Mathematical Foundations for Quantitative Investing

Qiao Zhou*

July 13, 2026

Contents

*Haas School of Business, University of California, Berkeley. Email address: qiao_zhou@mfe.berkeley.edu.

1 Linear Algebra

Linear algebra provides the language for representing data, linear transformations, covariance matrices, and optimization problems. This section summarizes the core concepts used throughout the subsequent notes.

1.1 Vectors, Matrices, and Dimensions

A scalar is a single number. A column vector with p elements is written as $x \in \mathbb{R}^{p\times 1}$, while its transpose $x^\top \in \mathbb{R}^{1\times p}$ is a row vector. A matrix with N rows and p columns is written as $X \in \mathbb{R}^{N\times p}$.

In a typical data matrix, rows represent observations and columns represent variables or features:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{bmatrix} \in \mathbb{R}^{N\times p}.$$

If $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$, then their product is well defined and $AB \in \mathbb{R}^{m\times k}$. The inner dimensions must agree:

$$(m \times n)(n \times k) = m \times k.$$

For example, if $X \in \mathbb{R}^{N\times p}$ and $\beta \in \mathbb{R}^{p\times 1}$, then $X\beta \in \mathbb{R}^{N\times 1}$.

The (i, j) entry of a matrix product is the inner product between row i of the first matrix and column j of the second matrix:

$$(AB)_{ij} = \sum_{\ell=1}^{n} A_{i\ell} B_{\ell j}.$$

Matrix multiplication is associative but generally not commutative:

$$(AB)C = A(BC), \qquad AB \neq BA.$$

1.2 Transpose, Inner Products, and Norms

The transpose of a matrix exchanges its rows and columns. If $A \in \mathbb{R}^{m\times n}$, then $A^\top \in \mathbb{R}^{n\times m}$. Important transpose identities are

$$(A^\top)^\top = A, \qquad (A + B)^\top = A^\top + B^\top, \qquad (AB)^\top = B^\top A^\top.$$

For vectors $x, y \in \mathbb{R}^{p\times 1}$, their inner product is the scalar

$$x^\top y = \sum_{j=1}^{p} x_j y_j.$$

The Euclidean norm of x is $\|x\|_2 = \sqrt{x^\top x}$. A unit vector satisfies $x^\top x = 1$, and two vectors are orthogonal when $x^\top y = 0$.

For two nonzero vectors x and y, the cosine of the angle between them is

$$\cos(\theta) = \frac{x^\top y}{\|x\|_2 \|y\|_2}.$$

This quantity is also called cosine similarity[1]. It measures similarity in direction while ignoring vector magnitude. Cosine distance is commonly defined as

$$d_{\cos}(x, y) = 1 - \cos(\theta).$$

If the columns of $Q \in \mathbb{R}^{p \times k}$ are orthonormal, then

$$Q^\top Q = I_k.$$

If Q is square, so that $Q \in \mathbb{R}^{p \times p}$, then it is an orthogonal matrix and

$$Q^\top Q = Q Q^\top = I_p, \qquad Q^{-1} = Q^\top.$$

1.3 Linear Transformations

A matrix $A \in \mathbb{R}^{m \times p}$ defines a linear transformation from $\mathbb{R}^p$ to $\mathbb{R}^m$:

$$y = Ax,$$

where $x \in \mathbb{R}^{p \times 1}$ and $y \in \mathbb{R}^{m \times 1}$.

The transformation is linear because, for scalars a and b,

$$A(ax_1 + bx_2) = aAx_1 + bAx_2.$$

Depending on its structure, a matrix may represent rescaling, rotation, projection, dimension reduction, or a change of basis.

If A is square and invertible, the transformation can be reversed:

$$x = A^{-1} y.$$

[1] Cosine similarity is widely used to compare high-dimensional representations, such as text embeddings.

1.4 Linear Independence, Rank, and Linear Systems

Vectors $v_1, \dots, v_k$ are linearly independent if

$$a_1 v_1 + \cdots + a_k v_k = 0$$

implies $a_1 = \cdots = a_k = 0$. Otherwise, at least one vector can be expressed as a linear combination of the others.

