariance Matrices and Dimensionality Reduction

Shisode Nikita Vijaysing
Research Scholar, Department of Mathematics, Malwanchal University, Indore
Dr. Shoyeb Ali Sayyed
Supervisor, Department of Mathematics, Malwanchal University, Indore

Abstract
Eigenvalue–eigenvector decomposition of covariance matrices lies at the heart of modern statistical analysis, machine learning, and signal processing. This review paper provides a comprehensive survey of the mathematical foundations, computational algorithms, and practical applications of eigenvalue–eigenvector methods in the context of covariance matrices and dimensionality reduction. We systematically trace the development of Principal Component Analysis (PCA) from its inception by Karl Pearson in 1901 to its modern extensions, including Kernel PCA, Sparse PCA, Robust PCA, and Randomized PCA. The review covers the spectral theorem, singular value decomposition (SVD), and their connections to covariance structure learning. Algorithmic approaches—including power iteration, the QR algorithm, Lanczos methods, and randomized numerical linear algebra—are evaluated for their computational complexity and numerical stability. We further examine applications in face recognition, natural language processing, genomic data analysis, image compression, and graph-based learning. Comparative analysis of algorithms across varying data dimensionalities and sample sizes is provided. The paper concludes by identifying open research challenges and emerging directions, including eigenvalue estimation in the high-dimensional regime, federated PCA, and quantum-accelerated decomposition methods.
Keywords: Eigenvalue decomposition; Covariance matrix; Principal Component Analysis; Dimensionality reduction; Singular Value Decomposition; Spectral methods; High-dimensional statistics; Randomized linear algebra.

1. Introduction

The proliferation of high-dimensional data across scientific disciplines—ranging from genomics and neuroimaging to financial modelling and computer vision—has placed dimensionality reduction at the forefront of data science research. Among the most fundamental and extensively studied methods in this domain are those based on the eigenvalue–eigenvector decomposition of covariance matrices. These techniques provide a rigorous mathematical framework for identifying the principal axes of variation within data, enabling compact representation, noise suppression, feature extraction, and latent structure discovery.

Eigenvalue problems have occupied a central role in applied mathematics since the eighteenth century, with contributions from Euler, Lagrange, Cauchy, and Sylvester laying the theoretical groundwork. The landmark reformulation by Karl Pearson (1901) introduced Principal Component Analysis (PCA) as a method to find the "line of closest fit" to a system of points

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 6.4 Website: https://ijarmt.com ISSN No.: 3048-9458

in high-dimensional space, a framework later extended by Harold Hotelling (1933) to the correlation matrix setting. Since then, the field has evolved dramatically, driven by advances in numerical linear algebra, statistical theory, and computational hardware.

The importance of eigenvalue–eigenvector methods extends well beyond PCA. Factor analysis, canonical correlation analysis (CCA), linear discriminant analysis (LDA), spectral graph theory, and quantum mechanics all rely on spectral decomposition as a foundational operation. In machine learning, the covariance matrix and its spectral properties underpin algorithms for anomaly detection, data whitening, manifold learning, and neural network analysis.

Despite the maturity of this field, several challenges persist. The classical assumption of low-dimensional signal embedded in noise becomes problematic when the number of variables p approaches or exceeds the sample size n (the high-dimensional regime). Moreover, outlier contamination, non-Gaussianity, missing data, and scalability to streaming or distributed datasets pose ongoing algorithmic challenges. Recent developments in random matrix theory, compressive sensing, and randomized numerical linear algebra have opened new avenues for addressing these challenges.

This review aims to consolidate the mathematical theory, computational methods, and applied domains of eigenvalue–eigenvector methods in covariance analysis and dimensionality reduction. We provide a structured synthesis of classical and modern literature, critically evaluate algorithmic trade-offs, and chart promising directions for future research. The paper is organized as follows: Section 2 covers mathematical foundations; Section 3 reviews PCA and its variants; Section 4 discusses computational algorithms; Section 5 surveys applications; Section 6 presents comparative analysis; Section 7 outlines open challenges; and Section 8 concludes.

