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:** $d \in \mathbb{Z}^+$, finite but it could be large.

- **Set:** $\mathbb{R}^d \overset{\text{def}}{=} \{ \mathbf{x} \overset{\text{def}}{=} (x_1, \ldots, x_d) : x_i \in \mathbb{R}, i = 1, \ldots, d \}$.

- **Linearity:** $(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d)(\forall c \in \mathbb{R})$, 
  
  $\mathbf{x} + c \mathbf{y} \overset{\text{def}}{=} (x_1 + cy_1, \ldots, x_d + cy_d) \in \mathbb{R}^d$.

- **Norm:** $\| \mathbf{x} \| \overset{\text{def}}{=} \sqrt{x_1^2 + \cdots + x_d^2} \geq 0$.
  
  $\| \mathbf{x} \| = 0 \iff \mathbf{x} = \mathbf{0} \overset{\text{def}}{=} (0, \ldots, 0)$.

- **Inner product:** $\langle \mathbf{x}, \mathbf{y} \rangle \overset{\text{def}}{=} x_1y_1 + \cdots + x_dy_d$. 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?
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: $\bigcup_{\alpha \in I} G_\alpha \in \mathcal{T}$.
- For any finite collection $\{G_1, \ldots, G_N\} \subset \mathcal{T}$, the intersection is open: $\bigcup_{i=1}^N G_i \in \mathcal{T}$.

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 $\mathcal{T}_Y \overset{\text{def}}{=} \{G \cap Y : G \in \mathcal{T}\}$. 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 \(\{G_1, G_2, \ldots\} \subset \mathcal{T}\), 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** \(B \subset \mathcal{T}\) that **generates** \(\mathcal{T}\), namely every \(G \in \mathcal{T}\) is a union of elements of \(B\).

**Exercise**: (a) Second countable implies first countable. (b) First countable and separable implies 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 separable, then every open cover of \(X\) has a countable subcover.
Metric Topology

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

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

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

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

**Exercise**: $(X, \mathcal{T})$ is a first countable Hausdorff topological space.
Finite Dimensional Euclidean Space

$E^d$ is a metric space with $d(x, y) \overset{\text{def}}{=} \|x - y\|$. Metric topology $\mathcal{T}$ for $E^d$ contains all finite intersections of open balls: Put $G = B(x, r)$ and $H = B(y, s)$. Then

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

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

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

$E^d$ is second countable: $\mathcal{B} \overset{\text{def}}{=} \{B(x, r) : x \in Q^d, r \in Q^+\}$ is a countable set of open balls that generates $\mathcal{T}$. 
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**: \(\phi\) 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, \(\phi\) 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 \(d \in \mathbb{Z}^+\) such that \(\phi_\alpha(G_\alpha)\) is an open subset of \(d\)-dimensional Euclidean space \(E^d\);
- if \(G = G_\alpha \cap G_\beta\), then \(\phi \overset{\text{def}}{=} \phi_\alpha^{-1} \circ \phi_\beta\) is a homeomorphism of metric subspace \((G, \mathcal{T}_G)\) to itself.

A manifold is said to be locally homeomorphic to \(E^d\), and \(d\)-dimensional if \(d\) is constant. Map \(\phi_\alpha\) gives coordinates for \(G_\alpha\) while \(\phi_\alpha^{-1}\) is a parametrization of \(G_\alpha\).

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 \(G \overset{\text{def}}{=} G_\alpha \cap G_\beta\) is nonempty, define the transition function

\[
\tau_{\alpha\beta} \overset{\text{def}}{=} \phi_\alpha \circ \phi_\beta^{-1} : U \to U.
\]

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

\[
\tau_{\beta\alpha} \overset{\text{def}}{=} \phi_\beta \circ \phi_\alpha^{-1} : U \to U.
\]

Remark. \(\phi_\alpha(G_\alpha) \subset E^d\) 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 \( f : \mathbb{E}^n \rightarrow \mathbb{E}^m \) is a function defined on an open set \( U \subset \mathbb{E}^n \). It may be written in standard coordinates as

\[
f(x) = (f_1(x), \ldots, f_m(x)) \in \mathbb{E}^m, \quad x \in U \subset \mathbb{E}^n.
\]

Call \( f \) **differentiable** if all partial derivatives are continuous on \( U \). Its derivative at \( x \in U \) is the linear transformation

\[
Df(x) \overset{\text{def}}{=} \begin{pmatrix}
\frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\
\vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x)
\end{pmatrix},
\]

a matrix with respect to the standard bases of \( \mathbb{E}^n \) and \( \mathbb{E}^m \).
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 $U = \phi_\alpha(G_\alpha \cap G_\beta) = \phi_\beta(G_\alpha \cap G_\beta) \subset \mathbb{E}^d$.

