Manifolds

Mladen Victor WICKERHAUSER

Washington University in St. Louis, Missouri
victor@wustl.edu
http://www.math.wustl.edu/~victor

Dimensionality Reduction and Manifold Estimation
PMF — University of Zagreb
Winter, 2022

Euclidean Vector Spaces

Euclidean d-space, $\mathbf{E}^d$, has these properties:

- Dimension: $d \in \mathbf{Z}^+$, finite but it could be large.
- Set $\mathbf{R}^d \stackrel{\text{def}}{=} \{\mathbf{x} \stackrel{\text{def}}{=} (x_1, \ldots, x_d) : x_i \in \mathbf{R}, i = 1, \ldots, d\}$.
- Linearity: $(\forall \mathbf{x}, \mathbf{y} \in \mathbf{R}^d)(\forall c \in \mathbf{R})$,
  $\mathbf{x} + c\mathbf{y} \stackrel{\text{def}}{=} (x_1 + cy_1, \ldots, x_d + cy_d) \in \mathbf{R}^d$.
- Norm: $\|\mathbf{x}\| \stackrel{\text{def}}{=} \sqrt{x_1^2 + \cdots x_d^2} \geq 0$.
  $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0} \stackrel{\text{def}}{=} (0, \ldots, 0)$.
- Inner product: $\langle \mathbf{x}, \mathbf{y} \rangle \stackrel{\text{def}}{=} x_1 y_1 + \cdots + x_d y_d$. Then $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.

Exercise: $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$. When is there equality?

Topology

A *topological space* is a set X with a *topology* $\mathcal{T}$, a collection of subsets called *open*, satisfying:

- For any index set I and collection $\{G_\alpha : \alpha \in I\} \subset \mathcal{T}$, the union is open: $\cup_{\alpha \in I} G_\alpha \in \mathcal{T}$.
- For any *finite* collection $\{G_1, \ldots, G_N\} \subset \mathcal{T}$, the intersection is open: $\cup_{i=1}^{N} G_i \in \mathcal{T}$.

Also, $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$, so $\mathcal{T}$ is nonempty.

Write $(X, \mathcal{T})$ to indicate the topology $\mathcal{T}$, since topological space X may have more than one.

If $Y \subset X$, then $(Y, \mathcal{T}_Y)$ is a topological space with the convention $\mathcal{T}_Y \stackrel{\text{def}}{=} \{G \cap Y : G \in \mathcal{T}\}$. This $\mathcal{T}_Y$ is called the *relative topology*.

Concepts from Topology

Let $(X, \mathcal{T})$ be a topological space.

- *Dense subset:* $Y \subset X$ is *dense* if $X \setminus Y$ contains no open sets.
- *Separable space:* X contains a *countable* dense subset.
- *Hausdorff space:* For any $x, y \in X$ with $x \neq y$, there exist disjoint $G, H \in \mathcal{T}$ with $x \in G$ and $y \in H$.
- *Neighborhood of $x \in X$:* subset $V \subset X$ with $x \in V$ and $(\exists G \in \mathcal{T})\ x \in G \subset V$.
- *First countable space:* For each $x \in X$, there exist $\{G_1, G_2, \dots\} \subset \mathcal{T}$, such that for every neighborhood V of x, there is some i such that $x \in G_i \subset V$.
- *Second countable:* There exists a countable *base* $\mathcal{B} \subset \mathcal{T}$ that *generates* $\mathcal{T}$, namely every $G \in \mathcal{T}$ is a union of elements of $\mathcal{B}$.

Exercise: (a) Second countable implies first countable. (b) Second countable implies separable.

Metric Topology

Metric space: set X with *distance function* $\mathrm{d} : X \times X \to \mathbf{R}$ satisfying:

- $\mathrm{d}(x, y) \geq 0$;
- $\mathrm{d}(x, y) = 0 \iff x = y$;
- $\mathrm{d}(x, y) = \mathrm{d}(y, x)$;
- $\mathrm{d}(x, z) \leq \mathrm{d}(x, y) + \mathrm{d}(y, z)$.

Open balls: $B(x, r) \stackrel{\text{def}}{=} \{y \in X : \mathrm{d}(x, y) < r\}$, $x \in X$ and $r > 0$.

Metric topology $\mathcal{T}$ is all open balls and all unions of open balls.

Exercise: (a) A metric space is a first countable Hausdorff topological space. (b) A separable metric space is second countable.

Open Covers and Compactness

Let $(X, \mathcal{T})$ be a topological space.

- An *open cover* of X is a collection of open sets $\{G_\alpha : \alpha \in I\} \subset \mathcal{T}$ such that $X \subset \bigcup_I G_\alpha$.
- A *subcover* of $\{G_\alpha : \alpha \in I\}$ is given by $I' \subset I$ satisfying $X \subset \bigcup_{I'} G_\alpha$.
- A subcover $\{G_\alpha : \alpha \in I'\}$ is called *countable* if I' is countable, and *finite* if I' is finite.

Definition

Topological space X is *compact* iff every open cover of X has a finite subcover.

Exercise: (Lindelöf) If X is a separable metric space, then every open cover of X has a countable subcover.

Finite Dimensional Euclidean Space

$\mathbf{E}^d$ is a metric space with $\mathrm{d}(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \|\mathbf{x} - \mathbf{y}\|$.

Metric topology $\mathcal{T}$ for $\mathbf{E}^d$ contains all finite intersections of open balls: Put $G = B(x, r)$ and $H = B(y, s)$. Then

$$\begin{aligned} G \cap H &= \{z \in X : \|z - x\| < r, \|z - y\| < s\} \\ &= \bigcup_{z \in G \cap H} B(z, t_z), \end{aligned}$$

where $t_z \stackrel{\text{def}}{=} \min(r - \|z - x\|, s - \|z - y\|)$ for each $z \in G \cap H$.

$\mathbf{E}^d$ is separable: $\mathbf{Q}^d$, the d-tuples of rational numbers, is a countable dense subset.

$\mathbf{E}^d$ is second countable: $\mathcal{B} \stackrel{\text{def}}{=} \{B(\mathbf{x}, r) : \mathbf{x} \in \mathbf{Q}^d, r \in \mathbf{Q}^+\}$ is a countable set of open balls that generates $\mathcal{T}$.

Homeomorphisms

Two topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ are *homeomorphic* if there exists a map $\phi : X \to Y$ satisfying:

- bijectivity: ϕ is 1-1 and onto.
- continuity: if $\phi(x) = y$, then for every $G_Y \in \mathcal{T}_Y$ with $y \in G_Y$ there exists $G_X \in \mathcal{T}_X$ with $x \in G_X$ such that $\phi(G_X) \subset G_Y$.
- openness: if $G_X \in \mathcal{T}_X$, then $\phi(G_X) \in \mathcal{T}_Y$.

