Calculus On Normed Vector Spaces

MTP REPORT

July 10, 2018

Submitted by
Joydeep Medhi
Entry No. 2013MT60599

Supervisor
Prof. Amit Priyadarshi

Department of Mathematics
Indian Institute of Technology Delhi,
New Delhi, INDIA

1 Introduction

The aim of the project is to study the notion of derivatives on general normed vector spaces and do Calculus on them. We will also look at some applications of these concepts. Till **section 5** was done for Mid-Term presentation and from **section 6 to section 10** is done after Mid-Term Presentations.

Norm: We will suppose that all vector spaces are real. Let E be a vector space. A mapping $\|.\| : E \to \mathbb{R}$, is said to be a *norm* if, for all $x,\ y \in E$ and $\lambda \in \mathbb{R}$

$\|x\| \geq 0$
$\|x\| = 0 \Leftrightarrow x = 0$
$\|\lambda x\| = |\lambda|\, \|x\|$
$\|x + y\| \leq \|x\| + \|y\|$

The pair $(E, \|.\|)$ is called a *normed vector space* and we say that $\|x\|$ is the norm of x.

Continuity: Suppose now that we have two normed vector spaces, $(E, \|.\|_E)$ and $(F, \|.\|_F)$. Let A be a subset of E, f a mapping of A into F and $a \in A$. We say that f is *continuous* at a if the following condition is satisfied:

for all $\epsilon > 0$, there exists $\delta > 0$ such that, if $x \in A$ and $\|x - a\|_E < \delta$,
then $\|f(x) - f(a)\|_F < \epsilon$
If f is *continuous* at every point $a \in A$, then we say that f is *continuous* on A.

Proposition 1.1. *The norm on a normed vector space is a continuous function.*

Proof. We have $\|x\| = \|x - y + y\| \leq \|x - y\| + \|x\| \Rightarrow \|x\| - \|y\| \leq \|x - y\|$
In similar way, $\|y\| - \|x\| \leq \|y - x\|$. As $\|y - x\| = \|x - y\|$, We get

$$|\ \|x\| - \|y\|\ | \leq \|x - y\|$$

And hence the continuity.

Proposition 1.2. *Let E and F be normed vector spaces, $A \subseteq E, a \in A$, f and g are mappings from E into F and $\lambda \in \mathbb{R}$.*

- If f and g are continuous at a, then so is $f + g$.
- If f is continuous at a, then so is λf.
- If α is a real-valued function defined on E and both f and α are continuous at a, then so is αf.

2 Differentiation

In this section we will be primarily concerned with extending the derivative defined for real-valued functions defined on an interval of $\mathbb{R}$. We will also consider minima and maxima of real-valued functions defined on a normed vector space.

2.1 Directional Derivatives

Let O be an open subset of a normed vector space E, f a real-valued function defined on O, $a \in O$ and u a nonzero element of E. The function $f_u : t \to f(a + tu)$ is defined on an open interval containing 0. If the derivative $\frac{df_u}{dt}(0)$ is defined, i.e., if the limit

$$\lim_{t\to 0} \frac{f(a + tu) - f(a)}{t}$$

exists, then it is called the *directional derivative* of f at a in the direction of u, i.e. $\partial_u f(a)$.

If $E = \mathbb{R}_n$ and e_i is its standard basis, then the directional derivative $e_i f(a)$ is called the i th partial derivative of f at a, or the derivative of f with respect to x_i at a.

$$\frac{\partial f}{\partial x_i} = \lim_{t\to 0} \frac{f(a_1, .., a_i + t, ..., a_n) - f(a_1, ..., a_n)}{t}$$

If for every point $x \in O$, the partial derivative $\dfrac{\partial f}{\partial x_i}(x)$ is defined, then we obtain the function i th partial derivative defined on O. If these functions are defined and continuous for all i , then we say that the function f is of class C^1.

Suppose now that O is an open subset of $\mathbb{R}^n$ and f a mapping defined on O with image in $\mathbb{R}^m$. f has m coordinate mappings $f_1, ..., f_m$. If $a \in O$ and the partial derivatives $\dfrac{\partial f_i}{\partial x_j}$ of a, for $1 \leq i \leq m$ and $1 \leq j \leq n$, are all defined, then the $m \times n$ matrix

$$J_f(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ . & . & . \\ . & . & . \\ . & . & . \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

is called the *Jacobian Matrix* of f at a.

2.2 The Differential

Let E and F be normed vector spaces, O an open subset of E containing 0, and g a mapping from O into F such that $g(0) = 0$. If there exists a mapping $\epsilon, defined on a neighbourhood of 0$ $\in E$ and with image in F , such that $\lim_{h\to 0}\epsilon(h) = 0$ and

$$g(h) = \|h\|_E\,\epsilon(h),$$

then we write $g(h) = o(h)$ and say that g is *"small o of h"*.

The condition $g(h) = o(h)$ is independent of the norms we choose for two spaces E and F.

Let O be an open subset of a normed vector space E and f a mapping from O into a normed vector space F . If $a \in O$ and there is a continuous linear mapping $\phi : E \to F$ such that

$$f(a + h) = f(a) + \phi(h) + o(h)$$

when h is close to 0, then we say that f is *differentiable* at a.

