MTH 995-003: Intro to CS and Big Data Spring 2014

Lecture 21 — Mar 18th, 2014

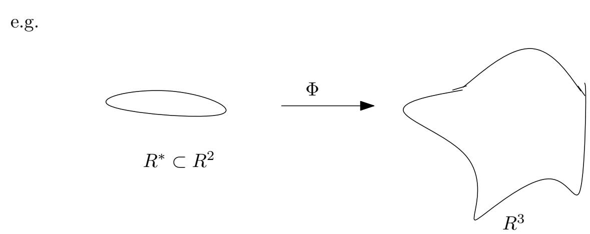
Inst. Mark Iwen *Scribe: Xianfeng Hu*

1 Manifold models for Data in $\mathbb{R}^D$

– A more general model for "intrinsically simple", intrinsically low-dimensional data. Sparsity is a special case.

– Consider a $\mathcal{C}^2$ and $1-1$ function $\Phi : \mathbb{R}^d \to \mathbb{R}^D$, $\Phi = (\Phi_1, \cdots, \Phi_D)$, where $\Phi_j : \mathbb{R}^d \to \mathbb{R}$, $\forall j \in [D]$.

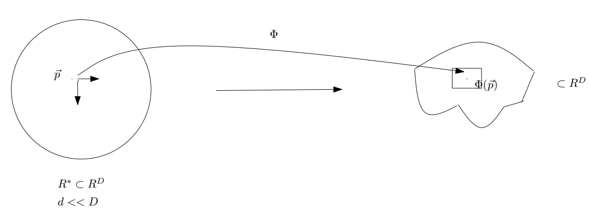
– The domains of each Φ will be always be a "regular region" $R^* \subset \mathbb{R}^d$ ("regular" means here that the boundary of R^* is $\mathcal{C}^2$, and that R^* is convex). We will call $\Phi(R^*) \subset \mathbb{R}^D$ a *simple d-dimensional manifold*.



Definition 1. *Recall the derivative of* $\Phi : \mathbb{R}^d \to \mathbb{R}^D$ *at* $\vec{p} \in R^*$ *is*

$$D\Phi|_{\vec{p}} := \begin{pmatrix} \frac{\partial \Phi_1}{\partial x_1}(\vec{p}) & \frac{\partial \Phi_1}{\partial x_2}(\vec{p}) & \cdots & \frac{\partial \Phi_1}{\partial x_d}(\vec{p}) \\ \frac{\partial \Phi_2}{\partial x_1}(\vec{p}) & \frac{\partial \Phi_2}{\partial x_2}(\vec{p}) & \cdots & \frac{\partial \Phi_2}{\partial x_d}(\vec{p}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \Phi_D}{\partial x_1}(\vec{p}) & \frac{\partial \Phi_D}{\partial x_2}(\vec{p}) & \cdots & \frac{\partial \Phi_D}{\partial x_d}(\vec{p}) \end{pmatrix} \in \mathbb{R}^{D \times d}$$

The columns of $D\Phi|_{\vec{p}}$ *span the tangent space to the d-dimensional manifold* $\Phi(R^*)$ *at* $\Phi(\vec{p})$.



The column span$\{D\Phi|_{\vec{p}}\}$ is a d-dim subspace and the affine subspace parallel to it passing through $\Phi(\vec{p})$ is tangent to $\Phi(R^*)$ at $\Phi(\vec{p})$.

Definition 2. *The d-dimensional volume element of* $\Phi(\mathbb{R}^*)$ *is*

$$dV|_{\vec{p}} := \sqrt{\det\left(D\Phi|_{\vec{p}}{}^T D\Phi|_{\vec{p}}\right)} = \prod_{j=1}^{d} \sigma_j(D\Phi|_{\vec{p}}).$$

–The d-dimensional volume of $\Phi(R^*)$ is $\int_{R^*} dV$

1.1 Examples

Example 1. *Suppose $\vec{c}: [0,1] \to \mathbb{R}^D$ parametrizes a path. We can calculate the length of the path by $\int_0^1 \|\vec{c}\,'(t)\|_2\, dt$. The area (i.e., arc length) element is $dV = \|\vec{c}\,'(t)\|$, and $R^* = [0,1]$, since*

$$\vec{c}\,' = \begin{pmatrix} \frac{\partial c_1}{\partial t} \\ \vdots \\ \frac{\partial c_D}{\partial t} \end{pmatrix}$$

where $\vec{c} = (c_1, \cdots, c_D)$; $c_j : [0,1] \to \mathbb{R}$.

