spectral decomposition

Author and editors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

A spectral decomposition writes a symmetric matrix as a sum of orthogonal projections onto perpendicular lines, each scaled by a real eigenvalue. It is an eigenvalue decomposition (EVD) whose eigenvectors form an orthonormal basis, so the matrix is diagonal in a rotated coordinate system. A real matrix admits one exactly when it is symmetric, and over the complex numbers the corresponding class is the normal matrices. The decomposition follows from maximizing the quadratic form $\mathbf{x}^{\top}\mathbf{A}\mathbf{x}$ over the unit sphere: a maximizer exists because the sphere is compact, symmetry makes it an eigenvector, and symmetry again makes the vectors orthogonal to it an invariant subspace, so the argument repeats one size down. The positive semi-definite (psd) property enters only at the end, where it amounts to every eigenvalue being nonnegative. The decomposition yields the spectral norm, the rank and the square root of a matrix, and it underlies principal component analysis (PCA), kernel methods and the convergence analysis of gradient descent (GD).

Definition

Stretching the plane by a factor 3 along one line, and leaving a perpendicular line untouched, turns a circle into an ellipse with its axes along those two lines. One matrix that acts this way is

$$\mathbf{A} = \begin{pmatrix} 5/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & 3/2 \end{pmatrix},$$

whose two lines sit at 30° and 120° (see Fig. 1). A spectral decomposition states that every symmetric matrix acts this way, scaling each of d perpendicular directions by its own real factor.

A spectral decomposition is an eigenvalue decomposition (EVD) whose eigenvectors are orthonormal. For a real symmetric matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$ it reads

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top} = \sum_{j=1}^{d} \lambda_j \mathbf{u}^{(j)} \big(\mathbf{u}^{(j)}\big)^{\top}, \tag{1}$$

with an orthogonal matrix $\mathbf{U} = \big(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(d)}\big)$ of eigenvectors and real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d$ on the diagonal of $\mathbf{\Lambda}$. The converse holds as well, since

$\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ equals its own transpose: a real matrix with an orthonormal basis of eigenvectors is symmetric. That equivalence is known as the spectral theorem (Axler, 2024, 7.29). Over the complex numbers the same statement holds for a normal matrix, with the conjugate transpose in place of the transpose and a unitary matrix in place of an orthogonal matrix, and the eigenvalues are real exactly when $\mathbf{A}$ is Hermitian.

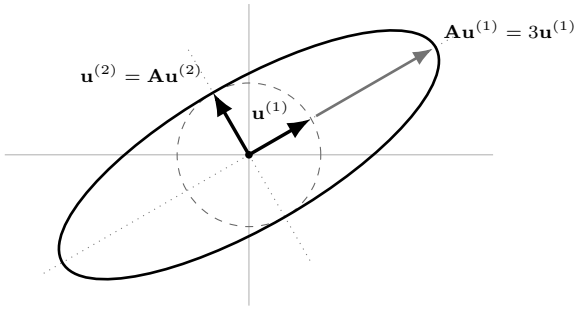

Figure 1: The matrix $\mathbf{A}$ of the display above maps the dashed unit circle onto the solid ellipse. The dotted lines carry the two eigenvectors $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$, at 30° and 120°, and the ellipse has its axes along them with half-lengths $\lambda_1 = 3$ and $\lambda_2 = 1$. Neither direction is turned into another: $\mathbf{A}$ stretches along $\mathbf{u}^{(1)}$ and leaves $\mathbf{u}^{(2)}$ where it is

One derivation of the symmetric case rests on a single analytic fact and produces the eigenvectors instead of assuming that they exist. It starts from the quadratic form $\mathbf{x}^\top\mathbf{A}\mathbf{x}$ restricted to the unit sphere $\mathbb{S}^{(d-1)} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\}$, where it coincides with the Rayleigh quotient $\mathbf{x}^\top\mathbf{A}\mathbf{x}/\mathbf{x}^\top\mathbf{x}$ (Meyer, 2000, Sect. 7.5, Eq. (7.5.4)). The sphere is closed and bounded, hence compact, and the quadratic form is continuous, so it attains a maximum on the sphere at some $\mathbf{u}^{(1)}$ (Rudin, 1976, Thm. 4.16). Write λ_1 for that maximum. This is the whole analytic input, and the remaining steps are algebra.

The maximizer is an eigenvector. Take a vector $\mathbf{v}$ of unit norm orthogonal to $\mathbf{u}^{(1)}$, so that the curve $\mathbf{x}(t) = \big(\mathbf{u}^{(1)} + t\mathbf{v}\big)/\sqrt{1+t^2}$ stays on the sphere, and abbreviate $c = \big(\mathbf{u}^{(1)}\big)^\top\mathbf{A}\mathbf{v}$ and $e = \mathbf{v}^\top\mathbf{A}\mathbf{v}$. Along that curve the quadratic form is $f(t) = \big(\lambda_1 + 2tc + t^2e\big)/\big(1+t^2\big)$, a differentiable function with a maximum at $t = 0$, so $f'(0) = 2c = 0$. Symmetry enters here: $c = \big(\mathbf{A}\mathbf{u}^{(1)}\big)^\top\mathbf{v}$, so the vector $\mathbf{A}\mathbf{u}^{(1)}$ is orthogonal to every vector orthogonal to $\mathbf{u}^{(1)}$, and is therefore a multiple of $\mathbf{u}^{(1)}$. Multiplying $\mathbf{A}\mathbf{u}^{(1)} = \lambda\mathbf{u}^{(1)}$ by $\big(\mathbf{u}^{(1)}\big)^\top$ identifies $\lambda = \lambda_1$.

