Calculation of manifold's tangent space at a given point from noisy data

by

Moldir Toleubek

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Master of Applied Mathematics

at the

NAZARBAYEV UNIVERSITY

Apr 2020

Author . . . .
Department of Mathematics
May 4, 2020

Certified by . . . .
Rustem Takhanov
Associate Professor
Thesis Supervisor

Accepted by . . . .
Daniel Pugh
Dean, School of Science and Humanities

Calculation of manifold's tangent space at a given point from noisy data

by

Moldir Toleubek

Submitted to the Department of Mathematics
on May 4, 2020, in partial fulfillment of the
requirements for the degree of
Master of Applied Mathematics

Abstract

Recently, several studies have been conducted in a field of machine learning to construct manifolds from data in a complex multidimensional space. Therefore manifold learning becomes remarkably attractable among researchers. One of the main tools to identify manifold's structure is tangent space. In this work, first, we simulate a method for finding tangent space of manifold at some point from noisy data by Principal Component Analysis. In fact, Principal Component Analysis(PCA) provides dimension reduction by its 'principal components'. Then we introduce concurrent method to PCA that is called Maximum Mean Discrepancy distance. It is based on measuring the distance between smooth distributions.

Thesis Supervisor: Rustem Takhanov
Title: Associate Professor

Acknowledgments

In this section, I want to mention some important people who have contributed to this research. First of all, I would like to express my very great gratitude to my supervisor Professor Rustem Takhanov, who guided me throughout this research process. Indeed, he inspires with his knowledge, and I would like to express special gratitude for his patience. Because my supervisor taught me not only how to conduct research, but also how to write it.

I am also extremely grateful to each of the members of Thesis Committee, my second reader Professor Zhenisbek Assylbekov and my external reader Professor Nurlan Ismailov, for their comments and suggestions needed for improvements of my thesis.

Contents

Chapter 1

Overview

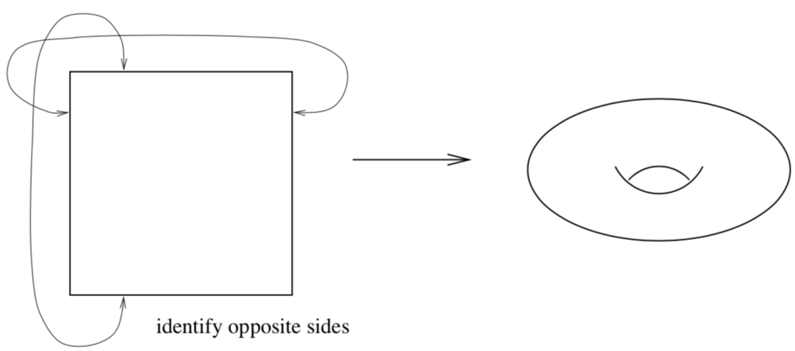
1.1 Introduction

In recent years, manifold learning has become more widely known due to effective solutions in machine learning problems. Manifold learning is an approach that based on constructing manifolds from data that embedded high-dimensional space. Sometimes in the process of fitting data points, it gets complicated to work with them. The reason behind this is the complexity of interpreting high-dimensional data. Therefore the main benefit of manifold learning is to simplify the problem by reducing the dimensionality of the dataset. Thus, manifold learning becomes remarkably attractive among researchers. Several works have been done in this field [6] [4]. One of the main tools to identify manifold's structure is tangent space. This study focuses on dimension reduction by finding a tangent space to the manifold.

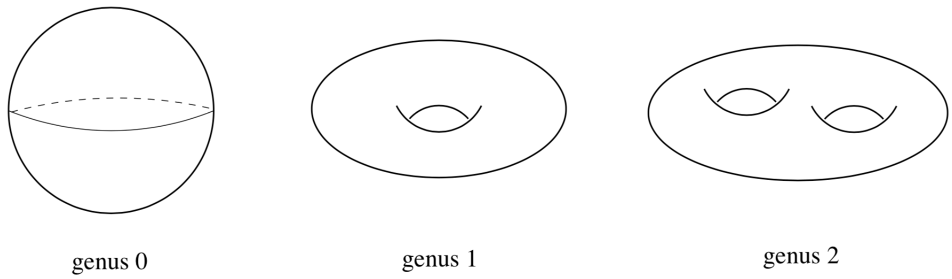
1.2 Manifold Examples

Before going to the tangent space of a manifold, it is import to understand the structure and description of a manifold. A manifold is one of the basic ideas in science, precisely, in mathematics and physics. Manifold resembles the n- dimensional Euclidean space in local regions while the global structure of $\mathbb{R}^n$ has a complex topology structure. Examples of manifolds are listed below:

- the simplest one is $\mathbb{R}^n$ whose special cases are a line in 1 - dimension or plane in 2- dimension and so on.
- another one is n - dimensional sphere. Imagine the Earth planet, globally it is sphere nevertheless locally it resembles a plane.
- the most absorbing example in topology is n - torus that obtained by converting n - dimensional cube to the figure like a doughnut. It was done by taking opposite sides of the cube and it becomes T^2 surface.



- the next example is similar to the previous example even though with two holes(formally, genus g) in torus which has the name 'Riemann surface'. It is denoted as S^2 if it is Riemann surface with genus 0.



- some abstract things as a rotation of $\mathbb{R}^n$ construct manifold and other continuous transformations could forge the manifold. On top of that, the Lie group also form manifolds.

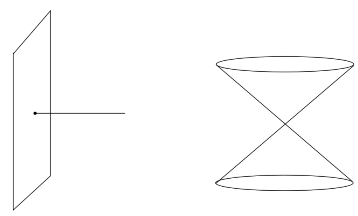
- the product of two manifolds establishes the manifold, more specifically, if we have two manifolds A and B with dimensions n and m respectively, then their direct product $A \times B$ is a manifold with $n + m$ dimension which contains elements (s, s') where $s \in A$ and $s' \in B$.

From the given examples, the topic is starting to be stirring and the question 'what doesn't belong to the manifolds?' appears. Therefore, let's move to the non-manifolds. There are certain non-manifold examples and demonstration of some of them explained below:

- one-dimensional line collapses to the plane.

- imagine two cones glued together by the sharp corner.

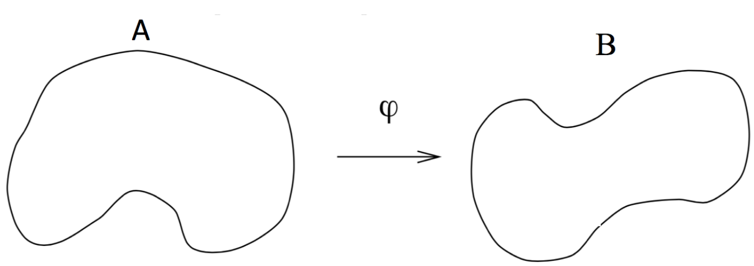
The reason for being non-manifold is discontent of resembling to $\mathbb{R}^n$ in some local regions.



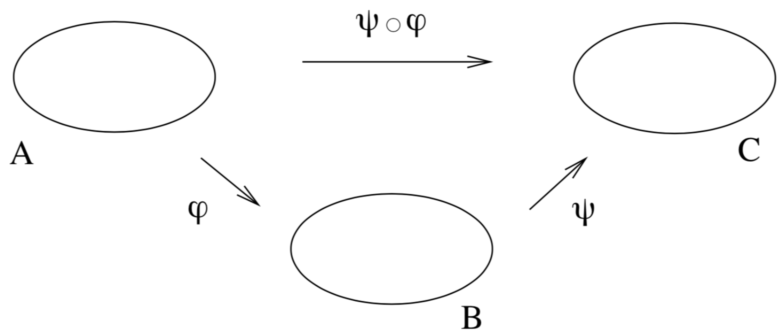
In order to briefly understand the structure of manifolds, it is recommended to study the definitions of diffeomorphism, chain rule, Jacobian, chart, open sets and tangent space in general.