Example 2. *Find the surface area of $A = \{(x,y,z) | x^2 + y^2 + z^2 = 1, z \geq 0\}$.*

– $A = \Phi(R^*)$ where R^* is the 2-dimensional rectangle $[0, 2\pi] \times [0, \pi/2]$, and $\Phi(\theta, \phi) = (\cos\theta \sin\phi, \sin\theta\sin\phi, \cos\phi)$. Thus, the upper half of the sphere is a simple 2-dimensional manifold according to our definition, and dV can be computed by

$$D\phi = \begin{pmatrix} -\sin\theta\sin\phi & \cos\theta\cos\phi \\ \cos\theta\sin\phi & \sin\theta\cos\phi \\ 0 & -\sin\phi \end{pmatrix}.$$

Thus,

$$D\Phi^T D\Phi = \begin{pmatrix} \sin^2\phi & 0 \\ 0 & 1 \end{pmatrix}.$$

We can now see that $dV = \sin\phi \Rightarrow$, and the surface area of the upper half of the sphere is

$$\int_{R^*} dV = \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \sin\phi \, d\phi \, d\theta = 2\pi.$$

– Simple manifolds are a bit too simple, so we will combine several simple manifolds to parametrize more complicated subsets of $\mathbb{R}^D$.

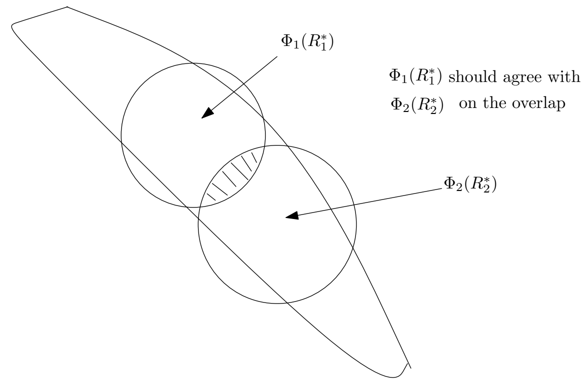
Definition 3. *When I say that $\mathcal{M} \subset \mathbb{R}^D$ is a d-dimensional manifold, I mean that $\exists I \subset Z$ (finite) such that*

1 $R_i^* \subset \mathbb{R}^d$, is a regular region $\forall i \in I$

2 $\Phi_i : R_i^* \to \mathbb{R}^D$ that are $\mathcal{C}^2$, 1-1 functions on R_i^* s.t.

3 $\Phi_i(R_i^* \cap R_j^*) = \Phi_j(R_i^* \cap R_j^*)$, $\forall i, j \in I$ with $R_i^* \cap R_j^* \neq \emptyset$ and

4 $\mathcal{M} = \cup_{i \in I} \Phi_i(R_i^*)$



I will call $(\Phi_i, \Phi_j)_{i \in I}$ an atlas for $\mathcal{M} \subset \mathbb{R}^D$.

– I will also generally assume that $\mathcal{M}$ is path connected (i.e $\exists$ a $\mathcal{C}^2$-path, $\vec{p} : [0,1] \to \mathcal{M}$, for any $\vec{x}, \vec{y} \in M$ s.t $\vec{p}(0) = \vec{x}$ and $\vec{p}(1) = \vec{y}$)

Definition 4. *Given a d-dimensional $\mathcal{M} \subset \mathbb{R}^D$, we define the geodesic distance $d_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^+$ by* $d_M(\vec{x}, \vec{y}) := \inf_{\substack{path\, \vec{p}:[0,1]\to M \\ with\, \vec{p}(0)=\vec{x} \\ and\, \vec{p}(1)=\vec{y}}} \int_0^1 \|\vec{p}\,'(t)\|_2 \, dt$ *(i.e., the shortest distance from $\vec{x}$ to $\vec{y}$ on M)*

1.2 Homework

Show that $\left\{\vec{z} \in R^D, \|\vec{z}\|_0 = d\right\} \subset R^D$ is a d-dimensional manifold by constructing an atlas.

References

[1] Partha Niyogi, Stephen Smale, Shmuel Weinberger. Finding the Homology of Submanifolds with High Condence from Random Samples. *J. Discrete Comput. Geom.*, 39(1): 419–441, 2008.

[2] Mira Bernstein, Vin De Silva, John C. Langford, Joshua B. Tenenbaum. Graph Approximations to Geodesics on Embedded Manifolds. 2000.

[3] Richard G. Baraniuk, Michael B. Wakin. Random Projections of Smooth Manifolds. *J. Found. Comput. Math.*, 9(1): 51–77, 2009.

[4] Mark A. Iwen, Mauro Maggioni. Approximation of Points on Low-Dimensional Manifolds Via Random Linear Projections. *J. CoRR*, 1204.3337, 2012.

[5] Armin Eftekhari, Michael B. Wakin. New Analysis of Manifold Embeddings and Signal Recovery from Compressive Measurements. *J. CoRR* , 1306.4748, 2013.