Two-Dimensional Linear Maps, Spectral Geometry, and Singular Value Decomposition

N-20260303

§1.1 Concept map of the chapter

The source note begins with a map of the main linear-algebra concepts used later in the machine-learning-oriented discussion. Determinants test invertibility and also appear before eigenvalues; eigenvalues determine eigenvectors; eigenvectors construct orthogonal matrices in favorable cases; orthogonal matrices are used in diagonalization; diagonalization and singular value decomposition then become the computational language for dimensionality reduction.

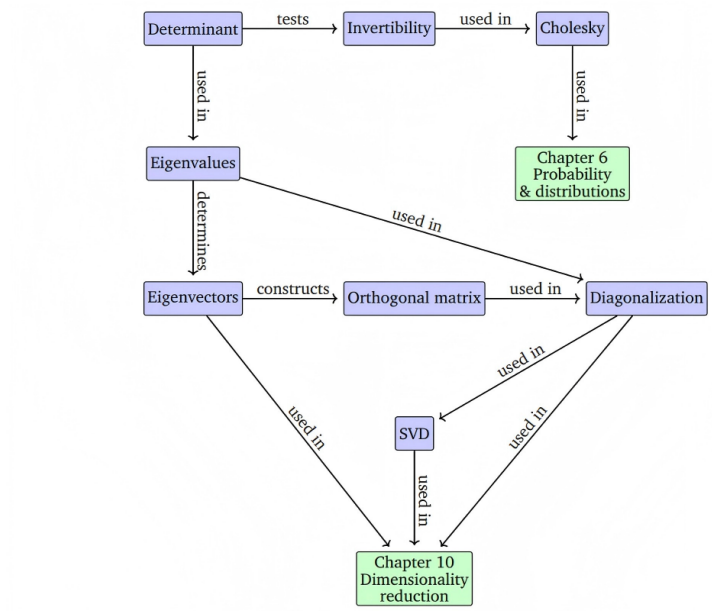


Figure 4.1 A mind map of the concepts introduced in this chapter, along with where they are used in other parts of the book.

Figure 1.1: A concept map linking determinant, invertibility, eigenvalues, eigenvectors, orthogonal matrices, diagonalization, Cholesky factorization, SVD, and dimensionality reduction.

The diagram should not be read as a strict dependency graph. It is a guide to how these objects are reused: determinant and eigenvalues are algebraic invariants; eigenvectors and orthogonal matrices give coordinate systems; SVD is the robust replacement when diagonalization is unavailable or when the matrix is rectangular.

§1.2 Five model transformations

A 2×2 matrix is not only an array of four numbers; it is a rule that moves every point of the plane. The determinant records signed area scaling, while eigenvectors record

directions that survive the transformation without turning:

$$|A(e_1), A(e_2)| = \det A, \qquad Av = \lambda v.$$

The source figure compares five concrete linear maps by applying each to a grid of colored points. The left column is the input, the right column is the transformed output, and the middle column marks eigenvector directions and eigenvalue data where they are real.

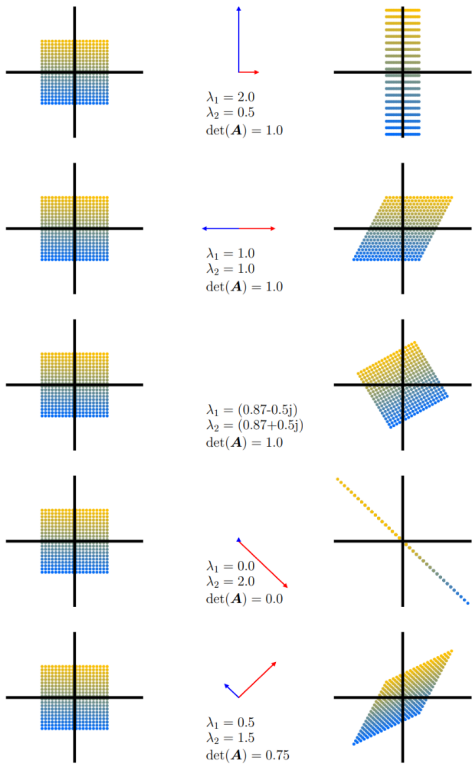


Figure 4.4 Determinants and eigenspaces. Overview of five linear mappings and their associated transformation matrices $\boldsymbol{A}_i \in \mathbb{R}^{2\times 2}$ projecting 400 color-coded points $\boldsymbol{x} \in \mathbb{R}^2$ (left column) onto target points $\boldsymbol{A}_i\boldsymbol{x}$ (right column). The central column depicts the first eigenvector, stretched by its associated eigenvalue λ_1, and the second eigenvector stretched by its eigenvalue λ_2. Each row depicts the effect of one of five transformation matrices $\boldsymbol{A}_i$ with respect to the standard basis.