The rank of a matrix is the number of linearly independent columns, equivalently the number of linearly independent rows. For $A \in \mathbb{R}^{m \times n}$,

$$\operatorname{rank}(A) \leq \min(m, n).$$

A square matrix $A \in \mathbb{R}^{p \times p}$ is invertible if there exists A^{-1} such that

$$A^{-1}A = AA^{-1} = I_p.$$

A square matrix is invertible if and only if it has full rank:

$$\operatorname{rank}(A) = p.$$

If A is invertible, the linear system $Ax = b$ has the unique solution

$$x = A^{-1}b.$$

In numerical work, it is generally preferable to solve $Ax = b$ directly rather than explicitly computing A^{-1}.

1.5 Span, Basis, and Subspaces

The span of vectors $v_1, \dots, v_k$ is the set of all their linear combinations:

$$\operatorname{span}(v_1, \dots, v_k) = \left\{ \sum_{j=1}^{k} a_j v_j : a_j \in \mathbb{R} \right\}.$$

A basis is a linearly independent set of vectors that spans a vector space. The number of vectors in a basis is the dimension of the space.

For $A \in \mathbb{R}^{m \times n}$, the column space is the subspace spanned by the columns of A. Its dimension is $\operatorname{rank}(A)$.

The null space of A is

$$\operatorname{Null}(A) = \left\{ x \in \mathbb{R}^{n \times 1} : Ax = 0 \right\}.$$

The rank–nullity relationship is

$$\text{rank}(A) + \text{nullity}(A) = n.$$

1.6 Orthogonal Projections

Suppose the columns of $Q \in \mathbb{R}^{p\times k}$ are orthonormal, so that $Q^\top Q = I_k$. The orthogonal projection of $x \in \mathbb{R}^{p\times 1}$ onto the column space of Q is

$$\hat{x} = QQ^\top x.$$

The projection matrix is

$$P = QQ^\top \in \mathbb{R}^{p\times p}.$$

An orthogonal projection matrix is symmetric and idempotent:

$$P^\top = P, \qquad P^2 = P.$$

The residual component is $(I_p - P)x$ and is orthogonal to the projected subspace.

If $X \in \mathbb{R}^{N\times p}$ has full column rank but does not necessarily have orthonormal columns, the projection matrix onto its column space is

$$P_X = X(X^\top X)^{-1}X^\top \in \mathbb{R}^{N\times N}.$$

This projection is the geometric foundation of ordinary least squares.

1.7 Quadratic Forms and Positive Semidefinite Matrices

For $x \in \mathbb{R}^{p\times 1}$ and $A \in \mathbb{R}^{p\times p}$, the scalar

$$x^\top Ax$$

is called a quadratic form. It is quadratic because it is a homogeneous polynomial of degree two in the elements of x:

$$x^\top Ax = \sum_{i=1}^{p}\sum_{j=1}^{p} A_{ij}x_i x_j.$$

For example, when $p = 2$,

$$x^\top Ax = A_{11}x_1^2 + (A_{12} + A_{21})x_1x_2 + A_{22}x_2^2.$$

A symmetric matrix A is **positive semidefinite** if

$$x^\top Ax \geq 0$$

for every x. It is **positive definite** if

$$x^\top Ax > 0$$

for every nonzero x.

A matrix of the form $X^\top X$ is always symmetric because

$$\left(X^\top X\right)^\top = X^\top X.$$

It is also positive semidefinite because, for any $z \in \mathbb{R}^{p\times 1}$,

$$\begin{aligned} z^\top X^\top Xz &= (Xz)^\top (Xz) \\ &= \|Xz\|_2^2 \\ &\geq 0. \end{aligned}$$

Therefore, $X^\top X$ **is symmetric positive semidefinite**.

The main implications are:

- a positive definite matrix is invertible;
- $X^\top X$ is positive definite when the columns of X are linearly independent;
- covariance matrices are positive semidefinite;
- quadratic forms appear in variance calculations, least squares, and optimization.

1.8 Eigenvalues and Eigenvectors

For a square matrix $A \in \mathbb{R}^{p\times p}$, a nonzero vector v_k is an eigenvector if

$$Av_k = \lambda_k v_k,$$

where λ_k is the corresponding eigenvalue.

An eigenvector keeps the same direction under the transformation A; and with magnitude scaled (merely elongated or shrunk) by a factor of λ known as eigenvalue.

