Mathematics for Neural Nets and Especially Field Computation

Bruce MacLennan

Computer Science Department
University of Tennessee, Knoxville

August 9, 2003

Chapter 2

Banach Spaces

2.1 Linear spaces

2.1.1 Definitions

Definition 2.1.1 (Field) *Briefly, a* field *is a set on which sum and product operations are defined that satisfy the usual properties (both are commutative and associative, product distributes over sum, additive identity 0, additive inverses, multiplicative inverses, multiplicative identity 1).*

Example 2.1.1 *The rationals, reals and complex numbers are the best-known examples of fields (and the ones we will have most use of).*

Definition 2.1.2 (Linear space) *A* linear space over K *comprises a field K (of "scalars"), a set L (of "vectors"), and two operations, a* sum $\oplus : L \times L \longrightarrow L$ *and a (scalar) product $\otimes : K \times L \longrightarrow L$. The operators satisfy the following properties (for $a, b \in K$ and $\mathbf{x}, \mathbf{y}, \mathbf{z} \in L$):*

associative sum: $(\mathbf{x} \oplus \mathbf{y}) \oplus \mathbf{z} = \mathbf{x} \oplus (\mathbf{y} \oplus \mathbf{z})$.

zero element: *There is a* zero element $\mathbf{0} \in L$ *satisfying*

$$\mathbf{0} \oplus \mathbf{x} = \mathbf{x} = \mathbf{x} \oplus \mathbf{0}.$$

additive inverses: *For every $\mathbf{x} \in L$ there is a $-\mathbf{x} \in L$ such that*

$$-\mathbf{x} \oplus \mathbf{x} = \mathbf{0} = \mathbf{x} \oplus -\mathbf{x}.$$

(This is implied by the additive inverses in the field.)

commutative sum: $\mathbf{x} \oplus \mathbf{y} = \mathbf{y} \oplus \mathbf{x}$.

right distribution: $a \otimes (\mathbf{x} \oplus \mathbf{y}) = (a \otimes \mathbf{x}) \oplus (a \otimes \mathbf{y})$.

left distribution: $(a + b) \otimes \mathbf{x} = (a \otimes \mathbf{x}) \oplus (b \otimes \mathbf{x})$, *where* $+$ *is the addition operation of the field.*

associative product: $a \otimes (b \otimes \mathbf{x}) = (a \times b) \otimes \mathbf{x}$, *where* $\times$ *is the product operation of the field.*

unit product: $1 \otimes \mathbf{x} = \mathbf{x}$, *where* 1 *is the unitary element (multiplicative identity) of the field. (It is easy to show* $0 \otimes \mathbf{x} = \mathbf{0}$ *and* $-\mathbf{x} = (-1) \otimes \mathbf{x}$*; do it!)*

Remark 2.1.1 *The terms "scalar" and "vector" are meant to be suggestive, but it should not be assumed that the elements of a linear space are scalars and vectors of the familiar kind (they need not even be numbers); we will sometimes say "generalized vector" to stress this point. Indeed any set of objects may form a linear space if they can be "scaled" and "summed" in accord with the preceding definition.*

Remark 2.1.2 K *is called the* basic field *or* ground field *of the linear space. A linear space over the real numbers is called a* real linear space, *and one over the complex numbers a* complex linear space.

Notation 2.1.1 *For convenience, when confusion is unlikely, we write* $a\mathbf{x}$ *for* $a \otimes \mathbf{x}$ *and* $\mathbf{x} + \mathbf{y}$ *for* $\mathbf{x} \oplus \mathbf{y}$*. Further, we will write* $-\mathbf{x}$ *for* $-1 \otimes \mathbf{x}$ *and* $\mathbf{x} - \mathbf{y}$ *for* $\mathbf{x} \oplus -\mathbf{y}$.

