C Topology and Metric Spaces

C.1 Topology

OPERATORS act on elements of sets. In appendix B, the set complement operator was defined with respect to a superset S. Furthermore, the definitions of open and closed sets were predicated on one definition: the open neighborhood. Now we are going to reverse things. An open neighborhood will be defined in terms of open sets, and a topological space will be defined in terms of its set of elements and its open sets. This appendix is meant to be introductory. See, e.g., [9] for a complete discussion of these topics.

DEFINITION C.1.1 (Topology) *A* topological space *is a set S together with a collection $\mathcal{O}$ of subsets called* open sets *such that*

- $\emptyset \in \mathcal{O}$ *and* $S \in \mathcal{O}$,
- *if* $U_1, U_2 \in \mathcal{O}$, *then* $U_1 \bigcap U_2 \in \mathcal{O}$,
- *the union of any collection of open sets is an open set.*

Open sets can be arbitrarily designed as long as they satisfy the above three properties. The *standard topology* on $\mathbb{R}^m$ has $S = \mathbb{R}^m$ with $\mathcal{O}$ containing $\mathbb{R}^m$, the empty set $\emptyset$, all open rectangles, and their unions. An example is the real line with open intervals, i.e., $S = \mathbb{R}$, with $\mathcal{O}$ consisting of any open interval, the union of open intervals, $\mathbb{R}$, and $\emptyset$. To show this we look to the three conditions in definition C.1.1:

- $\mathbb{R}, \emptyset \in O$ by definition,
- $(a, b) \in O$ and $(c, d) \in O$, so

$$(a, b) \bigcap (c, d) = \begin{array}{rl} (c, b) \in O & \text{or,} \\ (a, d) \in O & \text{or,} \\ \emptyset \in O, & \end{array}$$

- any finite or infinite union of open intervals is an open interval.

The *trivial topology* on a set S consists of $O = \{\emptyset, S\}$. The *discrete topology* of a set S is defined by $O = \{A \mid A \subset S\}$. That is, the open sets are everything.

Now the definition of the open neighborhood stems from the definition of open sets. The definitions of closed sets, closure, boundary, interior, and denseness remain the same.

DEFINITION C.1.2 *A* neighborhood *of a point x, denoted* nbhd(x)*, is an open set that contains x.*

C.2 Metric Spaces

The open sets of a topological space can be constructed using a distance function. In $\mathbb{R}^m$, the standard Euclidean distance function

$$d(x, y) = \left(\sum_{i=1}^{m} (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

defines open sets that are open balls. More generally,

DEFINITION C.2.1 (Metric Space) *Let M be a set. A* metric *on M is a function $d : M \times M \to \mathbb{R}^{\geq 0}$ such that for all $m_1, m_2, m_3 \in M$,*

1. *(Definiteness) $d(m_1, m_2) = 0$ if and only if $m_1 = m_2$*
2. *(Symmetry) $d(m_1, m_2) = d(m_2, m_1)$, and*
3. *(Triangle inequality) $d(m_1, m_3) \leq d(m_1, m_2) + d(m_2, m_3)$.*

A metric space *is the pair (M, d).*

Note that the intuitive notion that distance must be non-negative follows directly from the three conditions above. Specifically, condition 3 allows us to write $d(m_1, m_1) \leq d(m_1, m_2) + d(m_2, m_1)$. The left-hand side of this expression is zero by condition 1 and the right-hand side is $2d(m_1, m_2)$ by condition 2, yielding $d(m_1, m_2) \geq 0$.

For $\epsilon > 0$ and $m \in M$, the *open ball* centered at m is defined to be

$$B_\epsilon(m) = \{n \in M \mid d(m, n) < \epsilon\}.$$

The set of all open balls and the union of open balls forms the *metric topology* on the metric space (M, d).

There are many distance functions other than the standard Euclidean metric. For example, the *Manhattan distance metric* is defined to be

$$d(x, y) = \sum_{i=1}^{m} |x_i - y_i|.$$

This metric is so named because it measures how far a taxicab must drive in a city grid to get from one location to another. Different metrics can be used to induce the same topology. The Manhattan and standard Euclidean metrics induce the same topology. Two metrics induce the same topology if, for any open ball at x by the first metric, there is an open ball by the second metric contained completely in the first ball, and vice versa.

C.3 Normed and Inner Product Spaces

A metric space is a special case of a topological space. Next we introduce a normed space, which is a special case of a metric space. We also introduce an inner product space, which is a special case of a normed space.