2. Mathematical Foundations

2.1 Covariance Matrices: Definition and Properties

Let $X = \{x_1, x_2, ..., x_n\}$ denote a dataset of n observations, where each $x_i \in \mathbb{R}^p$. The sample covariance matrix S is defined as:

$$S = (1/(n-1)) \cdot \Sigma_i (x_i - \bar{x})(x_i - \bar{x})^T \in \mathbb{R}^{p\times p}$$

where $\bar{x} = (1/n) \Sigma_i x_i$ is the sample mean vector. The covariance matrix S is symmetric positive semi-definite (SPSD): it satisfies $S = S^T$ and $v^TSv \geq 0$ for all $v \in \mathbb{R}^p$. These properties are fundamental because they guarantee real, non-negative eigenvalues and a complete set of orthogonal eigenvectors.

The population covariance matrix $\Sigma = E[(X - \mu)(X - \mu)^T]$ encodes the linear correlation structure of the data-generating distribution. Estimating Σ reliably from finite samples is a core problem in multivariate statistics, with sample complexity heavily dependent on the ratio $\gamma = p/n$. When $\gamma \to 0$ (classical regime), standard results apply; when $\gamma \to c > 0$ (high-dimensional regime), classical estimators break down and random matrix theory becomes indispensable.

2.2 The Spectral Theorem

The spectral theorem for real symmetric matrices states that any symmetric matrix $S \in \mathbb{R}^{p\times p}$ admits the eigendecomposition:

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 6.4 Website: https://ijarmt.com ISSN No.: 3048-9458

$$S = Q \Lambda Q^T$$

where $Q = [q_1 \mid q_2 \mid ... \mid q_p] \in \mathbb{R}^{p \times p}$ is an orthogonal matrix whose columns are the eigenvectors of S, and $\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_p)$ is the diagonal matrix of corresponding real eigenvalues, conventionally ordered such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$. The eigenpair (λ_i, q_i) satisfies $Sq_i = \lambda_i q_i$, with $q_i^T q_j = \delta_{ij}$ (the Kronecker delta). The eigenvalues represent the variance explained along each principal direction, while the eigenvectors define the orthogonal axes of the transformed coordinate system.

2.3 Singular Value Decomposition (SVD)

Closely related to eigendecomposition, the SVD of the (centered) data matrix $X \in \mathbb{R}^{n \times p}$ is:

$$X = U \Sigma V^T$$

where $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{p \times p}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{n \times p}$ is a rectangular diagonal matrix of singular values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$ ($r = \min(n, p)$). The connection to eigendecomposition is explicit: the right singular vectors V are the eigenvectors of $X^TX = (n-1)S$, and the singular values satisfy $\sigma_i^2 = (n-1)\lambda_i$. SVD provides a numerically superior pathway to PCA compared to forming the covariance matrix explicitly, particularly for ill-conditioned or large-scale problems.

2.4 The Rayleigh Quotient and Variational Characterization

The leading eigenvector q_1 of S solves the constrained optimization:

$$q_1 = \text{argmax}\_\{\|v\|=1\} \; v^TSv = \text{argmax}\_\{\|v\|=1\} \; R(v, S)$$

where $R(v, S) = v^TSv / v^Tv$ is the Rayleigh quotient. The maximum value equals λ_1. Subsequent components are found by projecting out previously found directions (deflation). This variational characterization is the basis for power iteration, inverse iteration, and related eigensolver algorithms, and also underpins the theoretical analysis of PCA consistency and perturbation bounds (Davis–Kahan theorem).

3. Principal Component Analysis and Its Extensions

3.1 Classical PCA

PCA seeks a low-dimensional linear subspace that maximally preserves the variance of the original data. Given the eigendecomposition $S = Q\Lambda Q^T$, the first k principal components are defined as the projections $z = Q_k^T x$, where $Q_k = [q_1 \mid ... \mid q_k]$ comprises the top-k eigenvectors. The proportion of total variance retained by the k-dimensional projection is:

$$PVE(k) = (\Sigma_{i=1}^{k} \lambda_i) / (\Sigma_{i=1}^{p} \lambda_i)$$

PCA has been studied extensively in terms of consistency (Ahn & Horenstein, 2013), optimality under Gaussian assumptions (Anderson, 1963), and connections to information geometry. The choice of k is typically guided by scree plot analysis, cross-validation, or by thresholding cumulative PVE (e.g., retaining 95% of variance). PCA's key limitation is its sensitivity to outliers and its restriction to linear subspaces.