Definition 2.1.3 (Linear dependence and independence) *A set of vectors* $\{\mathbf{x}_1, \ldots \mathbf{x}_n\}$ *is* linearly independent *if* $a_1\mathbf{x}_1 + \cdots a_n\mathbf{x}_n = 0$ *implies* $a_1 = \cdots = a_n = 0$*. Conversely, they are* linearly dependent *if there are* a_k*, not all zero, such that* $a_1\mathbf{x}_1 + \cdots a_n\mathbf{x}_n = 0$.

Remark 2.1.3 *The significance, of course, of linear dependence is that the vectors can be written as linear combinations of the other vectors. For example, if* $a_1\mathbf{x}_1 + \cdots a_n\mathbf{x}_n = 0$ *with* $a_1 \neq 0$*, then* $\mathbf{x}_1 = -a_1^{-1}(a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n)$.

2.1.2 Examples

Example 2.1.2 *The most familiar example of a linear space is the vector space of* n*-dimensional vectors (either real or complex).*

Example 2.1.3 *The preceding example can be extended to the linear space of* infinite sequences *of elements of a field* K*.* (Give definitions for the sum and product operations of this linear space.)

Example 2.1.4 *The set of* K*-valued functions is a linear space over* K*.* (Define the sum and product operations of this linear space.)

Example 2.1.5 *Linear spaces can be defined for various important subsets of the space of* K*-valued functions, for example continuous functions and differentiable functions.*

Remark 2.1.4 *Thus the set of* spatially continuous images *(*fields*, in the sense of field computation) are a linear space. Describe the sum and product operations in intuitive terms (e.g., what is their effect on visual images).*

Exercise 2.1.1 *Show that the preceding are linear spaces. (The major problem is to show that the operations are closed, e.g., that the sum of two continuous functions is continuous.)*

Exercise 2.1.2 *Define the linear space of* n*-term polynomials with coefficients in* K*.*

Exercise 2.1.3 *Show that the space of* $m \times n$ *matrices forms a linear space with the usual definitions of scalar product and matrix addition.*

Definition 2.1.4 (Linear function) *A real function* $f : X \longrightarrow \mathcal{R}$ *is called* linear *if for all* $x, y \in X$, $f(ax + by) = af(x) + bf(y)$.

Exercise 2.1.4 *Show that the set of all linear functions on* X *forms a linear space.*

Remark 2.1.5 *There are few purely linear systems in nature. Usually some quantity cannot be increased without bound, either because some "substance" will run out, or because it will begin to behave nonlinearly — as when a high voltage arcs through a capacitor or a high current vaporizes a conductor.*

As a rule of thumb, it's best to treat any *linear model as an approximation. Further, it is unwise to depend on any properties of linear models that do not hold (at least approximately) for approximately linear models. One very useful class of approximately linear model is the* saturating linear *model, which is linear for moderate quantities, but approaches limits for extreme quantities. An example is a sigmoid function.*

2.2 Normed linear spaces

2.2.1 Norms

Definitions

A *norm* measures the "size" or "magnitude" of an element of a linear space.

Definition 2.2.1 (Norm) *An operation $N : L \longrightarrow \mathcal{R}_*$ on a real or complex linear space L is a* norm *if and only if:*

zero norm: $N(\mathbf{x}) = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

scaling: $N(a \otimes \mathbf{x}) = |a|N(\mathbf{x})$.

triangle inequality: $N(\mathbf{x} \oplus \mathbf{y}) \leq N(\mathbf{x}) + N(\mathbf{y})$.

Notation 2.2.1 $N(\mathbf{x})$ *is usually written* $\|\mathbf{x}\|$*, possibly with subscripts to distinguish different norms.*

Remark 2.2.1 *The absolute value is a norm on the reals.*

Exercise 2.2.1 *Show that for any norm $\|\cdot\|$ on $\mathcal{R}$, there is a positive η such that $\|x\| = \eta|x|$ for all $x \in \mathcal{R}$. What is the value of η?*

Exercise 2.2.2 *Show that a norm is a continuous mapping.*

Examples

Definition 2.2.2 (L_p norms) *The L_p norms of n-dimensional real or complex vectors are defined:*

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{k=1}^{n} |x_k|^p}.$$

More concisely, $\|\mathbf{x}\|_p^p = \sum x_k$. As usual, $\|\mathbf{x}\|_\infty = \max_k |x_k|$. L_2 is called the Euclidean norm.