Chart $(G, \phi)$ 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 $\phi(S \cap G) \subset \mathbb{E}^d$ 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 $f : \mathcal{M} \to \mathbb{E}^m$ is *differentiable at $x$* iff, for every chart $(G, \phi) \in \mathcal{A}$ with $x \in G$, the composition

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

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

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

An example differentiable manifold to keep in mind:

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

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

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

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

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

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

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

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

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

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

Warm-up exercise:

Theorem

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

- \( g \circ f(x') = x' \) for all \( x' \) sufficiently near \( x \), and
- \( f \circ g(y') = y' \) for all \( y' \) sufficiently near \( y \).

Furthermore, \( Dg(y) = Df(x)^{-1} \) is nonsingular, and

\[
Dg(y') = Df(g(y'))^{-1}
\]

for all \( y' \) sufficiently near \( y \).
Inverse Function Theorem (proof sketch, part 1)

For each \( y' \) near \( y = f(x) \), define a sequence by \( x_0 \overset{\text{def}}{=} x \) and

\[
x_{n+1} = x_n - Df(x)^{-1}[f(x_n) - y'] \overset{\text{def}}{=} K(x_n), \quad n = 0, 1, 2, \ldots.
\]

Use the differentiability of \( f \) near \( x \) to compare \( K \) at \( u, v \) near \( x \):

\[
K(u) - K(v) = u - v - Df(x)^{-1}[f(u) - f(v)]
\]

\[
= \left[ I - Df(x)^{-1}Df(v) \right] (u - v) + o(\|u - v\|).
\]

Since \( Df(v) \to Df(x) \) as \( v \to x \), so \( I - Df(x)^{-1}Df(v) \to 0 \). Thus \( K \) is a contraction near \( x \).

By a similar estimate: if \( y' \) is near \( y \), then \( \{x_n\} \) stays near \( x \).
Inverse Function Theorem (proof sketch, part 2)

By the contraction mapping theorem, $x_n = K^n(x) \rightarrow x'$, the unique fixed point $x' = K(x')$. Then by the definition of $K$,

$$0 = x' - K(x') = Df(x)^{-1}[f(x') - y'],$$

$\implies f(x') = y'$.

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

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

$$I = D[f \circ g](y') = Df(g(y'))Dg(y') = Df(x')Dg(y').$$

Conclude that $Df(x')$ is nonsingular, so $Dg(y') = Df(x')^{-1}$.

Details may be found in the supplement 01extra.pdf.
Newton-Raphson Iteration

For \( y' \) near \( y \), it is faster to find \( x' = g(y') \) by solving \( f(x') = y' \) for \( x' \) using Newton-Raphson iteration from \( x_0 = x \):

\[
x_{n+1} = x_n - Df(x_n)^{-1}[f(x_n) - y'] \overset{\text{def}}{=} K'(x_n), \quad n = 0, 1, 2, \ldots
\]

Note the similarity with \( K \) used in the existence proof: \( Df(x)^{-1} \) is simply replaced with \( Df(x_n)^{-1} \).

But \( f(x') = y' \) and \( f \) is also differentiable at \( x' \), so

\[
f(x' + h) = y' + Df(x')h + o(\|h\|), \quad \text{as} \quad h \to 0,
\]

\[
\implies K'(x' + h) = x' + h - Df(x' + h)^{-1}[Df(x')h + o(\|h\|)]
\]

\[
= x' + [I - Df(x' + h)^{-1}Df(x')]h + o(\|h\|).
\]

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

**Exercise:** \( K' \) iteration converges to the same unique root \( x' \) 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 \( n, m \in \mathbb{Z}^+ \) and define

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

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

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

for the first $n$ coordinates, and

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

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

**Exercise:** Linear $F$ implies $D_x F(x, y)$ and $D_y F(x, y)$ are constant.
Theorem

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

$F(a, b) = 0$, and

$D_y F(a, b)$ is invertible (as an $m \times m$ matrix).

