Unit 17: Taylor approximation

Lecture

17.1. Given a function $f : \mathbb{R}^m \to \mathbb{R}^n$, its derivative $df(x)$ is the Jacobian matrix. For every $x \in \mathbb{R}^m$, we can use the matrix $df(x)$ and a vector $v \in \mathbb{R}^m$ to get $D_v f(x) = df(x)v \in \mathbb{R}^m$. For fixed v, this defines a map $x \in \mathbb{R}^m \to df(x)v \in \mathbb{R}^n$, like the original f. Because D_v is a map on $\mathcal{X} = \{$ all functions from $\mathbb{R}^m \to \mathbb{R}^n$ $\}$, one calls it an **operator**. The **Taylor formula** $f(x+t) = e^{Dt} f(x)$ holds in arbitrary dimensions:

Theorem: $f(x+tv) = e^{D_v t} f = f(x) + \frac{D_v t f(x)}{1!} + \frac{D_v^2 t^2 f(x)}{2!} + \ldots$

17.2. Proof. It is the single variable Taylor on the line $x+tv$. The directional derivative $D_v f$ is there the usual derivative as $\lim_{t\to 0}[f(x+tv) - f(x)]/t = D_v f(x)$. Technically, we need the sum to converge as well: like functions built from polynomials, sin, cos, exp.

17.3. The Taylor formula can be written down using successive derivatives $df, d^2 f, d^3 f$ also, which are then called **tensors**. In the scalar case $n = 1$, the first derivative $df(x)$ leads to the gradient $\nabla f(x)$, the second derivative $d^2 f(x)$ to the **Hessian matrix** $H(x)$ which is a bilinear form acting on pairs of vectors. The third derivative $d^3 f(x)$ then acts on triples of vectors etc. One can still write as in one dimension

Theorem: $f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x-x_0)^2}{2!} + \cdots$

if we write $f^{(k)} = d^k f$. For a polynomial, this just means that we first write down the constant, then all linear terms then all quadratic terms, then all cubic terms etc.

17.4. Assume $f : \mathbb{R}^m \to \mathbb{R}$ and stop the Taylor series after the first step. We get

$$L(x_0 + v) = f(x_0) + \nabla f(x_0) \cdot v \ .$$

It is custom to write this with $x = x_0 + v, v = x - x_0$ as

$$L(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0)$$

This function is called the **linearization** of f. The kernel of $L - f(x_0)$ is a linear manifold approximating the surface $\{x \mid f(x) - f(x_0) = 0\}$. If $f : \mathbb{R}^m \to \mathbb{R}^n$, then the just said can be applied to every component f_i of f, with $1 \leq i \leq n$. One can not stress enough the importance of this linearization. [1]

[1] Again: the linearization idea is utmost important because it brings in linear algebra.

17.5. If we stop the Taylor series after two steps, we get the function $Q(x+v) = f(x) + df(x) \cdot v + v \cdot d^2 f(x) \cdot v/2$. The matrix $H(x) = d^2 f(x)$ is called the **Hessian matrix** at the point x. It is also here custom to eliminate v by writing $x = x_0 + v$.

$$Q(x) = f(x_0) + \nabla f(x_0) \cdot (x - x_0) + (x - x_0) \cdot H(x_0)(x - x_0)/2$$

is called the **quadratic approximation** of f. The kernel of $Q - f(x_0)$ is the **quadratic manifold** $Q(x) - f(x_0) = x \cdot Bx + Ax = 0$, where $A = df$ and $B = d^2 f/2$. It approximates the surface $\{x \mid f(x) - f(x_0) = 0\}$ even better than the linear one. If $|x - x_0|$ is of the order ϵ, then $|f(x) - L(x)|$ is of the order ϵ^2 and $|f(x) - Q(x)|$ is of the order ϵ^3. This follows from the exact **Taylor with remainder formula**. [2]

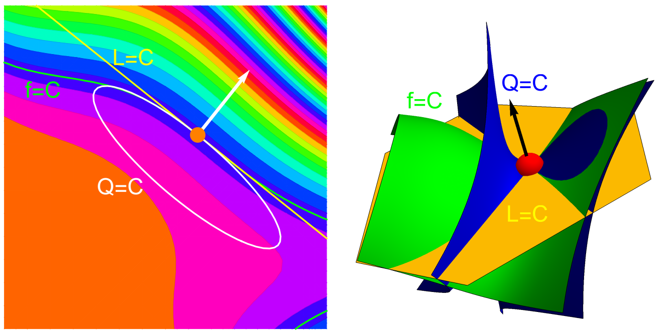


FIGURE 1. The manifolds $f(x, y) = C, L(x, y) = C$ and $Q(x, y) = C$ for $C = f(x_0, y_0)$ pass through the point (x_0, y_0). To the right, we see the situation for $f(x, y, z) = C$. We see the best linear approximation and quadratic approximation. The gradient is perpendicular.