Figure 1.2: Five planar linear mappings and their eigenspace behavior. Each row shows colored input points, eigenvalue/eigenvector information, and the transformed output points.

The five rows illustrate the standard possibilities: axis-aligned stretching, shear, rotation with complex eigenvalues, collapse onto a line, and symmetric stretching. The determinant and the eigenvalues answer different questions. The determinant says how area changes; eigenvectors say which directions are preserved.

§1.3 A spectral example: a biological neural network

The note then moves from geometry to data. For the *C. elegans* neural network, one may form a connectivity matrix. Since the directed connection matrix is generally not symmetric, one often studies a symmetrized matrix, for example

$$A_{\text{sym}} = A + A^T.$$

The eigenvalues of this matrix describe dominant modes of connectivity.

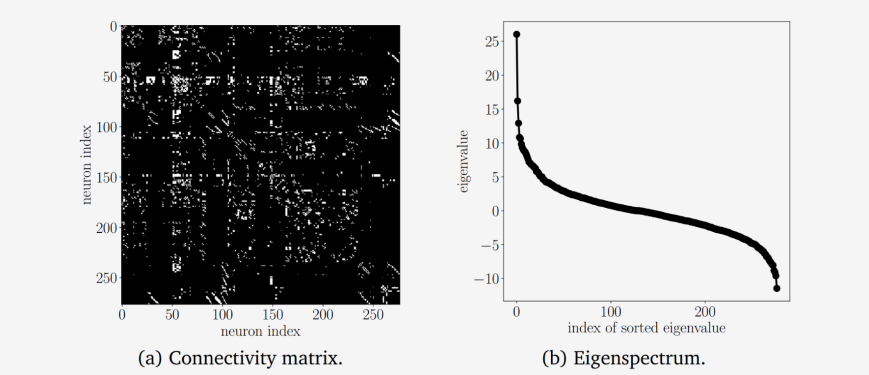


Figure 4.5 Caenorhabditis elegans neural network (Kaiser and Hilgetag, 2006).(a) Symmetrized connectivity matrix; (b) Eigenspectrum.

Figure 1.3: The symmetrized connectivity matrix of the C. elegans neural network and its sorted eigenspectrum.

This example is a useful warning: eigenvectors are not only arrows in the plane. In a network or a dataset, an eigenvector can indicate a coherent global mode, and the corresponding eigenvalue measures its strength.

§1.4 SVD as a sequence of geometric operations

Eigenvalue decomposition is powerful but fragile: it is best behaved for square matrices with enough eigenvectors, and especially for symmetric matrices. Singular value decomposition is more stable. Every real matrix $A \in \mathbb{R}^{m\times n}$ admits

$$A = U\Sigma V^T,$$

where U and V are orthogonal and Σ has nonnegative diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$. Geometrically, V^T changes the input coordinates, Σ stretches by the singular values, and U changes the output coordinates.

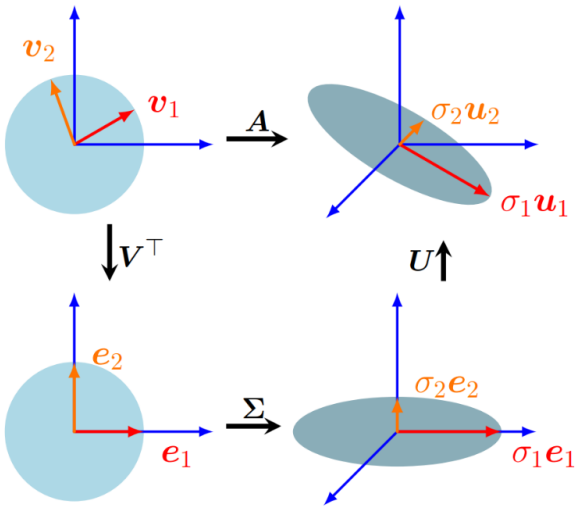


Figure 1.4: Geometric interpretation of SVD: the unit disk is first expressed in the right-singular-vector coordinates, then stretched by the singular values, and finally rotated into the output coordinates.

Thus SVD does not ask for directions that remain invariant under A. Instead, it asks for orthogonal input directions that are sent to orthogonal output axes. The unit circle becomes an ellipse, and the semi-axis lengths are the singular values.