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

$F(x, f(x)) = 0$

for all $x$ sufficiently near $a$. In addition, $f$ is differentiable at $a$ with

$Df(a) = -D_y F(a, b)^{-1} D_x F(a, b)$, and

$Df(x) = -D_y F(x, f(x))^{-1} D_x F(x, f(x))$

for all $x$ sufficiently near $a$. 
Linear Implicit Function Theorem

Special case: \( F : \mathbb{E}^{n+m} \to \mathbb{E}^m \) is a linear function. Then

\[
F(x, y) = L_x x + L_y y, \quad x \in \mathbb{E}^n, y \in \mathbb{E}^m,
\]

where \( L_x \in \mathbb{R}^{m \times n} \) and \( L_y \in \mathbb{R}^{m \times m} \) are matrices. Thus

1. \( L_x = D_x F(x, y) = D_x F(a, b), \) all \( (x, y) \in \mathbb{E}^{n+m}. \)
2. \( L_y = D_y F(x, y) = D_y F(a, b), \) all \( (x, y) \in \mathbb{E}^{n+m}. \)
3. \( D_y F(a, b) \) is invertible iff \( L_y \) is invertible.
4. \( F(x, f(x)) = L_x x + L_y f(x) = 0 \iff f(x) = -L_y^{-1} L_x x. \)
5. \( Df(x) = -L_y^{-1} L_x = -D_y F(x, f(x))^{-1} D_x F(x, f(x)). \)

The proof is simple: all derivatives are constant matrices.
Implicit Function Theorem (proof sketch 1)

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

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

$$y_0 = b; \quad y_{k+1} = y_k - D_y F(x, b)^{-1}F(x, y_k) \overset{\text{def}}{=} H(y_k), \quad k \geq 0.$$  

But $H$ is a contraction in a neighborhood of $b$:

$$H(u) - H(v) = u - v - D_y F(x, b)^{-1}[F(x, u) - F(x, v)]$$

$$= \left[ I - D_y F(x, b)^{-1}D_y F(x, v) \right](u - v) + o(||u - v||).$$

$$\rightarrow 0 \text{ as } v \rightarrow b$$

Hence $y_k \rightarrow y = H(y)$, the unique fixed point, so $F(x, y) = 0$.

Put $f(x) \overset{\text{def}}{=} y$ to get $F(x, f(x)) = 0$ for all $x$ sufficiently near $a$. 
Implicit Function Theorem (proof sketch 2)

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

$$0 = D_x F(x, f(x)) + D_y F(x, f(x)) Df(x)$$

$$\implies Df(x) = -D_y F(x, f(x))^{-1} D_x F(x, f(x)),$$

for all $x$ sufficiently near $a$. \hfill \square

**Remark.** Faster convergence $y_k \rightarrow y = f(x)$ is obtained with Newton-Raphson iteration:

$$y_{k+1} = y_k - D_y F(x, y_k)^{-1} F(x, y_k) \overset{\text{def}}{=} H'(y_k),$$

which differs from $H$ by using $D_y F(x, y_k)^{-1}$ instead of $D_y F(x, b)^{-1}$. 
Suppose $F: E^{n+m} \to E^m$ is differentiable and let $M$ be the differentiable variety

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

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

$$F(x, f(x)) = 0$$

for all $x$ sufficiently near $a$. Hence for some $r > 0$,

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

is a neighborhood of $(a, b)$ in $M$ given by a graph.
Local Coordinate Charts

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

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

This is obviously continuous. The inverse is local parametrization

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

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

$$D \left[ \psi \circ \phi^{-1} \right](x) = D_x \psi(x, f(x)) + D_y \psi(x, f(x)) Df(x),$$

so $\psi$ restricted to $G$ is differentially compatible with $\phi$. 

Parametrizations Elsewhere

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

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

Fix $z_0 \in M$ and suppose $DF(z_0)$ has maximal rank $m$.

**Lemma**

*There exists a coordinate system $z = U(x, y)$, $x \in E^n$, $y \in E^m$, with $z_0 = U(x_0, y_0)$, such that $D_yF(U(x_0, y_0))$ has rank $m$.*

**Proof sketch:** Find $m$ pivot columns by reducing matrix $DF(z_0)$ to row echelon form. Let $y$ be coordinates with respect to a basis for the pivot column space, and let $x$ be the coordinates for a basis of the orthogonal complement. □
Say that $F : E^{n+m} \to E^m$ gives a nonsingular differentiable variety $\mathcal{M} = \{z : F(z) = 0\}$ if $DF(z)$ has maximal rank $m$ for all $z \in \mathcal{M}$.