Exercise 2.2.3 *Show that that L_p norms are in fact norms.*

Exercise 2.2.4 *What is the relation between L_p norms and L_p metrics?*

Definition 2.2.3 (L_p product norms) *If L is a cartesian product of normed linear spaces L_k, that is, $L = L_1 \times \cdots L_n$, then an L_p norm can be defined as follows:*

$$\|X\|_p = \sqrt[p]{\sum_{k=1}^{n} \|X_k\|_k^p},$$

where $\|X_k\|_k$ represents the norm on L_k (possibly different in each case).

Exercise 2.2.5 *Show that the L_p product norm is in fact a norm.*

Definition 2.2.4 *$\mathcal{C}[a,b]$ is the space of all continuous real-valued functions on $[a,b]$.*

Definition 2.2.5 (Uniform norm) *For $f \in \mathcal{C}[a,b]$, the* uniform norm *is defined:*

$$\|f\| = \max_{a \le x \le b} |f(x)|$$

Exercise 2.2.6 *Show that the uniform norm is in fact a norm.*

2.2.2 Normality

Definition 2.2.6 (Normalized) *A (generalized) vector* $\mathbf{x}$ *is* normalized *if its norm is 1,* $\|\mathbf{x}\| = 1$. *A nonzero vector can always be normalized by dividing by its norm, that is,* $\mathbf{x}/\|\mathbf{x}\|$ *is normalized.*

Exercise 2.2.7 *Use the definition of a norm to show that* $\mathbf{x}/\|\mathbf{x}\|$ *is normalized (for* $\mathbf{x} \neq \mathbf{0}$*).*

Remark 2.2.2 *Normalization considers (generalized) vectors independently of their absolute size, that is independently of scaling.*

Exercise 2.2.8 *For various values of* $p \geq 1$*, describe the effect of* L_p*-normalization on two-dimensional vectors.*

Exercise 2.2.9 *Describe the effect of normalization on two-dimentional visual images. Describe the effect on acoustic signals of finite duration.*

2.2.3 Normed linear space

Definition 2.2.7 (Normed linear space) *A* normed linear space *is a real or complex linear space together with a norm on that space.*

Exercise 2.2.10 *Show that* $\mathcal{C}[a, b]$ *with the uniform norm is a normed linear space.*

Exercise 2.2.11 *If you know some numerical analysis, then show that* $\mathcal{C}[a, b]$ *is separable. Hint: Apply the Weierstrass approximation theorem.*

Proposition 2.2.1 *A normed linear space is a metric space under the* norm-metric $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.

Exercise 2.2.12 *Prove this proposition.*

Definition 2.2.8 (Translation and scaling invariance) *A metric is* translation invariant *if*

$$d(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z}) = d(\mathbf{x}, \mathbf{y}).$$

It is scaling invariant *if*

$$d(a\mathbf{x}, a\mathbf{y}) = |a| d(\mathbf{x}, \mathbf{y}).$$

Exercise 2.2.13 *Draw diagrams to illustrate the concepts of translation and scaling invariant metrics.*

Proposition 2.2.2 *The norm-metric is translation and scaling invariant.*

Exercise 2.2.14 *Prove this proposition.*

Definition 2.2.9 (Topologically equivalent norms) *Two norms, $\|\cdot\|$ and $\|\cdot\|'$, are* topologically equivalent *if there are positive real numbers ζ, η such that for all $\mathbf{x} \in L$,*

$$\zeta\|\mathbf{x}\| \leq \|\mathbf{x}\|' \leq \eta\|\mathbf{x}\|.$$

Exercise 2.2.15 *Explain why such norms should be called "topologically equivalent." Hint: Suppose a sequence converges by one norm. What can you say about convergence under topologically equivalent norms?*

