Lecture 7
Normed Spaces and Integration

7.1 Norms for Vector Spaces.

Recall that a metric space is a very simple mathematical structure. It is given by a set X together with a distance function for X, which is a mapping ρ from $X \times X \to \mathbf{R}$ that satisfies three properties:

- Positivity: $\rho(x_1, x_2) \geq 0$ with equality if and only if $x_1 = x_2$.
- Symmetry: $\rho(x_1, x_2) = \rho(x_2, x_1)$.
- Triangle Inequality: $\rho(x_1, x_3) \leq \rho(x_1, x_2) + \rho(x_2, x_3)$.

(You should think of the triangle inequality as saying that "things close to the same thing are close to each other".)

For the metric spaces that turn up in practice, X is usually some subset of a vector space V and the distance function ρ has the form $\rho(x_1, x_2) = N(x_1 - x_2)$ where the function $N : V \to \mathbf{R}$ is what is called a "norm" for V,

7.1.1 Definition. A real-valued function on a vector space V is called a *norm* for V if it satisfies the following three properties:

- Positivity: $N(v) \geq 0$ with equality if and only if $v = 0$.
- Positive Homogeneity: $N(\alpha v) = |\alpha| N(v)$.
- Triangle Inequality: $N(x_1 + x_2) \leq N(x_1) + N(x_2)$.

If N is a norm for V then we call $\rho_N(x_1, x_2) := N(x_1 - x_2)$ the *associated distance function (or metric) for* V. A vector space V together with some a choice of norm is called a *normed space*, and the norm is usually denoted by $\| \ \|$. If V is complete in the associated metric (i.e., every Cauchy sequence converges), then V is called a *Banach spacee*.

▷ **7.1—Exercise 1.** Show that if N is a norm for V then ρ_N really defines a distance function for V.

7.1.2 Remark. We have seen that for an inner-product space V, we can define a norm for V by $\|v\| := \sqrt{\langle v, v\rangle}$. Moreover, we could then recover the inner-product from this norm by using the so-called polarization identity: $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle$. It is natural to wonder if every norm "comes from an inner product" in this way, or if not which ones do. The answer is quite interesting. If we replace y by $-y$ in the polarization identity and then add the result to the original polarization identity, we get $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$. This has a nice interpretation: it says that the sum of the squares of the two diagonals of any parallelogram is equal to the sum of the squares of the four sides—and so it is called the *Parallelogram Law*. It is a pretty (and non-obvious!) fact that the Parallelogram Law is not only a necessary but also a sufficient condition for a norm to come from an inner product.

There is a remarkable fact about linear maps between normed spaces: if they have even a hint of continuity then they are very strongly continuous. To be precise:

7.1.3 Theorem. *Let V and W be normed spaces and $T : V \to W$ a linear map from V to W. Then the following are equivalent:*

1) *There is at least one point $v_0 \in V$ where T is continuous.*
2) *T is continuous at 0.*
3) *There is a constant K such that $\|Tv\| \le K \|v\|$ for all $v \in V$.*
4) *T satisfies a Lipschitz condition.*
5) *T is uniformly continuous.*

PROOF. First note that since $\|(x_n + y) - (x + y)\| = \|x_n - x\|$ it follows that $x_n \to x$ is equivalent to $(x_n + y) \to (x + y)$, a fact we will use several times

Assume 1) and let $v_n \to 0$. Then $(v_n + v_0) \to v_0$, so by 1) $T(v_n + v_0) \to Tv_0$, or by linearity, $Tv_n + Tv_0 \to Tv_0$; which is equivalent to $Tv_n \to 0$, hence 1) implies 2). Assuming 2), we can find a $\delta > 0$ such that if $\|x\| \le \delta$ then $\|T(x)\| < 1$. Now $\left\| \frac{\delta v}{\|v\|} \right\| = \delta$, so $\left\| T\left(\frac{\delta v}{\|v\|}\right) \right\| < 1$, hence $\|Tv\| < K\|v\|$ where $K = \frac{1}{\delta}$, so 2) implies 3). Since $Tv_1 - Tv_2 = T(v_1 - v_2)$, 3) implies that K is a Lipschitz constant for T, and finally 4) $\Rightarrow$ 5) $\Rightarrow$ 1) is trivial.