Proposition 2.1. *Let f be a mapping defined on an open subset O of a normed vector space E with image in the cartesian product $F = F_1 \times ... \times F_p$. Then f is differentiable at $a \in O$ if and only if the coordinate mappings f_i , for $i = 1, ..., p$, are differentiable at a.*

$$f'(a) = (f_1'(a), ..., f_p'(a))$$

Suppose that dim $E = n < \infty$ and that e_i is a basis of E. If $x = \sum_{i=1}^n x_i e_i$, then

$$f'(a)x = \sum_{i=1}^n x_i f'(a)e_i = \sum_{i=1}^n \partial_{e_i} f(a)e_i^*(x),$$

where (e_i^*) is the dual basis of (e_i). We thus obtain the expression. If $E = \mathbb{R}^n$ and (e_i) is its standard basis, then we usually write dx_i for e_i^*. This gives us the expression

$$f'(a)x = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a)dx_i.$$

Differentiability at a given point: If we wish to determine whether a real-valued function f defined on an open subset of $\mathbb{R}^n$ is differentiable at a given point a, then first we determine whether all its partial derivatives at a exist. If this is not the case, then f is not differentiable at a.

If all the partial derivatives exist, then we know that the only possibility for $f'(a)$ is the linear function $\phi = \sum_{i=1}^{n} frac\partial f \partial x_i(a) dx_i$. We consider the expression,

$$\frac{f(a+h) - f(a) - \phi(h)}{\|h\|} = \epsilon(h)$$

If $\lim_{h\to 0} \epsilon(h) = 0$, then f is differentiable at a, otherwise it is not.

2.3 Differentials of Compositions

Let E, F and G be normed vector spaces, O an open subset of E, U an open subset of F and $f : O \to F$, $g : U \to G$ be such that $f(O) \subset U$. Then the mapping $g \circ f$ is defined on O.

Theorem 2.1. *If f is differentiable at a and g is differentiable at f(a), then $g \circ f$ is differentiable at a and*

$$(g \circ f)'(a) = g'(f(a)) \circ f'(a).$$

This expression is referred to as Chain Rule.

Corollary 2.1. *If in the above theorem the normed vector spaces are euclidean spaces, then*

$$J_{g\circ f}(a) = J_g(f(a)) \circ J_f(a).$$

2.4 Differentiability of the Norm

If E is a normed vector space with norm $\|.\|$, then $\|.\|$ is itself a mapping from E into $\mathbb{R}$ and we can study its *differentiability*. We will write $Df(\|.\|)(x)$ for differentiability of the norm at x (if exists).

Proposition 2.2. *Norm is not differentiable at the origin.*

Proof. Suppose $Df(\|.\|)$ exists. Then for small non-zero values of h, we have

$$\|h\| = Df(\|.\|)(0)h + o(h) \Rightarrow \lim_{h\to 0}\left(1 - Df(\|.\|)\frac{h}{\|h\|}\right) = 0$$

And

$$\|h\| = \|-h\| = -Df(\|.\|)(0)h + o(h) \Rightarrow \lim_{h\to 0}\left(1 + Df(\|.\|)\frac{h}{\|h\|}\right) = 0$$

3 Mean Value Theorem

Theorem 3.1. *Let f be a real-valued function defined on a closed bounded interval $[a, b] \subset \mathbb{R}$. If f is continuous on $[a, b]$ and differentiable on $(a, b0$ then there is a point $c \in (a, b)$ such that*

$$f(b) - f(a) = \dot{f}(c)(b - a)$$

3.1 Generalization of Mean Value Theorem

Theorem 3.2. *Let O be a open subset of a normed vector space E and $a, b \in E$ with $[a, b] \in O$. If $f : O \to \mathbb{R}$ is differentiable , then there is a element $c \in (a, b)$ such that*

$$f(b) - f(a) = f'(c)(b - a)$$

Remark 3.1. *If $E = \mathbb{R}^n$ then this result can be written as*

$$f(b) - f(a) = \textstyle\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(c)(b_i - a_i)$$

Theorem 3.3. *Let $[a, b]$ be an interval of $\mathbb{R}$, F a normed vector space and $f : [a, b] \to F$ and $g : [a, b] \to \mathbb{R}$ both continuous and differentiable on (a, b). If $\left\|\dot{f}(t)\right\| \leq \dot{g}(t)$ for all $t \in (a, b)$, then* $$\|f(b) - f(a)\|_F \leq g(b) - g(a).$$

Corollary 3.1. *Let E and F are normed vector spaces, O an open subset of E and $f : O \to F$ differentiable on O. If the segment $[a, b] \subset O$,then*

$$\|f(b) - f(a)\|_F \leq sup_{x \in (a,b)} |f'(x)|_{L(E,F)} \|b - a\|_E$$

3.2 Partial Differentials

In this section we will generalize the notion of partial derivatives and its results.

Let $E_1, E_2, ..., E_n$ and F be normed vector spaces. We set
$E = E_1 \times .... \times E_n$ and define a norm on E

$$\|(x_1, ..., x_n)\|_E = \max_k \|x_k\|_{E_k}.$$

Now let O be an open subset of E and f a mapping from O into F . If we take a point $a \in O$ and let the kth coordinate vary and fix the others, then we obtain a mapping $f_{a,k}$ from E_k into F , defined on an open subset of E_k containing a_k.