17.6. To get the **tangent plane** to a surface $f(x) = C$ one can just look at the linear manifold $L(x) = C$. However, there is a better method:

The tangent plane to a surface $f(x, y, z) = C$ at (x_0, y_0, z_0) is $ax + by + cz = d$, where $[a, b, c]^T = \nabla f(x_0, y_0, z_0)$ and $d = ax_0 + by_0 + cz_0$.

17.7. This follows from the **fundamental theorem of gradients**:

Theorem: The gradient $\nabla f(x_0)$ of $f : \mathbb{R}^m \to \mathbb{R}$ is perpendicular to the surface $S = \{f(x) = f(x_0) = C\}$ at x_0.

Proof. Let $r(t)$ be a curve on S with $r(0) = x_0$. The chain rule assures $d/dt f(r(t)) = \nabla f(r(t)) \cdot r'(t)$. But because $f(r(t)) = c$ is constant, this is zero assuring $r'(t)$ being perpendicular to the gradient. As this works for any curve, we are done.

EXAMPLES

17.8. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given as $f(x, y) = x^3 y^2 + x + y^3$. What is the quadratic approximation at $(x_0, y_0) = (1, 1)$? We have $df(1, 1) = [4, 5]$ and

$$\nabla f(1,1) = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, H(1,1) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 6 & 8 \end{bmatrix}.$$

[2] If $f \in C^{n+1}$, $f(x+t) = \sum_{k=0}^{n} f^{(k)}(x) t^k/k! + \int_0^t (t-s)^n f^{(n+1)}(x+s) ds/n!$ (prove this by induction!)

The linearization is $L(x,y) = 4(x-1)+5(y-1)+3$. The quadratic approximation is $Q(x,y) = 3+4(x-1)+5(y-1)+6(x-1)^2/2+12(x-1)(y-1)/2+8(y-1)^2/2$. This is the situation displayed to the left in Figure (1). For $v = [7,2]^T$, the directional derivative $D_v f(1,1) = \nabla f(1,1) \cdot v = [4,5]^T \cdot [7,2] = 38$. The Taylor expansion given at the beginning is a finite series because f was a polynomial: $f([1,1]+t[7,2]) = f(1+7t,1+2t) = 3+38t+247t^2+1023t^3+1960t^4+1372t^5$.

17.9. For $f(x,y,z) = -x^4+x^2+y^2+z^2$, the gradient and Hessian are

$$\nabla f(1,1,1) = \begin{bmatrix} f_x \\ f_y \\ f_z \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, H(1,1,1) = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} = \begin{bmatrix} -10 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

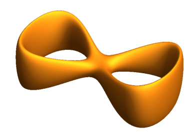
The linearization is $L(x,y,z) = 2-2(x-1)+2(y-1)+2(z-1)$. The quadratic approximation

$$Q(x,y,z) = 2-2(x-1)+2(y-1)+2(z-1)+(-10(x-1)^2+2(y-1)^2+2(z-1)^2)/2$$

is the situation displayed to the right in Figure (1).

17.10. What is the tangent plane to the surface $f(x,y,z) = 1/10$ for $f(x,y,z) =$

$$10z^2-x^2-y^2+100x^4-200x^6+100x^8-200x^2y^2+200x^4y^2+100y^4 = 1/10$$



at the point $(x,y,z) = (0,0,1/10)$? The gradient is $\nabla f(0,0,1/10) = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$. The tangent plane equation is $2z = d$, where the constant d is obtained by plugging in the point. We end up with $2z = 2/10$. The linearization is $L(x,y,z) = 1/20+2(z-1/10)$.

17.11. P.S. The following remark should maybe be skipped as many objects have not been properly introduced. The exterior derivative d for example will appear in the form of grad,curl,div later on and $d^2 = 0$ in the form curl(grad(f)) = 0. The quite deep remark illustrates **how important** the topic of Taylor series is if it is taken seriously.