3.2 Kernel PCA

Kernel PCA (Schölkopf et al., 1998) extends PCA to nonlinear feature spaces via the kernel trick. Given a positive semi-definite kernel function $\kappa(x_i, x_j) = \langle\varphi(x_i), \varphi(x_j)\rangle$, the kernel matrix $K \in \mathbb{R}^{n \times n}$ with $K_{ij} = \kappa(x_i, x_j)$ is centred and decomposed spectrally. The principal components in feature space are extracted without explicit computation of φ. Common kernels include the

Gaussian (RBF), polynomial, and sigmoid functions. Kernel PCA has proven effective for manifold learning, texture analysis, and bioinformatics, though it inherits $O(n^2)$ memory and $O(n^3)$ runtime costs for dense kernels.

3.3 Sparse PCA

Classical PCA eigenvectors are dense, complicating interpretation in high-dimensional settings. Sparse PCA (Zou et al., 2006; d'Aspremont et al., 2007) introduces sparsity penalties on the loading vectors, yielding components that depend on a small subset of original variables. Formulations include LASSO-penalized regression (SCoTLASS), semidefinite programming relaxations, and greedy forward-selection algorithms. Sparse PCA has been particularly impactful in genomics, where identifying sparse gene expression signatures is scientifically meaningful.

3.4 Robust PCA

Robust PCA (Candès et al., 2011) decomposes a matrix M into a low-rank component L and a sparse corruption component S: $M = L + S$. This is formulated as a convex optimization problem (Principal Component Pursuit) solvable via alternating direction method of multipliers (ADMM). Theoretical recovery guarantees hold under incoherence conditions on L and sparsity conditions on S. Applications include background subtraction in video surveillance, face recognition under occlusion, and anomaly detection in network traffic.

3.5 Randomized and Incremental PCA

For large-scale datasets, exact eigendecomposition of S is computationally prohibitive ($O(p^3)$ time, $O(p^2)$ storage). Randomized PCA (Halko et al., 2011) leverages random projections to compute an approximate rank-k SVD in $O(npk)$ time, with provable spectral approximation guarantees. Incremental PCA (Ross et al., 2008) processes data in mini-batches, updating the subspace estimate without storing the full dataset. These methods enable PCA on streaming, distributed, or privacy-constrained data and have been implemented in major machine learning libraries including scikit-learn, Apache Spark, and TensorFlow.

4. Computational Algorithms for Eigenvalue Decomposition

4.1 Power Iteration

Power iteration is the conceptually simplest eigensolver. Starting from a random vector v_0, the update $v_{k+1} = Sv_k / \|Sv_k\|$ converges to the dominant eigenvector at rate $|\lambda_2/\lambda_1|^k$. While simple and memory-efficient, it is slow when the eigen-gap $(\lambda_1 - \lambda_2)$ is small, and extensions (deflation, inverse iteration, Rayleigh quotient iteration) are needed for computing multiple eigenpairs.

4.2 The QR Algorithm

The QR algorithm (Francis, 1961; Kublanovskaya, 1961) remains the workhorse of dense eigensolvers. It iteratively applies QR factorization: $A_k = Q_kR_k \rightarrow A_{k+1} = R_kQ_k$. With implicit shifts (e.g., Wilkinson shifts) and Householder reduction to Hessenberg form as preprocessing, the algorithm achieves cubic convergence for simple eigenvalues. LAPACK's DSYEV routine, based on the tridiagonal QR algorithm, is the standard reference implementation for symmetric matrices and underlies most scientific computing workflows.

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 6.4 Website: https://ijarmt.com ISSN No.: 3048-9458

4.3 Lanczos and Arnoldi Methods

For large sparse or structured matrices, Krylov subspace methods are preferred. The Lanczos algorithm (Lanczos, 1950) constructs a tridiagonal matrix T_m whose eigenvalues (Ritz values) approximate the extreme eigenvalues of S. With $m \ll p$ iterations, extreme eigenvalues converge rapidly. The implicitly restarted Lanczos method (IRLM), implemented in ARPACK, is the standard for large-scale symmetric eigenproblems. The Arnoldi process generalises this to non-symmetric matrices. Lanczos methods require only matrix-vector products, making them suitable for implicit covariance representations.

4.4 Randomized Numerical Linear Algebra

The framework of randomized algorithms (Mahoney, 2011; Woodruff, 2014) has revolutionised large-scale eigendecomposition. The randomized SVD algorithm proceeds in two stages: (1) form a sketch $Y = X\Omega$ with Ω a random Gaussian or subsampled randomized Hadamard transform (SRHT) matrix, then compute an orthonormal basis Q for the range of Y; (2) project $B = Q^T X$ and compute the SVD of the small matrix B. The result is a rank-k approximation with error bounded by $(1 + \varepsilon)\sigma_{k+1}$ in expectation. This approach achieves $O(np \log k + (n + p)k^2)$ complexity, enabling PCA on billion-scale datasets.