Equivalently, ϕ is a bijection between X and Y (as a point map) and a bijection between $\mathcal{T}_X$ and $\mathcal{T}_Y$ (as a set map). This uses:

Exercise: If $\phi : X \to Y$ is bijective and continuous, then for each $G_Y \in \mathcal{T}_Y$ there exists $G_X \in \mathcal{T}_X$ such that $\phi(G_X) = G_Y$.

Abstract Manifolds

A manifold $(\mathcal{M}, \mathcal{T})$ is a separable metric space together with an open cover $\{G_\alpha : \alpha \in I\} \subset \mathcal{T}$ and a corresponding collection of homeomorphisms $\{\phi_\alpha : \alpha \in I\}$, satisfying:

- for each $\alpha \in I$ there is some $d \in \mathbf{Z}^+$ such that $\phi_\alpha(G_\alpha)$ is an open subset of d-dimensional Euclidean space $\mathbf{E}^d$;
- if $G = G_\alpha \cap G_\beta$, then $\phi \stackrel{\text{def}}{=} \phi_\alpha^{-1} \circ \phi_\beta$ is a homeomorphism of metric subspace $(G, \mathcal{T}_G)$ to itself.

A manifold is said to be locally homeomorphic to $\mathbf{E}^d$, and d-dimensional if d is constant. Map ϕ_α gives *coordinates* for G_α while ϕ_α^{-1} is a *parametrization* of G_α.

Collection $\{(G_\alpha, \phi_\alpha) : \alpha \in I\}$ is an *atlas of charts* for $(\mathcal{M}, \mathcal{T})$. Every $\mathcal{M}$ has a countable atlas; compact $\mathcal{M}$ has a finite atlas.

Transition Functions

Suppose that $(\mathcal{M}, \mathcal{T})$ is a manifold with atlas $\{(G_\alpha, \phi_\alpha) : \alpha \in I\}$. For $\alpha, \beta \in I$ such that $G \stackrel{\text{def}}{=} G_\alpha \cap G_\beta$ is nonempty, define the *transition function*

$$\tau_{\alpha\beta} \stackrel{\text{def}}{=} \phi_\alpha \circ \phi_\beta^{-1} : U \to U.$$

Here $U \stackrel{\text{def}}{=} \phi_\alpha(G) = \phi_\beta(G)$ is an open subset of E^d. Compositions of homeomorphisms are homeomorphisms, so $\tau_{\alpha\beta}$ is a homeomorphism with inverse

$$\tau_{\beta\alpha} \stackrel{\text{def}}{=} \phi_\beta \circ \phi_\alpha^{-1} : U \to U.$$

Remark. $\phi_\alpha(G_\alpha) \subset \mathbf{E}^d$ is a parameter space for $G_\alpha \subset \mathcal{M}$. $\tau_{\alpha\beta}$ and $\tau_{\beta\alpha}$ are reparametrizations of G on parameter space U.

Differentiable Functions

Suppose $\mathbf{f} : \mathbf{E}^n \to \mathbf{E}^m$ is a function defined on an open set $U \subset \mathbf{E}^n$. It may be written in standard coordinates as

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \in \mathbf{E}^m, \qquad \mathbf{x} \in U \subset \mathbf{E}^n.$$

Call $\mathbf{f}$ *differentiable* if all partial derivatives are continuous on U. Its derivative at $\mathbf{x} \in U$ is the linear transformation

$$D\mathbf{f}(\mathbf{x}) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix},$$

a matrix with respect to the standard bases of $\mathbf{E}^n$ and $\mathbf{E}^m$.

Differentiable Atlases

Atlas $\mathcal{A} = \{(G_\alpha, \phi_\alpha) : \alpha \in I\}$ for manifold $(\mathcal{M}, \mathcal{T})$ is *differentiable* if every transition function $\tau_{\alpha\beta}$, $\alpha, \beta \in I$, is differentiable on the overlap domain $U = \phi_\alpha(G_\alpha \cap G_\beta) = \phi_\beta(G_\alpha \cap G_\beta) \subset \mathbf{E}^d$.

Chart (G, ϕ) is *differentially compatible* with $\mathcal{A}$ iff $\mathcal{A} \cup (G, \phi)$ is again a differentiable atlas for $(\mathcal{M}, \mathcal{T})$.

Differentiable atlas $\mathcal{A}$ is *differentially maximal* if any chart that is differentially compatible with $\mathcal{A}$ already belongs to $\mathcal{A}$.

Remark. Coordinate maps from a differentially maximal atlas $\mathcal{A}$ are used like test functions: $S \subset \mathcal{M}$ is nice iff $\phi(S \cap G) \subset \mathbf{E}^d$ is nice for every chart $(G, \phi) \in \mathcal{A}$.

Differentiable Manifolds

A *differentiable manifold* is a manifold with a maximal differentiable atlas $\mathcal{A}$. It may be denoted by $(\mathcal{M}, \mathcal{T}, \mathcal{A})$.

Note that the underlying topological space $(\mathcal{M}, \mathcal{T})$ is separable, second countable, and Hausdorff.

Say that $f : \mathcal{M} \to \mathbf{E}^m$ is *differentiable at* x iff, for every chart $(G, \phi) \in \mathcal{A}$ with $x \in G$, the composition

$$f \circ \phi^{-1} : \mathbf{E}^d \to \mathbf{E}^m$$

is a differentiable function on $U = \phi(G) \subset \mathbf{E}^d$.

Say that f is differentiable on G if it is differentiable at every $x \in G$.

Linear Manifolds

An example differentiable manifold to keep in mind:

- $\mathcal{M} = \mathbf{E}^d$,
- $\mathcal{T}$ is the metric topology,
- $\mathcal{A}$ is all charts with coordinate functions ϕ differentially compatible with the identity $\mathrm{I} : \mathbf{E}^d \to \mathbf{E}^d$.

Exercise: (G, ϕ) is differentially compatible with (G, I) iff $\phi : \mathbf{E}^d \to \mathbf{E}^d$ is differentiable on G.

Diffeomorphisms

Differentiable manifolds $(\mathcal{M}, \mathcal{T}, \mathcal{A})$ and $(\mathcal{M}', \mathcal{T}', \mathcal{A}')$ are *diffeomorphic* iff there exists a bijection $\Delta : \mathcal{M} \to \mathcal{M}'$ such that

- $\Delta : \mathcal{T} \to \mathcal{T}'$ is a bijection, so Δ is a homeomorphism of topological spaces $(\mathcal{M}, \mathcal{T})$ and $(\mathcal{M}', \mathcal{T}')$;
- $f : \mathcal{M}' \to \mathbf{E}^m$ is differentiable on $G' \in \mathcal{T}'$ iff $f \circ \Delta : \mathcal{M} \to \mathbf{E}^m$ is differentiable on $G = \Delta^{-1}(G') \in \mathcal{T}$.