If $f_{a,k}$ is differentiable at a_k, then we call the differential $f'_{a,k}(a_k) \in L(E_k, F)$ the *kth partial differential of f at a* and write it as $\partial_k f(a)$ for $f'_{a,k}(a_k)$.

4 Higher Derivatives and Differentials

Let $O \subset \mathbb{R}^n$ be open and f a real valued function defined on O. If the function $\frac{\partial f}{\partial x_i}$ is defined on O, then we can consider the existance of its partial derivatives. If $\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_i})(a)$ exists, then we write for this derivative $\frac{\partial^2 x}{\partial x_j \partial x_i}(a)$ if $i \neq j$ and $\frac{\partial^2 x}{\partial^2 x_i}(a)$ if $i = j$.

If these functions are defined and continuous for all pairs (j, i), then we say that f is of class C^2.

We say that continuous functions are of class C^0 . If a function is of class C^K for all $K \in \mathbb{N}$, then we say that f is of class C^∞ , or smooth.

4.1 Schwarz's Theorem

Theorem 4.1. *Let $O \subset \mathbb{R}^2$ be open and $f : O \to \mathbb{R}$ be such that the second partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ are defined on O. If these functions are continuous at $(a, b) \in O$, then*

$$\frac{\partial^2 f}{\partial x \partial y}(a, b) = \frac{\partial^2 f}{\partial y \partial x}(a, b).$$

4.2 Multilinear Mapping

Second and higher differentials are more difficult to define than second and higher derivatives. The natural way of defining a second differential would be to take the differential of the mapping $x \mapsto f'(x)$. Unfortunately, if E and F are normed vector spaces and f a differentiable mapping from an open subset of E into F , then the image of f' lies not in F but in $L(E, F)$. This means that the differential of f' lies in $L(E, L(E, F))$.

4.3 Higher Differentials

Let E and F be normed vector spaces and O an open subset of E. If $f : O \to F$ is differentiable on an open neighbourhood V of $a \in O$, then the mapping

$$f' : V \mapsto L(E, F), x \mapsto f'(x) \text{ is defined.}$$

As we said in the previous section, if f' is differentiable at a, then we would be tempted to define the second differential $f^{(2)}(a)$ of f at a as $f''(a) = (f')'(a)$. However, in this way $f^{(2)}(a) \in L_2(E, F)$ and it is difficult to work with these higher order spaces. Hence we proceed in a different way.

We will define linear continuous mappings Φ_k from $L_k(E, F) \to L(E^k, F)$.

we define k-differentiability and the kth differential $f^{(k)}(a)$ for higher values of k. We will sometimes write $f^{(1)}$ for f'. To distinguish the differential in $L_k(E;F)$ corresponding to $f^{(k)}(a)$, we will write $f^{[k]}$ for it, i.e., $\Phi_k(f^{[k]}(a)) = f^{(k)}(a)$.

4.4 Taylor Formulas

Some useful notation

1) Let E and F be normed vector spaces, O an open subset of E containing 0 and g a mapping from O into F such that $g(0) = 0$. If there exists a mapping ϵ, defined on a neighbourhood of $0 \in E$ and with image in F, such that $lim_{h \to 0} \epsilon(h)$ and

$$g(h) = \|h\|_E^k \, \epsilon(h),$$

then we will write $g(h) = o(\|h\|_E^k)$ or $g(h) = o(\|h\|^k)$ when the norm is understood. If k =1, then $o(\|h\|) = o(h)$

2) If E is a normed vector space and h is a vector in E, then we will write h^k for the vector $(h, ..., h) \in E^k$.

Theorem 4.2. ***(Taytor's formula, asymptotic form)***. *Let E and F be normed vector spaces, O an open subset of E and $a \in O$. If $f : O \to F$ is $(k-1)$-differentiable and $f^{(k)}(a)$ exists, then for x is sufficiently small*
$f(a+x) = f(a) + f^{(1)}(a)(x) + \frac{1}{2} f^{(2)}(a)(x^2) + ... + \frac{1}{k!} f^{(k)}(a)(x^k) + o(\|h\|^k).$

4.5 Extrema: Second Order Condition

A local extremum of a differentiable function is always a critical point. We now suppose that the function is 2-differentiable at the critical point. We will discuss local minima; analogous results for local maxima can be easily obtained by slightly modifying the argument

Proposition 4.1. *Let O be an open subset of a normed vector space E, $a \in O$ and f a real-valued function defined on O having a second differential at a. If a is a local minimum, then for $h \in E$*

$$f^{(2)}(a)(h,h) \geq 0$$

5 Hilbert Space

A vector space E over the field F together with an inner product, i.e., with a map

$$< .,. > : E \text{ x } E \to F$$

which follows the three axioms for all vectors $\in E$ Positive-definiteness, Linearity in first argument, Conjugate symmetry

Then the pair $(E, <,>)$ is called an inner product space, for $x \in E$, $\|x\|$ is defined as $\sqrt{< x, x >}$.

If E is an inner product space and complete with the norm derived from the inner product, Then E is said to be a **Hilbert Space**.

5.1 The Riesz Representation Theorem

If H is a Hilbert space, then by the Riesz representation theorem, we may associate an element of H to a continuous linear form.