Table 1: Comparison of Eigenvalue Algorithms for Covariance Analysis

Algorithm	Time Complexity	Memory	Scalability	Best Use-Case
Power Iteration	$O(p^2 \cdot \text{iter})$	$O(p)$	Low	Top-1 eigenpair
QR Algorithm	$O(p^3)$	$O(p^2)$	Low	Dense, small p
Lanczos / IRLM	$O(p^2 \cdot k)$	$O(pk)$	Medium	Sparse, top-k
Randomized SVD	$O(npk)$	$O(nk + pk)$	High	Large n, p
Incremental PCA	$O(nk^2)$	$O(pk)$	High	Streaming data

Source: Compiled from Golub & Van Loan (2013), Halko et al. (2011), Saad (2011)

5. Applications of Eigenvalue–Eigenvector Methods

5.1 Face Recognition and Computer Vision

The "Eigenfaces" method (Turk & Pentland, 1991) pioneered the application of PCA to face recognition by representing face images as linear combinations of principal face images (eigenvectors of the face covariance matrix). Subsequent work combined PCA with LDA (Fisherfaces; Belhumeur et al., 1997), and kernel extensions (KernelFaces). In modern deep learning pipelines, PCA is used for whitening intermediate feature representations, initialising convolutional filters, and compressing feature descriptors.

5.2 Natural Language Processing

Latent Semantic Analysis (LSA; Deerwester et al., 1990) applies truncated SVD to the term–document matrix, revealing latent semantic structure. GloVe word embeddings exploit covariance structure of co-occurrence matrices. More recently, PCA-based post-processing of

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 6.4 Website: https://ijarmt.com ISSN No.: 3048-9458

contextual word embeddings (e.g., BERT representations) has been shown to remove anisotropy and improve downstream NLP tasks. Topic modelling methods such as Probabilistic LSA and Latent Dirichlet Allocation also draw on spectral intuitions.

5.3 Genomics and Bioinformatics

PCA has become a standard quality control and population stratification tool in genome-wide association studies (GWAS). Price et al. (2006) demonstrated that the top principal components of the genotype matrix effectively capture continental ancestry, enabling correction for population structure in regression analyses. In single-cell RNA sequencing (scRNA-seq), PCA is a key preprocessing step before clustering (e.g., Seurat pipeline), graph construction (UMAP, t-SNE), and trajectory inference (Monocle).

5.4 Image Compression

PCA-based image compression exploits the energy compaction property: for natural images, most variance concentrates in the first few principal components. Storing only the top-k eigenvectors and scores yields compression ratios proportional to $(p - k)/p$ with controlled reconstruction error. Though superseded by DCT-based standards (JPEG) for practical use, PCA compression remains important in satellite imaging, medical imaging (MRI, CT), and hyperspectral remote sensing.

5.5 Finance and Econometrics

In quantitative finance, PCA of the asset return covariance matrix reveals systematic risk factors (analogous to the Fama–French factors). The first principal component typically captures market beta; subsequent components represent sector or style factors. PCA-based covariance estimation (regularised via shrinkage, e.g., Ledoit–Wolf estimator) improves portfolio optimisation stability compared to sample covariance, particularly in the high-dimensional, small-sample regime typical of financial data.

5.6 Signal Processing and Sensor Networks

In array signal processing, the MUSIC (Multiple Signal Classification) algorithm (Schmidt, 1986) partitions the covariance matrix of observed signals into signal and noise subspaces via eigendecomposition, enabling super-resolution direction-of-arrival estimation. In sensor networks and IoT, distributed PCA algorithms allow nodes to collaboratively estimate covariance structure without centralising data, addressing privacy and communication constraints.

6. High-Dimensional Perspectives: Random Matrix Theory

When the data dimension p grows proportionally to the sample size n (i.e., $p/n \to \gamma \in (0, \infty)$), the classical theory of sample eigenvalues breaks down. The Marchenko–Pastur law (1967) describes the limiting spectral distribution of $(1/n)X^TX$ when X has i.i.d. entries with variance σ^2. The law predicts that eigenvalues spread over the interval $[\sigma^2(1-\sqrt{\gamma})^2, \sigma^2(1+\sqrt{\gamma})^2]$—a phenomenon called eigenvalue spreading—even when the population covariance is a multiple of the identity.