§1.5 SVD and latent structure in data

A rating table can be viewed as a matrix. Rows may represent movies and columns may represent users. The raw entries are ratings, but the singular vectors can reveal latent directions such as genre preference or taste profile.

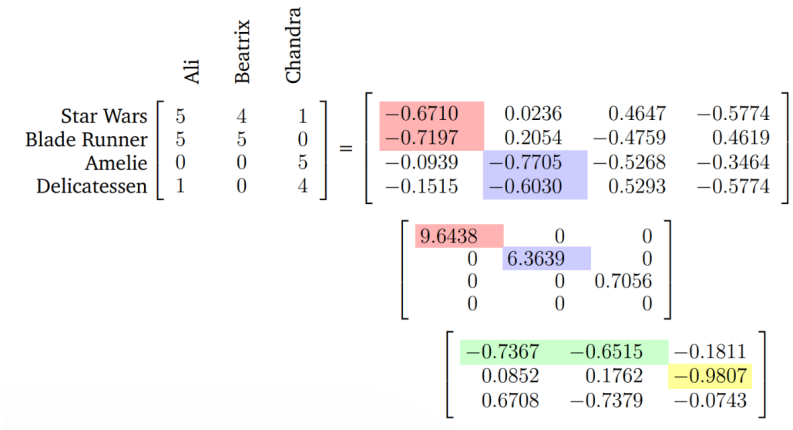


Figure 4.10 Movie ratings of three people for four movies and its SVD decomposition.

Figure 1.5: A movie-rating matrix and its SVD decomposition. Dominant singular directions highlight latent structure in user and movie preferences.

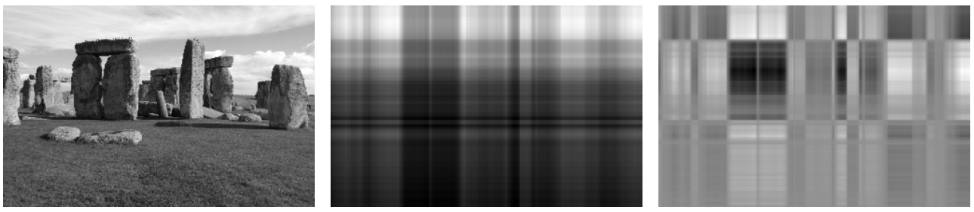
The labels placed on singular directions are interpretations, not part of the theorem. The theorem says that the first singular directions give the most efficient orthogonal explanation of the matrix energy.

§1.6 Rank-one blocks and image compression

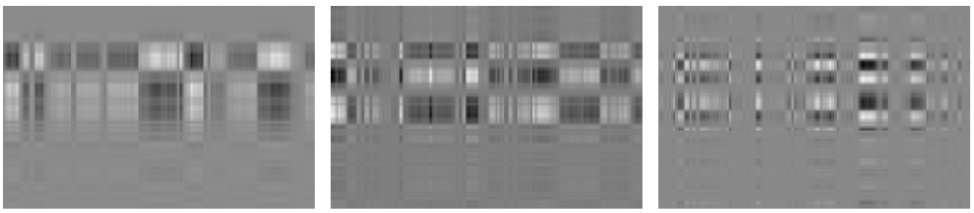
SVD decomposes a matrix into rank-one pieces:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

For an image matrix, each term $\sigma_i u_i v_i^T$ is a weighted rank-one visual pattern.



(a) Original image $\boldsymbol{A}$. (b) $\boldsymbol{A}_1,\ \sigma_1 \approx 228,052$. (c) $\boldsymbol{A}_2,\ \sigma_2 \approx 40,647$.



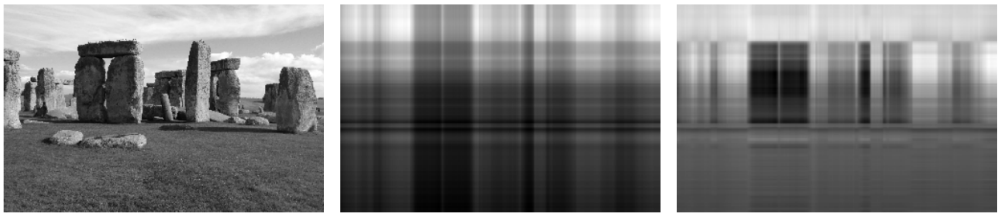
(d) $\boldsymbol{A}_3,\ \sigma_3 \approx 26,125$. (e) $\boldsymbol{A}_4,\ \sigma_4 \approx 20,232$. (f) $\boldsymbol{A}_5,\ \sigma_5 \approx 15,436$.