Special case: $\mathcal{M} = \mathcal{M}'$, same $\mathcal{T}$ and $\mathcal{A}$. Then the identity $x \mapsto x$ is a diffeomorphism, but there may be many others, and they form the *group of diffeomorphisms*.

Differentiable Varieties

Goal: Construct an n-dimensional differentiable manifold as a subset of $\mathbf{E}^{n+m}$.

Method: For differentiable $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ with $\mathbf{F} = (F_1, \ldots, F_m)$, define the *differentiable variety*

$$\mathcal{M} \stackrel{\text{def}}{=} \{\mathbf{z} \in \mathbf{E}^{n+m} : \mathbf{F}(\mathbf{z}) = \mathbf{0}\} = \bigcap_{i=1}^{m} \{\mathbf{z} \in \mathbf{E}^{n+m} : F_i(\mathbf{z}) = 0\}.$$

Define $\mathcal{T}$ to be the *relative (metric) topology*, the restrictions of open $\mathbf{E}^{n+m}$ subsets to $\mathcal{M}$.

Apply the Implicit Function Theorem (see below) to find charts.

Inverse Function Theorem

Warm-up exercise:

Theorem

Suppose that $\mathbf{f} : \mathbf{E}^d \to \mathbf{E}^d$ *is differentiable near* $\mathbf{x} \in \mathbf{E}^d$ *with nonsingular* $D\mathbf{f}(\mathbf{x})$ *(iff* $\det D\mathbf{f}(\mathbf{x}) \neq 0$*, iff matrix* $D\mathbf{f}(\mathbf{x})$ *is invertible). Then there exists a function* $\mathbf{g} : \mathbf{E}^d \to \mathbf{E}^d$*, differentiable near* $\mathbf{y} \overset{\text{def}}{=} \mathbf{f}(\mathbf{x})$*, such that:*

- $\mathbf{g} \circ \mathbf{f}(\mathbf{x}') = \mathbf{x}'$ *for all* $\mathbf{x}'$ *sufficiently near* $\mathbf{x}$*, and*
- $\mathbf{f} \circ \mathbf{g}(\mathbf{y}') = \mathbf{y}'$ *for all* $\mathbf{y}'$ *sufficiently near* $\mathbf{y}$*.*

Furthermore, $D\mathbf{g}(\mathbf{y}) = D\mathbf{f}(\mathbf{x})^{-1}$ *is nonsingular, and*

$$D\mathbf{g}(\mathbf{y}') = D\mathbf{f}(\mathbf{g}(\mathbf{y}'))^{-1}$$

for all $\mathbf{y}'$ *sufficiently near* $\mathbf{y}$*.*

Inverse Function Theorem (proof sketch, part 1)

For each $\mathbf{y}'$ near $\mathbf{y} = \mathbf{f}(\mathbf{x})$, define a sequence by $\mathbf{x}_0 \stackrel{\text{def}}{=} \mathbf{x}$ and

$$\mathbf{x}_{n+1} = \mathbf{x}_n - D\mathbf{f}(\mathbf{x})^{-1}[\mathbf{f}(\mathbf{x}_n) - \mathbf{y}'] \stackrel{\text{def}}{=} K(\mathbf{x}_n), \qquad n = 0, 1, 2, \ldots.$$

Use the differentiability of $\mathbf{f}$ near $\mathbf{x}$ to compare K at $\mathbf{u}, \mathbf{v}$ near $\mathbf{x}$:

$$\begin{aligned} K(\mathbf{u}) - K(\mathbf{v}) &= \mathbf{u} - \mathbf{v} - D\mathbf{f}(\mathbf{x})^{-1}\,[\mathbf{f}(\mathbf{u}) - \mathbf{f}(\mathbf{v})] \\ &= \left[\mathrm{I} - D\mathbf{f}(\mathbf{x})^{-1} D\mathbf{f}(\mathbf{v})\right](\mathbf{u} - \mathbf{v}) + o(\|\mathbf{u} - \mathbf{v}\|). \end{aligned}$$

Since $D\mathbf{f}(\mathbf{v}) \to D\mathbf{f}(\mathbf{x})$ as $\mathbf{v} \to \mathbf{x}$, so $\mathrm{I} - D\mathbf{f}(\mathbf{x})^{-1} D\mathbf{f}(\mathbf{v}) \to 0$.
Thus K is a contraction near $\mathbf{x}$.
By a similar estimate: if $\mathbf{y}'$ is near $\mathbf{y}$, then $\{\mathbf{x}_n\}$ stays near $\mathbf{x}$.

Inverse Function Theorem (proof sketch, part 2)

By the contraction mapping theorem, $\mathbf{x}_n = K^n(\mathbf{x}) \to \mathbf{x}'$, the unique fixed point $\mathbf{x}' = K(\mathbf{x}')$. Then by the definition of K,

$$\mathbf{0} = \mathbf{x}' - K(\mathbf{x}') = D\mathbf{f}(\mathbf{x})^{-1}[\mathbf{f}(\mathbf{x}') - \mathbf{y}'], \qquad \Longrightarrow \mathbf{f}(\mathbf{x}') = \mathbf{y}'.$$

This defines the inverse function $\mathbf{g}(\mathbf{y}') \stackrel{\text{def}}{=} \mathbf{x}'$ at all $\mathbf{y}'$ near $\mathbf{y}$.

Since $\mathbf{y}' = \mathbf{f} \circ \mathbf{g}(\mathbf{y}')$, apply the chain rule to compute

$$\mathrm{I} = D[\mathbf{f} \circ \mathbf{g}](\mathbf{y}') = D\mathbf{f}(\mathbf{g}(\mathbf{y}'))D\mathbf{g}(\mathbf{y}') = D\mathbf{f}(\mathbf{x}')D\mathbf{g}(\mathbf{y}').$$

Conclude that $D\mathbf{f}(\mathbf{x}')$ is nonsingular, so $D\mathbf{g}(\mathbf{y}') = D\mathbf{f}(\mathbf{x}')^{-1}$. $\square$

Details may be found in the supplement `01extra.pdf`.

