27 Differential Calculus

27.1 Introduction

This course is mostly about function spaces and linear operators acting on them. Linear operators on function spaces is the natural and quite satisfactory framework for the majority of the linear equations that arise in applied mathematics. Many interesting problems, however, lead to *nonlinear* equations, formulated in terms of a nonlinear map $F : X \to Y$, between two topological spaces X and Y. There is no general theory of nonlinear maps that is as powerful as, e.g., spectral theory for linear operators. But there is a procedure, called *linearization* , which allows one to address at some of the interesting questions about nonlinear problems in terms of a linear operator or a set of linear operators. This is very similar to looking for extrema of nonlinear functions by studying its differential, which can be considered as a linear map.

For concreteness, we first mention a couple of examples. Unless otherwise indicated F is a map from X to Y, where X and Y are toplogical spaces.

1. For fixed $y_0 \in Y$, we are interested in the solutions $x \in X$ of a nonlinear equation of the form
$$F(x) = y_0 \tag{27.1}$$
A common problem is to determine the *stability* of a given solution x_0 under small perturbations of y_0. Let us assume that X and Y are topological linear spaces, and ϵ is to be thought of as a small real parameter. If y is close to y_0 we can write y as
$$y = y_0 + \epsilon y_1,$$
and ask whether (27.1) has a solution $x(\epsilon)$ close to x_0. E.g., we look for
$$x(\epsilon) = x_0 + \epsilon x_1 + \mathbf{o}(\epsilon^2)$$
such that
$$F(x(\epsilon)) = y_0 + \epsilon y_1 \quad .$$
The form (1) for the solution implies that $x(\epsilon)$ is differentiable, where the derivative of a funtion $x : \mathbb{R} \to X$, where X is a topological linear space is defined by the standard formula:
$$\frac{d}{d\epsilon}x(\epsilon) = \lim_{h\to 0} \frac{x(\epsilon + h) - x(\epsilon)}{h},$$
where the limit is in the topology of X. If F is also differentibale in the sense that there is a map $L : X \to Y$ such that
$$F(x_0 + \epsilon x_1) = F(x_0) + \epsilon L x_1 + \mathbf{o}(\epsilon^2),$$

we can expect to have solutions of the form (1) with

$$x_1 = L^{-1} y_1$$

at least of L is invertible. In the finite-dimensional situation the fact that this procedure works is the content of the *inverse function theorem* (see, e.g., Marsden, *Elementary classical analysis*). We will see how one can define L in more general situations.

2. Many evolution equations are of the form

$$u_t = F(u)$$

where $X = Y$ is the state space of the dynamical system, which we assume to be a topological linear space, and F is a map $X \to X$ (in some situations F may only be defined of a subset of X). u_t denotes the derivative with respect to t of $u : \mathbb{R} \to X$ (sometimes solutions are only for $t \geq 0$, or up to some time T), and usually one has specified an initial condition: $u(0) = u_0 \in X$. Partial differential equations can be put into this form. One important type of problem concerns the stability of stationary solutions, i.e., $u^* \in X$ such that

$$F(u^*) = 0$$

Again we may try to linearize the nonlinear equation as follows:

$$F(u^* + \epsilon v) = F(u^*) + \epsilon L v + \mathbf{o}(\epsilon^2)$$

where L is a linear transformation on X, and we look for solutions of the form $u(t) = u^* + \epsilon v(t)$. Then, v satisfies

$$v_t = Lv$$

which is a linear equation with (formal) solution

$$v(t) = e^{tL} v(0) \quad .$$

$e^{tL} v(0)$ will be well-defined, e.g., by the spectral theorem, if L is bounded above. The stationary point u^* is called *linearly (asymptotically) stable* if

$$\lim_{t \to +\infty} u^* + e^{tL} v(0) = u^*$$

which happens if there exists $\gamma > 0$ such that for all $\lambda \in \sigma(L)$ (the spectrum of L), Re $\lambda \leq -\gamma$. If Re $\lambda = 0$ for some $\lambda \in \sigma(L)$, and Re $\lambda \leq 0$ for all $\lambda \in \sigma(L)$, the stability cannot be decided from the linearization. Purely imaginary eigenvalues correspond to periodic solutions of the linear problem, which do not decay nor diverge. If Re $\lambda > 0$ for some $\lambda \in \sigma(L)$, the soltuions of the linear problem blow up exponentially. u^* is then *linearly unstable* .