1.3 Diffeomorphism: a natural way to relate manifolds

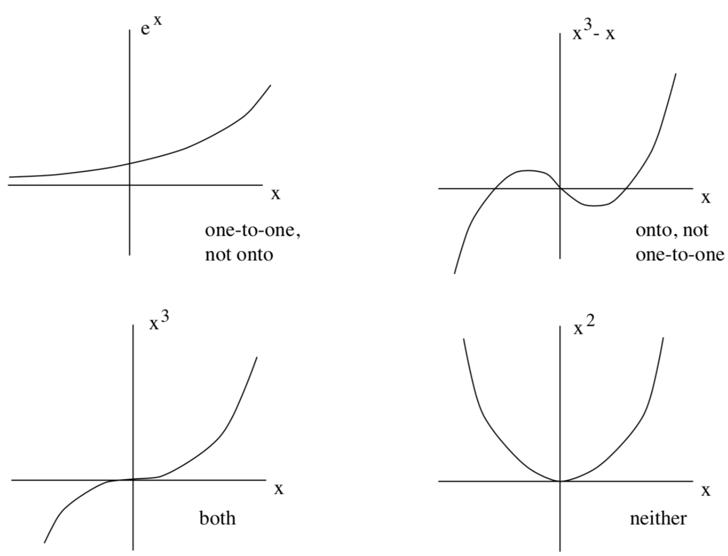
Let's begin to learn diffeomorphism by introducing the notion of a map. If we are given with two sets A and B, the **map** $\varphi : A \to B$ matches every element of A to only one element of B. Here the illustration of the map:



If two maps are known, their composition is defined, e.g. $\varphi : A \to B$ and $\psi : B \to C$ are given then $(\varphi \circ \psi)(t) = \psi(\varphi(t))$ where $t \in A$ and $\varphi(t) \in B$, consequently, $(\varphi \circ \psi)(t) \in C$. The arrangement of maps is significant owing to the fact that the right map behaves first.

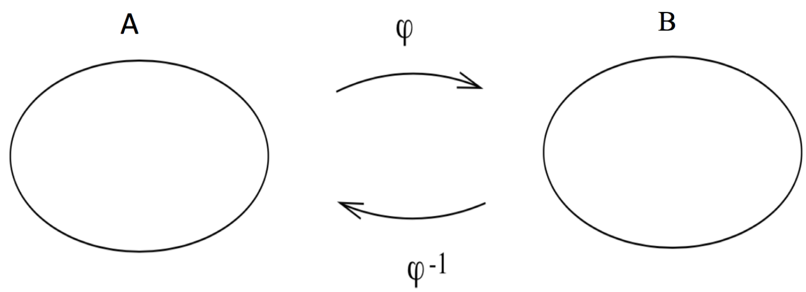


By acting conduction, the map has two types: *injective* or *one-to-one* and *surjective* or *onto*. In injective type for every element of A maps to exactly one element of B whereas in surjective type for every element of B there exists at least one element from A. Consider the example where $\phi : \mathbb{R} \to \mathbb{R}$ and $\phi(x) = e^x$ it is one-to-one although not onto in contrast to the example $\phi(x) = x^3 - x$ which is onto nevertheless not one-to-one. In addition, if $\phi(x) = x^3$ is both one-to-one and onto when $\phi(x) = x^2$ neither of them. The visualization of these examples is demonstrated in the next figure.



The set A is called the **domain** of the map ϕ and the points where elements of A that mapped to B are known as **image** of the map ϕ. The subset of B, let's denote it S and write as $S \subset B$. And for this subset S, the set of elements of A which have connected to the S is called **preimage** of S under ϕ and denoted as $\phi^{-1}(S)$.

There exists a mapping which belongs to both one-to-one and onto and it is known as **invertible** or **bijective**. In bijective type we can find **inverse mapping** between two sets write as $\varphi^{-1} : A \to B$ and $(\varphi^{-1} \circ \varphi)(a) = a$



It is necessary to indicate the **continuity** between topological spaces,also manifolds in our case. If the given map $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ where $(x^1, x^2, ..., x^n)$ goes to $(y^1, y^2, ..., y^m)$, is a family of m functions with n variables:

$$y^1 = \varphi^1(x^1, x^2, ..., x^n)$$

$$y^2 = \varphi^2(x^1, x^2, ..., x^n)$$

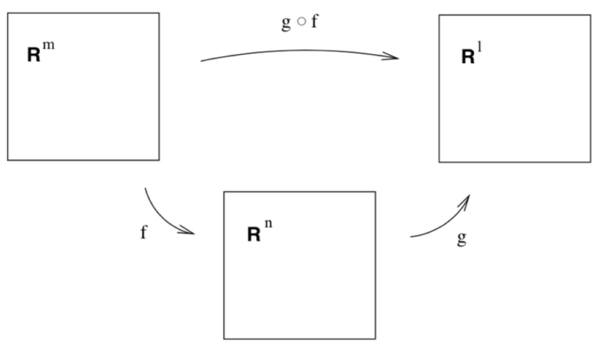
$$\vdots$$

$$y^m = \varphi^m(x^1, x^2, ..., x^n)$$

Let's determine p-differentiability and continuity of every function at the same time and call it C^p. If all elements of functions are no more than C^p then plead it to the map $\varphi : \mathbb{R}^n \to \mathbb{R}^m$. It can be seen that C^∞ is either continuous and infinitely differentiable whereas C^0 is just continuous though not differentiable. Hence it can be designated new type of mapping, that is **diffeomorphism** where two sets are given and C^∞ map with $\varphi : A \to B$ can be established, additionally its inverse $\varphi^{-1} : B \to A$ also defines.

1.4 Chain rule & Jacobian

To go further, it is needed to propose a **chain rule**. Consider two maps $f : \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^l$, therefore their composition would be $(g \circ f) : \mathbb{R}^m \to \mathbb{R}^l$



From the given information, we can take elements from each space, more precisely, let $x^{m'}$ on $\mathbb{R}^m$, $y^{n'}$ on $\mathbb{R}^n$ and $z^{l'}$ on $\mathbb{R}^l$ here m', n', l' are ranges of corresponding values. Here the partial derivatives of the composition and functions itself associate and it defines the chain rule:

$$\frac{\partial}{\partial x^{m'}}(g \circ f)^{l'} = \sum_{n'} \frac{\partial f^{n'}}{\partial x^{m'}} \frac{\partial g^{l'}}{\partial z^{m'}}$$

This equation can be written in the form:

$$\frac{\partial}{\partial x^{m'}} = \sum_{n'} \frac{\partial f^{n'}}{\partial x^{m'}} \frac{\partial}{\partial z^{m'}}$$

From Calculus III, it is known that the Jacobian of transformation $\mathbf{h} : \mathbb{R}^m \to \mathbb{R}^n$ at point x has the following form:

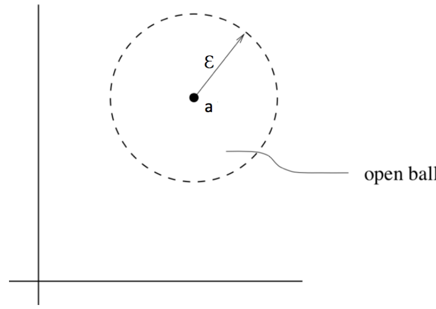
$$\frac{\partial h}{\partial x} = \begin{vmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \dots & \frac{\partial h_1}{\partial x_m} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \dots & \frac{\partial h_2}{\partial x_m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial h_n}{\partial x_1} & \frac{\partial h_n}{\partial x_2} & \dots & \frac{\partial h_n}{\partial x_m} \end{vmatrix}$$

In addition, the Jacobian has other representation:

$$\frac{\partial h}{\partial x} = \left[\frac{\partial h_i}{\partial x_j}\right]_{1 \le i \le m, 1 \le j \le n} = \begin{vmatrix} \frac{\partial h_1}{\partial x} \\ \frac{\partial h_2}{\partial x} \\ \dots \\ \frac{\partial h_n}{\partial x} \end{vmatrix} = \left[\frac{\partial h}{\partial x_1}, \frac{\partial h}{\partial x_2}, \dots, \frac{\partial h}{\partial x_m}\right]$$

1.5 Open ball & open set

The next main fundamental theory is a definition of an **open ball**. It defines for $a \in \mathbb{R}$ and $\epsilon \in \mathbb{R}$ as $B_\epsilon(a) = \{x \in \mathbb{R}^n : d(x, a) < \epsilon\}$. Notice that, it could be visualized as a sphere with the center a and radius ϵ.