Newton-Raphson Iteration

For $\mathbf{y}'$ near $\mathbf{y}$, it is faster to find $\mathbf{x}' = \mathbf{g}(\mathbf{y}')$ by solving $\mathbf{f}(\mathbf{x}') = \mathbf{y}'$ for $\mathbf{x}'$ using Newton-Raphson iteration from $\mathbf{x}_0 \stackrel{\text{def}}{=} \mathbf{x}$:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - D\mathbf{f}(\mathbf{x}_n)^{-1}[\mathbf{f}(\mathbf{x}_n) - \mathbf{y}'] \stackrel{\text{def}}{=} K'(\mathbf{x}_n), \qquad n = 0, 1, 2, \ldots.$$

Note the similarity with K used in the existence proof: $D\mathbf{f}(\mathbf{x})^{-1}$ is simply replaced with $D\mathbf{f}(\mathbf{x}_n)^{-1}$.
But $\mathbf{f}(\mathbf{x}') = \mathbf{y}'$ and $\mathbf{f}$ is also differentiable at $\mathbf{x}'$, so

$$\begin{aligned} \mathbf{f}(\mathbf{x}' + \mathbf{h}) &= \mathbf{y}' + D\mathbf{f}(\mathbf{x}')\mathbf{h} + o(\|\mathbf{h}\|), \quad \text{as } \mathbf{h} \to \mathbf{0}, \\ \Longrightarrow \quad K'(\mathbf{x}' + \mathbf{h}) &= \mathbf{x}' + \mathbf{h} - D\mathbf{f}(\mathbf{x}' + \mathbf{h})^{-1}[D\mathbf{f}(\mathbf{x}')\mathbf{h} + o(\|\mathbf{h}\|)] \\ &= \mathbf{x}' + [\mathrm{I} - D\mathbf{f}(\mathbf{x}' + \mathbf{h})^{-1}D\mathbf{f}(\mathbf{x}')]\mathbf{h} + o(\|\mathbf{h}\|). \end{aligned}$$

Now $\mathrm{I} - D\mathbf{f}(\mathbf{x}' + \mathbf{h})^{-1}D\mathbf{f}(\mathbf{x}') \to 0$ as $\mathbf{h} \to \mathbf{0}$, so K' is a contraction map near $\mathbf{x}'$.

Exercise: K' iteration converges to the same unique root $\mathbf{x}'$ as K.

Defining Functions Implicitly

Goal: given $\mathbf{F}(\mathbf{x},\mathbf{y}) = \mathbf{0}$, find $\mathbf{f}(\mathbf{x}) = \mathbf{y}$ such that $\mathbf{F}(\mathbf{x},\mathbf{f}(\mathbf{x})) = \mathbf{0}$.

Method: iteration, contraction, and implicit differentiation.

Notation: fix $n, m \in \mathbf{Z}^+$ and define

$$(\mathbf{x},\mathbf{y}) \stackrel{\text{def}}{=} (x_1,\ldots,x_n,y_1,\ldots,y_m) \in \mathbf{E}^n \times \mathbf{E}^m = \mathbf{E}^{n+m},$$

Write $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ in this notation as

$$\mathbf{F}(\mathbf{x},\mathbf{y}) = \begin{pmatrix} F_1(\mathbf{x},\mathbf{y}) \\ \vdots \\ F_m(\mathbf{x},\mathbf{y}) \end{pmatrix} = \begin{pmatrix} F_1(x_1,\ldots,x_n,y_1,\ldots,y_m) \\ \vdots \\ F_m(x_1,\ldots,x_n,y_1,\ldots,y_m) \end{pmatrix}.$$

Partial Derivative Matrices

Suppose $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ is differentiable at $(\mathbf{x}, \mathbf{y})$. Then

$$D_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial F_1}{\partial x_n}(\mathbf{x}, \mathbf{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial F_m}{\partial x_n}(\mathbf{x}, \mathbf{y}) \end{pmatrix} \in \mathbf{R}^{m \times n}$$

for the first n coordinates, and

$$D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial F_1}{\partial y_1}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial F_1}{\partial y_m}(\mathbf{x}, \mathbf{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1}(\mathbf{x}, \mathbf{y}) & \cdots & \frac{\partial F_m}{\partial y_m}(\mathbf{x}, \mathbf{y}) \end{pmatrix} \in \mathbf{R}^{m \times m}$$

for the last m. The second matrix is square so it can be invertible.

Exercise: Linear $\mathbf{F}$ implies $D_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \mathbf{y})$ and $D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{y})$ are constant.

Implicit Function Theorem

Theorem

Let $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ *be differentiable on an open set* $U \subset \mathbf{E}^{n+m}$*. Suppose that there is some point* $(\mathbf{a}, \mathbf{b}) \in U$ *such that*

- $\mathbf{F}(\mathbf{a}, \mathbf{b}) = \mathbf{0}$, *and*
- $D_{\mathbf{y}}\mathbf{F}(\mathbf{a}, \mathbf{b})$ *is invertible (as an* $m \times m$ *matrix).*

Then there exists $\mathbf{f} : \mathbf{E}^n \to \mathbf{E}^m$*, with* $\mathbf{f}(\mathbf{a}) = \mathbf{b}$*, such that*

$$\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}$$

for all $\mathbf{x}$ *sufficiently near* $\mathbf{a}$*. In addition,* $\mathbf{f}$ *is differentiable at* $\mathbf{a}$ *with* $D\mathbf{f}(\mathbf{a}) = -D_{\mathbf{y}}\mathbf{F}(\mathbf{a}, \mathbf{b})^{-1} D_{\mathbf{x}}\mathbf{F}(\mathbf{a}, \mathbf{b})$*, and*

$$D\mathbf{f}(\mathbf{x}) = -D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x}))^{-1} D_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x}))$$

for all $\mathbf{x}$ *sufficiently near* $\mathbf{a}$*.*

Linear Implicit Function Theorem

Special case: $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ is a linear function. Then

$$\mathbf{F}(\mathbf{x}, \mathbf{y}) = L_x \mathbf{x} + L_y \mathbf{y}, \qquad \mathbf{x} \in \mathbf{E}^n, \mathbf{y} \in \mathbf{E}^m,$$

where $L_x \in \mathbf{R}^{m \times n}$ and $L_y \in \mathbf{R}^{m \times m}$ are matrices. Thus

- $L_x = D_\mathbf{x}\mathbf{F}(\mathbf{x}, \mathbf{y}) = D_\mathbf{x}\mathbf{F}(\mathbf{a}, \mathbf{b})$, all $(\mathbf{x}, \mathbf{y}) \in \mathbf{E}^{n+m}$.
- $L_y = D_\mathbf{y}\mathbf{F}(\mathbf{x}, \mathbf{y}) = D_\mathbf{y}\mathbf{F}(\mathbf{a}, \mathbf{b})$, all $(\mathbf{x}, \mathbf{y}) \in \mathbf{E}^{n+m}$.
- $D_\mathbf{y}\mathbf{F}(\mathbf{a}, \mathbf{b})$ is invertible iff L_y is invertible.
- $\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = L_x \mathbf{x} + L_y \mathbf{f}(\mathbf{x}) = \mathbf{0} \iff \mathbf{f}(\mathbf{x}) = -L_y^{-1} L_x \mathbf{x}$.
- $D\mathbf{f}(\mathbf{x}) = -L_y^{-1} L_x = -D_\mathbf{y}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x}))^{-1} D_\mathbf{x}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x}))$.

The proof is simple: all derivatives are constant matrices.