A critical consequence is eigenvalue upward bias: sample eigenvalues systematically overestimate population eigenvalues (Johnstone, 2001). The BBP (Baik–Ben Arous–Péché) phase transition reveals that a spike in the population covariance (a large eigenvalue θ) is

International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)
An International Open Access, Peer-Reviewed Refereed Journal
Impact Factor: 6.4 Website: https://ijarmt.com ISSN No.: 3048-9458

detectable in the sample spectrum only when $\theta > \sigma^2(1 + \sqrt{\gamma})$; below this threshold, the spiked eigenvector becomes asymptotically orthogonal to the sample eigenvector (no information recovery). These results have profound implications for the interpretation of PCA in high-dimensional settings and have motivated shrinkage estimators for eigenvalues (Ledoit & Wolf, 2004; 2012) and consistent estimators of the population spiked eigenvalues (Nadler, 2008; Donoho & Gavish, 2014).

The Spiked Covariance Model (Johnstone, 2001) assumes $\Sigma = I + \Sigma_r$, where Σ_r is low-rank. Under this model, consistent eigenvector estimation is possible only when the spike-to-noise ratio exceeds the BBP threshold. Adaptive thresholding of sample eigenvalues (hard/soft thresholding of singular values; Gavish & Donoho, 2014) provides a principled approach to estimating the rank r and the spiked eigenvectors from data. These methods are now implemented in the OptShrink and RMTstat R packages.

7. Open Challenges and Future Directions

7.1 Non-Euclidean and Manifold Settings

Classical PCA assumes data lie in Euclidean space. Extensions to Riemannian manifolds (e.g., Principal Geodesic Analysis; Fletcher et al., 2004) and matrix manifolds (e.g., the manifold of symmetric positive definite matrices, relevant for covariance estimation in brain connectivity analysis) remain active areas. Developing computationally efficient, theoretically grounded spectral methods for non-Euclidean data is an open research problem.

7.2 Federated and Privacy-Preserving PCA

Centralising sensitive data is often legally or ethically infeasible. Federated PCA algorithms (Grammenos et al., 2020) allow distributed computation of global principal components from local covariance estimates, typically via gradient-based or gossip-protocol approaches. Incorporating differential privacy guarantees (Dwork et al., 2014) into eigenvector computation—while maintaining utility—is a technically challenging and practically important open problem, especially for medical and financial applications.

7.3 Tensor and Multilinear Extensions

Real-world data are often naturally multi-way arrays (tensors): videos (pixels × pixels × frames), fMRI data (x × y × z × time), and relational datasets (users × items × context). Multilinear PCA (Lu et al., 2008) and Tucker decomposition extend eigendecomposition to higher-order arrays. Robust tensor PCA and scalable tensor decomposition algorithms are areas of active development, with challenges in uniqueness guarantees, computational complexity, and statistical theory.

7.4 Quantum Computing Perspectives

Quantum algorithms for linear algebra (HHL algorithm; Harrow et al., 2009) offer exponential speedups for certain matrix inversion and eigenvalue problems, though the practical conditions for quantum advantage—requiring quantum RAM and fault-tolerant hardware—remain distant. Variational quantum eigensolvers (VQE) and quantum PCA (Lloyd et al., 2014) represent near-term approaches but face challenges of noise, qubit counts, and readout overhead. The eventual impact of quantum computation on covariance analysis remains a compelling open question.

7.5 Interpretability and Causality

Principal components are linear combinations of all original variables, making interpretation difficult. Sparse PCA partially addresses this, but a principled framework linking eigendecomposition to causal structure remains underdeveloped. Integrating causal discovery methods (e.g., PC algorithm, LiNGAM) with spectral analysis could yield dimensionality reductions that respect underlying data-generating mechanisms, a direction with potential impact in scientific applications.

8. Conclusion

This review has traced the development, mathematical foundations, computational methods, and applications of eigenvalue–eigenvector approaches to covariance matrices and dimensionality reduction. Starting from the spectral theorem and the variational characterization of PCA, we surveyed extensions including Kernel PCA, Sparse PCA, Robust PCA, and Randomized PCA, each addressing specific limitations of the classical formulation. Computational algorithms—from the QR algorithm to Krylov subspace methods and randomized linear algebra—were evaluated in terms of their computational complexity, memory requirements, and practical suitability.

