AI4PDE

A Bit Of Differential Geometry and Manifold Learning

Yunfeng Liao

23 August,2024

Presentation Overview

Tutorials about Solving PDEs on manifolds

Tutorials: Differential Geometry

- An Introduction to Manifolds. Loring W. Tu.
- Differential Geometry. Loring W. Tu.

Tutorials: Geometric Analysis

- Riemannian geomeotry and geometric analysis. Jurgen Jost.
- Notes for Analysis on Manifolds via the Laplacian. Yaiza Canzani.

Tutorials:Variation

- Introductory Variational Calculus on Manifolds. Ivo Terek.

Numerical Methods

- (ODE)Solving Differential Equations on Manifolds. Ernst Hairer.

Section Overview

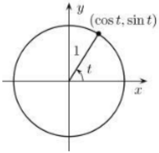
Manifold

Definition

A **n-dimension manifold** $\mathscr{M}$ is a Hausdorff topological space that **locally looks like** $\mathbb{R}^n$. It has a set of one-to-one **coordinate mappings**, $\phi : \mathbb{R}^n \to \mathscr{M}$.

Example

A circle in $\mathbb{R}^2$ is a 1d manifold. It can be equipped with two coordinates map: $\phi_1 : (x) \to (x, y), y \geq 0$ $\phi_2 : (x) \to (x, y), y \leq 0$



Sorts of Manifold

Intuitive Manifolds

A sphere S^2, any graph of a smooth two variables function $\{(x,y,f(x,y))\}$, any curve or surface and so on are manifolds. These topology can be embedded into an Euclidean space, which are studied in **extrinsic geometry**.

Somewhat *Weird* Manifolds

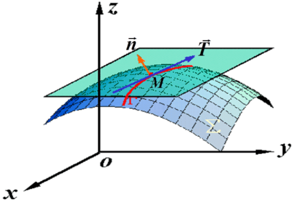
- Many **matrix groups** can be viewed as a manifold. For instance, the n-d general linear group:

$$GL(\mathbb{R}^n) := \{A \in M_n : \det A \neq 0\} \tag{1}$$

  It can be showed that it is a n^2-d manifold.
- **Level set**. Sometimes a system parametrized with θ satisfies $f(\theta_1, ...\theta_n) = 0$, then the solution of θ's is termed the **zero level set**, which also constitutes a manifold.

What is Tangent Space

Locally looks like $\mathbb{R}^n$? For a 2d surface manifold embedded in $\mathbb{R}^3$, its tangent space **at a point** $p \in$ can be imagined as its **tangent plane**. One can easily find two basis vectors in $\mathbb{R}^3$, which is $\mathbf{e}_1 = (a, b, 0)^T, \mathbf{e}_2 = (c, d, 0)^T$ after a proper coordinates transformation.



Since its tangent space is simply 2d, it is enough to write $\mathbf{e}_1 = (a, b)^T, \mathbf{e}_2 = (c, d)^T$. To jump out of what the Euclidean space limits, we use $\frac{\partial}{\partial x}, \frac{\partial}{\partial y}$ as the basis vectors, i.e.

$$e_1 = a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y} \equiv a\partial_x + b\partial_y; e_2 = c\frac{\partial}{\partial x} + d\frac{\partial}{\partial y} \equiv c\partial_x + d\partial_y \tag{2}$$

Notations

Definition

A n-d $\mathscr{M}$ at any $x \in \mathscr{M}$ has a n-d **tangent vector space** $T_x\mathscr{M}$. *Assembly* these spaces and yield the tangent bundle $T\mathscr{M} : \mathscr{M} \times \mathbb{R}^n$ with a natural projection $\pi : T\mathscr{M} \to \mathscr{M}$.

Einstein Summation Convention

$$a_i c_j d_k b^i := (\sum_i a_i b^i) c_j d_k \tag{3}$$

Definition

A **function** or **scalar field** on $\mathscr{M}$ is a map $f : \mathscr{M} \to \mathbb{R}$. A **vector field** on $\mathscr{M}$ is a map $X : \mathscr{M} \to T\mathscr{M}$. A **curve** on $\mathscr{M}$ is a map $\gamma : \mathbb{R} \to \mathscr{M}$

Definition

If we endow those tangent spaces with an inner product, i.e. ,assume them **Hilbert**, then we write $(x,y) = x^T G_p y, x, y \in T_p M, p \in M$. If $\forall p, G_p$ is **positive-definite**, such $G_p := g(p)$ is called a **Riemannian metric** on M.

Remark

g can be viewd as a matrix, whose (i,j) element g_{ij}, are all functions on M. Its **inverse matrix** is g^{ij} s.t. $g_{ij} g^{jk} = \delta_i^k$.

Remark

All basis vectors $\mathbf{e}_1, \mathbf{e}_2, ... \mathbf{e}_n$ on a smooth manifold can constitute n vector fields with a proper permutation.

Christoffel Symbol

Since the basis vectors in each tangent space will vary, we must use Christofel symbols to depict such changes.

$$\partial_{x^j} \mathbf{e}_i = \Gamma_{ij}^k \mathbf{e}_k \tag{4}$$

or equivently, where $g_{mk,l} := \frac{\partial}{\partial x^l} g_{mk}$

$$\Gamma_{ij}^k = \frac{1}{2} g^{im} (g_{mk,l} + g_{ml,k} - g_{kl,m}) \tag{5}$$

The **connection**, i.e., the generalized version of the **gradient** ∇ is unique in the sense of being compatible with Riemannian manifold (M, g). So the Christofel symbol is somewhat the components of a *directional derivative*.
Here $\nabla : T_x M \times T_x M \to T_x M, \nabla_X Y \mapsto Z$

$$\nabla_{\partial_i} \mathbf{e}_j = \Gamma_{ij}^k \mathbf{e}_k \tag{6}$$

Section Overview

ODEs on Manifolds

Problem Settings

Usually, the solution of an ODE on a manifold is a **curve**(also **flow** in some literature) $\gamma(t)$.

Example

Geodesic. $\gamma(t)$ where $\dot{\gamma} := \frac{d\gamma^i}{dt}\frac{\partial}{\partial x^i}$ is a geodesic iff

$$\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = 0 \Leftrightarrow \frac{\mathrm{d}^2}{\mathrm{dt}^2}\gamma^{\mathrm{i}} + \Gamma^i_{jk}\frac{d\gamma^j}{dt}\frac{d\gamma^k}{dt} = 0 \tag{7}$$

A more generalized version is:

$$\dot{y} = f(y), y(0) = y_0 \tag{8}$$

where $y \in M, f : M \to T_x M$

FDM with Projection

If the manifold is embedded in $\mathbb{R}^n$, then one can take the tangent space on M as a subspace in $\mathbb{R}^n$ and use FDM directly.

$$\dot{y} = f(y) \Rightarrow y^{n+1} - y^n = f(y^n)h \tag{9}$$

However, since it always happens that the obtained $y^{n+1} \notin M$ even if $y^n \in M$, a projection step is required.

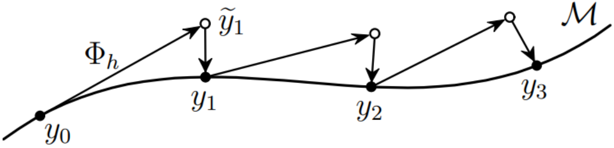


Fig. 1: Naive FDM

Section Overview

PDEs on Manifolds

Generalized Grad,Div,LBO on Manifolds

Use the generalizations below and one can develop the corresponding PDE on manifolds, where u, X are a function and a vector field on M, resp.

$$\nabla u := g^{ij}\frac{\partial u}{\partial x^i}\frac{\partial}{\partial x^j} \tag{10}$$

$$\operatorname{div}_g X := \frac{1}{\sqrt{|\det g|}}\frac{\partial}{\partial x^i}(b^i\sqrt{|\det g|}), X = b^i\frac{\partial}{\partial x^i} \tag{11}$$

$$\Delta_g u := \operatorname{div}_g \nabla u = \frac{1}{\sqrt{|\det g|}}\frac{\partial}{\partial x^i}(g^{ij}\frac{\partial u}{\partial x^i}\sqrt{|\det g|}) \tag{12}$$

Example

The eigenproblems of Laplacian-Beltramin Operator.

$$\Delta_g u(x) = \lambda u, x \in M; u(x) = 0, x \in \partial M \tag{13}$$

A Possible Task Scenario

Point Clouds On a Manifold.

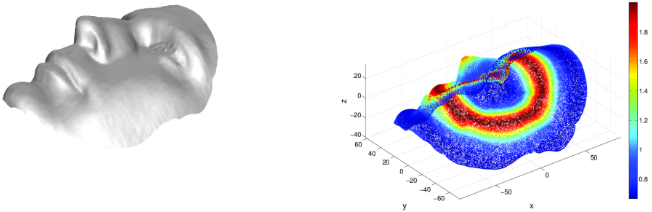


Fig. 2: Solving The Possion Equation On A Point Cloud

Challenges

- How to find a proper chart(or more exactly, the coordinate system) for the manifold?
- How to reconstruct the manifold by the point cloud?
- How to solve the PDE given a manifold?
- Another way: can we solve u without estimating g directly?
- Difference(or derivative) is not easily to realize on a point cloud, but integral is much easier.

Challenge I: Find a proper coordinate system

Possible coordinate systems are:

- The original coordinate system. Not recommended. i) Nobody know how to **segment** an unknown manifold properly; ii) it can be ill-conditioned;
- If M is known of $n-1$-dimension, i.e., **a hyper surface**, then one may compute its normal vector field via **SVD**. A simple BFS can generate a set of well-behaved coordinates.
- **Harmonic Coordinates**. Widely utilized in harmonic analysis and determined only by the metric itself.
- **Coordinates used in manifold learning**. However, these coordinates don't seem to involve much useful geometric info about the manifold.

Find An Euclidean Chart Of A Hypersurface N

Algorithm 1: Find An Euclidean Chart Of A Hypersurface N

```
Find k-nearest neighbors of each point k > n−1
for Each Point x_i do
    displacement u_j = x_j − x_i, x_i ∈ N(x_j)
    A = concat(u_j)
    U, Σ, V^T = SVD(A)
    normal vector n_i = lastRowOf(V^T)
while Not All Points Have A Coordinate System do
    Choose a uncoordinated point x_i with n_i.
    Establish an Euclidean coordinates using projection orthogonal to n_i.
    repeat
        BFS, add new neighboring point x_j
    until n_i · n_j > ϵ;
```

Harmonic Coordinates

The harmonic coordinate $(x^1, x^2..., x^n)$ on manifold M is n linearly independent harmonic functions $x^{1:n}$ s.t.

$$\Delta_g x_i = 0 \tag{14}$$

Equivalently

$$2g^{ij} g_{jk,i} = g^{ij} g_{ij,k} \tag{15}$$

Fortunately, it can be showed that:

$$-\int_M \Delta_g u(y) R'_t(x,y) d\mu_y \approx \frac{1}{t} \int_M R_t(x,y)(u(x) - u(y)) d\mu_y - 2\int_{\partial M} R'_t(x,y) \frac{\partial u}{\partial n}(y) d\tau_y \tag{16}$$

where R_t is Gaussian kernel and R'_t is its error function, i.e.

$$R_t(x,y) = C_t R(\frac{|x-y|^2}{4t}), R'_t(x,y) = C_t \int_{\frac{|x-y|^2}{4t}}^{+\infty} R(s) ds \tag{17}$$

$R(s)$ has a compact support within $[0,1]$ and is usually chosen as $\exp -s^2$ in engineering.

Section Overview

Manifold Learning

- Global Methods
- Local Methods

Overview

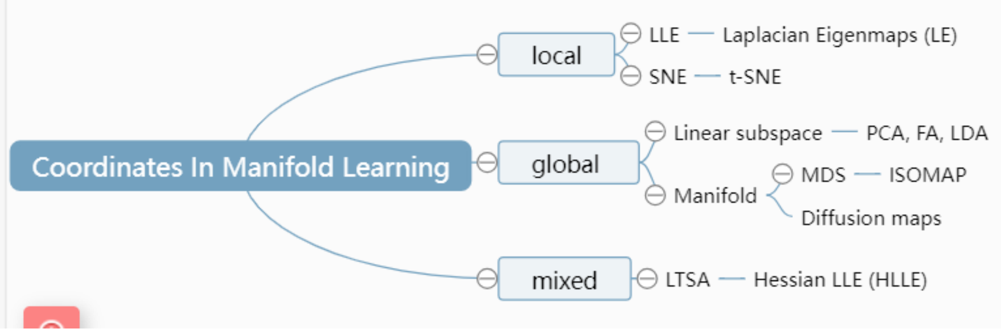
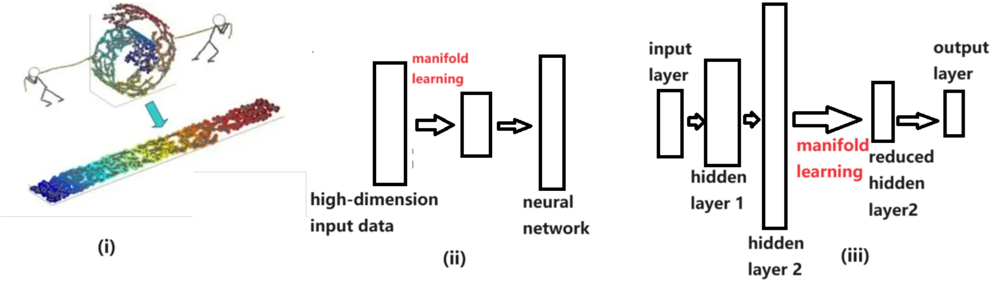


Fig. 3: Coordinate Representation Methods

Why Manifold Learning?

The essential spirit of manifold learning is dimension reduction(RD), or namely **feature extraction** to be high-brow. It helps to avoid **curse of dimensionality**.



How to use manifolds? **Coordinates is all you need** .

The Powerful SVD Technique

For any matrix $A_{d\times D}$, we have:

$$A = U_{d\times d}\Sigma_{d\times D}V^T_{D\times D} \tag{18}$$

where $UU^T = I$, $VV^T = I$, Σ without the last $(D-d)$ columns is diagonal, denoted by $\operatorname{diag}\{\lambda_1,...\lambda_d\}$. It is of great use in the following aspects:

- **Dimension Reduction**. SVD induces low-rank decomposition, since $A = \sum_{i=1}^{d}\lambda_i u_i v_i^T$. One can cast away terms where the corresponding λ_i is too small.
- **As the optimization solution**. The optimization target $\min_{X^TX=I}\operatorname{tr}X^TAX$ is often met in manifold learning. Its solution is just about the eigenpairs (the eigenvalue and eigenvector) of A.
- **Discrete spectral method**. We use its eigenvectors to generate the coordinates.

If the manifold is assumed to be a vector subspace...

Then we have classical dimension-reduction methods, like PCA, LDA and FA. LDA turns out to be linear in the end and FA only differs a little bit from eq.(20)

PCA: Principal component analysis

It has two equivalent optimization target: minimize reconstruction error and maximize the principal variance. $X, P_{d\times D}, Y = PX$ denotes the input data matrix,output matrix and linear transformation. In the sense of **maximize the principal variance**, $\operatorname{Cov} Y := \frac{1}{n} YY^T$ and we wish to maximize $\operatorname{tr} \operatorname{Cov} Y$, i.e.

$$\max_P \operatorname{tr}(P \operatorname{Cov}[X] P^T) \tag{19}$$

Apply SVD on $\operatorname{Cov}[X]$ and X is the first d columns of V. It is equivalent to minimize reconstruction error by

$$\min_P (Y' - X)^2 \tag{20}$$

Now the manifold is a non-trivial manifold

MDS:Multidimensional Scaling [1]

MDS gives a framework for ISOMAP. Given the distance matrix $L_{ij} := d(x_i, x_j)$, d here can be any distance function. Let d be l_2 and yield

$$G := XX^T \equiv -\frac{1}{2}JLJ, J = I_n - \frac{1}{n}11^T \tag{21}$$

Notice that G is symmetric, positive semi-definite (p.s.d.),so $\sqrt{G}$ is well-defined.

$$G = (\sqrt{\Lambda}P)^T\sqrt{\Lambda}P, X \approx \sqrt{G} := \sqrt{\Lambda}P \tag{22}$$

SVD here acts as low-rank decomposition. One can cast away small eigenvalues in Λ for RD.

What if d is the geodesic distance? It is followed with ISOMAP.

ISOMAP [2]

Suppose the manifold M is given in the form of **point cloud**. One can construct a KNN(k-nearest neighbors graph) onwards.

Spirit

It is unrealistic to solve a geodesic on M. Instead, we use the virtue of M that

- It looks locally like $\mathbb{R}^d$
- Its parameter space is Euclidean.
- Geodesic is the shortest curve on M.

ISOMAP:Isometric Mapping

- Approximate geodesic distance via Euclidean distance. $d(x_i, x_j) := \|x_i - x_j\|_2, x_i \in N(x_j)$
- Evaluate the distance matrix L_{ij} by a shortest path algorithm.

Local Embedding Methods:LLE/Eigenmaps

Since M **looks locally like** $\mathbb{R}^d$, the linear approximation holds water:

$$x_i = \sum_{j \neq i} w_{ij} x_j, x_j \in N(x_i) \tag{23}$$

Thus the embedded coordinate Y also enjoys:

$$y_i = \sum_{j \neq i} w_{ij} y_j, y_j \in N(y_i) \tag{24}$$

Generally, we require Y to be translation-invariant and orthogonal normal:

$$Y\mathbf{1} = 0, Y^T Y = I \tag{25}$$

LLE only gives the relative coordinates.

LLE: Local Linear Embedding [3]

- Yield weights W by minimizing $|x_i - \sum_{j \neq i} w_{ij} x_j|^2 + \alpha \sum_j w_{ij}^2$
- Yield $Y = \arg\min_{Y^T Y = I} |Y - WY|^2 = \arg\min_{Y^T Y = I} Y^T [(I-W)^T (I-W)] Y$ by SVD.

Laplacian LE/Laplacian Eigenmaps

What if we wanna get something like harmonic coordinates? i.e. $\Delta f = 0$. We may not as well impose the homogenous Neumann condition on M. Then by Green's first equation

$$f = \arg\min_{f} \int_M |\nabla f|^2 = \arg\min_{f} \int_M f\Delta f \tag{26}$$

Use the integration where G_t is the Neumann heat kernel

$$e^{-t\Delta} = \int_M G_t(x,y)f(y), G_t(x,y) \approx (4\pi t)^{-\frac{d}{2}} e^{-\frac{|x-y|^2}{4t}} \tag{27}$$

Given $\Delta = -\lim_{t\to 0^+} \frac{I-e^{-t\Delta}}{t}$, a discretized version is:

$$\Delta f(x_i) = \frac{1}{t}(f(x_i) - (4\pi t)^{-\frac{d}{2}} \sum_{x_j \in N(x_i)} \exp -\frac{|x_i - x_j|^2}{4t} f(x_j)) =: \frac{1}{t}(d_i f(x_i) - \sum_j w_{ij} f(x_j)) \tag{28}$$

Whence here comes the graph Laplacian matrix $L := D - W$, the approximation of Δ.

Laplacian LE/Laplacian Eigenmaps

Normalize L symmetrically as what we do in GNN, $\mathscr{L} := D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$

Laplacian Eigenmaps [4],chap. 12

As required in LLE,

$$Y\mathbf{1} = 0, Y^T Y = I \tag{29}$$

The variation and SVD gives Y:

$$\arg\min_f \int_M f \Delta f \approx \arg\min_Y \operatorname{tr} Y^T \mathscr{L} Y =: Y \tag{30}$$

LTSA:Local Tangent Space Alignment

There is a mapping between the coordinates on the manifold and in the Euclidean space, $f(y)=x$. Hence we have a push-forward df s.t.

$$x_j-\overline{x}=\mathrm{df}(\overline{y})(y_j-\overline{y}) \tag{31}$$

Hence its inverse $\mathrm{dh}:=\mathrm{df}^{-1}$ gives:

$$y_j-\overline{y}=\mathrm{dh}(\overline{x})(x_j-\overline{x}) \tag{32}$$

H denotes the centralized matrix, and we have:

$$Y_iH=\mathrm{dh}(\overline{x})X_iH=\mathrm{dh}(\overline{x})U\Sigma V^T \tag{33}$$

Then we have:

$$Y_iH(I-VV^T)=O \tag{34}$$

Define $W_i=H(I-VV^T)$ and we have $Y_iW_i=O$. Extend W_i,Y_i to the whole graph and yield W^i,Y. And it is expected that

$$YW^i=O,\forall i\Rightarrow YK=O,K=\sum W^i \tag{35}$$

LTSA [4],chap.11

Since K is s.p.d. , the optimization problem can be:

$$Y = \arg_Y \min_{Y^TY=I} \operatorname{tr} Y^T K Y \tag{36}$$

The solution is the 2nd to $(n+1)st$ eigenvectors because K has a trivial eigenvector $\mathbf{1}$ s.t. $K\mathbf{1} = \mathbf{0}$.

HLLE: Hessian LLE [4],chap.13

We may define the **Hessian operator** $H[f] := \frac{\partial^2}{\partial y^i \partial y^j} f$ and must have:

$$H[y^i] = 0, H[c] = 0 \tag{37}$$

where c is the constant function on M. If we can find an approximation of H at x_i and replace the W_i in LTSA, then we have the so-called HLLE.

Ref.

1. https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec9mds.pdf
2. https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec10ISOmap.pdf
3. https://www.stat.cmu.edu/ cshalizi/350/lectures/14/lecture-14.pdf
4. Geometric Structure of High-Dimensional Data and Dimensionality Reduction,Jianzhong Wang.
5. Manifold Learning: What, How, and Why. Marina Meila and Hanyu Zhang. Annual Review of Statistics and Its Applications.