3. Another class of problems are formulated in terms of the minimization of a functional $F : X \to \mathbb{R}$. We are then looking for $x^* \in X$ such that

$$F(x^*) = \inf_{x \in X} F(x)$$

A maybe naive but sometimes useful appraoch may be to look for critical points, i.e., x^* such that

$$\frac{d}{dt} F(x^* + ty)\Big|_{t=0} = 0, \qquad \text{for all } y \in X$$

or for all y in some suitable subset.

In the above examples we tried to demonstrate the need for a notion of derivative for maps between topological linear spaces. Once again our problem is to generalize a notion which is well-understood in the finite-dimensional case (the differential of a function of finitely many variables) to the infinite-dimensional situation (differential calculus for nonlinear transformations on function spaces). We will introduce the Fréchet and Gâteaux derivatives. The Fréchet derivative generalizes the notion of differential of a vector-valued function of several variables. The Gâteaux derivative generalizes the notion of gradient or directional derivative of a scalar function on a linear space or a manifold.

27.2 The Fréchet derivative

Let X and Y be Banach spaces, and let $U \subset X$ be an open set. Consider a map $F : U \to Y$, which, in general is nonlinear.

Definition 27.1 (Fréchet derivative) *If there exists a dense linear subspace $\mathcal{D}(L) \subset X$, and a linear operator $L : \mathcal{D}(L) \to Y$, such that*

$$\lim_{x \to 0} \frac{\|F(x_0 + x) - F(x_0) - Lx\|}{\|x\|} = 0$$

L is called the Fréchet derivative *of F at x_0. If $\mathcal{D}(L) = X$ and L is bounded, F is called* Fréchet differentiable *at x_0.*

Note that if X and Y are finite-dimensional, then $F : X \to Y$ is differentiable in the standard sense iff F is Fréchet differentiable and the linear map L is determined by the matrix of the partial derivatives of the components of F evaluated in the point x_0.

Here is an infinite-dimensional example:

Example:
Let $X = H^1(\Omega)$ and $Y = L^1(\Omega)$, with Ω a regular domain $\Omega \subset \mathbb{R}^d$, and define $F : X \to Y$ by

$$F(u) = |\nabla u|^2 + |u|^2$$

In order to identify the Fréchet derivative of F at $u \in X$, consider functions $u + \epsilon v$. Then

$$\begin{aligned} F(u+\epsilon v) &= (\nabla u + \epsilon \nabla v)^2 + (u+\epsilon v)^2 \\ &= |\nabla u|^2 + 2\epsilon \nabla u \cdot \nabla v + \mathbf{o}(\epsilon^2) + |u|^2 + 2\epsilon uv + \mathbf{o}(\epsilon^2) \end{aligned}$$

Hence, if it exists, the Fréchet derivative at u is given by

$$Lv = (2\nabla u \cdot \nabla + 2u)v$$

By using the properties of the Sobolev norm, one can show that $L : H^1 \to L^1$, is a bounded linear map, and that L is the Fréchet derivative of F.

In the following exercise you are asked to generalize the property that the derivative of linear functions are constant.

Exercise 27.1 *Prove that if $F : X \to Y$ is a bounded linear map itself, then F is Fréchet differentiable and its Fréchet derivative is independent of the point $x_0 \in X$, and given by F itself.*

Exercise 27.2 *Suppose F and G are two Fréchet differentiable maps $X \to Y$. Show that $F + G$ is Fréchet differentiable with the Fréchet derivative given by the sum of the Fréchet derivatives of F and G.*

27.3 Linearization of nonlinear PDE by means of the Fréchet derivative

We will discuss linearization of nonlinear PDE by considering a class of important examples: *reaction-diffusion equations*. . In general this is a set of nonlinear PDE describing the dynamics of space-dependent chemical concentrations and temperature in the presence of chemical reactions and diffusion of reactants and heat. Suppose there are n such variables. Then, the system is described by a function $u : \Omega \times \mathbb{R} \to \mathbb{R}^n$, where $\Omega \subset \mathbb{R}^d$ is the spatial domain we wish to consider. The diffusion is described by the Laplacian and a $n \times n$ diffsuion matrix D. The reactions are described by a sources-and-sinks term $f : \mathbb{R}^n \to \mathbb{R}^n$.