From unions of open balls it could be generated **open sets**, more formally, consider some subset $S \subset \mathbb{R}^n$ and we claim that S is open, if for $a \in S$ there is an open ball near a that belongs to S.

1.6 Chart & Atlas

Given arbitrary point x in topological space T and some open subset of $\mathbb{R}^n$ such that x has a neighborhood S where exists homeomorphism ϕ from S to open subset of $\mathbb{R}^n$. Such topological space is called *locally Euclidean of dimension n*. The pair $(S, \phi : S \to \mathbb{R}^n)$ is called **chart** [3].

Definition 1. Given two charts $(M, \phi : S \to \mathbb{R}^n)$ and $(N, \phi : S \to \mathbb{R}^n)$ are called $C^\infty - compatible$ if:

$$\varphi_\alpha \circ {\varphi_\beta}^{-1} : \varphi_\beta(M \cap N) \to \varphi_\alpha(M \cap N), \quad \varphi_\beta \circ {\varphi_\alpha}^{-1} : \varphi_\alpha(M \cap N) \to \varphi_\beta(M \cap N)$$

where the maps $\varphi_\alpha \circ {\varphi_\beta}^{-1}$ and $\varphi_\beta \circ {\varphi_\alpha}^{-1}$ are C^∞.

Definition 2. Let $\{M_\alpha, \varphi_\alpha\}$ is an indexed collection of charts such that its union covers a given set L, more precisely, $L = \cup M_\alpha$. This collection is called C^∞ *atlas* of a set L.

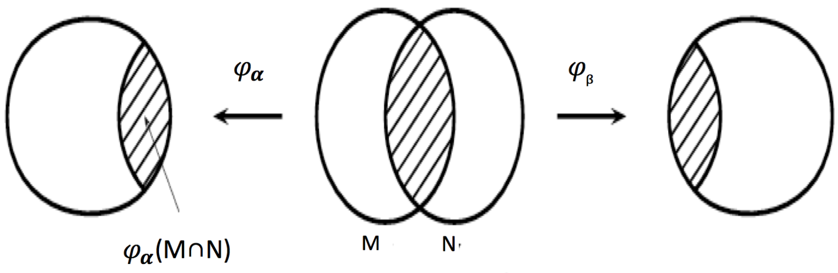


Figure 1-1: Illustration of maps

If atlas $\mathfrak{N}$ of a set L is not contained in any larger atlas of L, then this type of atlas is known as *maximal*.

Definition 3. A set N with maximal atlas is C^∞ **manifold** or *smooth*. If all elements of manifold have dimension n, then it is said to be $n - manifold$.

To check the smoothness of manifold, it is not needed to define maximal atlas. Recommend way is to follow the next proposition.

Definition 4. Any atlas $\mathfrak{M} = (M_\alpha, \varphi_\alpha)$ in a locally Euclidean space exists in a unique maximal atlas.

Proof. Let's collect all charts $(N_ß, \varphi_i)$ that are compatible with $\mathfrak{M}$ to the atlas $\mathfrak{M}$. All charts are compatible with each other. Thus, the whole collection is an atlas. If we take any chart that compatible with the new atlas, also compatible with $\mathfrak{M}$. And this chart belongs to a new atlas. Therefore, it can be concluded that the new atlas is maximal.

In order to prove its uniqueness, let's take a maximal atlas $\mathfrak{N}$ that contains atlas $\mathfrak{M}$. If there is exists another maximal atlas $\mathfrak{N}'$ that also contains $\mathfrak{M}$ atlas. Although, all charts in $\mathfrak{N}'$ are compatible with $\mathfrak{M}$ and it must belong to $\mathfrak{N}$. Therefore, $\mathfrak{N}' < \mathfrak{N}$. Furthermore, we assumed $\mathfrak{N}$ and $\mathfrak{N}'$ are maximal, then $\mathfrak{N} = \mathfrak{N}'$. It implies that maximal atlas which consists of $\mathfrak{M}$ is unique.

Why was it essential to scrupulously deal with charts and overlapping of them, rather than cover a manifold with only one chart? To answer this question, let's consider some examples.

Example 1. For instance, take T^1 and let $\mu : T^1 \to \mathbb{R}$, where $\mu = 0$ turns around to 2π. By definition of the chart, it is compulsory to have an open image $\mu(T^1)$ in $\mathbb{R}$. If we involve $\mu = 0$ or $\mu = 2\pi$, then we get closed interval instead of open interval. By eliminating both points, it couldn't be fully covered. Thus, it is needed to take at least 2 charts.

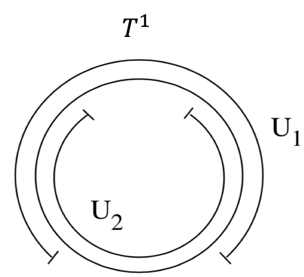


Figure 1-2: Illustration of Example 1

Example 2. Another example that shows the impossibility of covering a manifold with a single chart is T^2. Consider Mercator projection, also known as a cylindrical map, that passes through North and South poles. Assume T^2 is a collection of points that lies in $\mathbb{R}^3$ and satisfies this equation: $(x^1)^2+(x^2)^2+(x^3)^2 = 1$. By "stereographic projection" there can be created a chart that belongs to open set M_1, which identifies by subtracting north pole from the sphere.

Therefore, we take a line from the north pole and direct it to a plane that determines by $x^3 = -1$. Then, impose a point on T^2 with a Cartesian coordinate (y^1, y^2) that

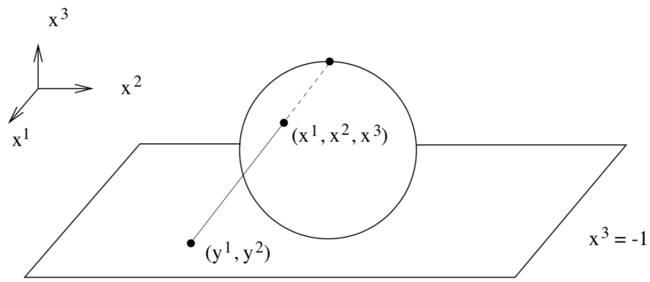


Figure 1-3: Illustration of Example 2

appeared from the intersection of line and plane. The map is defined as:

$$\varphi_1(x^1, x^2, x^3) = (y^1, y^2) = \left(\frac{2x^1}{1 - x^3}, \frac{2x^2}{1 - x^3}\right)$$

Second chart could be obtained by projecting from south pole to the plane $x^3 = +1$. Similarly with a case with north pole, the map can be defined as:

$$\varphi_2(x^1, x^2, x^3) = (z^1, z^2) = \left(\frac{2x^1}{1 + x^3}, \frac{2x^2}{1 + x^3}\right)$$

Consequently, φ_1 and φ_2 together can cover whole manifold and they overlap in $-1 < x^3 < 1$. In addition to this, we can check the composition of two maps $\varphi_1 \circ \varphi_2^{-1}$:

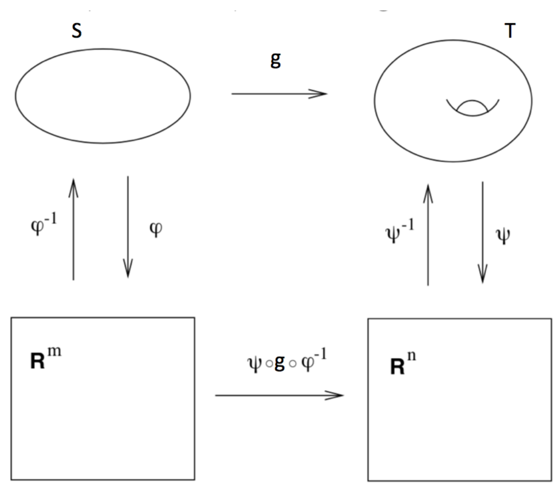
$$z^i = \frac{4y^i}{[(y^1)^2 + (y^2)^2]}$$

and this is C^∞ in covered region.

By summarising these examples, we can say that some manifolds can not be overlapped with only one chart or in our case, with one coordinate system.