DEFINITION C.3.1 *A* normed space *E is a subset of a metric space M that has an operator $\|\cdot\| : E \to \mathbb{R}$ such that*

- $\|e\| \geq 0$ *for all $e \in E$, and $\|e\| = 0$ if and only if e is the zero vector (positive definiteness),*
- $\|\lambda e\| = |\lambda| \|e\|$ *for all $e \in E$ and $\lambda \in \mathbb{R}$ (homogeneity),*
- $\|e_1 + e_2\| \leq \|e_1\| + \|e_2\|$ *for all $e_1, e_2 \in E$ (triangle inequality).*

The norm can be used to define the open sets and induce a metric. A sequence $\{x_1, x_2, x_3, \ldots\}$ is said to be a *Cauchy sequence* if for any $\epsilon > 0$ there exists an integer k such that $\|x_i - x_j\| < \epsilon$ for all $i, j > k$. When a normed space has a corresponding metric for which every Cauchy sequence converges to a point in the space, we term this space a *Banach space.*

DEFINITION C.3.2 *An* inner product *on a real vector space E is a mapping $\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$ such that*

- $\langle e, e_1 + e_2 \rangle = \langle e, e_1 \rangle + \langle e, e_2 \rangle$,
- $\langle e, \alpha e_1 \rangle = \alpha \langle e, e_1 \rangle$,
- $\langle e_2, \alpha e_1 \rangle = \langle e_1, \alpha e_2 \rangle$,
- $\langle e, e \rangle \geq 0$ *and* $\langle e, e \rangle = 0$ *if and only if e is zero.*

An inner product induces the norm $\|e\| = \langle e, e \rangle$, and a norm in turn induces a metric. When an inner product space has a corresponding metric for which every Cauchy sequence converges, we call this space a *Hilbert space*.

C.4 Continuous Functions

Paths are defined in terms of a continuous function. Let $f : S \to T$ be a mapping from the *domain* S to the *range* T. The points $f(s)$ are the *values* of f, where $s \in S$. If $U \subset S$, then the *image* of U under f is denoted $f(U) = \{f(x) \in T \mid x \in U\}$. If $V \subset T$, then the *preimage* of V under f is denoted $f^{-1}(V) = \{x \in S \mid f(x) \in V\}$. First, we introduce an abstract notion of a continuous function and then specialize it for metric spaces.

DEFINITION C.4.1 *Let S and T be topological spaces and $f : S \to T$ be a mapping. f is* continuous *at $u \in S$ if for every $V = \mathrm{nbhd}(f(u))$ there is a $U = \mathrm{nbhd}(u)$ such that $f(U) \subset V$. The mapping f is continuous if for every open subset $V \subset T$, $f^{-1}(V) = \{u \in S \mid f(u) \in V\}$ is open in S.*

Essentially, a continuous function is a function where the preimage of an open set is an open set. Now we introduce the standard “delta-epsilon” method for defining continuous functions on metric spaces: The function f is continuous at s if for every $\epsilon > 0$ there exists a $\delta > 0$ where

(C.1) $$d(x, s) < \delta \text{ implies } d(f(x), f(s)) < \epsilon.$$

EXAMPLE C.4.2 (Continuous Function) *Let* $f : \mathbb{R} \to \mathbb{R}$ *be defined as* $f(x) = x^2$. *In order to show that* f *is continuous at a point* s, *we must find a* $\delta > 0$ *such that* $d(f(x), f(s)) < \epsilon$ *for arbitrarily small* $\epsilon > 0$. *Note that in* $\mathbb{R}$ *the distance function is* $d(x, s) = |x - s|$. *First, we study the quantity* $|f(x) - f(s)|$*:*

$$\begin{aligned} |f(x) - f(s)| &= |x^2 - s^2| \\ &= |x - s||x + s| \\ &= |x - s||x - s + 2s| \end{aligned}$$

Using the triangle inequality, we get

$$|f(x) - f(s)| \leq |x - s|(|x - s| + 2|s|).$$

Now we can subsitute $|x - s| < \delta$ *to see that* $|f(x) - f(s)|$ *will be less than* ϵ *if*

$$\delta(\delta + 2|s|) < \epsilon.$$

Using the quadratic formula, we see that this inequality can be satisfied for

$$\delta < -|s| + \sqrt{s^2 + \epsilon}.$$

The term on the right-hand side of this inequality is positive, so we can find a suitable δ. *This proves that the function* $f(x) = x^2$ *is continuous at any point* $s \in \mathbb{R}$. *Note the that choice of* δ *depends on both* s *and* ϵ.