Before looking at the general Theorem, let us see what happens in $\mathbb{R}^n$. If l is a linear form defined on $\mathbb{R}^n$, (e_i) its standard basis and $x = \sum_{i=1}^{n} x_i e_i$ then

$$l(x) = \sum_{i=1}^{n} x_i l(e_i) = x.w, \text{ where } w = (l(e_1), ..., l(e_n)).$$

If $\overline{w}$ is such that $l(x) = x.\overline{w}$ for all $x \in \mathbb{R}^n$, then $x.(w - \overline{w}) = 0$ for all $x \in \mathbb{R}^n$, it follows that $w - \overline{w} = 0$. Hence, the element w such that $l(x) = x.w$ for all x is **unique**.

Theorem 5.1. *Let l be a continuous linear form defined on a Hilbert space H . Then there is a unique element $a \in H$ such that*

$$l(x) =< x, a >$$

for all $x \in H$. In addition, $|l|_{H^*} = \|a\|$.

5.2 Convex Functions

Let X be a convex subset of a vector space V . We say that $f : X \to \mathbb{R}$ is **convex** if for all $x, y \in X$ and $\lambda \in (0, 1)$ we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

If this inequality is strict when $x \neq y$, then we say that f is **strictly convex**.

5.3 Convex Hull

Let E be a vector space and $x_1, ..., x_n \in E$. We say that $y \in E$ is a *convex combination* of the points $x_1, ..., x_n$ if there exists $\lambda_1, ..., \lambda_n \in [0, 1]$ with $\sum_{i=1}^{n} \lambda_i = 1$ such that $y = \sum_{i=1}^{n} \lambda_i x_i$.

If X is a nonempty subset of E, then we define *co* X , the convex hull of X , to be the set of points $y \in E$ which are convex combinations of points in X.

Definition. A subset X of a vector space E is convex if and only if **co** $X = X$.

5.4 Continuity of Convex functions

Lemma 5.1. *If P is a bounded nonempty polyhedron in a finite-dimensional normed vector space E and f is a convex function defined on E, then f has an upper bound on P.*

Theorem 5.2. *Let X be a finite dimensional normed vector space E and $f : X \to \mathbb{R}$ is convex. If $x \in intX$, then f is continuous at x.*

Corollary 5.1. *If a convex function is defined on an open subset of a finite-dimensional normed vector space, then it is continuous.*

5.5 Differentiable Convex Functions

Let O be an open subset of a normed vector space E and f a real-valued differentiable functions defined on O. If $X \subset O$ is convex and $x, y \in X$, then the following are equivalent:

- f is convex on X.
- $f(y) - f(x) \geq f'(x)(y - x)$
- $(f'(y) - f'(x))(y - x) \geq 0$

Remark 5.1. *From third condition we deduce that, if f is differentiable on an open interval I of $\mathbb{R}$, then f is convex (resp. strictly convex) if and only if the derivative of f is increasing (resp. strictly increasing) on I.*

6 The Inverse Mapping Theorem

Proposition 6.1. *Let E and F be Banach spaces, $O \subset E$ and $U \subset F$ open sets and $f : O \to U$ a differentiable homeomorphism. If $a \in O$ and $f'(a)$ is invertible then f^{-1} is differentiable at $b = f(a)$.*

In addition, if f is of class C^1 then there is a open neighbourhood O' such that $f_{|O'}$ is a C^1-diffeomorphism onto its image.

Theorem 6.1. *Let E and F be Banach spaces, $O \subset E$ and $f : O \to F$ of class C^1. If $a \in O$ and $f'(a)$ is invertible, then there is an open neighbourhood O' of such that $f_{|O'}$ is a C^1-diffeomorphism onto its image.*

Remark 6.1. *Under the conditions of the theorem,*

- *$f_{|O'}$ is a C^1-diffeomorphism onto its image. In fact if f is of class C^k, then $f_{|O'}$ is a C^k-diffeomorphism.*

- *If a mapping f is such that each point in its domain has an open neighbourhood O such that f restricted to O defines a diffeomorphism onto its image, then we say that f is a **local diffeomorphism**.*

7 Vector Fields

Let E be a normed vector space and O an open subset of E. A continuous mapping X from O into E is called a vector field. If X is of class C^k , with $k \geq 1$, then we refer to X as a vector field of class C^k . If I is an open interval of $\mathbb{R}$ and ϕ a differentiable mapping from I into O such that

$$\dot{\phi}(t) = X(\phi(t))$$

for all $t \in I$, then we say that ϕ is an integral curve of X or a solution of the first-order differential equation $\dot{x} = X(x)$. As X and ϕ are both continuous, so is $X \circ \phi$ and it follows that ϕ is of class C^1.

An integral curve ϕ defined on an open interval I is a maximal integral curve if it cannot be extended to an integral curve defined on an open interval strictly containing I .

Notation. Let $a \in R$. We will write T a for the translation mapping of R into R defined by

$$T_a(t) = t - a$$

The integral curves of a vector field have an important elementary property, namely they are invariant under translation.