$$\begin{aligned} &u_t = D\Delta u + \lambda f(u) \\ &u(x,t) = 0, \text{ for } x \in \partial\Omega \\ &u(\cdot, 0) = u_0 \end{aligned}$$

where, for concreteness, we have chosen Dirichlet boundary conditions, and u_0 is the initial condition. The nonlinearity is due to the function f which is often a polynomial in the chemical concentrations, with coefficients which may depend on temperature as exponential functions. In order to elucidate the role of the nonlinearity we introduced a real parameter λ. For simplicity of our discussion, assume $D = \mathbb{1}$. As we discussed before, an important problem is to determine the stability of a stationary solution u^* satisfying

$$\Delta u^* + \lambda f(u^*) = 0$$
$$u^*(x) = 0, \text{ for } x \in \partial\Omega$$

To study the linear stability of u^*, we have to compute the Fréchet derivative of the map $u \mapsto \Delta u + \lambda f(u)$. Due to the exercises of the previous paragraph we only need to compute the derivative of $u \mapsto f(u)$. The following lemma gives the result as well as sufficient conditions on f.

Lemma 27.2 *Suppose $\Omega \subset \mathbb{R}^d$ is a bounded regular open domain, and $f : \mathbb{R}^n \to \mathbb{R}^n$ is twice continuously differentiable with differential $Df : \mathbb{R}^n \to \mathbb{R}^{n\times n}$. Then, $u \mapsto f(u)$ as a map from $H^k(\Omega) \to L^2(\Omega)$, for any $k > d/2$, is Fréchet differentiable and its derivative at $u \in H^k(\Omega)$, is given by $v \mapsto Df(u)v$.*

Proof: Let $|u|$ denote the Euclidean norm of $u \in \mathbb{R}^n$. As f is twice continuously differentiable, by Taylor's theorem we have the following inequality: for all $R > 0$, there is constant $C(R)$ such that

$$|f(u+v) - f(u) - Df(u)v| \leq C(R)|v|^2, \quad \text{for all} u, v \in \mathbb{R}^n, |u| \leq R, |u+v| \leq R$$

By the Sobolev embedding theorem (Theorem 26.3) and the assumptions, there is a constant M such that

$$\|u\|_\infty \leq M\|u\|_{H^n}$$

By combining these inequalities and intergating over Ω, we obtain

$$\begin{aligned} \|f(u+v) - f(u) - Df(u)v\|_{L^2} &\leq (C(\|u\|_\infty + \|v\|_\infty))^2\|v^2\|_{L^2} \\ &\leq C'\|v\|_\infty\|v\|_{L^2} \\ &\leq C''\|v\|^2_{H^n} \end{aligned}$$

where C', and C'' are constants depending on the bounded domain Ω and u, but not on $\|v\|_\infty$, as long as $\|v\|_{H^n} \leq 1$. The result follows by dividing by $\|v\|_{H^n}$ and taking the limit $\|v\|_{H^n} \to 0$. ∎

In conclusion, we have found that the Fréchet derivative of $u \mapsto \Delta u + \lambda f(u)$, in the point u, is given by

$$Lv := \Delta v + \lambda Df(u)v \quad .$$

If the functions are defined on a bounded domain Ω with suitable boundary conditions, and $Df(u)$ satisfies some reasonable conditions (cfr. Sturm-Liouville operators), L will have a normal compact inverse, and therefore, L will have a spectrum consisting entirely of eigenvalues, which will depend on the parameter λ. E.g., if $Df(u)$ is a symmetric matrix with constant entries, the eigenvalues will depend linearly on λ, and it is then easy to decide the linear stability properties of u in terms of the eigenvalues of the Laplacian on Ω.

27.4 The Gâteaux derivative

The notion of *Gâteaux derivative* or *directional derivative* generalizes the directional derivative of functions on $\mathbb{R}^n$. If the function is not differentiable, the directional derivative may exists in certain directions, but not in others, and/or it may not depend linearly on the vector specifying the direction.