The set of continuous functions is denoted C^0. If the derivative of a continuous function f is continuous, then f is said to be differentiable and belongs to a set denoted C^1. If c is k-wise differentiable, then it belongs to a set denoted C^k. If all derivatives of f exist, then f belongs to C^∞ and f is said to be *smooth*. While a path is only required to be of class C^0, a trajectory must belong to C^k, $k > 0$, to allow the definition of velocity, acceleration, etc., at all points where the system is moving.

The following are equivalent statements:

$$\begin{aligned} f : S \to T \text{ is continuous.} \quad &\Longleftrightarrow \quad f(\mathrm{cl}(A)) \subset \mathrm{cl}(f(A)) \text{ for } A \subset S \\ &\Longleftrightarrow \quad f^{-1}(\mathrm{int}(B)) \subset \mathrm{int}(f^{-1}(B)) \text{ for } B \subset T \end{aligned}$$

Finally, another useful property of continuous functions is that things "change" continuously. Specifically, if the scalar functions f and g are continuous at x and $f(x) < g(x)$, then there exists a nbhd(x) such that for all $y \in$ nbhd(x), $f(y) < g(y)$.

C.5 Jacobians and Gradients

Consider a vector-valued function $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ where f can be written

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix},$$

where $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$ for $i \in \{1, 2, \ldots, n\}$.

We define the *differential* of f to be the matrix[1]

$$Df = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}.$$

The matrix Df is denoted in a number of different ways. It is sometimes called the *Jacobian* of f and denoted J (see chapter 3). It is sometimes called the *tangent map* of f and denoted Tf. Sometimes it is necessary to specify which variables are used in the differentiation. Hence the differential can also be denoted $\frac{\partial f}{\partial x}$. Putting the variable name in the subscript serves a similar purpose. The symbols $D_x f$, J_x, and $T_x f$ all denote the differential of f with respect to the variable x.

Given a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, the *gradient* of g is defined to be

$$\nabla g = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \vdots \\ \frac{\partial g}{\partial x_n} \end{bmatrix}.$$

As in the case of the differential, the notation $\nabla_x g$ is sometimes used to make explicit the fact that g is differentiated with respect to x. The vector $\nabla g(x)$ points in the direction that maximally increases the function at the point x. Note that by this defintition $\nabla f(x) = Df^T$. The decision as to whether the gradient should be a row vector or a column vector is somewhat arbitrary. In this book we define it as a column vector because that is the convention commonly used in the robotics community when discussing planning algorithms based on artificial potential fields.

1. To be technically accurate, the differential is actually a map from the tangent space of $\mathbb{R}^m$ (which happens to also be $\mathbb{R}^m$) to the tangent space of $\mathbb{R}^n$ (which is $\mathbb{R}^n$). For the purposes of this appendix, we simply represent Df as an $n \times m$ matrix.

Let $c(t)$ be a *smooth curve* in $\mathbb{R}^n$, i.e., c is a C^∞, vector-valued map $c : \mathbb{R} \to \mathbb{R}^n$. If t is time, the derivative

$$\dot{c}(t) = \frac{dc}{dt}(t) = \begin{bmatrix} \frac{dc_1}{dt}(t) \\ \frac{dc_2}{dt}(t) \\ \vdots \\ \frac{dc_n}{dt}(t) \end{bmatrix}$$

can be thought of as the velocity of a point moving along $c(t)$.

For a real-valued function $g : \mathbb{R}^n \to \mathbb{R}$, one is often interested in how the value of the function g changes as the state follows the trajectory $c(t)$. This is the same as finding the derivative of $g(c(t))$ with respect to t, $\dot{g}(t) = \frac{d}{dt}(g \circ c)(t)$, where $(g \circ c)(t) = g(c(t))$ is called the *composition* of g and c. To calculate $\dot{g}$ we can use the *chain rule,* which can be stated in a number of different ways:

$$\begin{aligned} \frac{d}{dt}(g \circ c)(t) &= \sum_{i=1}^{n} \frac{\partial g}{\partial c_i} \frac{dc}{dt}(t) \\ &= \frac{\partial g}{\partial c} \dot{c}(t) \\ &= D_c g \dot{c}(t) \end{aligned}$$

Note here that $\frac{\partial g}{\partial c_i}$ denotes the partial derivative of g with respect to x_i evaluated at $x_i = c_i(t)$. Likewise, $\frac{\partial g}{\partial c}$ and $D_c g$ denote the differential of g with respect to x evaluated at $x = c(t)$. When it is necessary to be explicit, these quantities are sometimes denoted as $\frac{\partial g}{\partial x}|_{x=c(t)}$ or $D_x g|_{x=c(t)}$.