▷ **7.1—Exercise 2.** Show that if V and W are finite dimensional inner-product spaces then any linear map $T : V \to W$ is automatically continuous. Hint: Choose bases, and look at the matrix representation of T.

7.1.4 Remark. This theorem shows that if $T : V \to W$ is continuous, then T is bounded on the unit sphere of V. Thus $|||\, T\, ||| := \sup_{\|v\|=1} \|Tv\|$ is well-defined and is called the norm of the linear map T. It is clearly the smallest Lipschitz constant for T.

7.1.5 Extension Theorem For Continuous Linear Maps. *Let V be a normed space, U a linear subspace of V, W a Banach space, and $T : U \to W$ a continuous linear map. Then T has a unique extension to a linear map $\tilde{T} : \bar{U} \to W$ (where as usual $\bar{U}$ denotes the closure of U in V). Moreover $|||\, \tilde{T}\, ||| = |||\, T\, |||$.*

PROOF. Suppose $v \in \bar{U}$. Then there is a sequence $u_n \in U$ such that $u_n \to v$. Then u_n is a Cauchy sequence, and since T is Lipschitz it follows that Tu_n is Cauchy in W, and since W is by assumption a Banach space (complete) it follows that Tu_n converges to a limit $w \in W$, and we define $\tilde{T}v := w$. We leave further details as an exercise. ∎

▷ **7.1—Exercise 3.** If u'_n is a second sequence in U that converges to v show that Tu'_n has the same limit as Tu_n. (Hint: Consider the sequence $u_1, u'_1, u_2, u'_2, \ldots$.)

▷ **7.1—Exercise 4.** Why does $|||\, \tilde{T}\, ||| = |||\, T\, |||$?

7.2 The Space $B([a,b],V)$ of Bounded Functions on V.

7.2.1 Definition. Let V be a finite dimensional inner-product space and $[a,b] \subseteq \mathbf{R}$ a closed interval of real numbers. We denote by $B([a,b],V)$ the vector space of all bounded functions $\sigma : [a,b] \to V$, with pointwise vector operations, and we define a function $\sigma \mapsto \|\sigma\|_\infty$ on $B([a,b],V)$ by $\|\sigma\|_\infty = \sup_{a \le t \le b} \|\sigma(t)\|$.

▷ **7.2—Exercise 1.** Show that $\| \ \|_\infty$ really defines a norm for $B([a,b],V)$.

Henceforth we shall always regard $B([a,b],V)$ as a metric space with the distance function defined by the norm $\| \ \|_\infty$. But what does convergence mean in this metric space?

▷ **7.2—Exercise 2.** Let $\sigma_n : [a,b] \to V$ be a sequence in $B([a,b],V)$ and $\sigma \in B([a,b],V)$. To say that the sequence σ_n converges to σ of course means by definition that the sequence of real numbers $\|\sigma_n - \sigma\|_\infty$ converges to zero. Show that this is the case if and only if the sequence of functions $\sigma_n(t)$ converges **uniformly** to the function $\sigma(t)$ on the interval $[a,b]$, i.e., if and only if for every $\epsilon > 0$ there is a positive integer N such that the inequality $\|\sigma_n(t) - \sigma(t)\|$ holds **for all** $t \in [a,b]$ provided $n > N$.

7.2.2 Remark. A similar proof shows that the sequence σ_n is Cauchy in $B([a,b],V)$ if and only if the sequence of functions $\sigma_n(t)$ is uniformly Cauchy on $[a,b]$, i.e., if and only if for every $\epsilon > 0$ there is a positive integer N such that the inequality $\|\sigma_n(t) - \sigma_m(t)\|$ holds for all $t \in [a,b]$ provided both m and n are greater than N. Now if $\sigma_n(t)$ is uniformly Cauchy on $[a,b]$, then *a fortiori* σ_n is a Cauchy sequence in V for each $t \in [a,b]$, and since V is complete, $\sigma_n(t)$ converges to some element $\sigma(t)$ in V, and then it is easy to see that σ is in $B([a,b],V)$ and that $\sigma - \sigma_n \to 0$. This proves that $B([a,b],V)$ is a Banach space, i.e., any Cauchy sequence in $B([a,b],V)$ converges to an element of $B([a,b],V)$.