Figure 1.6: The first rank-one SVD components of an image. Each component captures one weighted visual pattern rather than a recognizable full image.

Truncated SVD keeps only the first k terms:

$$\widehat{A}(k) = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

The Eckart–Young theorem says that this is the best rank-k approximation in the spectral norm and the Frobenius norm.



(a) Original image $\boldsymbol{A}$. (b) Rank-1 approximation $\widehat{\boldsymbol{A}}(1)$. (c) Rank-2 approximation $\widehat{\boldsymbol{A}}(2)$.

(d) Rank-3 approximation $\widehat{\boldsymbol{A}}(3)$. (e) Rank-4 approximation $\widehat{\boldsymbol{A}}(4)$. (f) Rank-5 approximation $\widehat{\boldsymbol{A}}(5)$.

Figure 1.7: Low-rank image reconstruction from the first few singular components. As the rank increases, the approximation recovers more recognizable structure.

This example explains why SVD is useful beyond pure linear algebra: it gives a principled compression method by separating large-scale structure from smaller-scale details.

§1.7 A taxonomy of matrix properties

The final source figure is a taxonomy of real matrices. It distinguishes square from nonsquare matrices, singular from regular matrices, defective from nondefective matrices, normal from non-normal matrices, and the special positions of symmetric, positive definite, orthogonal, diagonal, and identity matrices.

Figure 4.13 A functional phylogeny of matrices encountered in machine learning.

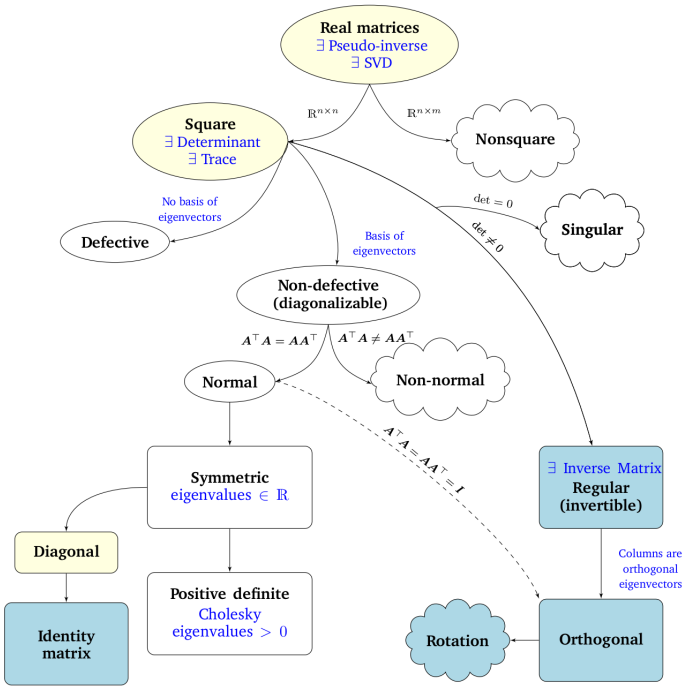


Figure 1.8: A functional phylogeny of real matrices encountered in machine learning, indicating which tools apply to which matrix classes.

The diagram prevents a common error: applying the wrong theorem to the wrong class. Determinants and eigenvalues belong naturally to square matrices; orthogonal diagonalization belongs to symmetric real matrices; pseudoinverses and SVD work for arbitrary real matrices.

§1.8 Closing perspective

The unifying theme is that a matrix should be read through the invariant appropriate to the question. The determinant measures signed volume scaling; the trace and eigenvalues measure intrinsic square-matrix behavior; rank measures the dimension of the image; singular values measure metric distortion; and the pseudoinverse measures least-squares inverse behavior when an exact inverse is unavailable.

Similarity

$$A' = P^{-1}AP$$

preserves the linear operator but changes coordinates on the same space. A general matrix representation

$$\widetilde{A} = T^{-1}AS$$

may change both input and output coordinates. SVD chooses these input and output coordinates orthogonally, which is why it is numerically stable.

For the examples in this note, the practical reading rule is:

question	tool
area/orientation	$\det A$
invariant directions	eigenvectors
collapse or information loss	$\operatorname{rank} A$
principal stretching	singular values
best low-rank compression	truncated SVD

This is why the images are not decorative. They show the same matrix simultaneously as an algebraic object, a geometric transformation, and a data-processing device.