Implicit Function Theorem (proof sketch 1)

General case: $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ is differentiable.

$D_\mathbf{y}\mathbf{F}(\mathbf{a}, \mathbf{b})$ is invertible and continuous, so $D_\mathbf{y}\mathbf{F}(\mathbf{x}, \mathbf{b})$ is invertible for all $\mathbf{x} \in \mathbf{E}^n$ sufficiently near $\mathbf{a}$. Given such $\mathbf{x}$, define $\{\mathbf{y}_k\} \subset \mathbf{E}^m$ by

$$\mathbf{y}_0 = \mathbf{b}; \quad \mathbf{y}_{k+1} = \mathbf{y}_k - D_\mathbf{y}\mathbf{F}(\mathbf{x}, \mathbf{b})^{-1}\mathbf{F}(\mathbf{x}, \mathbf{y}_k) \stackrel{\text{def}}{=} H(\mathbf{y}_k), \; k \geq 0.$$

But H is a contraction in a neighborhood of $\mathbf{b}$:

$$\begin{aligned} H(\mathbf{u}) - H(\mathbf{v}) &= \mathbf{u} - \mathbf{v} - D_\mathbf{y}\mathbf{F}(\mathbf{x}, \mathbf{b})^{-1}[\mathbf{F}(\mathbf{x}, \mathbf{u}) - \mathbf{F}(\mathbf{x}, \mathbf{v})] \\ &= \underbrace{\left[\mathrm{I} - D_\mathbf{y}\mathbf{F}(\mathbf{x}, \mathbf{b})^{-1} D_\mathbf{y} F(\mathbf{x}, \mathbf{v})\right]}_{\to 0 \text{ as } \mathbf{v} \to \mathbf{b}} (\mathbf{u} - \mathbf{v}) + o(\|\mathbf{u} - \mathbf{v}\|). \end{aligned}$$

Hence $\mathbf{y}_k \to \mathbf{y} = H(\mathbf{y})$, the unique fixed point, so $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$.
Put $\mathbf{f}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{y}$ to get $\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}$ for all $\mathbf{x}$ sufficiently near $\mathbf{a}$.

Implicit Function Theorem (proof sketch 2)

Apply the chain rule to $\mathbf{x} \mapsto \mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}$ to get

$$\begin{aligned} \mathbf{0} &= D_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) + D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x}))D\mathbf{f}(\mathbf{x}) \\ \implies D\mathbf{f}(\mathbf{x}) &= -D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x}))^{-1} D_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})), \end{aligned}$$

for all $\mathbf{x}$ sufficiently near $\mathbf{a}$. □

Remark. Faster convergence $\mathbf{y}_k \to \mathbf{y} = \mathbf{f}(\mathbf{x})$ is obtained with Newton-Raphson iteration:

$$\mathbf{y}_{k+1} = \mathbf{y}_k - D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{y}_k)^{-1}\mathbf{F}(\mathbf{x}, \mathbf{y}_k) \stackrel{\text{def}}{=} H'(\mathbf{y}_k),$$

which differs from H by using $D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{y}_k)^{-1}$ instead of $D_{\mathbf{y}}\mathbf{F}(\mathbf{x}, \mathbf{b})^{-1}$.

Local Parametrizations

Suppose $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ is differentiable and let $\mathcal{M}$ be the differentiable variety

$$\mathcal{M} = \{(\mathbf{x}, \mathbf{y}) \in \mathbf{E}^{n+m} : \mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\},$$

with the relative metric topology $\mathcal{T}$ inherited from $\mathbf{E}^{n+m}$. For $(\mathbf{a}, \mathbf{b}) \in \mathcal{M}$ where $D_{\mathbf{y}}\mathbf{F}(\mathbf{a}, \mathbf{b})$ is nonsingular, there exists differentiable $\mathbf{f} : \mathbf{E}^n \to \mathbf{E}^m$ such that

$$\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}$$

for all $\mathbf{x}$ sufficiently near $\mathbf{a}$. Hence for some $r > 0$,

$$G \stackrel{\text{def}}{=} \{(\mathbf{x}, \mathbf{f}(\mathbf{x})) : \mathbf{x} \in B(\mathbf{a}, r) \subset \mathbf{E}^n\} \subset \mathcal{M}$$

is a neighborhood of $(\mathbf{a}, \mathbf{b})$ in $\mathcal{M}$ given by a graph.

Local Coordinate Charts

The graph $G = \{(\mathbf{x}, \mathbf{f}(\mathbf{x})) : \mathbf{x} \in B(\mathbf{a}, r) \subset \mathbf{E}^n\} \subset \mathcal{M}$ has a coordinate chart

$$\phi : G \to B(\mathbf{a}, r) \subset \mathbf{E}^n; \qquad \phi(\mathbf{x}, \mathbf{f}(\mathbf{x})) \stackrel{\text{def}}{=} \mathbf{x}.$$

This is obviously continuous. The inverse is local parametrization

$$\phi^{-1}(\mathbf{x}) = (\mathbf{x}, \mathbf{f}(\mathbf{x})).$$

If $\psi : \mathbf{E}^{n+m} \to \mathbf{E}^n$ is differentiable, then by the chain rule:

$$D\left[\psi \circ \phi^{-1}\right](\mathbf{x}) = D_{\mathbf{x}}\psi(\mathbf{x}, \mathbf{f}(\mathbf{x})) + D_{\mathbf{y}}\psi(\mathbf{x}, \mathbf{f}(\mathbf{x}))D\mathbf{f}(\mathbf{x}),$$

so ψ restricted to G is differentially compatible with ϕ.

Parametrizations Elsewhere

Suppose $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ is differentiable and let $\mathcal{M}$ be the differentiable variety

$$\mathcal{M} = \{\mathbf{z} \in \mathbf{E}^{n+m} : \mathbf{F}(\mathbf{z}) = \mathbf{0}\},$$

Fix $\mathbf{z}_0 \in \mathcal{M}$ and suppose $D\mathbf{F}(\mathbf{z}_0)$ has maximal rank m.

Lemma

There exists a coordinate system $\mathbf{z} = U(\mathbf{x}, \mathbf{y})$, $\mathbf{x} \in \mathbf{E}^n$, $\mathbf{y} \in \mathbf{E}^m$, *with* $\mathbf{z}_0 = U(\mathbf{x}_0, \mathbf{y}_0)$, *such that* $D_{\mathbf{y}}\mathbf{F}(U(\mathbf{x}_0, \mathbf{y}_0))$ *has rank* m.

Proof sketch: Find m pivot columns by reducing matrix $D\mathbf{F}(\mathbf{z}_0)$ to row echelon form. Let $\mathbf{y}$ be coordinates with respect to a basis for the pivot column space, and let $\mathbf{x}$ be the coordinates for a basis of the orthogonal complement. $\square$

Graph Parametrizations for Nonsingular Varieties

Say that $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$ gives a *nonsingular* differentiable variety $\mathcal{M} = \{\mathbf{z} : \mathbf{F}(\mathbf{z}) = \mathbf{0}\}$ if $D\mathbf{F}(\mathbf{z})$ has maximal rank m for all $\mathbf{z} \in \mathcal{M}$.

For each $\mathbf{w} \in \mathcal{M}$, let $\mathbf{z} = U_{\mathbf{w}}(\mathbf{x}, \mathbf{y})$ be change of variables such that

$$D_{\mathbf{y}}\mathbf{F}(U_{\mathbf{w}}(\mathbf{x}, \mathbf{y})) \text{ is nonsingular (has rank } m).$$

By Implicit Function Theorem, there exists $\mathbf{f}_{\mathbf{w}} : \mathbf{E}^n \to \mathbf{E}^m$, differentiable on some neighborhood $G_{\mathbf{w}} \subset \mathbf{E}^n$, such that

$$\mathbf{F} \circ U_{\mathbf{w}}(\mathbf{x}, \mathbf{f}_{\mathbf{w}}(\mathbf{x})) = \mathbf{0}, \quad \mathbf{x} \in G_{\mathbf{w}}.$$

This $\mathbf{f}_{\mathbf{w}}$ gives a graph parametrization of $\mathcal{M}$ near $\mathbf{w}$.

Manifold Estimation Application

Write $\mathbf{F} = (F_1, \ldots, F_m)$, for $\mathbf{F} : \mathbf{E}^{n+m} \to \mathbf{E}^m$, where $F_i(\mathbf{z}) \in \mathbf{R}$ measures some undesirable property of $\mathbf{z}$.

Then the variety

$$\mathcal{M} \stackrel{\text{def}}{=} \{\mathbf{z} \in \mathbf{E}^{n+m} : \mathbf{F}(\mathbf{z}) = \mathbf{0}\}$$

is a set of points without those undesirable properties.

If $\mathbf{F}$ is differentiable and $D\mathbf{F}(\mathbf{z})$ has rank m near some $\mathbf{z} \in \mathcal{M}$, then the graph parametrization generates nearby samples of desirable points.

Curves on a Manifold

Suppose $(\mathcal{M}, \mathcal{T})$ is a manifold with $x \in \mathcal{M}$. A *curve through* x is a continuous function $\gamma : (-1, 1) \to \mathcal{M}$ with $\gamma(0) = x$.

For every chart (G, ϕ) with $x \in G$ and $\phi : G \to \mathbf{E}^d$, the composition

$$\phi \circ \gamma : (-1, 1) \to \mathbf{E}^d$$

is a parametrized curve in $\mathbf{E}^d$ in the ordinary sense, with $\phi \circ \gamma(t)$ defined in some open interval near $t = 0$.

For differentiable manifold $(\mathcal{M}, \mathcal{T}, \mathcal{A})$, the curve γ is *differentiable* iff

$$\frac{d}{dt}[\phi \circ \gamma(t)] \text{ exists and is continuous at } t = 0$$

for every chart $(G, \phi) \in \mathcal{A}$ with $x \in G$.

Directional Derivatives

Given:

- differentiable function $f : \mathcal{M} \to \mathbf{R}$;
- differentiable curve $\gamma : (-1, 1) \to \mathcal{M}$ through $x = \gamma(0)$.

Define the *directional derivative* at x of f along γ to be

$$d_\gamma f(x) \stackrel{\text{def}}{=} \frac{d}{dt}[f \circ \gamma(t)]\Big|_{t=0} \in \mathbf{R}.$$

For coordinate function $\phi : \mathcal{M} \to \mathbf{E}^d$ with $\phi = (\phi_1, \ldots, \phi_d)$, the directional derivative is $\mathbf{E}^d$-valued:

$$d_\gamma \phi(x) \stackrel{\text{def}}{=} \frac{d}{dt}[\phi \circ \gamma(t)]\Big|_{t=0} = (d_\gamma \phi_1(x), \ldots, d_\gamma \phi_d(x)) \ \in \mathbf{E}^d$$

In general, differentiable $\mathbf{F} : \mathcal{M} \to \mathbf{E}^m$ has $d_\gamma \mathbf{F}(x) \in \mathbf{E}^m$.

Tangent Vectors

Define direction vectors at $x \in \mathcal{M}$ uniquely using equivalence classes of curves through x:

Definition

γ and η are equivalent curves through x iff

$$d_\gamma \phi(x) = \frac{d}{dt}[\phi \circ \gamma(t)]\Big|_{t=0} = \frac{d}{dt}[\phi \circ \eta(t)]\Big|_{t=0} = d_\eta \phi(x)$$

for every x-containing chart in the maximal differentiable atlas.

Each equivalence class of such curves defines a unique *tangent vector* to $\mathcal{M}$ at x.

Call the set of such tangent vectors the *tangent space* to $\mathcal{M}$ at x and denote it by $T_x\mathcal{M}$.

Tangent Space Homeomorphisms

Coordinate chart (G, ϕ), with homeomorphism $\phi : G \to \mathbf{E}^d$, "pushes forward" to a map $d\phi(x) : T_x\mathcal{M} \to \mathbf{E}^d$ at each $x \in G$:

$$d\phi(x)(v) \stackrel{\text{def}}{=} d_\gamma\phi(x) = \frac{d}{dt}[\phi \circ \gamma(t)]\Big|_{t=0},$$

where γ is any curve through x in the equivalence class $v \in T_x\mathcal{M}$. This is well-defined precisely because of the equivalence relation.

Theorem

(a) $T_x\mathcal{M}$ is a vector space.

(b) $d\phi(x)$ is a linear homeomorphism of $T_x\mathcal{M}$ onto $\mathbf{E}^d$.

Proof.

Represent $u + cv \leftrightarrow \phi^{-1}(\phi \circ \gamma(t) + \phi \circ \eta(ct))$ to push forward from curves γ, η on $\mathcal{M}$ to tangent vectors u, v in $T_x\mathcal{M}$.
See the notes at `01tange.pdf` for details. □

Tangent Space of a Linear Manifold

Special case: linear manifold $\mathcal{M} = \mathbf{E}^d$, tangent vector $v \in T_x\mathcal{M}$ represented by curve γ through $\gamma(0) = x \in \mathcal{M}$, and differentiable function $f : \mathcal{M} \to \mathbf{R}$. Then by the chain rule, $df(x)(v)$ is

$$d_\gamma f(x) = \frac{d}{dt}[f \circ \gamma(t)]\Big|_{t=0} = \sum_{k=1}^{d} \gamma_k'(0)\partial_k f(x) = \langle \gamma'(0), Df(x) \rangle$$

the inner product of gradient $Df(x) = (\partial_1 f(x), \ldots, \partial_d f(x))$ with direction vector $\gamma'(0) = (\gamma_1'(0), \ldots, \gamma_d'(0))$.