Definition 27.3 *Let X be a linear space and Y a Banach space. A map $F : \mathcal{D}(F) \to Y$ is called* Gâteaux differentiable *at $x_0 \in X$, in the direction $x \in X$, if the following limit exists:*

$$\lim_{t\to 0} \frac{F(x_0 + tx - F(x_0)}{t} =: F'_x(x_0) \quad .$$

F is called Gâteaux differentiable *or* weakly differentiable *at x_0 if this limit exists for all $x \in X$.*

Note that if F is Fréchet differentiable then F is Gâteaux differentiable, and the map

$$x \mapsto \lim_{t\to 0} \frac{F(x_0 + tx - F(x_0)}{t}$$

is linear and bounded and equal to the Fréchet derivative. However, th converse is not generally true. If F is Gâteaux differentiable, and the map 27.4 is linear and bounded, one can still not conclude that F is Fréchjet differentiable (in x_0). Observe that the limit required for the Fréchet derivative is as $\|x\| \to 0$, which implies that the limits $t \to 0$ in the Gâteaux derivative are uniform in the direction specified by x.

Note that in the case $X = \mathbb{R}^d, Y = \mathbb{R}$, and F differentiable, the directional derivative is given by the gradient of F:

$$F'_{x_0}(x) = \nabla F(x_0) \cdot x$$

As an example in the infinite-dimensional case we will compute the Gâteaux derivative of the L^p norm.

Example:
Let $X = L^p(\Omega, \mu)$, $1 \leq p < +\infty$, and consider $F : X \to \mathbb{R} : f \mapsto \|f\|_p^p$. Then

$$F'_f(g) = \frac{d}{dt} \int |f + tg|^p$$

First, we show that we can interchange the derivative and the integral by the dominated convergence theorem (Theorem 24.6).

$$\frac{d}{dt}|f(x) + tg(x)|^p = \lim_{t \to 0} \frac{1}{t} \left(|f(x) + tg(x)|^p - |f(x)|^p\right)$$

As the function $x \mapsto |x|^p$, for $p \geq 1$, is convex, we have, for $0 < t \leq 1$:

$$|f(x) + tg(x)|^p \leq t|f(x) + g(x)|^p + (1-t)|f(x)|^p$$

and hence

$$\frac{1}{t} \left(|f(x) + tg(x)|^p - |f(x)|^p\right) \leq |f(x) + g(x)|^p - |f(x)|^p$$

and similarly

$$|f(x)|^p - |f(x) - g(x)|^p \leq \frac{1}{t} \left(|f(x) + tg(x)|^p - |f(x)|^p\right)$$

These two inequalities imply

$$\frac{1}{t}\left||f(x) + tg(x)|^p - |f(x)|^p\right| \leq |f(x)|^p + |f(x) + g(x)|^p + |f(x) - g(x)|^p$$

The three terms in the RHS are in L^1. Therefore, we can apply the dominated convergence theorem if the integrand converges pointwise a.e. as $t \to 0$, which is the case:

$$\frac{d}{dt}|f(x) + tg(x)|^p\Big|_{t=0} = p|f(x)|^{p-2}\frac{1}{2}(\overline{f(x)}g(x) + f(x)\overline{g(x)})$$

Note that we were careful and wrote $|z|^p$ as $|z\overline{z}|^{p/2}$, before differentiating. Hence

$$F'_f(g) = \frac{p}{2} \int |f|^{p-2}(\overline{f}g + f\overline{g}) \tag{27.2}$$

This is a bounded real-linear functional on L^p due to Hölder's inequality: $|f|^{p-1} \in L^q$, with q the index dual to p.

We have proved that the L^p norm is Gâteaux differentiable on L^p and the Gâteaux derivative is given by the bounded real-linear functional given in (27.2).

Exercise 27.3 *Is the L^p norm Fréchet differentiable on L^p?*

We already noted that the notion of Gâteaux derivative is weaker than the Fréchet derivative. Its mere existence does not imply that is is continuous in the direction x or even linear in x, two necessary properties for Fréchet differentiability. One can prove that if the Gâteaux derivative $F'_x(x_1)$ exists as a bounded linear operator for all x_1 in an open neighborhood of x_0 and is, as a function $x_1 \to (\mathcal{B}(X,Y), \|\cdot\|)$, it is continuous at x_0, the F is Fréchet differentiable at x_0. This result generalize the theorem relating continuity of all first order partial derivatives to differentiability. As the proof of this statement requires the general Hahn-Banach theorem for Banach spaces, we will not prove this here. We also warn the student that there is little consistency in the literature in the usage of the words Gâteaux derivative, differential, differentiable etc..