7.3 Quick and Dirty Integration

We will now see that it is quite easy to define the integral, $\int_a^b f(t)\,dt$, of a continuous map $f : [a,b] \to V$, and in fact we shall define it for a class of f considerably more general than continuous. In all of the following we assume that V is a finite dimensional inner-product space.

7.3.1 Definition. A *partition* Π of $[a,b]$ is a finite sequence $a = t_0 \le t_1 \le \ldots \le t_n = b$, and we say that a function $f : [a,b] \to V$ is *adjusted* to the partition Π if f is constant on each of the sub-intervals (t_i, t_{i+1}). We call f a *step-function* if there exists a partition to which it is adjusted, and we denote by $\mathcal{S}([a,b],V)$ the set of all step functions. Clearly $\mathcal{S}([a,b],V) \subseteq B([a,b],V)$. If f_1 and f_2 are two step-functions, and Π_i is a partition adjusted to f_i, then any partition that contains all the points of Π_1 and Π_2 is adjusted to both f_1 and f_2 and hence to any linear combination of them. This shows that $\mathcal{S}([a,b],V)$ **is a linear subspace of** $B([a,b],V)$.

7.3.2 Definition. If $f : [a,b] \to V$ is a step-function and if f is adjusted to a partition $\Pi = t_0 \le t_1 \le \ldots \le t_n$, then we define its *integral*, $\mathcal{I}(f) \in V$, by $\mathcal{I}(f) := \sum_{i=1}^n (t_i - t_{i-1}) v_i$, where v_i is the constant value of f on the interval (t_{i-1}, t_i).

▹ **7.3—Exercise 1.** Show that the integral $\mathcal{I} : \mathcal{S}([a,b],V) \to V$ is a well-defined linear map and satisfies $\|\mathcal{I}(f)\| \le (b-a)\,\|f\|_\infty$, so that $\mathcal{I}$ is continuous and has norm $|||\,\mathcal{I}\,||| = (b-a)$. Hint: The only (slightly) tricky point is to show that $\mathcal{I}(f)$ does not depend on the choice of a partition to which f is adjusted. Reduce this to showing that when you subdivide one subinterval of a partition, the integral does not change.

We can now use our Extension Theorem For Continuous Linear Maps to conclude that $\mathcal{I}$ has a unique extension to a continuous linear map $\mathcal{I} : \bar{\mathcal{S}}([a,b],V) \to V$, where $\bar{\mathcal{S}}([a,b],V)$ denotes the closure of the step functions in the space $B([a,b],V)$ of bounded functions. For $f \in \bar{\mathcal{S}}([a,b],V)$ we will also denote its integral, $\mathcal{I}(f)$, by $\int_a^b f(t)\,dt$, and we will refer to elements of $\bar{\mathcal{S}}([a,b],V)$ as *integrable* functions. Note that according to the extension theorem, the inequality $\left\|\int_a^b f(t)\,dt\right\| \le (b-a)\,\|f\|_\infty$ continues to hold for any integrable function f.

7.3.3 Proposition. *Integration commutes with linear maps. That is, if $T : V \to W$ is linear and $f : [a,b] \to V$ is integrable, then $T \circ f : [a,b] \to W$ is integrable and $T(\int_a^b f(t)\,dt) = \int_a^b T(f(t))\,dt$.*

PROOF. The Proposition is obvious for step functions, and it then follows for integrable functions by the uniqueness of the extension.

It is natural to wonder just how inclusive the space of integrable functions is. We note next that it contains all continuous functions.

7.3.4 Proposition. *The space $C([a,b],V)$ of continuous maps of $[a,b]$ into V is a linear subspace of the space $\bar{\mathcal{S}}([a,b],V)$ of integrable functions.*

PROOF. By a standard result of analysis, if $f \in C([a,b],V)$ then f is uniformly continuous. That is, given $\epsilon > 0$, there is a $\delta > 0$ such that if t_1 and t_2 are in $[a,b]$ and $|t_1 - t_2| < \delta$, then $\|f(t_1) - f(t_2)\| < \epsilon$. Choose an integer N so large that $\frac{b-a}{N} < \delta$ and partition $[a,b]$ into N equal subintervals, and let ϕ be the step function that is constant on each subinterval of this partition and agrees with f at the left endpoint. Then clearly $\|f - \phi\|_\infty < \epsilon$, proving that f is in the closure of $\mathcal{S}([a,b],V)$ ∎