For each $w \in \mathcal{M}$, let $z = U_w(x, y)$ be change of variables such that $D_yF(U_w(x, y))$ is nonsingular (has rank $m$).

By Implicit Function Theorem, there exists $f_w : E^n \to E^m$, differentiable on some neighborhood $G_w \subset E^n$, such that

$$F \circ U_w(x, f_w(x)) = 0, \quad x \in G_w.$$ 

This $f_w$ gives a graph parametrization of $\mathcal{M}$ near $w$. 

Manifold Estimation Application

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

Then the variety

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

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 \(\gamma : (-1, 1) \rightarrow \mathcal{M}\) with \(\gamma(0) = x\).

For every chart \((G, \phi)\) with \(x \in G\) and \(\phi : G \rightarrow \mathbb{E}^d\), the composition

\[\phi \circ \gamma : (-1, 1) \rightarrow \mathbb{E}^d\]

is a parametrized curve in \(\mathbb{E}^d\) 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 \(\gamma\) 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 $f: \mathcal{M} \rightarrow \mathbb{R}$;
- differentiable curve $\gamma: (-1, 1) \rightarrow \mathcal{M}$ through $x = \gamma(0)$.

Define the directional derivative at $x$ of $f$ along $\gamma$ to be

$$d_\gamma f(x) \overset{\text{def}}{=} \left. \frac{d}{dt} [f \circ \gamma(t)] \right|_{t=0} \in \mathbb{R}.$$

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

$$d_\gamma \phi(x) \overset{\text{def}}{=} \left. \frac{d}{dt} [\phi \circ \gamma(t)] \right|_{t=0} = (d_\gamma \phi_1(x), \ldots, d_\gamma \phi_d(x)) \in \mathbb{E}^d.$$

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

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

**Definition**

$\gamma$ and $\eta$ are equivalent curves through $x$ iff

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

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

\[
d\phi(x)(v) \overset{\text{def}}{=} d\gamma \phi(x) = \left. \frac{d}{dt} \left[ \phi \circ \gamma(t) \right] \right|_{t=0},
\]

where \(\gamma\) 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 \(\mathbb{E}^d\).

**Proof.**

Represent \(u + cv \leftrightarrow \phi^{-1}(\phi \circ \gamma(t) + \phi \circ \eta(ct))\) to push forward from curves \(\gamma, \eta\) 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 $\mathcal{M} = \mathbb{E}^d$, tangent vector $v \in T_x\mathcal{M}$ represented by curve $\gamma$ through $\gamma(0) = x \in \mathcal{M}$, and differentiable function $f : \mathcal{M} \to \mathbb{R}$. Then by the chain rule, $df(x)(v)$ is

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

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 $x$:

$$v \overset{\text{def}}{=} \sum_{k=1}^{d} \gamma_k'(0) \partial_k \bigg|_x \implies v(f) = df(x)(v)$$
Tangent Vectors as Derivations

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

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

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

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

This generalizes to abstract differentiable manifold $M$:

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

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

\[
T\mathcal{M} \overset{\text{def}}{=} \bigcup_{x \in \mathcal{M}} \{x\} \times T_x\mathcal{M},
\]

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

\[
\Phi(x, v) \overset{\text{def}}{=} (\phi(x), d\phi(x)(v))
\]

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 \( f : \mathcal{M} \to \mathbb{R} \) has a differential \( df : T \mathcal{M} \to \mathbb{R} \), defined using directional derivatives:

\[
\begin{align*}
\text{df}(x, v) & \overset{\text{def}}{=} \frac{d}{dt}[f \circ \gamma(t)]_{t=0}, \quad \left\{ \begin{array}{l} 
\gamma : (-1, 1) \to \mathcal{M} \\
\gamma(0) = x, \quad \gamma \leftrightarrow v.
\end{array} \right.
\end{align*}
\]

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

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

using the chain rule with \( f \circ \phi^{-1} : \mathbb{E}^d \to \mathbb{R} \).
Differentials Between Manifolds

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

$$df(x, v) \overset{\text{def}}{=} (y, w);$$

$$\begin{cases} y = f(x) \in \mathcal{N}; \\
\gamma \leftrightarrow v \in T_x\mathcal{M}; \\
f \circ \gamma \leftrightarrow w \in T_y\mathcal{N}. \end{cases}$$

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.

$$\frac{d}{dt}[\psi \circ f \circ \gamma(t)]\bigg|_{t=0} = \frac{d}{dt}[\psi \circ f \circ \phi^{-1} \circ \phi \circ \gamma(t)]\bigg|_{t=0}$$

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

which is the same for all curves in the same equivalence class as $\gamma$. 
Vector Fields on $\mathbb{E}^d$

Special case: Linear manifold $\mathcal{M} = \mathbb{E}^d$, $T_x\mathcal{M} = \mathbb{E}^d$, $T\mathcal{M} = \mathbb{E}^{2d}$.