Alternative viewpoint: $v \in T_x\mathcal{M}$ is a first-order differential operator, evaluated at x:

$$v \stackrel{\text{def}}{=} \sum_{k=1}^{d} \gamma_k'(0)\partial_k \Big|_x \qquad \Longrightarrow \quad v(f) = df(x)(v)$$

Tangent Vectors as Derivations

Formally, for linear manifold $\mathcal{M} = \mathbf{E}^d$,

$$T_x\mathbf{E}^d = \mathrm{span}\,\{\partial_1, \ldots, \partial_d\}, \qquad \text{with "basis" } \{\partial_k\}.$$

First-order differential operators ∂ are *derivations*, linear but also obeying the product rule for functions f, g and $c \in \mathbf{R}$:

$$\partial(f + cg) = \partial f + c\partial g; \qquad \partial(fg) = f\partial g + g\partial f.$$

This generalizes to abstract differentiable manifold $\mathcal{M}$:

$$v(f + cg) = v(f) + cv(g); \quad v(fg) = g(x)v(f) + f(x)v(g),$$

for $v \in T_x\mathcal{M}$, differentiable $f, g : \mathcal{M} \to \mathbf{R}$, and $c \in \mathbf{R}$.

Tangent Bundles

If $x \neq y$ are distinct points in $\mathcal{M}$, then $T_x\mathcal{M}$ and $T_y\mathcal{M}$ have no points in common.

The *tangent bundle* of a differentiable manifold $\mathcal{M}$ is

$$T\mathcal{M} \stackrel{\text{def}}{=} \bigcup_{x \in \mathcal{M}} \{x\} \times T_x\mathcal{M},$$

For each chart (G, ϕ) in the maximal atlas for $\mathcal{M}$, the map $\Phi : T\mathcal{M} \to \mathbf{E}^d \times \mathbf{E}^d$ defined by

$$\Phi(x, v) \stackrel{\text{def}}{=} (\phi(x), d\phi(x)(v))$$

is a homeomorphism on the open set $\{\{x\} \times T_x\mathcal{M} : x \in G\}$, so $T\mathcal{M}$ is itself a manifold (of dimension $2d$).

Differentials

Differentiable $f : \mathcal{M} \to \mathbf{R}$ has a *differential* $df : T\mathcal{M} \to \mathbf{R}$, defined using directional derivatives:

$$df(x,v) \stackrel{\text{def}}{=} \frac{d}{dt}[f \circ \gamma(t)]\Big|_{t=0}, \quad \begin{cases} \gamma : (-1,1) \to \mathcal{M} \\ \gamma(0) = x, \ \gamma \leftrightarrow v. \end{cases}$$

Any other curve $\eta \leftrightarrow v$ (representing v) gives the same result:

$$\begin{aligned} \frac{d}{dt}[f \circ \eta(t)]\Big|_{t=0} &= \frac{d}{dt}[(f \circ \phi^{-1}) \circ \phi \circ \eta(t)]\Big|_{t=0} \\ &= D[f \circ \phi^{-1}](\phi(x)) \ \frac{d}{dt}[\phi \circ \eta(t)]\Big|_{t=0} \\ &= D[f \circ \phi^{-1}](\phi(x)) \ \frac{d}{dt}[\phi \circ \gamma(t)]\Big|_{t=0} \\ &= \frac{d}{dt}[f \circ \gamma(t)]\Big|_{t=0}. \end{aligned}$$

using the chain rule with $f \circ \phi^{-1} : \mathbf{E}^d \to \mathbf{R}$.

Differentials Between Manifolds

For $f : \mathcal{M} \to \mathcal{N}$, define $df : T\mathcal{M} \to T\mathcal{N}$ by:

$$df(x, v) \stackrel{\text{def}}{=} (y, w); \quad \left\{ \begin{array}{ccc} y & = & f(x) \in \mathcal{N}; \\ \gamma & \leftrightarrow & v \in T_x\mathcal{M}; \\ f \circ \gamma & \leftrightarrow & w \in T_y\mathcal{N}. \end{array} \right.$$

This df is well-defined, since for any charts $(G, \phi), (H, \psi)$ on $\mathcal{M}, \mathcal{N}$ with $x \in G, y \in H$, respectively.

$$\begin{aligned} \frac{d}{dt}[\psi \circ f \circ \gamma(t)]\Big|_{t=0} &= \frac{d}{dt}[\psi \circ f \circ \phi^{-1} \circ \phi \circ \gamma(t)]\Big|_{t=0} \\ &= D[\psi \circ f \circ \phi^{-1}](y) \frac{d}{dt}[\phi \circ \gamma(t)]\Big|_{t=0}, \end{aligned}$$

which is the same for all curves in the same equivalence class as γ.

Vector Fields on $\mathbf{E}^d$

Special case: Linear manifold $\mathcal{M} = \mathbf{E}^d$, $T_x\mathcal{M} = \mathbf{E}^d$, $T\mathcal{M} = \mathbf{E}^{2d}$.
Generalize vector $v = \sum_k c_k \partial_k \big|_x \in T_x\mathbf{E}^d$ to a *vector field*

$$\xi(x) \stackrel{\text{def}}{=} \sum_{k=1}^{d} c_k(x) \partial_k \Big|_x,$$

using coefficient functions $c_1(x), \ldots, c_d(x)$ instead of constants.

For each $x \in \mathcal{M}$, this sends a differentiable function $f : \mathbf{E}^d \to \mathbf{R}$ to its directional derivative at x in the $\xi(x)$ direction:

$$\xi(x)(f) = \sum_{k=1}^{d} c_k(x) \partial_k f(x).$$

It generalizes to vector valued f in the obvious componentwise way.

Vector Fields in General

For differentiable manifold $\mathcal{M}$, define a vector field $\xi : \mathcal{M} \to T\mathcal{M}$ by

$$\xi(x) \stackrel{\mathrm{def}}{=} (x, v), \qquad v \in T_x\mathcal{M},$$

where v is a tangent vector whose action on differentiable functions $f : \mathcal{M} \to \mathbf{R}$ is

$$v(f)(x) = df(x)(v) = d_\gamma f(x),$$

the directional derivative of f at x along any curve γ through x in the equivalence class of v at x.

Exercise: ξ is well-defined. Namely, explain why the directional derivatives of f agree for all of $v's$ equivalent curves through x.

Germs

Fix $x \in \mathcal{M}$ for differentiable manifold $(\mathcal{M}, \mathcal{T}, \mathcal{A})$.

Say that two differentiable functions $f_1, f_2 : \mathcal{M} \to \mathbf{E}^m$ are in the same *germ* at x iff

$$(\exists G \in \mathcal{T}) \Big(x \in G \text{ and } (\forall z \in G)\ f_1(z) = f_2(z) \Big).$$

(Without loss, G is part of a chart in $\mathcal{A}$.) Each germ at x is an equivalence class. Germs allow generalization to smooth manifolds.