▹ **7.3—Exercise 2.** The "standard result of analysis" says that, if X is a compact metric space then a continuous map $f : X \to Y$ is uniformly continuous. Prove this. Hint: One approach is to prove the "contrapositive" statement, i.e., show that if X is **not** uniformly continuous then there is at least one point x of X where f is not continuous. For this, pick an ϵ so that for each integer n there are two points x_n and x'_n with $\rho(x_n, x'_n) < \frac{1}{n}$ but $\rho(f(x_n), f(x'_n)) > \epsilon$. By compactness, a subsequence x_{n_k} converges to some point x. Show that x'_{n_k} also converges to x and deduce that f is not continuous at x.

7.3.5 Remark. The above proof generalizes to show that even a piecewise continuous $f : [a,b] \to V$ is integrable, where $f : [a,b] \to V$ is called piecewise continuous if there is a partition of $[a, b]$ such that f is continuous on each of its open subintervals, with a limit at each endpoint of the subinterval.

▷ **7.3—Exercise 3.** Show that if $f : [a, b] \to V$ is integrable then $\left\| \int_a^b f(t)\, dt \right\| \leq \int_a^b \|f(t)\|\, dt$. Hint: For a step function, this follows from the triangle inequality for norms.

▷ **7.3—Exercise 4.** . Show that if $f : [a, b] \to V$ is integrable, and $a < c < b$, then $\int_a^b f(t)\, dt = \int_a^c f(t)\, dt + \int_c^b f(t)\, dt$

▷ **7.3—Exercise 5.** Let $f : [a, b] \to V$ be continuous and define $F : [a, b] \to V$ by $F(t) := \int_a^t f(s)\, ds$. Show that F is differentiable and that $F' = f$.
Hint: $\frac{1}{h}[F(t_0 + h) - F(t_0)] - f(t_0) = \frac{1}{h} \int_{t_0}^{t_0+h} (f(s) - f(t_0))\, ds$.

We next prove the vector version of the Fundamental Theorem of Integral Calculus (allowing us to evaluate a definite integral if we know an anti-derivative for the integrand). As in the classic case it depends on knowing that only constant functions have a zero derivative.

7.3.6 Lemma. *If $f : [a, b] \to V$ is differentiable and f' is identically zero, then f is constant.*

PROOF. This can be reduced to the classic special case that $V = \mathbf{R}$ (proved in elementary Calculus as an easy consequence of Rolle's Theorem). In fact, if $\ell \in V^*$, then $(\ell \circ f)' = \ell \circ f' = 0$ so $(\ell \circ f)$ is constant by the special case. But since this holds for *every* $\ell \in V^*$ it follows easily that f itself must be constant. ∎

▷ **7.3—Exercise 6.** Prove the vector version of the Fundamental Theorem of Integral Calculus, That is, let $f : [a, b] \to V$ be continuous and let $\Phi : [a, b] \to V$ be an antiderivative of f, i.e., Φ is differentiable and $\Phi' = f$. Show that $\int_a^b f(t)\, dt = \Phi(b) - \Phi(a)$. Hint: If we define $F(t) := \int_a^t f(s)\, ds$, then we know F and Φ have the same derivative, so $F - \Phi$ has derivative 0, and by the lemma F and Φ differ by a constant vector.

7.3.7 Finite Difference Formula. *Let O be a convex open set in V and let $F : O \to W$ be differentiable. If $v_0, v_1 \in O$, then $F(v_1) - F(v_0) = \int_0^1 DF_{\sigma(t)}(v_1 - v_0)\, dt$, where $\sigma(t) = v_0 + t(v_1 - v_0)$, $0 \leq t \leq 1$ is the line segment joining v_0 to v_1.*

PROOF. By the chain rule, $(F \circ \sigma)'(t) = DF_{\sigma(t)}(\sigma'(t))$, and clearly $\sigma'(t) = v_1 - v_0$. Thus $(F \circ \sigma)$ is an anti-derivative for $DF_{\sigma(t)}(v_1 - v_0)$, and the result follows from the Fundamental Theorem.