Generalize vector $v = \sum_k c_k \partial_k \bigg|_x \in T_x\mathbb{E}^d$ to a vector field

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

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

For each $x \in \mathcal{M}$, this sends a differentiable function $f : \mathbb{E}^d \to \mathbb{R}$ 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

$$\xi(x) \overset{\text{def}}{=} (x, v), \quad v \in T_x\mathcal{M},$$

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

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

the directional derivative of $f$ at $x$ along the curve $\gamma$ through $x$ that represents $v$.

**Exercise:** $\xi$ is well-defined.
Fix \( x \in \mathcal{M} \) for differentiable manifold \((\mathcal{M}, \mathcal{T}, \mathcal{A})\).

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

\[
(\exists G \in \mathcal{T}) \left( x \in G \text{ and } (\forall z \in G) f_1(z) = f_2(z) \right).
\]

(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:** \( \mathcal{G}(x) \overset{\text{def}}{=} \{ \text{all germs at } x \} \) 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 \( g_k(z) \overset{\text{def}}{=} \phi_1(z)^k \), \( k = 0, 1, 2, \ldots \) are linearly independent polynomials in the first coordinate \( \phi_1 \).
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 \( \{ \rho_k : \mathcal{M} \to \mathbb{R} \} \) such that, for all \( k = 1, 2, \ldots \),

- \( \rho_k \) 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

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

(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.
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: $\Phi$ 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 $\mathbb{E}^N$ for all sufficiently large $N$.

**Theorem (Whitney 2)**

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

**Theorem (Whitney 3)**

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

A compact d-dimensional differentiable manifold has an embedding into $E^N$ 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 $\Phi : \mathcal{M} \to E^{nd+n}$ by

$$\Phi(x) \overset{\text{def}}{=} \left(\rho_1(x)\phi_1(x), \ldots, \rho_n(x)\phi_n(x), \rho_1(x), \ldots, \rho_n(x)\right),$$

with the convention that $\rho_k(x)\phi_k(x) = \rho_k(x) = 0$ for $x \not\in G_k$.

To prove that $\Phi$ is an embedding, it remains to show that $\Phi$ is injective and differentiable with injective differential.
Weaker Embedding Theorem, part 2

Φ is injective: if Φ(x₁) = Φ(x₂), then (∃k)ρ_k(x₁) = ρ_k(x₂) ≠ 0, so x₁, x₂ ∈ G_k. But then also

ρ_k(x₁)φ_k(x₁) = ρ_k(x₂)φ_k(x₂) ⇒ φ_k(x₁) = φ_k(x₂) ⇒ x₁ = x₂,

since φ_k is injective.

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

▶ φ_k ◦ φ⁻¹ : E^d → E^d is a differentiable transition function,
▶ ρ_k ◦ φ⁻¹ : E^d → R is differentiable by construction.

Thus every component of Φ is differentiable on M.
Weaker Embedding Theorem, part 3

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

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

\[
d\Phi(x)(v) = \left( v(\rho_1)\phi_1(x) + \rho_1(x)d\phi_1(x)(v), \ldots, v(\rho_n)\phi_n(x) + \rho_n(x)d\phi_n(x)(v), v(\rho_1), \ldots, v(\rho_n) \right) = 0
\]

\[\implies v(\rho_1) = \cdots = v(\rho_n) = 0\]

\[\implies \rho_1(x)d\phi_1(x)(v) = \cdots = \rho_n(x)d\phi_n(x)(v) = 0.\]

But \( (\exists k)\rho_k(x) \neq 0 \), so \( d\phi_k(x)(v) = 0 \), 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 $\mathbb{E}^d$
- Convex hull of a finite set
- Simplexes: convex hulls with nonempty relative interiors.
- Tessellations: unions of nonoverlapping simplexes.
References

- Jenny Wilson’s talk on manifolds: WOMPtalk-Manifolds.pdf
- Thurston’s notes on partitions of unity: PartOfUnityLocFiniteRefinements.pdf
- Whitney Embedding Theorem proofs: lecture10.pdf