Stat 6341.001: Numerical Linear Algebra and Statistical Computing, Fall 2026 **Topic 14**

Vector Spaces, Norms, and Dual Spaces

October 12, 2026

Lecturer: Yun Wei *Scribe: Zhiyuan Ning*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Introduction

The lecture begins by recalling principal component analysis (PCA) and Newton's method for equality-constrained minimization. We then review vector spaces and their bases, norms and inner products, matrix norms, and the dual space, following Chapter 1 of Beck's *First-Order Methods in Optimization* [1, Sections 1.1–1.4, 1.6–1.8, and 1.11]. Throughout these notes, scalars are real. Statements about a finite basis or the representation of every linear functional are made in the finite-dimensional setting.

2 Vector Spaces, Bases, and Dimension

Definition 2.1 (Real vector space)**.** A real vector space E is a nonempty set equipped with addition $E \times E \to E$ and scalar multiplication $\mathbb{R} \times E \to E$ satisfying the following properties. For all $x, y, z \in E$ and $\alpha, \beta \in \mathbb{R}$:

(i) Addition is commutative: $x + y = y + x$.

(ii) Addition is associative: $x + (y + z) = (x + y) + z$.

(iii) There is a zero vector $0 \in E$ such that $x + 0 = x$.

(iv) Each $x \in E$ has an additive inverse $-x \in E$ such that $x + (-x) = 0$.

(v) Scalar multiplication is associative: $\alpha(\beta x) = (\alpha\beta)x$.

(vi) Multiplication by one leaves a vector unchanged: $1x = x$.

(vii) Scalar multiplication distributes over vector addition: $\alpha(x + y) = \alpha x + \alpha y$.

(viii) Scalar multiplication distributes over scalar addition: $(\alpha + \beta)x = \alpha x + \beta x$.

The zero vector and the additive inverse of each vector are unique. For example, if 0 and $0'$ are both zero vectors, then $0 = 0 + 0' = 0'$. The familiar spaces $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ are vector spaces under entrywise addition and scalar multiplication. In particular, a "vector" in an abstract vector space may itself be a matrix.

2.1 Linear independence and span

Definition 2.2 (Linear independence)**.** A finite family $v_1, \ldots, v_k \in E$ is *linearly independent* if

$$\sum_{i=1}^{k} \alpha_i v_i = 0 \quad \Longrightarrow \quad \alpha_1 = \cdots = \alpha_k = 0.$$

Thus, the only linear combination of these vectors that equals zero is the trivial combination. Otherwise the family is linearly dependent.

Definition 2.3 (Span and basis)**.** The span of $v_1, \dots, v_k$ is

$$\operatorname{span}\{v_1, \dots, v_k\} = \left\{ \sum_{i=1}^{k} \beta_i v_i : \beta_1, \dots, \beta_k \in \mathbb{R} \right\}.$$

The vectors *span* E if every $x \in E$ can be written as $x = \sum_{i=1}^{k} \beta_i v_i$. A *basis* of E is a linearly independent family that spans E.

Spanning guarantees that a coordinate representation exists, while linear independence guarantees that it is unique. Indeed, if $\sum_i \alpha_i v_i = \sum_i \beta_i v_i$, then $\sum_i (\alpha_i - \beta_i) v_i = 0$, so $\alpha_i = \beta_i$ for every i. All bases of a finite-dimensional vector space have the same number of vectors. This number is the *dimension* of E, denoted $\dim(E)$.

Example 2.4 (Standard basis of $\mathbb{R}^n$)**.** Let e_i be the vector with a one in coordinate i and zeros elsewhere. Every $x = (x_1, \dots, x_n)^{\mathsf{T}}$ has the representation

$$x = \sum_{i=1}^{n} x_i e_i.$$

Moreover, $\sum_i \alpha_i e_i = 0$ implies $\alpha_i = 0$ by examining coordinate i. Hence $e_1, \dots, e_n$ form a basis and $\dim(\mathbb{R}^n) = n$.

Example 2.5 (Standard basis of $\mathbb{R}^{m \times n}$)**.** For $1 \le i \le m$ and $1 \le j \le n$, let E_{ij} denote the $m \times n$ matrix whose (i, j) entry is one and whose other entries are zero. Then

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} E_{ij}, \qquad A \in \mathbb{R}^{m \times n}.$$

If $\sum_{i,j} \alpha_{ij} E_{ij} = 0$, its (i, j) entry is α_{ij}, so every coefficient vanishes. Thus the mn matrices E_{ij} form a basis and

$$\dim\big(\mathbb{R}^{m \times n}\big) = mn.$$

3 Norms and Inner Products

3.1 Norms and balls

Definition 3.1 (Norm)**.** A norm on a real vector space E is a function $\|\cdot\| : E \to \mathbb{R}$ satisfying, for all $x, y \in E$ and $\lambda \in \mathbb{R}$:

(i) *Nonnegativity and definiteness:* $\|x\| \ge 0$, with $\|x\| = 0$ if and only if $x = 0$.