7.3.8 Corollary. *Let O be a convex open set in V and let $F : O \to W$ be continuously differentiable. Then F satisfies a Lipschitz condition on any closed, bounded subset S of O. In fact, if K is an upper bound for $|||\, DF_p \,|||$ for $p \in S$, then K is a Lipschitz bound for F.*

▷ **7.3—Exercise 7.** Use the Finite Difference Formula to prove the corollary.

7.4 Numerical Integration (or Quadrature Rules)

Since one usually cannot find an anti-derivative for an integrand in closed form, it is important to be able to "evaluate an integral numerically"—meaning approximate it with arbitrary precision. In fact, this is so important that there are whole books devoted the study of numerical integration methods (aka quadrature rules). We will consider only two such methods, one known as the Trapezoidal Rule and the other as Simpson's Rule. In what follows, we will assume that the integrand f is always at least continuous, but for the error estimates that we will mention to be valid, we will need f to have several continuous derivatives.

7.4.1 Definition. By a *quadrature rule* we mean a function M that assigns to each continuous function $f : [a, b] \to V$ (mapping a closed interval $[a, b]$ into an inner-product space V) a vector $M(f, a, b) \in V$—which is supposed to be an approximation of the integral, $\int_a^b f(t)\, dt$. A particular quadrature rule M is usually given by specifying a linear combination of the values of f at certain points of he interval $[a, b]$; that is, it has the general form $M(f, a, b) := \sum_{i=1}^{n} w_i f(t_i)$, where the points $t_i \in [a, b]$ are called the *nodes* of M and the scalars w_i are called its *weights*. The *error* of M for a particular f and $[a, b]$ is defined as $\mathrm{Err}(M, f, a, b) := \left\| \int_a^b f(t)\, dt - M(f, a, b) \right\|$.

7.4—Example 1. The Trapezoidal Rule: $M^T(f, a, b) := \frac{b-a}{2}[f(a) + f(b)]$.
In this case, there are two nodes, namely the two endpoints of the interval, and they have equal weights, namely half the length of the interval. Later we shall see the origin of this rule (and explain its name).

7.4—Example 2. Simpson's Rule: $M^S(f, a, b) := \frac{b-a}{6}[f(a) + 4f(\frac{a+b}{2}) + f(b)]$.
So now the nodes are the two endpoints, as before, and in addition the midpoint of the interval. And the weights are $\frac{b-a}{6}$ for the two endpoints and $\frac{2(b-a)}{3}$ for the midpoint.

7.4.2 Remark. Notice that in both examples the weights add up to $b - a$. This is no accident; any "reasonable" quadrature rule should have a zero error for a constant function, and this easily implies that the weights must add to $b - a$.

7.4.3 Proposition. *If $f : [a, b] \to V$ has two continuous derivatives, and $\|f''(t)\| < C$ for all $t \in [a, b]$ then $\mathrm{Err}(M^T, f, a, b) \le C\frac{(b-a)^3}{12}$. Similarly, if $f : [a, b] \to V$ has four continuous derivatives, and $\|f''''(t)\| < C$ for all $t \in [a, b]$ then $\mathrm{Err}(M^S, f, a, b) \le C\frac{(b-a)^5}{90}$*

7.4.4 Remark. The proof of this proposition is not difficult—it depends only the Mean Value Theorem—but it can be found in any numerical analysis text and will not be repeated here.

7.4.5 Definition. If M is a quadrature rule then we define a sequence M_n of *derived* quadrature rules by $M_n(f, a, b) := \sum_{i=0}^{n-1} M(f, a + ih, a + (i + 1)h)$ where $h = \frac{b-a}{n}$. We say that the rule M is *convergent* for f on $[a, b]$ if the sequence $M_n(f, a, b)$ converges to $\int_a^b f(t)\, dt$.

In other words, to estimate the integral $\int_a^b f(t)\,dt$ using the n-th derived rule M_n, we simply divide the interval $[a, b]$ of integration into n equal sub-intervals, estimate the integral on each sub-interval using M, and then add these estimates to get the estimate of the integral on the whole interval.