Proposition 7.1. *Let $\phi : I \to O$ be an integral curve of a vector field $X : O \to E$ and $a \in \mathbb{R}$. Then $\phi \circ T_a$ a is an integral curve of X defined on $I + a$. In addition, $\phi \circ T$ a is maximal if and only if ϕ is maximal.*

There is a slight variety of vector fields which is time-dependent vector fields. Suppose that E and O are as above, J an open interval of $\mathbb{R}$ and Y a continuous mapping from $J \times O \to E$. Then we say that Y is a **time-dependent vector field**. In this case, for each $t \in J$, the mapping $Y_t : x \mapsto Y(t, x)$ is a **vector field**.

An integral curve of Y is a differentiable mapping ϕ from an open interval $I \subset J$ into O such that

$$\dot{\phi}(t) = Y(t, \phi(t))$$

for $t \in I$. As $Y(t, \phi(t))$ is continuous, ϕ is of class C^1.

7.1 Existence of Integral Curves

Let $X : O \to E$ be a locally Lipschitz vector field and $x_0 \in O$. There exists a closed ball $\bar{B}(x_0, r)$, with $r > 0$, and a constant $K > 0$ such that X is restricted to $\bar{B}(x_0, r)$ is K-Lipschitz. As X is Lipschitz on $\bar{B}(x_0, r)$,X is bounded on $\bar{B}(x_0, r)$. We take the norm,

$$M = \sup_{x \in \bar{B}(x_0,r)} \|X(x)\|.$$

Theorem 7.1. *If E is a Banach space, $t_0 \in \mathbb{R}$ and $\epsilon > 0$, with $M < r$, then there is a unique integral curve ϕ defined on the interval $I_\epsilon = (t_0 - \epsilon, t_0 + \epsilon)$ satisfying the condition $\phi(t_0) = x_0$. The image of ϕ lies in the open ball $B(x_0, r)$.*

Proposition 7.2. *Let ϕ and ψ be integral curves of the vector field X defined on the same open interval I. If $t_0 \in I$ and $\phi(t_0) = \psi(t_0)$, then $\phi(t) = \psi(t)$ for all $t \in I$.*

Theorem 7.2. *Let $t_0 \in \mathbb{R}$ and $x_0 \in O$, the domain of the vector field X. Then there is a unique maximal integral curve ϕ with $\phi(t_0) = x_0$.*

Proof. From *Theorem 7.1* there exists an interval $\tilde{I}$ on which an integral curve $\tilde{\phi}$ is defined with

Let I be the union of all such open intervals. For $t \in I$ we take any one of these integral curves $\tilde{\phi}$ defined at t and set $\phi(t) = \tilde{\phi}(t)$. From the preceding proposition $\phi(t)$ does not depend on the integral curve Q which we have chosen, so ϕ is well-defined. Clearly ϕ is maximal and $\phi(t_0) = x_0$.

If $\psi : I_1 \to O$ is another maximal integral curve and $\psi(t_0) = x_0$,then $I_1 \subset I$ and, because ψ is maximal, $I_1 = I$. Using *Proposition 7.2* we see that $\psi = \phi$. Hence Proved. □

Now consider a class of vector fields which is particularly easy to study. Let E be a Banach space and $A \in L(E)$. If we set $X(x) = Ax$ for $x \in E$, then X is a smooth vector field defined on E. Such vector fields are called **linear vector fields**.

$$\|A(x) - A(y)\| = |A| \, \|x - y\|$$

X is Lipschitz. Let us consider the mapping $\alpha : \mathbb{R} \to L(E)$, $t \mapsto \exp(tA)$.

Proposition 7.3. $\dot{\alpha}(t) = \exp(tA) \circ A = A \circ \dot{\alpha}(t) = \exp(tA)$

Now let us take $x_0 \in E$ and consider the mapping $\phi : \mathbb{R} \to E$, $t \mapsto \exp(tA)x_0$

Theorem 7.3. *ϕ is the maximal integral curve of the linear vector field $X(x) = Ax$ with $\phi(0) = x_0$.*

Remark. If $t_1 \neq 0$ and we require the maximal integral curve with $\psi(t_1) = x_0$, then it is sufficient to set $\psi = \phi \circ T_{t_1}$.

We will often write $\Phi(t, t_0, x_0)$ for the maximal integral curve ϕ satisfying the condition $\phi(t_0) = x_0$) and I_{t_0,x_0} for the domain of ϕ. If O is the domain of the vector field and we set

$D = \cup_{t_0 \in \mathbb{R}, x_0 \in O} I_{t_0,x_0} \times t_0 \times x_0$,

then Φ defines a mapping from D into O. We will refer to Φ as the **flow of the vector field** X.

Example. Let E be a Banach space and $A \in L(E)$. If we set $X(x) = Ax$ for $x \in E$, then X is a smooth vector field defined on E.

Then, $D = \mathbb{R} \times \mathbb{R} \times E$ and $\Phi(t, t_0, x_0) = \exp\left((t - t_0)A\right)x_0$. Φ is continuous on D.

7.2 Initial Conditions

If x_0 is in domain of a vector field and $t_0 \in \mathbb{R}$, then there is a unique maximal integral curve ϕ with $\phi(t_0) = x_0$. The pair (t_0, s_0) is called the initial condition that satisfy ϕ.