(ii) *Absolute homogeneity:* $\|\lambda x\| = |\lambda| \, \|x\|$.

(iii) *Triangle inequality:* $\|x + y\| \le \|x\| + \|y\|$.

A norm measures the size of a vector, and $\|x - y\|$ measures the distance between two vectors. For example, the Euclidean norm on $\mathbb{R}^n$ is

$$\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

Given a center $c \in E$ and a radius $r > 0$, the open and closed balls in the chosen norm are, respectively,

$$B(c, r) = \{x \in E : \|x - c\| < r\}, \qquad B[c, r] = \{x \in E : \|x - c\| \le r\}.$$

Their shapes depend on the norm, so a ball need not look like a Euclidean sphere.

3.2 Inner products and their induced norms

Definition 3.2 (Real inner product)**.** An inner product on E is a function $\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$ with the following properties:

(i) *Symmetry:* $\langle x, y \rangle = \langle y, x \rangle$.

(ii) *Linearity in the first argument:*

$$\langle \alpha_1 x_1 + \alpha_2 x_2, y \rangle = \alpha_1 \langle x_1, y \rangle + \alpha_2 \langle x_2, y \rangle.$$

(iii) *Positive definiteness:* $\langle x, x \rangle \geq 0$, with $\langle x, x \rangle = 0$ if and only if $x = 0$.

Here $x, y, x_1, x_2 \in E$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. Symmetry and linearity in the first argument also give linearity in the second argument.

Example 3.3 (Euclidean inner product)**.** On $\mathbb{R}^n$, the usual dot product is

$$\langle x, y \rangle = x^{\mathsf{T}} y = \sum_{i=1}^{n} x_i y_i.$$

Its induced norm $\sqrt{\langle x, x \rangle}$ is the Euclidean norm $\|x\|_2$.

Proposition 3.4 (Norm induced by an inner product)**.** *Every real inner product induces a norm by*

$$\|x\| = \sqrt{\langle x, x \rangle}. \tag{1}$$

Proof. Nonnegativity and definiteness follow from the inner product axioms, and bilinearity gives $\|\lambda x\|^2 = \lambda^2 \|x\|^2$. To verify the triangle inequality, first note the Cauchy–Schwarz inequality

$$|\langle x, y \rangle| \leq \|x\| \, \|y\|.$$

For $y \neq 0$, it follows by expanding $0 \leq \|x - ty\|^2$ at $t = \langle x, y \rangle / \|y\|^2$; the case $y = 0$ is immediate. Consequently,

$$\|x + y\|^2 = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \leq (\|x\| + \|y\|)^2.$$

Taking square roots proves the claim. □

Example 3.5 (Q-inner product and Q-norm)**.** Let $Q \in \mathbb{S}^n_{++}$, where $\mathbb{S}^n_{++}$ denotes the set of real symmetric positive definite $n \times n$ matrices. Thus $Q^{\mathsf{T}} = Q$ and $x^{\mathsf{T}} Q x > 0$ for every $x \neq 0$. Define

$$\langle x, y \rangle_Q = x^{\mathsf{T}} Q y, \qquad \|x\|_Q = \sqrt{x^{\mathsf{T}} Q x}.$$

Verifying that this is an inner product was an exercise in the lecture. Linearity follows from matrix multiplication, and symmetry follows from

$$x^{\mathsf{T}} Q y = (x^{\mathsf{T}} Q y)^{\mathsf{T}} = y^{\mathsf{T}} Q^{\mathsf{T}} x = y^{\mathsf{T}} Q x.$$

Positive definiteness follows from the assumption on Q. Equation (1) then gives the stated Q-norm. For $Q = I_n$, these reduce to the ordinary dot product and Euclidean norm.

3.3 The p-norms on $\mathbb{R}^n$

For $1 \le p < \infty$, the p-norm of $x \in \mathbb{R}^n$ is

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

In particular, $\|x\|_1 = \sum_i |x_i|$, and $p = 2$ gives the Euclidean norm. The restriction $p \ge 1$ is necessary for this formula to define a norm: when $0 < p < 1$, the triangle inequality fails in general. For $p \ge 1$, its triangle inequality is Minkowski's inequality. The infinity norm is

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i|.$$

It is the limiting case of the p-norms:

$$\lim_{p \to \infty} \|x\|_p = \|x\|_\infty. \tag{2}$$

To see this, put $M = \max_i |x_i|$. At least one term in the sum is M^p, and all n terms are at most M^p. Therefore

$$M \le \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \le n^{1/p} M.$$

Since $n^{1/p} \to 1$, the squeeze theorem proves (2), including the case $x = 0$.

4 Norms on Matrix Spaces

4.1 The Frobenius inner product and norm

Because $\mathbb{R}^{m \times n}$ is a vector space, its elements can also be equipped with an inner product. For $A, B \in \mathbb{R}^{m \times n}$, define

$$\langle A, B \rangle_F = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij} = \operatorname{tr}(A^\mathsf{T} B). \tag{3}$$

Here $A^\mathsf{T} B$ is $n \times n$, and the trace is the sum of its diagonal entries. Indeed, $(A^\mathsf{T} B)_{jj} = \sum_i A_{ij} B_{ij}$, which proves the last equality. This is the ordinary Euclidean inner product applied to all mn matrix entries.

The associated norm is the *Frobenius norm*:

$$\|A\|_F = \sqrt{\langle A, A \rangle_F} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2 \right)^{1/2} = \sqrt{\operatorname{tr}(A^\mathsf{T} A)}. \tag{4}$$

Thus $\operatorname{tr}(A^\mathsf{T} A) = \|A\|_F^2$; the square root is essential when writing the norm itself.

4.2 Induced matrix norms

An $m \times n$ matrix can also be viewed as a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, sending x to Ax. Suppose the domain has a norm $\|\cdot\|_a$ and the codomain has a norm $\|\cdot\|_b$.

Definition 4.1 (Induced matrix norm)**.** The matrix norm induced by these two vector norms is

$$\|A\|_{a,b} = \max_{\|x\|_a \le 1} \|Ax\|_b = \max_{\|x\|_a = 1} \|Ax\|_b = \sup_{x \ne 0} \frac{\|Ax\|_b}{\|x\|_a}. \tag{5}$$

The first subscript specifies the input norm, and the second specifies the output norm. In finite dimensions the closed unit ball is compact and $x \mapsto \|Ax\|_b$ is continuous, so the maximum exists. The equivalent forms in (5) follow by scaling any nonzero vector to have unit norm.

In particular, the definition gives

$$\|Ax\|_b \le \|A\|_{a,b}\,\|x\|_a \qquad \text{for every } x \in \mathbb{R}^n. \tag{6}$$

For $x \neq 0$, apply the definition to $x/\|x\|_a$ and multiply by $\|x\|_a$; for $x = 0$, both sides vanish. Thus the induced norm measures the largest possible amplification of a vector's size by the linear map.

It is also a norm on the vector space of matrices. Absolute homogeneity is immediate. If $A \neq 0$, some Ax is nonzero, so $\|A\|_{a,b} > 0$. Finally, the vector triangle inequality gives

$$\|(A+B)x\|_b \le \|Ax\|_b + \|Bx\|_b \le (\|A\|_{a,b} + \|B\|_{a,b})\|x\|_a,$$

and maximizing over the unit ball proves the matrix triangle inequality.

4.3 Spectral norm

When both vector norms are Euclidean, we obtain the *spectral norm*:

$$\|A\|_2 := \|A\|_{2,2} = \max_{\|x\|_2 \le 1} \|Ax\|_2 = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^\mathsf{T}A)}. \tag{7}$$

Here $\sigma_{\max}(A)$ is the largest singular value of A, and $\lambda_{\max}(A^\mathsf{T}A)$ is the largest eigenvalue of the symmetric positive semidefinite matrix $A^\mathsf{T}A$.

Verification of the spectral-norm formula (lecture exercise). For a unit vector x, $\|Ax\|_2^2 = x^\mathsf{T}A^\mathsf{T}Ax$. Write $A^\mathsf{T}A = U\Lambda U^\mathsf{T}$, with U orthogonal and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_i \ge 0$. If $y = U^\mathsf{T}x$, then $\|y\|_2 = 1$ and

$$x^\mathsf{T}A^\mathsf{T}Ax = \sum_{i=1}^{n} \lambda_i y_i^2 \le \lambda_{\max} \sum_{i=1}^{n} y_i^2 = \lambda_{\max}.$$

Equality is attained at a unit eigenvector for $\lambda_{\max}$. Taking square roots gives (7). □

4.4 Maximum absolute column sum norm

For $a = b = 1$, the induced matrix norm is

$$\|A\|_1 := \|A\|_{1,1} = \max_{\|x\|_1 \le 1} \|Ax\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |A_{ij}|. \tag{8}$$

Thus the matrix 1-norm is the largest absolute column sum.