The argument then repeats on a smaller matrix. The subspace $\mathcal{W}$ of vectors orthogonal to $\mathbf{u}^{(1)}$ is invariant under $\mathbf{A}$, because $\big(\mathbf{u}^{(1)}\big)^\top\mathbf{A}\mathbf{w} = \big(\mathbf{A}\mathbf{u}^{(1)}\big)^\top\mathbf{w} = \lambda_1\big(\mathbf{u}^{(1)}\big)^\top\mathbf{w} = 0$ for every $\mathbf{w} \in \mathcal{W}$. Symmetry is used a second time, and this is where a general matrix fails: for $\mathbf{N} = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ the eigenvector $(1,0)^\top$ spans an invariant line, yet $\mathbf{N}(0,1)^\top = (1,0)^\top$ leaves the orthogonal complement,

which is why a general matrix need not be diagonalizable at all. Collecting an orthonormal basis of $\mathcal{W}$ into a matrix $\mathbf{Q} \in \mathbb{R}^{d\times(d-1)}$ and setting $\mathbf{M} = \mathbf{Q}^\top\mathbf{A}\mathbf{Q}$ gives a symmetric matrix one size smaller, so induction supplies an orthonormal basis of its eigenvectors, and $\mathbf{Q}$ lifts them back to eigenvectors of $\mathbf{A}$ orthogonal to $\mathbf{u}^{(1)}$. Each later maximum is taken over a smaller set, so the eigenvalues arrive in decreasing order.

Nonnegativity of the eigenvalues is a separate matter, and it is where the positive semi-definite (psd) property enters. Symmetry is an assumption rather than a consequence of $\mathbf{x}^\top\mathbf{A}\mathbf{x} \geq 0$: a skew-symmetric matrix $\mathbf{S}$ with $\mathbf{S}^\top = -\mathbf{S}$ satisfies $\mathbf{x}^\top\mathbf{S}\mathbf{x} = 0$ for every $\mathbf{x}$, so the quadratic form cannot see any skew part. For a symmetric matrix, expanding a vector in the eigenvector basis gives $\mathbf{x}^\top\mathbf{A}\mathbf{x} = \sum_{j=1}^{d} \lambda_j \big(\big(\mathbf{u}^{(j)}\big)^\top\mathbf{x}\big)^2$, which is nonnegative for every $\mathbf{x}$ exactly when every eigenvalue is. Each $\mathbf{u}^{(j)}\big(\mathbf{u}^{(j)}\big)^\top$ in (1) is the orthogonal projection onto a line, and these projections annihilate each other and sum to the identity, so a symmetric matrix resolves a vector along d perpendicular axes, scales each component, and adds the pieces back. The eigenvalues are determined by $\mathbf{A}$, while individual eigenvectors are not when an eigenvalue repeats: any orthonormal basis of the corresponding subspace serves, and what is determined is the orthogonal projection onto that subspace.

Several quantities are immediate from (1). The spectral norm is λ_1, by the same maximization the derivation started from, while $\operatorname{tr}(\mathbf{A})$ and $\det\mathbf{A}$ are the sum and the product of the eigenvalues. The rank of $\mathbf{A}$ counts the nonzero eigenvalues, and the eigenvectors belonging to the zero ones span the nullspace. A psd matrix has the square root $\mathbf{A}^{1/2} = \sum_{j=1}^{d} \sqrt{\lambda_j}\mathbf{u}^{(j)}\big(\mathbf{u}^{(j)}\big)^\top$, which is the one psd matrix squaring to $\mathbf{A}$, so $\mathbf{A} = \mathbf{B}^\top\mathbf{B}$ for $\mathbf{B} = \mathbf{A}^{1/2}$ (Axler, 2024, 7.39). Running the maximization over subspaces rather than one deflation at a time characterizes every eigenvalue the same variational way (Meyer, 2000, Sect. 7.5, Eq. (7.5.5)).

Machine learning (ML) meets symmetric matrices wherever a quadratic form appears. The sample covariance matrix of a dataset is symmetric and psd, and principal component analysis (PCA) keeps the eigenvectors of its largest eigenvalues, so the variational step above is what makes the first principal direction the direction of largest variance. A kernel matrix is psd by construction, and its square root exhibits it as a Gram matrix of feature vectors, which is what lets a kernel method work without those feature vectors. The training objective of linear regression has the constant Hessian $(2/m)\mathbf{X}^\top\mathbf{X}$, whose eigenvalues decide how fast gradient descent (GD) converges, and the Laplacian matrix of an undirected graph is symmetric and psd, which is what spectral clustering decomposes.

Two remarks on scope. Replacing the transpose by the conjugate transpose carries the derivation verbatim to a Hermitian matrix, for which the quadratic form $\mathbf{x}^H\mathbf{A}\mathbf{x}$ is real. In infinite dimensions the argument breaks at exactly one point, the existence of the maximizer, because the unit sphere of an infinite-dimensional Hilbert space is not compact; assuming instead that the operator is compact and self-adjoint restores that step and the rest of the derivation runs

unchanged (Young, 1988).

Cross References

EVD, normal matrix, symmetric matrix, eigenvector, eigenvalue, psd, orthonormal, orthogonal, orthogonal projection, spectrum, spectral norm, PCA, sample covariance matrix, spectral clustering

References

1. Axler (2024). Linear Algebra Done Right. 4th ed., Springer, Cham, Switzerland.

2. Meyer (2000). Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia, PA, USA.

3. Rudin (1976). Principles of Mathematical Analysis. 3rd ed., McGraw-Hill, New York, NY, USA.

4. Young (1988). An Introduction to Hilbert Space. Cambridge Univ. Press, New York, NY, USA.