Lemma 7.1. *Gronwell's Inequality.* *Let f and g are continuous functions defined on closed interval $I = [a, b]$ into $\mathbb{R}_+$. If there is a constant c such that for $t \in I$*

$$f(t) \leq c + \int_a^t f(s)g(s)ds,$$

then for $t \in I$, we have

$$f(t) \leq c \exp \int_a^t g(s)ds.$$

Corollary 7.1. *Let f is continuous functions defined on closed interval $I = [a, b]$ into $\mathbb{R}_+$. If there is a constant $c \geq 0$ and $k \geq 0$ such that for $t \in I$*

$$f(t) \leq c + \int_a^t f(s)g(s)ds,$$

then for $t \in I$, we have

$$f(t) \leq ce^{k(t-a)}.$$

Now we will see how the distance between two vector fields are defined.

Theorem 7.4. *Let O be an open subset of the Banach Space E and X a $K-$Lipschitz vector field defined on O. If ϕ and ψ are integral curves defined on the interval $I = [t_0, t_1]$, $\phi(t_0) = x_0$ and $\psi(t_0) = y_0$, then*

$$\|\phi(t) - \psi(t)\| \leq \|x_0 - y_0\| \, e^{K(t-t_0)}, \text{ for all } t \in I.$$

7.3 Complete Vector Field

A vector field is said to be complete if all its maximal integral curves are defined on $\mathbb{R}$. from previous theory we can say that, linear vector fields are complete.

Clearly, if a vector field is complete, then the domain of its flow is $\mathbb{R}^2 \times O$, where O is the domain of the vector field.

Let E be a normed vector space, O an open subset of E and X a vector field defined on O. If f is a real-valued C^1 -function defined on O such that, for any integral curve ϕ of X, $f \circ \phi$ is constant, then we say that ϕ is a first integral of X.

If f is a function of constant value, then f is a first integral; though, we will study non constant first integrals.

Proposition 7.4. *Let E be a Banach space, O an open subset of E and X a vector field defined on O. A C^1 function $f : O \to \mathbb{R}$ is a first integral of X if and only if $f'(x)X(x) = 0$ for all $x \in O$.*

Example. Consider the vector field X defined on $\mathbb{R}^3$ by

$$X(x, y, z) = (4xy - 1, -2x(1 + 2z), -2x(1 + 2y)).$$

If $f : \mathbb{R}^3 \to \mathbb{R}$ is defined by $f(x, y, z) = y + y^2 - z - z^2$, then

$$\nabla f(x, y, z) = (0, 1 + 2y, -1 - 2z).$$

Dot product of X and ∇f is 0. Hence, f is a first integral of X.

In case of Hilbert Space the condition can be modified with $< \nabla f, X >= 0$, for all $x \in O$. f is first integral if and only if X is orthogonal the gradient of f.

Theorem 7.5. *Suppose that a vector field X has a first integral f. If the inverse image f^{-1} is compact for all $c \in R$, then the vector field X is complete.*

Theorem 7.6. *If a vector field X is defined on an open subset of a finite- dimensional normed vector space E and is (K-)Lipschitz, then X is complete.*

8 The Flow of a Vector Field

In this chapter we will discuss continuity and differentiability of the flow.

8.1 Continuity of the Flow

Theorem 8.1. *The domain of the flow Φ of a vector field X defined on an open subset of a Banach space E is an open subset of $R^2 \times E$ and the flow is continuous on its domain.*

8.2 Differentiability of the Flow

In this section we will show that the flow is of class C^k if the vector field is of class C^k. Smooth vector fields generate smooth flows.

To prove that the flow is of class C^1 , we will show that the partial differentials are defined and continuous.

Proposition 8.1. *Let X be a vector field defined on an open subset of a Banach space E. If X is of class C^1, then the flow $\Phi(s,t,x)$ of X has partial differentials $\partial_1\Phi$, $\partial_2\Phi$ and $\partial_3\Phi$, which are continuous on the domain D of Φ.*

Using the proposition we have,

Theorem 8.2. *If X is a vector field of class C^1 defined on an open subset of a Banach space E, then the flow Φ of X is also of class C^1.*

8.3 Higher Differentiability of the Flow

From above, if a vector field X is of class C^1 , then its flow is also of class C^1. This result is also realizable for C^k.

Suppose that E_1, E_2 and F are normed vector spaces, O an open subset of $E_1 \times E_2$ and $f : O \to F$ a C^2 -mapping. If $a \in O$, then $\partial_{ij} f(a) \in L(E_i, L(E_j, F))$. We may consider $\partial_{ij} f(a)$ to be a continuous bi-linear mapping from $E_i \times E_j$ into F.

Theorem 8.3. *Let X be a vector field defined on an open subset O of a Banach space E. If X is of class C^k with $k \geq 1$, then the flow Φ of X is of class C^k.*

8.4 The Reduced Flow

Let E be a Banach space, X a vector field defined on an open subset of E, with flow $\Phi(s, t, x)$. If we fix the second variable t, then we obtain a mapping from a subset of $R \times E$ into E; in particular, if we take $t = 0$. The mapping

$$\phi(s, x) = \Phi(s, 0, x)$$

is called the **Reduced Flow** *(r-flow)* of X.