Exercise 2.2.16 *Show that in $\mathcal{E}^n$, the L_2-norm is topologically equivalent to the L_∞-norm.*

Exercise 2.2.17 *Show that in $\mathcal{E}^n$, all norms are topologically equivalent.*

Proposition 2.2.3 *Suppose $f : X \longrightarrow Y$, where X and Y are normed linear spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. If there are positive reals ζ, η such that for all $\mathbf{x}, \mathbf{y} \in X$,*

$$\zeta\|\mathbf{x}-\mathbf{y}\|_X \leq \|f(\mathbf{x}) - f(\mathbf{y})\|_Y \leq \eta\|\mathbf{x}-\mathbf{y}\|_X,$$

then f is a homeomorphism, and X and Y are homeomorphic.

Exercise 2.2.18 *Prove this. Hint: What does the inequality say about the continuity of f and f^{-1}?*

2.3 Banach spaces

Definition 2.3.1 (Banach space) *A* Banach space *is a complete normed linear space.*

Remark 2.3.1 *Thus the significance of a Banach space is that Cauchy sequences are guaranteed to converge to limits; that is, if* $\|\mathbf{x}_m - \mathbf{x}_n\| \to 0$*, then we know* $\lim_{n\to\infty} \mathbf{x}_n$ *exists. In particular, this means that operations such as differentiation can be defined.*

Exercise 2.3.1 *Show that* $\mathcal{E}^n$ *is a Banach space.*

2.4 Differentiation in Banach spaces

2.4.1 Fréchet derivative

Definition 2.4.1 (Fréchet differentiation) *Suppose* X *and* Y *are two Banach spaces and* U *is an open subset of* X*. Then* $T : U \to Y$ *is* Fréchet differentiable *at* ϕ *if there is a bounded linear operator* $D : X \to Y$ *such that the following holds. For all* $\alpha \in X$ *such that* $\phi + \alpha \in U$*, there is an* $E : X \to Y$ *such that*

$$T(\phi + \alpha) = T(\phi) + D(\alpha) + E(\alpha)$$

and

$$\lim_{\|\alpha\|\to 0} \frac{\|E(\alpha)\|}{\|\alpha\|} = 0.$$

Under these circumstances, D *is called the Fréchet derivative of* T *at* ϕ*; it is denoted by* $T'(\phi)$*. The Fréchet derivative is a locally linear approximation to* T*;* $T'(\phi)(\alpha) = D(\alpha)$ *is called the* Fréchet differential *of* T*.*

Remark 2.4.1 *Since a linear operator is continuous if and only if it is bounded, Fréchet derivatives are (by definition) continuous.*

Proposition 2.4.1 *The derivative of a linear operator is that operator:* $L'(\phi) = L$*.*

Remark 2.4.2 *Note that $T' : X \to \mathcal{L}(X,Y)$, where $\mathcal{L}(X,Y)$ is the space of all continuous (bounded) linear operators from X to Y.*

Remark 2.4.3 *Higher order derivatives are defined in the obvious way. Suppose $T : X \to Y$. Since $T' : X \to \mathcal{L}(X,Y)$, it is easy to see that the higher derivatives have the types:*

$$\begin{aligned}&T'' : X \to \mathcal{L}(X, \mathcal{L}(X,Y)),\\&T^{(3)} : X \to \mathcal{L}(X, \mathcal{L}(X, \mathcal{L}(X,Y))),\end{aligned}$$

and so forth. Note that each successive derivative is of "higher type" than its predecessor.