For a real symmetric matrix:

- all eigenvalues are real;
- eigenvectors associated with distinct eigenvalues are orthogonal;
- the eigenvectors can be chosen to form an orthonormal basis.

If A is positive semidefinite, then

$$v_k^\top A v_k = \lambda_k v_k^\top v_k \geq 0.$$

Since $v_k^\top v_k > 0$, it follows that

$$\lambda_k \geq 0.$$

Thus, **all eigenvalues of a positive semidefinite matrix are nonnegative**. For a positive definite matrix, all eigenvalues are strictly positive.

1.9 Eigendecomposition

The decomposition of a square matrix A into its eigenvectors and eigenvalues is called an **eigendecomposition**. A real symmetric matrix $A \in \mathbb{R}^{p\times p}$ admits the eigendecomposition

$$A = V\Lambda V^\top,$$

where

$$V = \begin{bmatrix} v_1 & \cdots & v_p \end{bmatrix} \in \mathbb{R}^{p\times p}$$

contains orthonormal eigenvectors and

$$\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^{p\times p}$$

contains the corresponding eigenvalues.

Because V is orthogonal, $V^\top V = I_p$. Equivalently,

$$A = \sum_{k=1}^{p} \lambda_k v_k v_k^\top.$$

This representation expresses A as a weighted sum of mutually orthogonal directions.

For any vector x, define its coordinates in the eigenvector basis as $z = V^\top x$. Then

$$x^\top A x = z^\top \Lambda z = \sum_{k=1}^{p} \lambda_k z_k^2.$$

Thus, the eigenvalues determine how strongly the quadratic form scales each orthogonal direction.

1.10 Trace

For a square matrix $A \in \mathbb{R}^{p\times p}$, the trace is the sum of its diagonal elements:

$$\operatorname{tr}(A) = \sum_{j=1}^{p} A_{jj}.$$

The trace is linear:

$$\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B), \qquad \operatorname{tr}(cA) = c\operatorname{tr}(A).$$

It also satisfies the cyclic property

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB),$$

provided that the products are well defined. The order may be rotated cyclically, but it cannot generally be rearranged arbitrarily.

Using the eigendecomposition $A = V\Lambda V^\top$,

$$\begin{aligned}\operatorname{tr}(A) &= \operatorname{tr}(V\Lambda V^\top) \\ &= \operatorname{tr}(\Lambda V^\top V) \\ &= \operatorname{tr}(\Lambda) \\ &= \sum_{j=1}^{p} \lambda_j.\end{aligned}$$

Therefore, the trace equals the sum of the eigenvalues. For a covariance matrix, the trace is the sum of the individual variable variances.

1.11 Singular Value Decomposition (SVD)

Any matrix $X \in \mathbb{R}^{N\times p}$ can be decomposed as

$$X = UDV^\top.$$

Using the reduced singular value decomposition,

$$U \in \mathbb{R}^{N\times r}, \qquad D \in \mathbb{R}^{r\times r}, \qquad V \in \mathbb{R}^{p\times r},$$

where $r = \operatorname{rank}(X)$.

The columns of U and V are orthonormal, and D is diagonal with positive singular values ordered as

$$d_1 \geq d_2 \geq \cdots \geq d_r > 0.$$

The singular value decomposition implies

$$X^\top X = V D^2 V^\top$$

and

$$X X^\top = U D^2 U^\top.$$

Therefore, the columns of V are eigenvectors of $X^\top X$, the columns of U are eigenvectors of $X X^\top$, and the corresponding nonzero eigenvalues are d_k^2.

Unlike eigendecomposition, singular value decomposition applies directly to rectangular matrices.

1.12 Summary of Key Identities

The following identities will be used repeatedly:

$$(AB)^\top = B^\top A^\top,$$

$$Q^\top Q = I \qquad \text{for a matrix with orthonormal columns,}$$

$$P = QQ^\top, \qquad P^\top = P, \qquad P^2 = P,$$

$$A = V \Lambda V^\top \qquad \text{for a real symmetric matrix } A,$$

$$\operatorname{tr}(A) = \sum_{j=1}^{p} \lambda_j,$$

and

$$X = U D V^\top.$$