Theorem 8.4. *Let X be a vector field defined on an open subset of a Banach space E. The r-flow ϕ of X is defined on an open subset Ω of $\mathbb{R} \times E$. If X is of class C^k with $k \geq 1$, then ϕ is also of class C^k.*

Let us now return to general vector fields. From the definition of the reduced flow ϕ we have,

$$\phi(0, x) = x$$

If we fix x and set $\phi_x(s) = \phi(s, x)$, then the mapping ϕ_x is defined on an open interval I_x of $\mathbb{R}$ containing 0.

9 The Calculus of Variations

If ξ is a normed vector space composed of mappings, for example, the space of continuous real-valued functions defined on a compact interval, then we refer to a real-valued mapping F defined on a subset S of ξ as a functional. The calculus of variations is concerned with the search for extrema of functionals.

In general, the set S is determined, at least partially, by constraints on the mappings and the functional F is defined by an integral. The elements of S are often said to be F-admissible.

9.1 The Space $C^1(I, E)$

Let E be a normed vector space and $I = [a, b]$ is a compact interval of $\mathbb{R}$. The space of continuous mappings $\gamma : I \to E$, which we note $C^o(I, E)$ or $C(I, E)$, is a vector space. The norm associated with it is

$$\|\gamma\|_{C^0} = \sup_{t \in I} \|\gamma(t)\|_E.$$

9.2 Lagrangian Mappings

Let O is an open subset of $\mathbb{R} \times E \times E$. A continuous real-valued function L defined on O is called a Lagrangian function (or Lagrangian). A curve $\gamma \in C^1(I, E)$ is said to be L-admissible if $(t, \gamma(t), \dot{\gamma}(t))$ belongs to O for all $t \in I$.

Proposition 9.1. *The set of L-admissible curves (Ω or Ω_L) is open in $C^1(I, E)$.*

If $\gamma \in \Omega$, then the function, $t \mapsto L(t, \gamma(t), \dot{\gamma}(t))$, is defined and continuous on $I = [a, b]$. We define the mapping $\mathcal{L} : \Omega_L \to \mathbb{R}$

$$\mathcal{L}(\gamma) = \textstyle\int_a^b L(t, \gamma(t), \dot{\gamma}(t))$$

$\mathcal{L}$ is called the *Lagrangian Functional* associated with L.

Proposition 9.2. *If L is of class C^1 then $\mathcal{L}$ is of class C^1.*

9.3 Fixed Endpoint Problems

Suppose now that we aim to minimize Land that at the same time we require that the endpoints of the curves $\gamma \in \Omega_L$ have certain values: $\gamma(a) = \alpha$ and $\gamma(b) = \beta$.

In this case we consider L restricted to the space

$$\Omega_L(\alpha, \beta) = \{\gamma \in C^1(I, E) | \gamma(a) = \alpha, \gamma(b) = \beta\}$$

i.e., we search for **minima** of the functional Lon the set $S = \Omega_L(\alpha, \beta)$.

Proposition 9.3. *Let I=[a,b] be a closed bounded interval of $\mathbb{R}$ and $A : I \to E^*$ is a continuous mapping such that*

$$\int_a^b A(t)v(t)dt = 0$$

for every continuous function $v : I \to E$ such that $\int_a^b v(t)dt = 0$. Then A is constant.

9.4 Euler-Lagrange Equations

Proposition 9.4. *Let I be an interval of $\mathbb{R}$ and $A : I \to E^*$ and $u : I \to E$ both differentiable at $t \in I$. Then $Au : I \to \mathbb{R}$ is differentiable at t and*

$$\tfrac{d}{dt}(A(t)u(t)) = \dot{A}(t)u(t) + A(t)\dot{u}(t).$$

The Euler-Lagrange Equation is

Theorem 9.1. *The curve $\gamma \in C^1(I, E)$ is an extremal of the fixed endpoint problem if and only if the mapping $\partial_3 L(t, \gamma(t), \dot{\gamma}(t))$ is differentiable and*

$$\frac{d}{dt}\partial_3 L(t, \gamma(t), \dot{\gamma}(t)) = \partial_2 L(t, \gamma(t), \dot{\gamma}(t)).$$

Example. Let L be the Lagrangian defined on $\mathbb{R}^3$ by

$$L(t, x, y) = 2tx - x^2 + y^2.$$

and $I = [0, 1]$.

We intend to find a curve $\gamma \in C^1(I, \mathbb{R})$, with $\gamma(0) = 1, \gamma(1) = 2$, which minimizes the Lagrangian functional

$$\mathcal{L}(\gamma) = \int_0^1 L(t, \gamma(t), \dot{\gamma}(t)) = \int_0^1 (2t\gamma(t) - \gamma(t)^2 + \dot{\gamma}(t)^2)dt.$$

We have, $\dfrac{dL}{dx} = 2(t - x)$ and $\dfrac{dL}{dy} = 2y$

The Euler-Lagrange equation is,

$$2\frac{d}{dt}\dot{\gamma}(t) = 2(t - \gamma(t) \text{ or}$$

$$\ddot{\gamma}(t) + \gamma(t) = t$$

Therefore, γ is a solution of the differential equation $\ddot{y} + y = t$,
The general solution is

$$y(t) = \alpha \cos t + \beta \sin t + t,\ \alpha, \beta \in \mathbb{R}$$

Taking endpoint values we get,

$$\gamma(t) = \frac{sint}{sin1} + t.$$

which is a unique extrema.