7.4.6 Remark. We next note an interesting relation between the errors of M and of M_n. Namely, with the notation just used in the above definition, we see that by the additivity of the integral, $\int_a^b f(t)\,dt = \sum_{i=0}^{n-1} \int_{a+ih}^{a+(i+1)h} f(t)\,dt$, hence from the definition of M_n and the triangle inequality, we have $\mathrm{Err}(M_n, f, a, b) \le \sum_{i=0}^{n-1} \mathrm{Err}(M, f, a+ih, a+(i+1)h)$. We can now use this together with Proposition 7.4.3 to prove the following important result:

7.4.7 Theorem. *If $f : [a, b] \to V$ has two continuous derivatives, and $\|f''(t)\| < C$ for all $t \in [a, b]$ then $\mathrm{Err}(M_n^T, f, a, b) \le C\frac{(b-a)^3}{12n^2}$. Similarly, if $f : [a, b] \to V$ has four continuous derivatives, and $\|f''''(t)\| < C$ for all $t \in [a, b]$ then $\mathrm{Err}(M_n^S, f, a, b) \le C\frac{(b-a)^5}{90n^4}$*

▷ **7.4—Exercise 1.** Fill in the details of the proof of this theorem.

7.4.8 Remark. This shows that both the Trapezoidal Rule and Simpson's Rule are convergent for any reasonably smooth function. But it also shows that Simpson's Rule is far superior to the Trapezoidal Rule. For just fifty per cent more "effort" (measured by the number of evaluations of f) one gets a far more accurate result.

Where did the formulas for the Trapezoidal Rule and Simpson's Rule come from? It helps to think of the classical case of a real-valued function f, so we can regard $\int_a^b f(t)\,dt$ as representing the area under the graph of f between a and b. Now, if f is differentiable and the interval $[a, b]$ is short, then the graph of f is well-approximated by the straight line segment joining $(a, (f(a))$ to $(b, f(b))$, so the area under of the graph of f should be well-approximated by the area between the x-axis and this line segment. The latter area is of course a trapezoid (whence the name) and it has the area given by the Trapezoidal Rule formula. Simpson's Rule arises if instead of interpolating f by a linear function that agrees with f at a and b we instead interpolate by a quadratic function $cx^2 + dx + e$ that agrees with f at a and b and also at the mid-point $\frac{a+b}{2}$.

▷ **7.4—Exercise 2.** Using the method of "undetermined coefficients", show that there is a unique choice of coefficients c, d, e such that the quadratic polynomial $cx^2 + dx + e$ agrees with the function f at the three points a, b and $\frac{a+b}{2}$. Find c, d, and e explicitly and integrate the polynomial from a to b, and check that this gives $M^S(f, a, b)$.

7.5 Second Matlab Project.

The second Matlab project is to develop Matlab code to implement the Trapezoidal Rule and Simpson's Rule, and then to do some experimentation with your software, checking that the error estimates of theorem 7.4.7 are satisfied for some test cases where the function f has a known anti-derivative and so can be evaluated exactly. In more detail:

1) Write a Matlab function M-file defining a function TrapezoidalRule(f,a,b,n). This should return the value of $M_n^T(f,a,b)$. Here of course the parameters a and b represent real numbers and the parameter n a positive integer. But what about the parameter f, i.e., what should it be legal to substitute for f when the TrapezoidalRule(f,a,b,n) is called? Answer: f should represent a function of a real variable whose values are arrays (of some fixed size) of real numbers. The function that you are permitted to substitute for f should either be a built-in Matlab function (such as sin) or an inline function in the Matlab Workspace, or a function that is defined in some other M-File.

2) Write a second Matlab function M-file defining a function SimpsonsRule(f,a,b,n) that returns $M_n^S(f,a,b)$.

3) Recall that $\int_0^t \frac{dx}{1+x^2} = \arctan(t)$, so that in particular $\int_0^1 \frac{4\,dx}{1+x^2} = 4\arctan(1) = \pi$. Using the error estimates for the Trapezoidal Rule and Simpson's Rule, calculate how large n should be to calculate π correct to d decimal places from this formula using Trapezoidal and Simpson. Set format long in Matlab and get the value of π to fifteen decimal places by simply typing pi. Then use your Trapezoidal and Simpson functions from parts 1) and 2) to see how large you actually have to choose n to calculate π to 5, 10, and 15 decimal places.

4) Be prepared to discuss your solutions in the Computer Lab.