Finally, we mention that is is the stronger notion of Fréchet derivative that is needed to generalize the basic theorems of calculus (such as the fundamental theorem of calculus and Taylord' theorem) to the infinite dimensional situation. See R.D. Milne, *Applied Functional Analysis*, for a more detailed account and applications. For other applications the Gâteaux derivative is sufficient. E.g., if the $f : X \to \mathbb{R}$, and F has a minimun or a maximum at x_0, then all Gâteaux derivatives of F at x_0 that exist, must vanish.

Exercise 27.4 *Let $g : \mathbb{R} \times [0,1] \to \mathbb{R}$ be a continuous function such that for each $t \in [0,1]$, the function $g(\cdot, t) \in C^1(\mathbb{R})$. Define a functional $\varphi : C([0,1]) \to \mathbb{R}$ by*

$$\varphi(f) = \int_0^1 g(f(t), t)dt$$

Determine the differentiability properties of φ.

27.5 Functional derivatives

The *functional derivative* of a functional on a function space, also called the *variational derivative*, can be regarded as a special case of the Gâteaux derivative. We limit ourselves to an example that illustrates the most important aspects of this notion.

***

27.6 The Euler-Lagrange equations

Langrangian mechanics can be formulated as a variational principle of the following type. One looks for a differentiable function $x : [0,1] \to \mathbb{R}^n$, satisfying boundary conditions, say $x(0) = x(1) = 0$, for simplicity. x must be a critical point of a Lagrangian functional $\mathcal{L}$ of the following form:

$$\mathcal{L}(x) = \int_0^1 L(x, \dot{x}, t)dt$$

where L is a function $\mathbb{R}^n \times \mathbb{R}^n \times [0,1] \to \mathbb{R}$, and $\dot{x}$ denotes the derivative of x with respect to t. If L has continuous partial derivatives one can show that the vanishing of the Gâteaux derivatives in all directions (or sufficiently many directions, e.g., taken from a test function space), implies that x satisfies -the so-called *Euler-Lagrange equations*:

$$\frac{d}{dt}L_{\dot{x}}(x,\dot{x},t) - L_x(x,\dot{x},t) = 0$$

27.7 Appendix: the inverse and implicit function theorems

Recall that we say that $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is C^1 iff its *differential*

$$Df = (\frac{\partial f_i}{\partial x_j})$$

exists and is continuous. In the case $n = m$, Df is a square matrix and its determinant is called the *Jacobian* of f.

Theorem 27.4 (Inverse Function Theorem) *Suppose $f : D \subset \mathbb{R}^n \to \mathbb{R}^n$, where D is an open set, is C^1 on D. Let $x_0 \in D$ be such that* $\det Df(x_0) \neq 0$. *Then, there exists an open neighborhood $V \subset D$ of x_0, and an open neighborhood $W \subset \mathbb{R}^n$ of $f(x_0)$, such that, f has an inverse on U, say f^{-1}, which is C^1 and one-to-one and onto $W \to V$.*

Under the same conditions one can also show that the differential of the inverse function is the inverse matrix of the differential of f, i.e., for every $y \in W$, and $x = f^{-1}(y)$, we have

$$Df^{-1}(y) = (Df(x))^{-1}$$

One can also prove that if f is C^p, $p \geq 1$, then so is f^{-1}.

Theorem 27.5 (Implicit Function Theorem) *Let $D \subset \mathbb{R}^n \times \mathbb{R}^m$ be an open set, and suppose $F : D \to \mathbb{R}^m$ is C^p, $p \geq 1$. Let $(x_0, y_0) \in D$ be a point such that*

$$\det(\frac{\partial}{\partial y_j}F_i(x,y))\Big|_{(x_0,y_0)} \neq 0$$

The, there exists an open neighborhood $V \subset \mathbb{R}^n$ of x_0, and an open neighborhood $W \subset \mathbb{R}^m$ of y_0, such that there is a unique function $V \to W$ satisfying

$$F(x, f(x)) = 0, \ \textit{for all } x \in V,$$

and f is C^p.

The proof of the inverse function theorem uses the contraction mapping principle (Theorem 4.2). The implicit function theorem can be proved by using the inverse function theorem. For the proofs and applications see Marsden & Hoffman, *Elemantary Classical Analysis.*