10 Applications

10.1 Brachistochrone curve

10.1.1 Lagrange Equation

Given two points A and B, find the path along which an object would slide (disregarding any friction) in the shortest possible time from A(0,0) to B(a,b), if it starts at A in rest and is only accelerated by gravity.

This is an optimization problem, we want to minimize travel time, but the minimization takes place over all possible paths from A to B. Thus we cannot expect this problem to fall directly into the finite dimensional setting.
We can formulate the brachistochrone problem as the minimization of the functional

$$F(y) = \int_0^a \frac{\sqrt{1+y'(x)^2}}{\sqrt{2gy(x)}}dx$$

where $y(x) \geq 0$ and $y(0) = 0$ and $y(a) = b$.
Euler-Lagrange Equation

$$F(y) = \int_0^a f(x, y(x), y'(x))dx$$

with $f : (0, a) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$

$$f(x, y, y') = \frac{\sqrt{1+y'^2}}{\sqrt{2gy}}$$

$$\frac{\partial f}{\partial y} = -\frac{1}{2}\sqrt{\frac{1+y^2}{2g}}\frac{1}{y^{3/2}}$$

$$\frac{\partial f}{\partial y'} = -\frac{1}{\sqrt{2gy}}\frac{y'}{\sqrt{1+y'^2}}$$

Now, Using Euler-Lagrange equation,

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial x}\frac{\partial f}{\partial y'}$$

which has a form of $y + yy' = C$ with some $C > 0$. Parametric solution is

$$t \mapsto (x(t), y(t)) = C(t - \frac{1}{2}\sin 2t, \frac{1}{2} - \frac{1}{2}\cos 2t)$$

It is such that $C > 0$ and it passes through (a, b). It is a **Cycloid**.

10.1.2 Optimization

Starting point $A(0,1)$ and Ending point $B(3,0)$. The code is written in **python** using Google's **tensorflow**. [8]

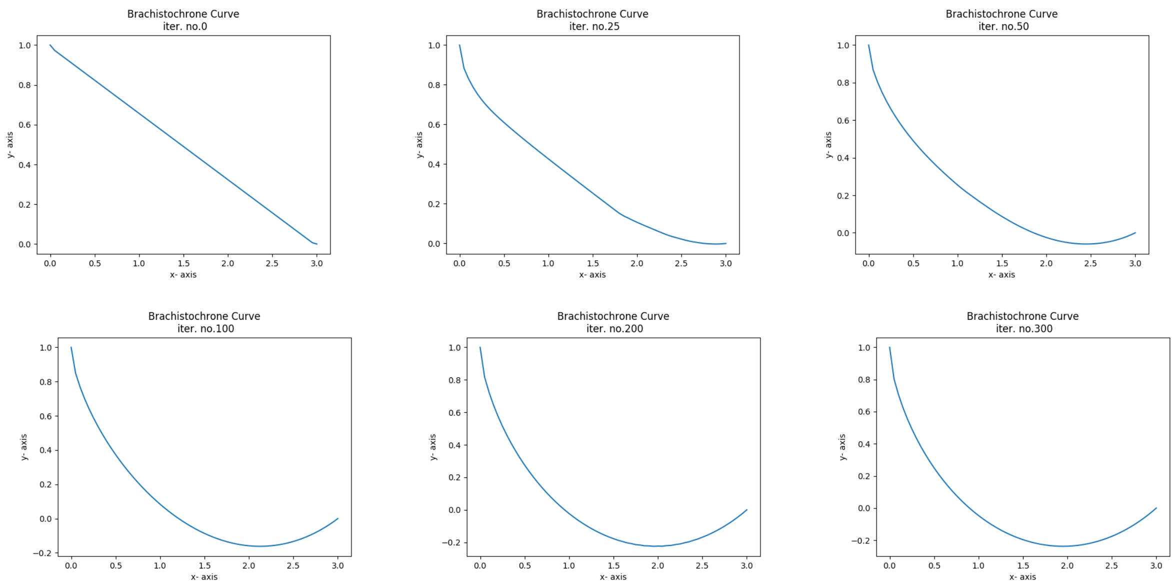

Figure 1: Optimization Steps and Curve Formation

Here, the Brachistochrone Curve problem is treated as a optimization problem with minimizing total time taken. Laws of motion and gravity $g = 9.81 m/s^2$ is used to minimize time.
Equations used,

$$s = ut + \frac{1}{2}at^2$$

$$v = u + at$$

$$v = \frac{s}{t}$$

$$a = \frac{v}{t}$$

Adam stands for Adaptive Moment Estimation. Adam is used here as optimizer. Adam is a method that computes adaptive learning rates for each parameter.[7]

* * * * *

11 References

[1] Avez, A.: Differential Calculus. J. Wiley and Sons Ltd, New York (1986)

[2] Dacorogna, B.: Introduction to the Calculus of Variations. Imperial College Press, London (2004)

[3] Markus Grasmair: Basics of Calculus of Variations

[4] Rodney Coleman: Calculus on Normed Vector Spaces (2012)

[5] Martin Burger: Infinite Dimensional Optimization and Optimal Design

[6] https://en.wikipedia.org/wiki/Brachistochrone_curve

[7] Diederik P. Kingma, Jimmy Ba : Adam: A Method for Stochastic Optimization

[8] TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems