LINEAR ALGEBRA AND VECTOR ANALYSIS

MATH 22B

Unit 17: Taylor approximation

Introduction

17.1. According to legend [1], Richard Feynman got into the challenge to compute the cube root of 1729.03 against an Abacus computation. By using linear approximation and a bit o luck, he could get 12.002384 using paper and pencil. The actual cube root is 12.00238378569171812305 7. How did Feynman do it? The secret is in linear approximation. This means that we approximate a function like $f(x) = x^{1/3}$ with a linear function. The same can be done with functions of several variables. The linear approximation if of the form $L(x) = f(a) + f'(a)(x - a)$.

Figure 1. The Abacus scene in the movie “Infinity”.

17.2. One can also do higher order approximations. The function $f(x) = e^x$ for example has the linear approximation $L(x) = 1 + x$ at $a = 0$ and the quadratic approximation $Q(x) = 1 + x + x^2/2$ at $a = 0$. To get the quadratic term, we just need to make sure that the first and second derivative at $x = a$ agree. This gives the formula $Q(x) = f(a)+f'(a)(x-a)+f''(a)(x-a)^2/2$. Indeed, you can check that $f(x)$ and $Q(x)$ have the same first derivatives and the same second derivatives at $x = a$. A

[1]”Feynmans book ”What do you care what other people think”

degree n approximation is then the **polynomial**

$$P_n(x) = \sum_{k=0}^{n} f^{(k)}(a)\frac{(x-a)^k}{k!} \ .$$

For the function e^x for example, we have the m'th order approximation

$$e^x = 1 + x + x^2/2! + x^3/3! + ... + x^n/n! \ .$$

17.3. The same can be done in higher dimensions. Everything is the same. We just have to use the derivative df rather than the usual derivative f'. We look here only at linear and quadratic approximation of functions $\mathbb{R}^n \to \mathbb{R}$ The linear approximation is then

$$L(x) = f(a) + \nabla f(a)(x-a)$$

where $\nabla f(a) = df(a) = [f_{x_1}(a), \cdots, f_{x_n}(a)]$ is the Jacobian matrix, which ii a row vector. Now, since we can see $df(x) : \mathbb{R}^n \to \mathbb{R}^n$ the second derivative is a matrix $d^2f(x) = H(x)$. It is called the Hessian. It encodes all the second derivatives $H_{ij}(x) = f_{x_i x_j}$.

Lecture

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

17.5. 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.

17.6. The Taylor formula can be written down using successive derivatives df, d^2f, d^3f 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 $d^2f(x)$ to the **Hessian matrix** $H(x)$ which is a bilinear form acting on pairs of vectors. The third derivative $d^3f(x)$ 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.

17.7. 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. [2]

17.8. 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**. [3]

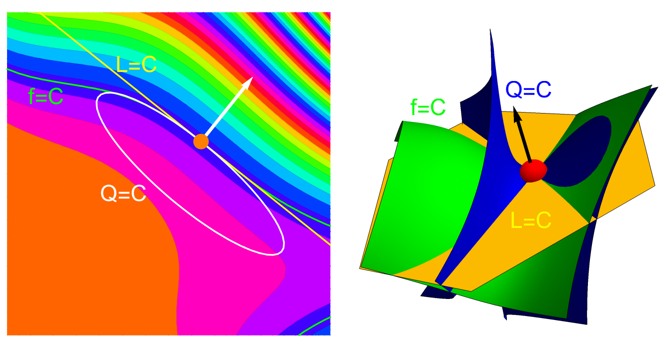

FIGURE 2. The manifolds $f(x, y) = C, L(x, y) = C$ and $Q(x, y) = C$ for $C = f(x_0, y_0)$ pass through the point (x_0, y_0). To the right, we see the situation for $f(x, y, z) = C$. We see the best linear approximation and quadratic approximation. The gradient is perpendicular.

17.9. 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 $d = ax_0 + by_0 + cz_0$.

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

[3] 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!)

17.10. 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

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

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

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 (2). 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: $f([1,1] + t[7,2]) = f(1+7t, 1+2t) = 3 + 38t + 247t^2 + 1023t^3 + 1960t^4 + 1372t^5$.

17.12. 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 (2).

17.13. 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 $L(x,y,z) = 1/20 + 2(z - 1/10)$.

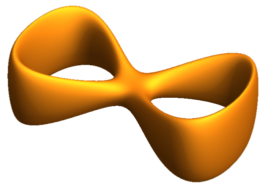

FIGURE 3.

Homework

Problem 16-17.1: Let $r(t) = [3t + \cos(t), t + 4\sin(t)]^T$ be a curve and $f([x,y]^T) = [x^3 + y, x + 2y + y^3]^T$ be a coordinate change.
a) Compute $v = r'(0)$ at $t = 0$, then $df(x,y)$ and $A = df(r(0))$ and $df(r(0))r'(0) = Av$.
b) Compute $R(t) = f(r(t))$ first, then find $w = R'(0)$. It should agree with a).

Problem 16-17.2: a) The surface

$$f(x,y,z) = x^2 + \frac{y^2}{4} + \frac{z^2}{9} = 4 + 1/4 + 1/9$$

is an ellipsoid. Compute $z_x(x,y)$ at the point $(x,y,z) = (2,1,1)$ using the implicit differentiation rule. (Use the formula).
b) Apply the Newton step 3 times starting with $x = 2$ to solve the equation $x^2 - 2 = 0$.

Problem 16-17.3: Evaluate without technology the cube root of 1002 using quadratic approximation. Especially look how close you are to the real value.

Problem 16-17.4: 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 16-17-5: Find the quadratic approximation $Q(x,y)$ of $f(x,y) = x^3 + x^2y + x^2 + y^2 - 2x + 3xy$ at the point $(1,2)$ by computing the gradient vector $\nabla f(1,2)$ and the Hessian matrix $H(1,2)$. The vector $\nabla f(1,2)$ is a 1×2 matrix (row vector) and the Hessian matrix $H(1,2)$ is a 2×2 matrix.

OLIVER KNILL, KNILL@MATH.HARVARD.EDU, MATH 22B, HARVARD COLLEGE, SPRING 2022