1.7 Tangent vector & tangent space

As we mentioned before, locally manifolds look like $\mathbb{R}^n$ that emerged in the process of building coordinate charts. Therefore, operations such as differentiation and integration could be analyzed in manifolds. Assume S is m-dimensional and T n-dimensional manifolds with coordinate charts φ and ψ respectively. Let $g : S \to T$ be a function: By considering S and T as sets, we can't differentiate g owing to the



fact that, operation's behavior is not known. However, we can use coordinate charts to build a map $(\psi \circ g \circ \varphi^{-1}) : \mathbb{R}^m \to \mathbb{R}^n$. It is a simple Euclidean map, it implies that all principles from calculus can be applied. For instance, g can be considered as function on S that has a differentiation $\frac{\partial g}{x^\nu}$ and x^ν introduces $\mathbb{R}^m$.

$$\frac{\partial g}{x^\nu} = \frac{\partial}{x^\nu}(\psi \circ g \circ \varphi^{-1})x^\nu$$

After building the above framework, we can go further to define the structure of manifolds. Thus, we start this process by identifying tangent vectors and tangent spaces. In this study, it is necessary to eliminate a general sense of vectors, i.e. it is suggested to consider vectors as a single point, rather than thinking that it stretches from one point of the manifold to the second one. According to this statement, we can say "the vector points in x direction".

Suppose we are dealing with a problem where we want, by using only inherent things of the manifold U to build tangent space at some point t in U. The first thing that comes to mind, it is objects "tangent vectors to curves" that lie in tangent space. Thus, we can take a collection of all parameterized curves that pass through t, more precisely, it is a space of all maps $\beta : \mathbb{R} \to U$ where our point t lies in an image of β. Here difficult thing is to determine tangent space by considering the space of all tangent vectors to mentioned curves at some point t. However, we still don't know what is "tangent vector to a curve", nevertheless, space that constructed from vectors at t point is considered to be tangent space T_t.

1.8 Coordinate-free definition of tangent space

Now we want to give a definition of tangent space without mentioning coordinates. To this purpose, assume B is a space of all smooth functions on a manifold U. Every curve that passes through t describes directional derivative with such mapping: $g \to \frac{dg}{d\mu}$. According to this, we can declare that, by space of directional derivatives on curves that passes through t there can be determined tangent space T_t. To accomplish this claim, we need to show two facts:

- set of directional derivatives constructs vector space
- mentioned vector space is targeted vector space

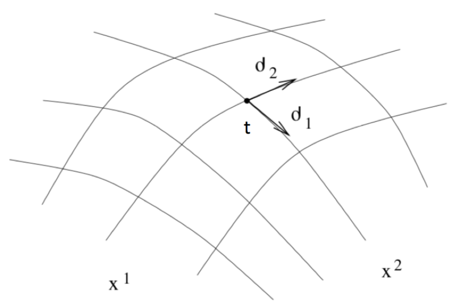
To demonstrate first item, suppose $\frac{d}{d\mu}$ and $\frac{d}{d\nu}$ are operators that define derivatives along curves across point t. It couldn't be an error if we do such changes to obtain another operator : $c\frac{d}{d\mu}+d\frac{d}{d\nu}$. On the one hand, this new operator is also an operator representing derivative. If operator linear on manifolds and performs usual Leibniz rule related to products, then such operator counts as good one. Let's check our operator.

$$(c\frac{d}{d\mu}+d\frac{d}{d\nu})(fg)=cf\frac{dg}{d\mu}+cg\frac{df}{d\mu}+df\frac{dg}{d\nu}+dg\frac{df}{d\nu}=$$

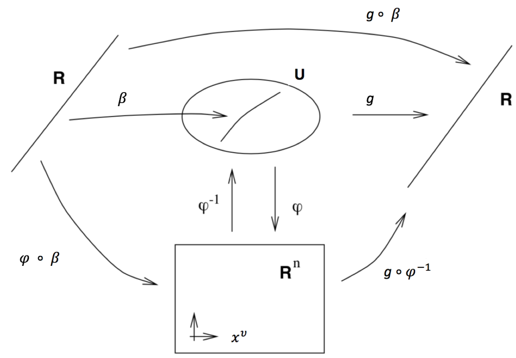
$$=(c\frac{df}{d\mu}+d\frac{df}{d\nu})g+(c\frac{dg}{d\mu}+d\frac{dg}{d\nu})f.$$

From the above equations, it is clear that Leibniz's rule satisfies. In fact, evidently, the new operator is linear. Therefore, the collection of directional derivatives constructs vector space. This learning confirms the first fact.

Then, we should answer to the second question. Whether the vector space we are looking for has been found. To be confident it is suggested to find a basis for vector space. Let x_ν coordinates of coordinate charts. At some point t there are exist partial derivatives $\partial\nu$ at t. Next, it is necessary to declare that, partial derivatives $\partial\nu$



at t construct a basis for T_t. In order to observe it, we should show that directional derivatives could be written as a product of summation of real numbers and partial derivatives. Suppose we have 4 main things : 1) n - dimensional manifold U, 2) curve $\beta:\mathbb{R}\to U$, 3) coordinate chart $\varphi:U\to\mathbb{R}^n$, 4) function $g:U\to\mathbb{R}$. Consider



μ is a parameter on a curve β, then we should extend the operator $\frac{d}{d\mu}$ by using ∂_ν. From Section 1.3, it could be written as:

$$\begin{aligned}
\frac{d}{d\mu}g &= \frac{d}{d\mu}(g \circ \beta) \\
&= \frac{d}{d\mu}[(g \circ \varphi^{-1}) \circ (\varphi \circ \beta)] \\
&= \frac{d(\varphi \circ \beta)^\nu}{d\mu}\frac{\partial(g \circ \varphi^{-1})}{\partial x^\nu} \\
&= \frac{dx^\nu}{d\mu}\partial_\nu g
\end{aligned}$$

So, let's explain these equations. In the first equation, we take expression on the left-hand side and rewrite it with a derivative of $(g \circ \beta)$. Owing to the associativity of composition we can do such operation as in the second equation. The third equation obtained by chain rule whereas the last one came from initial notations. There have taken arbitrarily g, thus observed results could be written as:

$$\frac{d}{d\mu} = \frac{dx^\nu}{d\mu}\partial_\nu$$

In fact, such partials ∂_ν introduce a basis for vector space that constructed from directional derivatives. And these directional derivatives determine tangent space.

∂_ν is called **coordinate basis** for T_t and it is a general formalization of concept where basis vectors to some point set on coordinate axes. And there is no explanation of using coordinate bases when some bases could be more advantageous.

1.9 Examples of manifolds in question

In all examples below, we first define $M \subseteq R^k$.

There is a standard topology τ_k on R^k which is defined as:

$$\tau_k = \{S | S \subseteq R^k, \textit{where } S \textit{ is an open set}\}$$

Given $M \subseteq R^k$, topology on M is simply induced by τ_k by the following rule:

$$\tau = \{S \cap M | S \in \tau_k\}$$

Let's consider some examples of tangent space according to mentioned statements.

Example 1. Suppose $f(x)$ is infinitely differentiable and continuous function, i.e. $f(x) \in C^\infty(\mathbb{R})$. And topology on the graph of $f(x)$ is $T = (M, \tau)$, where M is given by this rule:

$$M = \{(x, f(x)) | x \in (a, b)\}$$

which means that $M \subseteq \mathbb{R}^2$.

We should construct tangent space on M. For this purpose, assume $\bar{\tau} = (x^*, y^*) \in M$. From Calculus of Vector Valued Functions, it is known that, derivative of $f(x)$ is given by:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

and $\langle 1, f'(x) \rangle$ is a tangent vector to a graph of $f(x)$ at point x.

By shifting x^* by Δx we obtain

$$\bar{\tau}' = (x^* + \Delta x, f(x^* + \Delta x)).$$

Difference between $\bar{\tau}$ and $\bar{\tau}'$ is

$$\bar{\tau}' - \bar{\tau} = (\Delta x, f'(x) \cdot \Delta x) = \Delta x(1, f'(x))$$

From observed result, it can be concluded that, desired tangent space $T_{\bar{\tau}}$ is:

$$T_{\bar{\tau}} = span\{\bar{t}\}$$

where $\bar{t} = (1, f'(x))$.

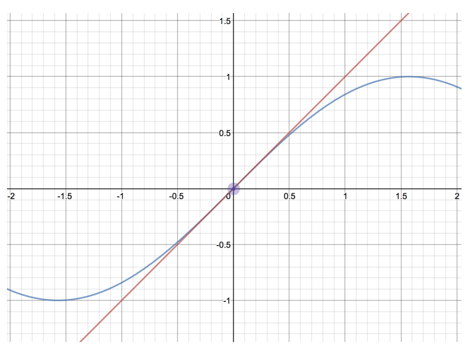


Figure 1-4: Illustration of tangent space of Example 1

Example 2. Assume $f(x, y) \in C^\infty(\Omega), \Omega \subseteq \mathbb{R}^2$ where Ω is open. As in Example

1, topology on $f(x,y)$ is $T = (M, \tau)$, where M is given by this rule:

$$M = \{(x, y, f(x,y)) | (x,y) \in \Omega\}$$

In order to find tangent space, lets define $\bar{\tau}$ as:

$$\bar{\tau} = (x^*, y^*, f(x^*, y^*))$$

Small shifting in x^* and y^* will give us following results:

$$\bar{\tau}_1 = (x^* + \Delta x, y^*, f(x^* + \Delta x, y^*))$$

$$\bar{\tau}_2 = (x^*, y^* + \Delta y, f(x^*, y^* + \Delta y))$$

Difference between main $\bar{\tau}$ and $\bar{\tau}_1, \bar{\tau}_2$ is respectively:

$$\bar{\tau} - \bar{\tau}_1 = (\Delta x, 0, \frac{\partial f(x^*, y^*)}{\partial x}\Delta x) = \Delta x(1, 0, \frac{\partial f(x^*, y^*)}{\partial x})$$

And

$$\bar{\tau} - \bar{\tau}_2 = (0, \Delta y, \frac{\partial f(x^*, y^*)}{\partial y}\Delta y) = \Delta y(0, 1, \frac{\partial f(x^*, y^*)}{\partial y})$$

Lets denote $f(x^*, y^*)$ as f. Then desired tangent space will be:

$$T_\tau = span\left\{ \begin{bmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial x} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \frac{\partial f}{\partial y} \end{bmatrix} \right\}$$

Example 3. Instead of considering one function lets consider 2 functions. Suppose

$f(x,y), g(x,y) \in C^{\infty}(\Omega)$ where $\Omega \subseteq \mathbb{R}^2$ and open.

To find tangent space, we should first define M.

$$M = \{(x, y, f(x,y), g(x,y) | (x,y) \in \Omega\}$$

Given $M \subseteq \mathbb{R}^4$, some point in M :

$$\bar{\tau} = (x^*, y^*, f(x^*, y^*), g(x^*, y^*))$$

To find tangent space, we change x^* and y^* a little by $\Delta x, \Delta y$ respectively. These changes obviously will affect on $\bar{\tau}$ and it gives following results:

$$\bar{\tau}_1 = (x^* + \Delta x, y^*, f(x^* + \Delta x, y^*), g(x^* + \Delta x, y^*))$$

$$\bar{\tau}_2 = (x^*, y^* + \Delta y, f(x^*, y^* + \Delta y), g(x^*, y^* + \Delta y))$$

Lets calculate the difference between $\bar{\tau}$ and changed $t\bar{a}u_1, t\bar{a}u_2$. And denote $f(x^*, y^*)$ as f and $g(x^*, y^*)$ as g for convenience.

$$\bar{\tau} - \bar{\tau}_1 = (\Delta x, 0, \frac{\partial f}{\partial x}\Delta x, \frac{\partial g}{\partial x}\Delta x) = \Delta x(1, 0, \frac{\partial f}{\partial x}, \frac{\partial g}{\partial x})$$

$$\bar{\tau} - \bar{\tau}_2 = (0, \Delta y, \frac{\partial f}{\partial y}\Delta y, \frac{\partial g}{\partial y}\Delta y) = \Delta y(0, 1, \frac{\partial f}{\partial y}, \frac{\partial g}{\partial y})$$

Tangent space is a set of tangent vectors. Thus, from above equations, it is seen that tangent space has this form:

$$T_{\bar{\tau}} = span\left\{ \begin{bmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial x} \\ \frac{\partial g}{\partial x} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial y} \end{bmatrix} \right\}$$

Example 4. Now lets increase number of parameters and functions. So we have $f_1(x_1, x_2, ..., x_n), f_2(x_1, x_2, ..., x_n), ..., f_m(x_1, x_2, ..., x_n)$, where $f_i \in C^\infty(\mathbb{R}^n), i = \overline{1, m}$. M is given by this rule:

$$M = \{(x_1, ..., x_n, f_1(x_1, ..., x_n), ..., f_m(x_1, ..., x_n)|(x_1, ..., x_n) \in \Omega\}$$

As in previous examples, to find tangent space, first we should take a point:

$$\bar{\tau} = (x_1^*, ..., x_n^*, f_1(x_1^*, ..., x_n^*), ..., f_m(x_1^*, ..., x_n^*))$$

If we make small shifting in $x_1^*, ..., x_n^*$, we get:

$$\bar{\tau}_1 = (x_1^* + \Delta x_1, ..., x_n^*, f_1(x_1^* + \Delta x_1, ..., x_n^*), ..., f_m(x_1^* + \Delta x_1, ..., x_n^*))$$

$$\bar{\tau}_2 = (x_1^*, x_2^* + \Delta x_2 ..., x_n^*, f_1(x_1^*, x_2^* + \Delta x_2, ..., x_n^*), ..., f_m(x_1^* x_2^* + \Delta x_2, ..., x_n^*))$$

$$\vdots$$

$$\bar{\tau}_n = (x_1^*, ..., x_n^* + \Delta x_n, f_1(x_1^*, ..., x_n^* + \Delta x_n), ..., f_m(x_1^*, ..., x_n^* + \Delta x_n))$$

It diverges from original $\bar{\tau}$ in this way:

$$\bar{\tau} - \bar{\tau}_1 = \Delta x_1 \big(1, 0, ..., 0, \frac{\partial f_1}{\partial x_1}, ..., \frac{\partial f_m}{\partial x_1}\big)$$

$$\bar{\tau} - \bar{\tau}_2 = \Delta x_2 \big(0, 1, ..., 0, \frac{\partial f_1}{\partial x_2}, ..., \frac{\partial f_m}{\partial x_2}\big)$$

$$\vdots$$

$$\bar{\tau} - \bar{\tau}_n = \Delta x_n \big(0, 0, ..., 1, \frac{\partial f_1}{\partial x_n}, ..., \frac{\partial f_m}{\partial x_n}\big)$$

Finally, tangent space is

$$T_{\bar{\tau}} = span \left\{ \begin{bmatrix} 1 \\ 0 \\ ... \\ 0 \\ \frac{\partial f_1}{\partial x_1} \\ ... \\ \frac{\partial f_1}{\partial x_1} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ ... \\ 0 \\ \frac{\partial f_1}{\partial x_2} \\ ... \\ \frac{\partial f_1}{\partial x_2} \end{bmatrix}, ..., \begin{bmatrix} 0 \\ 0 \\ ... \\ 1 \\ \frac{\partial f_1}{\partial x_n} \\ ... \\ \frac{\partial f_1}{\partial x_n} \end{bmatrix} \right\}$$

The columns of $T_{\bar{\tau}}$ form basis for R^n. Therefore it can be concluded that, $T_{\bar{\tau}}$ is n-dimensional vector space and $M \in \mathbb{R}^{n+m}$.