Verification of the column-sum formula (lecture exercise). Let $c = \max_j \sum_i |A_{ij}|$. For every x,

$$\|Ax\|_1 = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{ij}x_j \right| \le \sum_{j=1}^{n} |x_j| \sum_{i=1}^{m} |A_{ij}| \le c\|x\|_1.$$

Hence $\|A\|_1 \le c$. Choose a column j_* attaining c and let $x = e_{j_*}$. Then $\|x\|_1 = 1$, and Ax is that column, so $\|Ax\|_1 = c$. This proves equality. □

4.5 Maximum absolute row sum norm

For $a = b = \infty$, the induced norm is

$$\|A\|_\infty := \|A\|_{\infty,\infty} = \max_{\|x\|_\infty \le 1} \|Ax\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |A_{ij}|. \tag{9}$$

Thus the matrix infinity norm is the largest absolute row sum.

Verification of the row-sum formula (lecture exercise). Let $r = \max_i \sum_j |A_{ij}|$. Then

$$\|Ax\|_\infty = \max_i \left| \sum_j A_{ij} x_j \right| \le \max_i \sum_j |A_{ij}|\,|x_j| \le r\|x\|_\infty.$$

For equality, choose a row i_* attaining r, and set $x_j = 1$ if $A_{i_* j} \ge 0$ and $x_j = -1$ otherwise. This choice has $\|x\|_\infty = 1$ and

$$(Ax)_{i_*} = \sum_{j=1}^{n} A_{i_* j} x_j = \sum_{j=1}^{n} |A_{i_* j}| = r.$$

Therefore $\|A\|_\infty = r$. □

Remark 4.2 (Distinguishing matrix norms)**.** The Frobenius norm measures the Euclidean size of all matrix entries, whereas the spectral norm measures the largest Euclidean amplification of a vector. For example, $\|I_n\|_F = \sqrt{n}$ but $\|I_n\|_2 = 1$. Also, the induced matrix 1-norm is a maximum column sum, not the sum of the absolute values of all entries.

5 The Dual Space

Definition 5.1 (Linear functional and dual space)**.** A linear functional on E is a linear map $f : E \to \mathbb{R}$, meaning

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \qquad (x, y \in E,\ \alpha, \beta \in \mathbb{R}).$$

The dual space is the set of all such functionals:

$$E^* = \{f : E \to \mathbb{R} : f \text{ is linear}\}.$$

With pointwise operations $(f+g)(x) = f(x) + g(x)$ and $(\alpha f)(x) = \alpha f(x)$, the set E^* is itself a vector space.

Lemma 5.2 (Representation of linear functionals)**.** *Let E be a finite-dimensional real inner product space. For every $f \in E^*$, there is a unique vector $v \in E$ such that*

$$f(x) = \langle v, x\rangle \qquad \textit{for every } x \in E. \tag{10}$$

Proof. Choose an orthonormal basis $u_1, \dots, u_d$ of E, where $d = \dim(E)$, and set

$$v = \sum_{i=1}^{d} f(u_i) u_i.$$

For $x = \sum_i x_i u_i$, linearity and orthonormality imply

$$f(x) = \sum_{i=1}^{d} x_i f(u_i) = \left\langle \sum_{i=1}^{d} f(u_i) u_i, \sum_{j=1}^{d} x_j u_j \right\rangle = \langle v, x\rangle.$$

This proves existence. If v and w represent the same functional, then $\langle v - w, x\rangle = 0$ for every x. Setting $x = v - w$ gives $\|v - w\|^2 = 0$, so $v = w$. □

For each $v \in E$, the map $J(v) : x \mapsto \langle v, x\rangle$ is a linear functional. The lemma shows that $J : E \to E^*$ is a bijection, and bilinearity shows that J is linear. Therefore

$$E \cong E^*, \qquad \dim(E^*) = \dim(E).$$

The board notation $E^* = E$ expresses this identification: vectors and linear functionals are different objects, but the chosen inner product provides a one-to-one linear correspondence between them. Finite dimensionality is part of the statement here; the same claim about all linear functionals does not hold for arbitrary infinite-dimensional inner product spaces.

Example 5.3 (Coordinate form of the representation)**.** On $\mathbb{R}^n$ with the standard dot product, every functional is $f(x) = a^{\mathsf{T}}x$, where $a = (f(e_1), \dots, f(e_n))^{\mathsf{T}}$. If instead we use the Q-inner product from Example 3.5, the representing vector is $v = Q^{-1}a$, since

$$\langle Q^{-1}a, x\rangle_Q = (Q^{-1}a)^{\mathsf{T}}Qx = a^{\mathsf{T}}x.$$

Thus the functional remains the same, but its representing vector depends on the chosen inner product.

References

[1] A. Beck, *First-Order Methods in Optimization*, MOS-SIAM Series on Optimization, vol. 25, Society for Industrial and Applied Mathematics and Mathematical Optimization Society, Philadelphia, PA, 2017. doi: 10.1137/1.9781611974997.