2.4.2 Gâteaux derivative

Definition 2.4.2 (Gâteaux differentiation) *Suppose X and Y are Banach spaces, $U \subseteq X$ is open, and $T : U \to Y$. Then T has a Gâteaux derivative at $\phi \in U$ if, for all $\alpha \in U$ the following limit exists:*

$$\mathrm{d}T(\phi,\alpha) = \lim_{t\to 0} \frac{T(\phi + t\alpha) - T(\phi)}{t} = \frac{\mathrm{d}}{\mathrm{d}t} T(\phi + t\alpha)|_{t=0}.$$

We write $\mathrm{d}T(\phi,\alpha)$ for the Gâteaux derivative of T at ϕ in the "direction" α.

Proposition 2.4.2 *The Gâteaux derivative, if it exists, is unique.*

2.4.3 Properties of derivatives

Proposition 2.4.3 *Every Fréchet derivative is a Gâteaux derivative. Since the Gâteaux derivative is unique, the two derivatives are identical if the Fréchet exists.*

Definition 2.4.3 ("Uncurried" derivatives) *The spaces*

$$\mathcal{L}(X, \mathcal{L}(X, \cdots \mathcal{L}(X, \mathcal{L}(X,Y)) \cdots))$$

[with k nestings of $\mathcal{L}(X, --)$] are isomorphic to the spaces $\mathcal{L}(X^k, Y)$, and we will often make use of this fact. We use $\mathrm{d}^k T$ to denote that k-th order "uncurried" derivative of T:

$$\begin{aligned}&\mathrm{d}^k T : X \to \mathcal{L}(X^k, Y),\\&\mathrm{d}^k T(\phi)(\alpha_1, \ldots, \alpha_k) = T^{(k)}(\phi)(\alpha_1) \cdots (\alpha_k).\end{aligned}$$

Proposition 2.4.4 *The derivative of a composition is given by the following equation (shown in both curried and uncurried forms):*

$$\begin{aligned}(T \circ U)'(\phi)(\alpha) &= T'[U(\phi)][U'(\phi)(\alpha)],\\ \mathrm{d}(T \circ U)(\phi, \alpha) &= \mathrm{d}T[U(\phi), \mathrm{d}U(\phi, \alpha)].\end{aligned}$$

2.5 Taylor theorem in Banach spaces

We consider now a generalization of the familiar Taylor theorem from real analysis. It permits the expansion of a function on a Banach space in an infinite series about a point in that space. If the Banach space is a function space, this theorem permits expanding an operator around a particular function. (Thus, in field computation, it permits expanding a field transformation around a fixed field.)

Proposition 2.5.1 (Taylor) *Suppose U is any open subset of a banach space X and $T : X \to Y$ is a map which is C^n in U (that is, the first n derivatives of T exist). Let $\phi \in U$ and $\alpha \in X$ be such that $\phi + r\alpha \in U$ for all $r \in [0, 1]$. Then:*

$$T(\phi + \alpha) = \sum_{k=0}^{n-1} \frac{T^{(k)}(\phi)(\alpha)^k}{k!} + R_n(\phi, \alpha),$$

where

$$R_n(\phi, \alpha) = \int_0^1 \frac{(1-r)^{n-1} T^{(n)}(\phi + r\alpha)(\alpha)^n}{(n-1)!} \mathrm{d}r.$$

Here '$(\alpha)^k$' denotes k occurrences of the argument α. Also note that $T^{(0)} = T$.

In uncurried form the Taylor expansion is:

$$T(\phi + \alpha) = \sum_{k=0}^{n-1} \frac{\mathrm{d}^k T(\phi, \alpha, \ldots, \alpha)}{k!} + R_n(\phi, \alpha),$$

where

$$R_n(\phi, \alpha) = \int_0^1 \frac{(1-r)^{n-1} \mathrm{d}^n T(\phi + r\alpha, \alpha, \ldots, \alpha)}{(n-1)!} \mathrm{d}r,$$

and the appropriate number of α arguments (zero or more) must be supplied for $\mathrm{d}^k T$.

2.6 References

[EM] *Encyclopedic Dictionary of Mathematics*, 256A.

[M] Moore, Ramon E. (1985) *Computational Functional Analysis.* Chichester: Ellis Horwood, chs. 5, 16.