Example 5. We can complicate example by removing variables in topology, i.e. we have only functions in our topology $f(x, y), g(x, y) \in C^\infty(\mathbb{R}^2)$. Assume $\varphi : \Omega \to M$ which is one - to - one mapping. And

$$\varphi\Big(\begin{bmatrix} x \\ y \end{bmatrix} \Big) = \begin{bmatrix} f(x, y) \\ g(x, y) \end{bmatrix} \in C^\infty$$

Here

$$(\Omega, \varphi : \Omega \to M) - \text{chart}$$

In other words, M is given by single chart. Lets fix condition on $f(x, y)$ and $g(x, y)$:

$$\frac{\partial(f, g)}{\partial(x, y)} \neq 0, \text{ where } (x, y) \in \Omega$$

According to these statements, M is given by:

$$M = \{\varphi(x, y) | (x, y) \in \Omega, \varphi - \text{one-to-one mapping}\}$$

Take a point on M:

$$\bar{\tau} = (f(x^*, y^*), g(x^*, y^*))$$

We do same things as in previous examples, by shifting x^* and y^* we get:

$$\bar{\tau}_1 = (f(x^* + \Delta x, y^*), g(x^* + \Delta x, y^*))$$

And

$$\bar{\tau}_2 = (f(x^*, y^* + \Delta y), g(x^* +, y^* + \Delta y))$$

In order to find tangent vectors, we do such operation:

$$\bar{\tau} - \bar{\tau}_1 = \Delta x(\frac{\partial f}{\partial x}, \frac{\partial g}{\partial x})$$

$$\bar{\tau} - \bar{\tau}_2 = \Delta y(\frac{\partial f}{\partial y}, \frac{\partial g}{\partial y})$$

Set of tangent vectors :

$$T_{\bar{\tau}} = span\left\{ \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial g}{\partial x} \end{bmatrix}, \begin{bmatrix} \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial y} \end{bmatrix} \right\}$$

It is valid when $det \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix} \neq 0$. And due our assumptions tangent space is definable.

Chapter 2

Main reslut s

2.1 Problem formulation

As mentioned before, finding a tangent space of a manifold at a certain point contributes to an effective solution to the dimension reduction problem.

In real-world problems, some data could be looking like manifold, although it is not manifold yet. This study works with this kind of data.

Our manifold M is not given directly. And we assume that the manifold M is equipped with probabilistic space structure $\Omega = (M, \sigma, P)$ where $\sigma \subseteq 2^M$ is the sigma algebra of events and $P : \sigma \to \mathbb{R}$ is a probability function, such that $P(S) = 1$.

What does it mean that manifold M is not given directly? Probably, some noise appears in data. According to this statement, suppose that points $\mathbf{z}_1, \cdots, \mathbf{z}_N \overset{\text{iid}}{\sim} P$ are random points on the manifold M and $\epsilon_1, \cdots, \epsilon_N \overset{\text{iid}}{\sim} N(\mathbf{0}, \delta^2 I_n)$ are noise vectors.

Therefore, any point in our data could be written in this way:

$$x_i = z_i + \epsilon_i, \text{ where } i = 1, ..., N$$

From above statement, it is seen that points $\mathbf{z}_1, \cdots, \mathbf{z}_N \stackrel{\text{iid}}{\sim} P$ are not given directly, but with a little noise!

Now given a point $x^* \in M$ and data $x_1, ..., x_N$ our goal is to reconstruct the tangent space at the point x^*.

There are several ways to find tangent space to the manifold at some point. In this work, first, we are working with the Principal Component Analysis method.

2.2 Principal components analysis

Principal Component Analysis(PCA) is a one of the basic methods to reduce dimension of a complex data. Main useful side of PCA is losing less information about data when reducing dimension of the data.

The main target of PCA is to find *principal components* that construct collection of projections which are ordered in variance.

In this section, we represent principal components as linear manifolds approaching family of N points $x_i \in \mathbb{R}^m$.

Before going to PCA algorithm, first, recall that, a *linear subspace* $V \in \mathbb{R}^n$ is subset V that contains zero vector and closed under addition and scalar multiplication. While an *affine subspace* is a set $P = \{a + x | x \in Q = a + Q\}$ where $a \in \mathbb{R}^n$ and $Q \in \mathbb{R}^n$.

Let $x_1, x_2, ..., x_N$ be our observations and the linear model with rank n that corresponds to this observation is given by:

$$g(\mu) = \nu + V_n \mu$$

where μ stands for an m-dimensional vector of parameters. A matrix $V_n = \{v_1, v_2, ..., v_m\}$

is an orthogonal matrix with $n \times m$ dimension and ν is a location vector. Here $V_n\mu = \sum_{i=1}^{m} v_i\mu_i \in Col(V_n)$ and $dim\ Col(V_n) \leq m$. Therefore $V_n\mu$ is a linear subspace, whereas $\nu + V_n\mu$ is an affine subspace of rank n. The illustration of PCA is seen in Figure 1-5, where $n = 1$ and $n = 2$: According to Least Square Estimation,

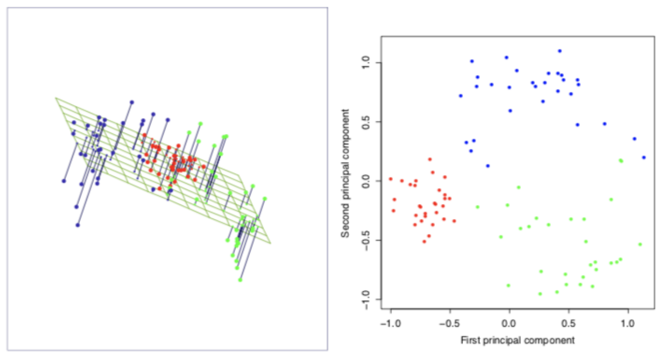


Figure 2-1: *Demonstration of the best linear approximation to the "half-sphere" data. In the second panel it can be seen the projected points of the given data. The rank of approximation is 2.*

our goal is to minimize the error:

$$\min_{\nu,\{\mu_i\},V_n} \sum_{i=1}^{N} ||x_i - \nu - V_n\mu_i||^2$$

From Calculus II, to minimize some function, we should take partial derivatives of above expression with respect to ν and μ_i:

$$\hat{\nu} = \bar{x},$$

$$\hat{\mu_i} = V_n^T(x_i - \bar{x}).$$

Then our goal is to find V_n:

$$\min_{V_n} \sum_{i=1}^{N} ||x_i - \bar{x}) - V_n V_n^T (x_i - \bar{x})||^2.$$

To simplify above expression we take $\bar{x} = 0$. Let's denote by S_n the expression $S_n = V_n V_n^T$. It is a "projection matrix" of dimension $m \times m$ that maps all x_i onto $S_n x_i$. By putting the transformed observations (recall, $\bar{x} = 0$) into a matrix X with dimension $N \times m$ as rows, we can form *singular value decomposition* of X:

$$X = UDV^T$$

where U has orthogonal columns u_j which are known as "left singular vectors". Therefore, U is an orthogonal matrix of size $n \times m$. And V is also orthogonal matrix of size $m \times m$ that has columns v_j which are known as "right singular vectors" and D is a diagonal matrix of size $m \times m$ that has positive nonzero *singular values* in its diagonal. From above equation, we say that the principal components of X are the columns of UD. The n columns of V contain solution of V_n [1].

In the figure below, it can be seen a demonstration of the one-dimensional principal component in $\mathbb{R}^2$. By projecting each of x_i onto the line $u_i d_1 v_1$ we can find the minimum distance between the point and line. The multiplication $u_i d_1$ is a distance between the origin and points along this line whereas v_i is direction vector of the line.

In fact, the first principal component gives the highest variance and the second one is the second highest and so on.

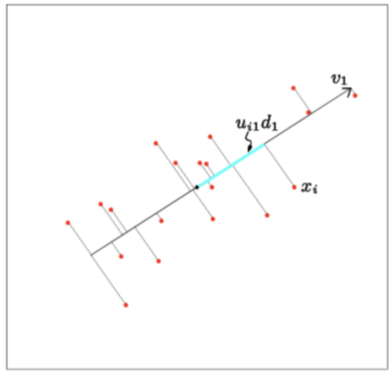


Figure 2-2: *The first principal component. The points are projected to the line and these distances are minimum.*

2.3 Baseline method (using PCA)

In this section, we present a Principal Components Analysis method. However, first we should find a tangent space to a manifold by using directional derivative to compare with the PCA method for efficiency.

The tangent space to a manifold is considered to be a set of all possible directional derivatives. In some sense, the tangent space $T_p(M)$ of a manifold M with dimension n at a point p is a hyperplane which approximates the manifold M in the best way.

Before finding tangent space on an n dimensional manifold, first, let's try to work with three-dimensional case.

Let $S \subseteq \mathbb{R}^3$ be a plane spanned by a set of vectors $\{\bar{u}_1, \bar{u}_2\}$, where $||u|| = 1$ and $\bar{u}_1, \bar{u}_2$ are orthogonal, i.e. $\bar{u}_1 \cdot \bar{u}_2 = 0$. Also O is a mapping $O : \mathbb{R}^3 \to \mathbb{R}^3$ with such operation $O(\bar{x}) = P\bar{x}$ where $P \in \mathbb{R}^{3\times 3}$. Our goal is to find this operator P.

Let $\bar{v} \in \mathbb{R}^3$ and we project this $\hat{v}$ onto S. From Linear Algebra, the projection of the vector to a plane $\bar{v} \to \bar{v}'$ will be $\bar{v} \to P\bar{v}$:

$$\bar{v}' = Proj_{\bar{u}_1}\bar{v} + Proj_{\bar{u}_2}\bar{v}$$

By using the formula of projection we get:

$$\begin{aligned}
\bar{v}' &= (\bar{v} \cdot \bar{u}_1)\bar{u}_1 + (\bar{v} \cdot \bar{u}_2)\bar{u}_2 \\
&= \bar{u}_1\bar{u}_1^T\bar{v} + \bar{u}_2\bar{u}_2^T\bar{v} \\
&= (\bar{u}_1\bar{u}_1^T + \bar{u}_2\bar{u}_2^T)\bar{v} \\
&= P\bar{v}
\end{aligned}$$

$P = \bar{u}_1\bar{u}_1^T + \bar{u}_2\bar{u}_2^T$ is a projection operator and $\bar{v}, \bar{u}_1, \bar{u}_2 \in \mathbb{R}^3$. Therefore we can generalize formula for $i = n$ as follows:

$$P = \sum_{i=1}^{n} \bar{u}_i\bar{u}_i^T$$

At the end, the multiplication $P \cdot \bar{v}$ gives us desired tangent space. As we mentioned before, this traditional method of finding tangent space is needed to check the results of the PCA method.

To use the PCA method our data should be close to $\mathbf{0}$ (Section 1.10). For this reason, we choose ϵ and find all the points $\mathbf{x}_i$ such that $||\mathbf{x}_i|| < \epsilon : \mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_l}$.

For deep understanding, let us do the PCA algorithm for a lower dimensional cases as in general way. Let $\bar{x}_1, \bar{x}_2, ..., \bar{x}_l \in \mathbb{R}^3$.

$$\bar{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix}$$

Let

$$X = [\bar{x}_1, ..., \bar{x}_l] = \begin{bmatrix} x_{11} & x_{21} & \cdots x_{l1} \\ x_{12} & x_{22} & \cdots x_{l2} \\ x_{13} & x_{23} & \cdots x_{l3} \end{bmatrix}$$

As it is seen, $X \in \mathbb{R}^{3\times l}$. From Section 1.10 we know that to find principal components, first, we should construct SVD by this formula $X = U\Sigma V^T$.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \quad U = [\bar{r}_1 \ \ \bar{r}_2 \ \ \bar{r}_3]$$

where $\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0$ and $||\bar{r}_i|| = 1$ and $\bar{r}_i \cdot \bar{r}_j, \ i \neq j$.

Let's find a two dimensional principal components space.

$$X = U\Sigma V^T \approx U\Sigma' V^T = X'$$

where Σ' is obtained by truncation. The last singular value σ_3 is very close to $\mathbf{0}$.

$$\Sigma' = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The $rankX = 3$, in other words, dimension of column space of X is equal to 3 while the $rankX' = 2$. It can be easily proved by property of the rank:

$$rankX' = rank(U\Sigma' V^T) = rank\Sigma' = 2$$

By taking the first two columns of U matrix we define first two principal components.

$$P' = \bar{r}_1\bar{r}_1^T + \bar{r}_2\bar{r}_2^T = U'(U')^T$$

By multiplying P' to elements of X we get projected .

So we have $X = [\bar{x}_1, \bar{x}_2, \cdots \bar{x}_l]$ and $X' = [\bar{x'}_1, \bar{x'}_2, \cdots \bar{x'}_l]$. Also, we have the correct P and estimated P' with $\bar{x} \to P\bar{x}$, $\bar{x} \to P'\bar{x}$, respectively.

To check performance of PCA algorithm, we take Frobenius norm:

$$d(P, P') = ||(P - P')||_F$$

where Frobenius norm is computed by this formula $||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{i=1}^{n}|a_{ij}|}$.

2.4 Maximum Mean Discrepancy distance

Let

$$k(x) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} e^{-\frac{|x|^2}{2\sigma^2}}$$

be the gaussian kernel on $\mathbb{R}^n$

Let $\mathcal{B}$ be a set of Borel probability measures μ defined on $\mathbb{R}^n$ such that their weighted Fourier transform is square integrable:

$$e^{-\frac{\sigma^2|x|^2}{2}}\mathcal{F}[\mu] = e^{-\frac{\sigma^2|x|^2}{2}} \int e^{-iy^Tx} d\mu(y) \in L_2(\mathbb{R}^n)$$

After applying Weierstrass transform we get :

$$\mu \in \mathcal{B} \to \phi(\mu) = \mathbb{E}_{y\sim\mu} k(x - y)$$

Then Maximum Mean Discrepancy (MMD) distance is defined as the distance induced by metrics on $L_2(\mathbb{R}^n)$, i.e. for $\mu, \nu \in \mathcal{B}$:

$$d_{MMD}(\mu, \nu) = ||\phi(\mu) - \phi(\nu)||_{L_2(\mathbb{R}^2}$$

Here, $\phi(\mu)$ depends on the distribution type μ: discrete or continuous.

$$\phi(\mu) = \frac{1}{N} \sum_{i=1}^{N} k(x - x_i) \quad \text{for discrete}$$

$$\phi(\mu) = \int k(x - y) \quad \text{for continuous}$$

Let $x_1, ..., x_N \in \mathbb{R}_n$ be our dataset points. This dataset defines the empirical probabilistic measure μ_{Data}:

$$\underset{Borel}{\forall A} : \mu_{Data}(A) = \frac{|x_i | x_i \in A|}{N}$$

As it is seen, by smoothing our empirical probability measure μ_{Data} we obtain:

$$\phi(\mu_{Data}) = \frac{1}{N} \sum_{i=1}^{N} k(x - x_i)$$

This result is the same as kernel density estimation.

Now we need to find ν with such restrictions:

- ν is close to empirical probability measure μ
- $\mathcal{G}_k = \{\mu \in \mathcal{B} | \exists v_1, ..., v_k \in \mathbb{R}^n \quad \underset{Borel}{\forall A} : \mu(A) = \mu(A \cap \text{ span } (v_1, ..., v_k))\}$

i.e. is a set of probability measures with all probability concentrated in some m-dimensional subspace span$(v_1, ..., v_k)$ where $m \leq k$.

We will study the method concurrent to PCA that is based on solving the problem:

$$I(\nu) = d_{MMD}(\mu_{Data}, \nu) = ||\phi(\mu_{Data}) - \phi(\nu)||_{L_2(\mathbb{R}^n)} \rightarrow \min_{\nu \in \mathcal{G}_k}$$

i.e. we want to approximate the empirical probabilistic measure μ_{Data} by another probabilistic measure ν which is supported in some k-dimensional subspace of $\mathbb{R}^n$.

Chapter 3

Experiments

3.1 Experiments with synthetic manifold data

All datasets in this section were created by examples that are given below.

Example 1. We choose a function $f(x)$ and generate points in the following way:

- Generate $x_1, x_2, \cdots, x_N$ uniformly in interval $[-2, 2]$;
- Calculate $y_i = f(x_i)$;
- Generate 2-dimensional gaussian error $\epsilon_i \sim N([0,0]^T, I_2)$ and find $\mathbf{x}_i = [x_i, y_i]^T + \epsilon_i$;

So we can generate points $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\} \subseteq \mathbb{R}^2$ and set $k = 1$. In other words, the problem is equivalent to finding tangent line to the curve $y = f(x)$ at point $x = 0$. Therefore, the right answer will be that $\mathcal{L} = span\{[1 \;\; f'(0)]^T\}$. Therefore we know with which vector to compare the result.

Example 2. Choose a function $f(x, y)$ and generate points in the following way:

- Generate $x_1, x_2, \cdots, x_N, y_1, y_2, \cdots, y_N$ uniformly in interval $[-2, 2]$;

- Calculate $z_i = f(x_i, y_i)$;
- Generate 3-dimensional gaussian error $\epsilon_i \sim N([0,0,0]^T, I_3)$ and find $\mathbf{x}_i = [x_i, y_i, z_i] + \epsilon_i$;

So we can generate points $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\} \subseteq \mathbb{R}^3$ and set $k = 2$. In other words, the problem is equivalent to finding tangent plane to the surface $z = f(x, y)$ at point $x = 0, y = 0$. Therefore, the right answer will be $\mathcal{L} = span\{[1\ \ 0\ \ f'_x(0)]^T, [0\ \ 1\ \ f'_y(0)]^T\}$. It was shown in Section 1.8. Then this result is needed to compare output from PCA. To construct datasets following functions was used:

1. $f(x) = sin(x)$
2. $f(x) = tan(x)$
3. $f(x) = x^2 + 4x$
4. $f(x) = x^5 - 3x$
5. $f(x, y) = tan(2x + y)$
6. $f(x, y) = sin(x + y)$
7. $f(x, y) = x^4 + 2y$
8. $f(x, y) = x^2 + y^2 - 5x$
9. $f(x, y) = e^{(x+y)} - 1$
10. $f(x, y) = x^3 - y^3 + 3y$

We take 20000 samples and add some noise to it according to the mentioned instructions. Then we can apply the PCA to created data.

The code was implemented in Python 2.7. Here the dataset was constructed in this way:

```
import random
X = []
for x in range(20000):
    X.append(random.uniform(-10,10))
print X
import math
Y = np.sin(np.array(X))
```

And add some noise to the data by following code:

```
mean = [0,0]
cov = [[0.03,0], [0,0.03]]
e=np.random.multivariate_normal(mean, cov, 20000)
```

After creating data points, we save them as ".txt" file for accessing data when implementing PCA algorithm.

```
np.savetxt("data1.txt",x, delimiter = ' ',newline='\n')
a = open("data1.txt", 'r')# open file in read mode
```

Then we can create Dataframe from above data points:

```
import pandas as pd
data = pd.read_csv('data1.txt',sep = ' ', header = None)
df = data.T
```

Our goal is to choose ϵ and find all points close to $\mathbf{0}$, i.e. $||x_i|| < \epsilon$.

```
norm = []
```

```
for x in range(len(data):
    normed = np.linalg.norm(df[x])
    norm.append(normed)

data_new = []
df = np.asarray(df)
for i in range(len(norm)):
    if norm[i]<=0.3: # epsilon
        data_new.append(df[:,i])
data_new = pd.DataFrame(data_new)
```

The next step is to find SVD of the data. And we set $k = 1$ or $k = 2$ according to a dimension of the data. First k columns are principal components.

```
u, s, v = np.linalg.svd(df_n)

r1 = u[:,0]
r2 = u[:,1]
r_1 = np.reshape(r1,(3,1))
r_2 = np.reshape(r2,(3,1))

P = r_1*r_1.T + r_2*r_2.T
```

The calculation of tangent spaces of the functions at point 0 was done by hand. In the Python shell, we wrote only the results of them.

```
P_exact = [[0.5,0.5],[0.5,0.5]]
difference = P - P_exact
Norm = np.linalg.norm(difference,'fro')
```

By implementing this code to all functions we get good results for PCA.

1	2	3	4	5	6	7	8	9	10
0.0199	0.0775	0.0148	0.0127	0.0202	0.0437	0.0199	0.0247	0.0671	0.0382

Table 3.1: Results of PCA for 10 functions

From the Table 3.1 it can be concluded that, we can use results of PCA in order to estimate concurrent method's(MMD) results.

		1	2	3	4	5
Local	PCA	0.0199	0.0775	0.0148	0.0127	0.0202
	MMD	0.0059	0.0109	0.0097	0.0105	0.0452
Global	PCA	0.6905	0.1904	0.3163	0.8183	0.6391
	MMD	0.0018	0.0029	0.0174	0.9933	0.9388

Table 3.2: Results of PCA and MMD.

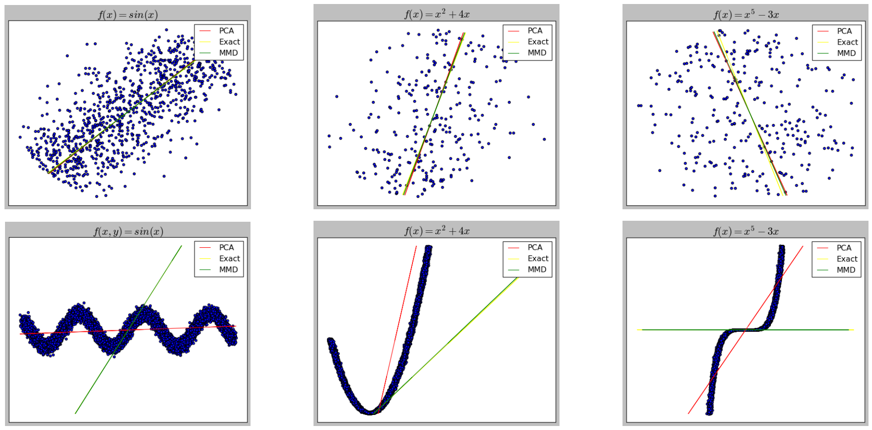


Table 3.3: Visualization of outputs of the PCA and MMD methods in two-dimensional case

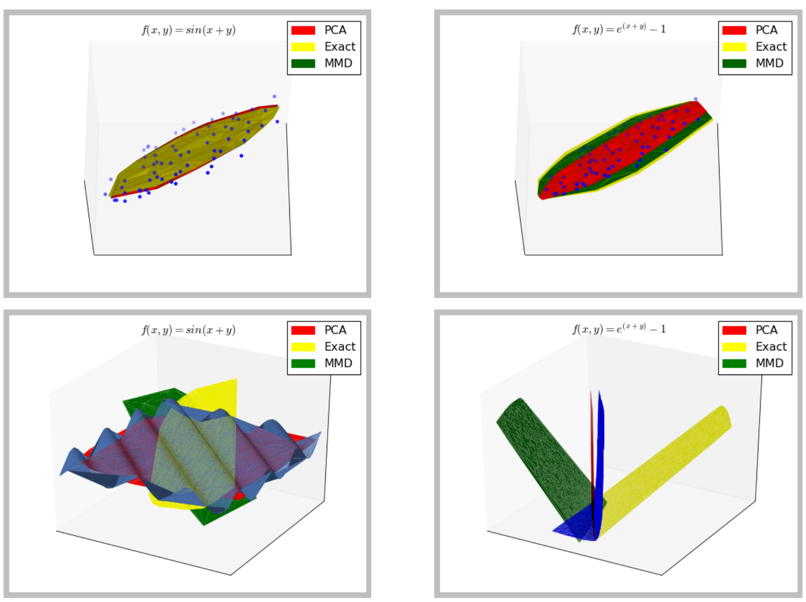


Table 3.4: Visualization of outputs of the PCA and MMD methods in three-dimensional case.

Chapter 4

Conclusion

4.1 Conclusion

From the results of the experiment it can be conluded that,the behaviour of PCA and MMD on the local dataset are almost identical, but they are significantly different on global one. MMD,unlike PCA, tried to catch ideal alignment of points rather than searching global alignment of points. This property of MMD makes it promising tool for the problem of the tangent space calculation to a data manifold at a given point.

On top of that, for two-dimensional data both methods give almost perfect results. However, for three-dimensional case, PCA tried to find a global pattern by taking all points into account .And the MMD focused on local data points and produced results, as shown in the figures above.

Bibliography

[1] Friedman J., Hastie T., Tibshirani R. The elements of statistical learning. – New York : Springer series in statistics, 2001. – . 1. – . 10.

[2] Milnor, John. "On manifolds homeomorphic to the 7-sphere." Annals of Mathematics (1956): 399-405.

[3] Loring, W. Tu. "An introduction to manifolds." (2008), Springer

[4] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 290: 2323–2326, 2000.

[5] Wang, Jing, Zhenyue Zhang, and Hongyuan Zha. "Adaptive manifold learning." Advances in neural information processing systems. 2005.

[6] Z. Zhang and H. Zha. Principal Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. SIAM J. Scientific Computing, 26:313–338, 2004.

[7] Zhang, Tianhao, et al. "Linear local tangent space alignment and application to face recognition." Neurocomputing 70.7-9 (2007): 1547-1553.