The derivative d acts on anti-symmetric tensors (= **forms**), where $d^2 = 0$. A vector field X then defines a **Lie derivative** $L_X = d\iota_X + \iota_X d = (d+\iota_X)^2 = D_X^2$ with **interior product** ι_X. For scalar functions and the constant field $X(x) = v$, one gets the **directional derivative** $D_v = \iota_X d$. The projection ι_X in a specific direction can be replaced with the transpose d^* of d. Rather than transport along X, the signal now radiates everywhere. The operator $d+\iota_X$ becomes then the **Dirac operator** $D = d+d^*$ and its square is the **Laplacian** $L = (d+d^*)^2 = dd^*+d^*d$. The **wave equation** $f_{tt} = -Lf$ can be written as $(\delta_t^2+D^2)f = (\delta_t - iD)(\delta_t+iD)f = 0$ which has the solution $ae^{iDt}+be^{-iDt}$. Using the **Euler formula** $e^{iDt} = \cos(Dt)+i\sin(Dt)$ one gets the explicit solutions $f(t) = f(0)\cos(Dt)+iD^{-1}f_t(0)\sin(Dt)$ of the wave equation. It gets more exciting: by packing the initial position and velocity into a **complex wave** $\psi(0,x) = f(0,x)+iD^{-1}f_t(0,x)$, we have $\psi(t,x) = e^{iDt}\psi(0,x)$. **The wave equation is solved by a Taylor formula, which solves a Schrödinger equation for** D and **the classical Taylor formula is the Schrödinger equation for** D_X. This works in any framework featuring a derivative d, like finite graphs, where Taylor resembles a **Feynman path integral**, a sort of Taylor expansion used by physicists to compute complicated particle processes.

The Taylor formula shows that the directional derivative D_v generates translation by $-v$. In physics, the operator $P = -i\hbar D_v$ is called the **momentum operator** associated to the vector v. The Schrödinger equation $i\hbar f_t = Pf$ has then the solution $f(x-tv)$ which means that the solution at time t is the initial condition translated by tv. This generalizes to the Lie derivative L_X given by **Cartan's magic formula** as $L_X = D_X^2$ acting on forms defined by a vector field X. For the analog $L = D^2$, the motion is not channeled in a determined direction X (this is a photon) but spreads (this is a wave) in all direction leading to the wave equation. We have just seen both the "photon picture" L_X as well as the "wave picture" L of light. **And whether it is particle or wave, it is all just Taylor.**

Homework

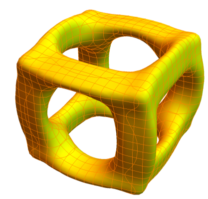
Problem 17.1: Evaluate without technology the cube root of 1002 using quadratic approximation. Especially look how close you are to the real value.

Problem 17.2: Compute without a computer the square root of 102 using quadratic approximation. Also here, look how close you get to the actual value.

Problem 17.3: Given $g(x,y) = (6y^2-5)^2(x^2+y^2-1)^2$, define the surface S by $f(x,y,z) = g(x,y) + g(y,z) + g(z,x) = 3$. The following equation could be derived with the chain rule. You can take this for granted:

$$\nabla f(1,-1,1) = \begin{bmatrix} g_x(1,-1) + g_y(1,1) \\ g_x(-1,1) + g_y(1,-1) \\ g_x(1,1) + g_y(-1,1) \end{bmatrix} .$$

Using this, find the tangent plane to S at $(1,-1,1)$.



Problem 17.4: a) Find the tangent plane to the surface $f(x,y,z) = \sqrt{xyz} = 60$ at $(x,y,z) = (100,36,1)$. b) Estimate $\sqrt{100.1 \cdot 36.1 \cdot 0.999}$ using linear approximation (compute $L(x,y,z)$ rather than $f(x,y,z)$.)

Problem 17.5: a) At which of the points $P, Q, R, S, T, \ldots, Y$ does $\nabla f(x)$ have maximal length?
b) At which of the points is $f_x > 0$ and $f_y = 0$?

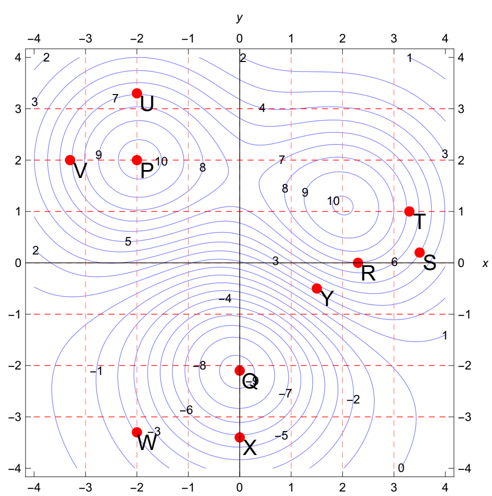


Figure 2.

Oliver Knill, knill@math.harvard.edu, Math 22a, Harvard College, Fall 2018