The applications reviewed span face recognition, natural language processing, genomics, finance, signal processing, and image compression, demonstrating the extraordinary breadth of impact that eigenvalue–eigenvector methods have achieved. The high-dimensional perspective, grounded in random matrix theory, has revealed fundamental statistical limitations and motivated the development of shrinkage estimators and phase-transition-aware inference methods.

Looking forward, open challenges in federated PCA, Riemannian generalizations, tensor extensions, quantum acceleration, and causal interpretability chart a rich agenda for future research. As data dimensionality and volume continue to grow, the development of theoretically principled and computationally efficient spectral methods will remain indispensable to the advance of data science, artificial intelligence, and applied mathematics.

Eigenvalue–eigenvector decomposition of covariance matrices occupies an irreplaceable position in the mathematical toolkit of the modern data scientist and applied mathematician. A deep understanding of these methods—from their algebraic foundations to their statistical properties in high dimensions—is essential for both advancing the theory and applying it responsibly to real-world problems.

References

[1] Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. Econometrica, 81(3), 1203–1227.

[2] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. Annals of Mathematical Statistics, 34(1), 122–148.

[3] Baik, J., Ben Arous, G., & Péché, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. Annals of Probability, 33(5), 1643–1697.

[4] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE TPAMI, 19(7), 711–720.

[5] Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis? Journal of the ACM, 58(3), 1–37.

[6] d'Aspremont, A., El Ghaoui, L., Jordan, M. I., & Lanckriet, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. SIAM Review, 49(3), 434–448.

[7] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. JASIS, 41(6), 391–407.

[8] Donoho, D. L., & Gavish, M. (2014). Minimax risk of matrix denoising by singular value thresholding. Annals of Statistics, 42(6), 2413–2440.

[9] Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in TCS, 9(3–4), 211–407.

[10] Fletcher, P. T., Lu, C., Pizer, S. M., & Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE TMI, 23(8), 995–1005.

[11] Francis, J. G. F. (1961). The QR transformation: A unitary analogue to the LR transformation. Computer Journal, 4(3), 265–271.

[12] Gavish, M., & Donoho, D. L. (2014). The optimal hard threshold for singular values is $4/\sqrt{3}$. IEEE Transactions on Information Theory, 60(8), 5040–5053.

[13] Golub, G. H., & Van Loan, C. F. (2013). Matrix Computations (4th ed.). Johns Hopkins University Press.

[14] Grammenos, A., Mendoza Smith, R., Crowcroft, J., & Mascolo, C. (2020). Federated principal component analysis. NeurIPS, 33.

[15] Halko, N., Martinsson, P.-G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Review, 53(2), 217–288.

[16] Harrow, A. W., Hassidim, A., & Lloyd, S. (2009). Quantum algorithm for linear systems of equations. Physical Review Letters, 103(15), 150502.

[17] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441.

[18] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. Annals of Statistics, 29(2), 295–327.

[19] Lanczos, C. (1950). An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. Journal of Research of the National Bureau of Standards, 45(4), 255–282.

[20] Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2), 365–411.

[21] Lloyd, S., Mohseni, M., & Rebentrost, P. (2014). Quantum principal component analysis. Nature Physics, 10(9), 631–633.

[22] Lu, H., Plataniotis, K. N., & Venetsanopoulos, A. N. (2008). MPCA: Multilinear principal component analysis of tensor objects. IEEE TNN, 19(1), 18–39.
[23] Marchenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. Matematicheskii Sbornik, 72(4), 507–536.
[24] Mahoney, M. W. (2011). Randomized algorithms for matrices and data. Foundations and Trends in ML, 3(2), 123–224.
[25] Nadler, B. (2008). Finite sample approximation results for principal component analysis. Annals of Statistics, 36(6), 2791–2817.
[26] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2(11), 559–572.
[27] Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909.
[28] Ross, D. A., Lim, J., Lin, R.-S., & Yang, M.-H. (2008). Incremental learning for robust visual tracking. IJCV, 77(1–3), 125–141.
[29] Saad, Y. (2011). Numerical Methods for Large Eigenvalue Problems (Revised ed.). SIAM.
[30] Schmidt, R. O. (1986). Multiple emitter location and signal parameter estimation. IEEE Transactions on Antennas and Propagation, 34(3), 276–280.
[31] Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 10(5), 1299–1319.
[32] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1), 71–86.
[33] Woodruff, D. P. (2014). Sketching as a tool for numerical linear algebra. Foundations and Trends in TCS, 10(1–2), 1–157.
[34] Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. Journal of Computational and Graphical Statistics, 15(2), 265–286.