Sometimes it is useful to be able to see how g changes when the curve c is represented in different coordinates than those for which g is defined. Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be the smooth change of coordinates that maps from the coordinates in which c is defined (y-coordinates) into the x-coordinates required by g, i.e., $x = g(y)$. So we are interested in seeing how $g(h(c(t))) = (g \circ h \circ c)(t)$ evolves with t. Again, the chain rule allows us to express this quantity in a number of different ways:

$$\begin{aligned} \frac{d}{dt}(g \circ h \circ c)(t) &= \left.\frac{\partial g}{\partial x}\right|_{x=(h\circ c)(t)} \left.\frac{\partial h}{\partial y}\right|_{y=c(t)} \dot{c}(t) \\ &= \underbrace{\frac{\partial g}{\partial h}}_{(1\times n)} \underbrace{\frac{\partial h}{\partial c}}_{(n\times n)} \underbrace{\dot{c}(t)}_{(n\times 1)} \\ &= D_h g D_c h \dot{c}(t) \\ &= D_c(f \circ g)\dot{c}(t) \end{aligned}$$

In the second of these expressions the dimension of each of the three objects on the right-hand side is written below the underbrace. This is to make it clear that the dimensions are suitable for matrix multiplication and that the resulting product is indeed a scalar.

EXAMPLE C.5.1 *Consider the curve $c : \mathbb{R} \to \mathbb{R}^2$ defined in polar coordinates*

$$c(t) = \begin{bmatrix} r(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} 2 \\ 2\pi t \end{bmatrix}.$$

Note that this curve corresponds to a point moving around a circle of constant radius 2 at a velocity of 4π, i.e., the point travels around the circle once every second. Now consider a function $g : \mathbb{R}^2 \to \mathbb{R}$ defined in Cartesian coordinates

$$g(x) = x_1.$$

In order to compute $\dot{g}$, we introduce a coordinate change $h : \mathbb{R}^2 \to \mathbb{R}^2$ that maps the vector $y = [y_1,\ y_2]^T = [r,\ \theta]^T$ in polar coordinates into Cartesian coordinates:

$$h(y) = \begin{bmatrix} y_1 \cos(y_2) \\ y_1 \sin(y_2) \end{bmatrix}$$

Using the chain rule we get

$$\begin{aligned} \frac{d}{dt}(g \circ h \circ c)(t) &= \frac{\partial g}{\partial h}\frac{\partial h}{\partial c}\dot{c}(t) \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \cos(y_2) & -y_1 \sin(y_2) \\ \sin(y_2) & y_1 \cos(y_2) \end{bmatrix}_{y=[2,2\pi t]^T} \begin{bmatrix} 0 \\ 2\pi \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \cos(2\pi t) & -2\sin(2\pi t) \\ \sin(2\pi t) & 2\cos(2\pi t) \end{bmatrix} \begin{bmatrix} 0 \\ 2\pi \end{bmatrix} \\ &= -4\pi \sin(2\pi t). \end{aligned} \tag{C.2}$$

This can be checked by differentiating $(g \circ h \circ c)(t) = 2\cos(2\pi t)$ directly to get the same answer.

The chain rule can be used in a similar manner to differentiate compositions of functions of any compatible dimension. For example, if we redefine the functions h and g so that $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$, then the chain rule gives the derivative of the composition

$$\underbrace{\frac{d}{dt}(g \circ h \circ c)(t)}_{(p\times 1)} = \underbrace{\frac{\partial g}{\partial h}}_{(p\times m)} \underbrace{\frac{\partial h}{\partial c}}_{(m\times n)} \underbrace{\dot{c}(t)}_{(n\times 1)}.$$

A remark about rows and columns: In mechanics, a force vector F is usually represented by a row vector as it is a member of the cotangent space (see chapter 12). Velocity vectors belong to the tangent space and are usually represented as column vectors, e.g., v. This allows us to easily take the product Fv to get power, which is a scalar value. Many mechanics texts use up-down indicial notation to facilitate this, but such notation is not required for this book.