Exercise: : $\mathcal{G}(x) \stackrel{\text{def}}{=} \{\text{all germs at } x\}$ is an algebra under pointwise addition and multiplication.

Remark. $\mathcal{G}(x)$ is infinite-dimensional: for $(G, \phi) \in \mathcal{A}$ with $x \in G$, the functions $g_k(z) \stackrel{\text{def}}{=} \phi_1(z)^k$, $k = 0, 1, 2, \ldots$ are linearly independent polynomials in the first coordinate ϕ_1.

Partitions of Unity

A *partition of unity* subordinate to a countable locally finite open cover $\{G_k\}$ for a manifold $(\mathcal{M}, \mathcal{T}, \mathcal{A})$ is a countable set of functions $\{\rho_k : \mathcal{M} \to \mathbf{R}\}$ such that, for all $k = 1, 2, \ldots,$

- ρ_k is differentiable on $\mathcal{M}$,
- $0 \leq \rho_k(x) \leq 1$ for all $x \in \mathcal{M}$,
- $\rho_k(x) = 0$ for all $x \notin G_k$,

and

$$\sum_{k=1}^{\infty} \rho_k(x) = 1, \qquad \text{for all } x \in \mathcal{M}.$$

(Note that only finitely many summands are nonzero.)

Remark. A finite cover is obviously locally finite, but in fact every (differentiable) manifold has a countable locally finite open cover and a partition of unity subordinate to that cover.

Immersions and Embeddings

Suppose that X and Y are differentiable manifolds with tangent bundles TX and TY, respectively.

Say that

- X is *immersed* in Y if there is a differentiable map $\Phi : X \to Y$ whose derivative $d\Phi : TX \to TY$ is injective. Note: Φ need not be injective.
- X is *embedded* in Y if the immersion $\Phi : X \to Y$ is also injective, so it is diffeomorphism between X and $\Phi(X) \subset Y$.

Lemma

If X is compact, then an injective immersion is an embedding.

Whitney Embedding Theorem

Roughly speaking, any abstract manifold can be realized as a differentiable variety. There are various versions:

Theorem (Whitney 1)

A compact d-dimensional differentiable manifold can be embedded into $\mathbf{E}^N$ for all sufficiently large N.

Theorem (Whitney 2)

A compact d-dimensional differentiable manifold can be embedded into $\mathbf{E}^{2d+1}$ and immersed into $\mathbf{E}^{2d}$.

Theorem (Whitney 3)

A d-dimensional smooth manifold can be embedded into $\mathbf{E}^{2d}$ and immersed into $\mathbf{E}^{2d-1}$.

Weaker Whitney Embedding Theorem, part 1

Theorem

A compact d-dimensional differentiable manifold has an embedding into $\mathbf{E}^N$ *for all sufficiently large* N*.*

Proof: Compact $\mathcal{M}$ has finite atlas $\mathcal{A} = \{(G_1, \phi_1), \ldots, (G_n, \phi_n)\}$. Let $\{\rho_1, \ldots, \rho_n\}$ be a differentiable partition of unity subordinate to $\{G_1, \ldots, G_n\}$.

Define $\Phi : \mathcal{M} \to \mathbf{E}^{nd+n}$ by

$$\Phi(x) \stackrel{\text{def}}{=} \Big(\rho_1(x)\phi_1(x), \ldots, \rho_n(x)\phi_n(x), \rho_1(x), \ldots, \rho_n(x)\Big),$$

with the convention that $\rho_k(x)\phi_k(x) = \rho_k(x) = 0$ for $x \notin G_k$.

To prove that Φ is an embedding, it remains to show that Φ is injective and differentiable with injective differential.

Weaker Embedding Theorem, part 2

Φ is injective: if $\Phi(x_1) = \Phi(x_2)$, then $(\exists k)\rho_k(x_1) = \rho_k(x_2) \neq 0$, so $x_1, x_2 \in G_k$. But then also

$$\rho_k(x_1)\phi_k(x_1) = \rho_k(x_2)\phi_k(x_2) \implies \phi_k(x_1) = \phi_k(x_2) \implies x_1 = x_2,$$

since ϕ_k is injective.

Φ is differentiable: for any differentially compatible chart (G, ϕ), and any $k = 1, \ldots, n$,

- $\phi_k \circ \phi^{-1} : \mathbf{E}^d \to \mathbf{E}^d$ is a differentiable transition function,
- $\rho_k \circ \phi^{-1} : \mathbf{E}^d \to \mathbf{R}$ is differentiable by construction.

Thus every component of Φ is differentiable on $\mathcal{M}$.

Weaker Embedding Theorem, part 3

$d\Phi$ is injective: suffices to prove $d\Phi(x,v) = (\Phi(y), \mathbf{0}) \implies v = 0$.

Fix x and evaluate $d\Phi(x)$ on $v \in T_x\mathcal{M}$ using the product rule:

$$\begin{aligned} d\Phi(x)(v) \;=\; \Big(& v(\rho_1)\phi_1(x) + \rho_1(x)d\phi_1(x)(v), \ldots \\ & \ldots, v(\rho_n)\phi_n(x) + \rho_n(x)d\phi_n(x)(v), \\ & v(\rho_1), \ldots, v(\rho_n)\Big) \;=\; \mathbf{0} \\ \implies \quad & v(\rho_1) = \cdots = v(\rho_n) = 0 \\ \implies \quad & \rho_1(x)d\phi_1(x)(v) = \cdots = \rho_n(x)d\phi_n(x)(v) = \mathbf{0}. \end{aligned}$$

But $(\exists k)\rho_k(x) \neq 0$, so $d\phi_k(x)(v) = \mathbf{0}$, which implies that $v = 0$ since $d\phi_k(x)$ is linear and injective. □

Piecewise Linear Manifolds

Idea: Replace "differentiable," or locally close to linear, with "piecewise linear."

Method:

- Require transition functions to be piecewise linear.
- Use only piecewise linear functions and germs.

Tools:

- Convex sets in $\mathbf{E}^d$
- Convex hull of a finite set
- Simplexes: convex hulls with nonempty relative interiors.
- Tesselations: unions of nonoverlapping simplexes.

References

- Louis Auslander and Robert E. MacKenzie, *Introduction to Differentiable Manifolds*, Dover, 1977.
- Jenny Wilson's talk on manifolds: `WOMPtalk-Manifolds.pdf`
- Thurston's notes on partitions of unity: `PartOfUnityLocFiniteRefinements.pdf`
- Brahim Abdenbi's paper on the Whitney Embedding Theorem: `Brahim-Abdenbi-presentation-Whitney.pdf`
- Whitney Embedding Theorem proofs: `lecture10.pdf`