MATH 2850 Final RC Part 3

By Yang Mo

Content

- Second Derivative
- Free Extrema
- Taylor Series

Second Derivative

Second Derivative

We can write

$$Df : x \mapsto (\nabla f(x))^T = Df|_x. \tag{14}$$

Hence, $Df = (\cdot)^T \circ \nabla f$ and we can differentiate Df by the chain rule. The derivative of the map

$$\nabla f : \mathbb{R}^n \to \mathbb{R}^n, \quad \nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}\Big|_x \\ \vdots \\ \frac{\partial f}{\partial x_n}\Big|_x \end{pmatrix} \tag{15}$$

can be easily calculated: assuming sufficient smoothness of f, it is just the Jacobian of ∇f.

Remark: We represent $Df|_x$ using $\nabla f(x)$ because we are familiar with how to differentiate a column vector.

We hence have

$$D(\nabla f)|_x = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}\Big|_x & \frac{\partial^2 f}{\partial x_2 \partial x_1}\Big|_x & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}\Big|_x \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}\Big|_x & \frac{\partial^2 f}{\partial x_2 \partial x_n}\Big|_x & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}\Big|_x \end{pmatrix} \in \mathrm{Mat}(n \times n; \mathbb{R}). \tag{16}$$

Where

$$\frac{\partial^2 f}{\partial x_i \partial x_j} := \frac{\partial}{\partial x_i} \frac{\partial f}{\partial x_j} \tag{17}$$

is the second partial derivative of *f* concerning x_j (first) and x_i (second). The matrix in equation is important enough to warrant a special name: It is called the *Hessian* of *f* and denoted by $Hess\ f(x)$.

Schwarz's Theorem

Let (X,V) be normed vector spaces and (\Omega\subset X) an open set. Let $f \in C^2(\Omega, V)$. Then $D^2 f|_x \in \mathcal{L}^{(2)}(X \times X, V)$ is symmetric for all $x \in \Omega$,

$$D^2 f(u,v) = D^2 f(v,u), \qquad \text{for all } u, v \in X. \tag{18}$$

which means $\mathbf{Hess}\ f(x) = (\mathbf{Hess}\ f(x))^T$,implying that the Hessian of f at x is a symmetric matrix. Writing out the components of $\mathbf{Hess}\ f(x)$, this means that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}. \tag{19}$$

In other words, if f is twice continuously differentiable, the order of differentiation in the second-order partial derivatives does not matter.

Exercise 2. Calculate the first derivative of the function

$$f\colon \operatorname{Mat}(n \times n; \mathbb{R}) \to \operatorname{Mat}(n \times n; \mathbb{R}), \qquad f(A) = A^3$$

Find the second derivative, i.e., the derivative of $Df|_x$ as a function of x. **(6 Marks)**

We have

$$\begin{aligned} f(A+H) &= (A+H)^3 = (A+H)(A+H)(A+H) = (A+H)(A^2+HA+AH+H^2) \\ &= A^3 + AHA + A^2H + HA^2 + o(H) \end{aligned}$$

(1 Mark) so

$$Df|_A H = AHA + A^2H + HA^2$$

(1 Mark) Then

$$\begin{aligned} Df|_{A+J}H &= (A+J)H(A+J) + (A+J)^2H + H(A+J)^2 \\ &= AHA + A^2H + HA^2 + JHA + AHJ + AJH + JAH + AHJ + AJH + o(J) \end{aligned}$$

(1 Mark) so

$$D^2 f|_A(H,J) = JHA + AHJ + AJH + JAH + AHJ + AJH$$

(1 Mark) and the second derivative is interpreted as a bilinear map.

Free Extrema

22.3. Definition. Let $A \in \mathrm{Mat}(n \times n, \mathbb{R})$. Then the ***quadratic form induced by A*** is defined as the map

$$Q_A := \langle \cdot \,, A(\cdot)\rangle, \qquad x \mapsto \langle x, Ax\rangle = \sum_{j,k=1}^{n} a_{jk} x_j x_k, \qquad x \in \mathbb{R}^n.$$

Clearly, $Q_A(\lambda x) = \lambda^2 Q_A(x)$ for any $\lambda \in \mathbb{R}$. Note also that Q_A is continuous, because it is a polynomial in $x_1, \ldots, x_n$.

22.4. Definition. A quadratic form Q_A induced by a matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ is called

- ***positive definite*** if $Q_A(x) > 0$ for all $x \neq 0$,
- ***negative definite*** if $Q_A(x) < 0$ for all $x \neq 0$,
- ***indefinite*** if $Q_A(x_0) > 0$ for some $x_0 \in \mathbb{R}^n$ and $Q_A(y_0) < 0$ for some $y_0 \in \mathbb{R}^n$.

A matrix A is said to be negative definite / positive definite / indefinite if the induced quadratic form Q_A has the corresponding property.

22.5. Remarks.

- It is easy to see that not all quadratic forms fall into one of the above three categories.
- If A is positive definite, then $-A$ is negative definite.

22.6. Example. The matrix

$$A = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}$$

is positive definite, since

$$\begin{aligned} Q_A(x) &= \left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\rangle \\ &= x_1(x_1 - 2x_2) + x_2(x_1 + x_2) \\ &= x_1^2 + x_2^2 - x_1x_2 = \frac{1}{2}(x_1^2 + x_2^2 + (x_1 - x_2)^2). \end{aligned}$$

This expression is strictly positive when $x \neq 0$.

Free Extrema

Do you find it is sounds easy for you to find an extrema? Yes! In VV186, you have already learned the meaning of derivatives and extreme points. You can say, all the saddle points and extreme points you find is the result of linear approximation. Furthermore, to check whether the point is an extrema, second derivative is a necessity, which I believe you have already known in your VV186. Second derivative, is actually an quadratic approximation, which means you knows more properties locally about the function.

For *Quadratic Approximation.* Let $\Omega \subset \mathbb{R}^n$ be an open set and f in $C^2(\Omega, \mathbb{R})$. Then, as $h \to 0$,

$$f(x+h) = f(x) + < \nabla f(x), h > + \frac{1}{2} < \text{Hess}\, f(x)h, h > + o(h^2). \tag{20}$$

You can think about what the case will be like if all the things is one dimention. It is Taylor Expansion.

22.13. Corollary. Let $\Omega \subset \mathbb{R}^2$ be open, $f \in C^2(\Omega)$ and $\xi \in \Omega$ with $\nabla f(\xi) = 0$. Set

$$\Delta := \det \operatorname{Hess} f|_\xi = \frac{\partial^2 f}{\partial x_1^2}\bigg|_\xi \frac{\partial^2 f}{\partial x_2^2}\bigg|_\xi - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\bigg|_\xi\right)^2$$

Then $f(\xi)$ is

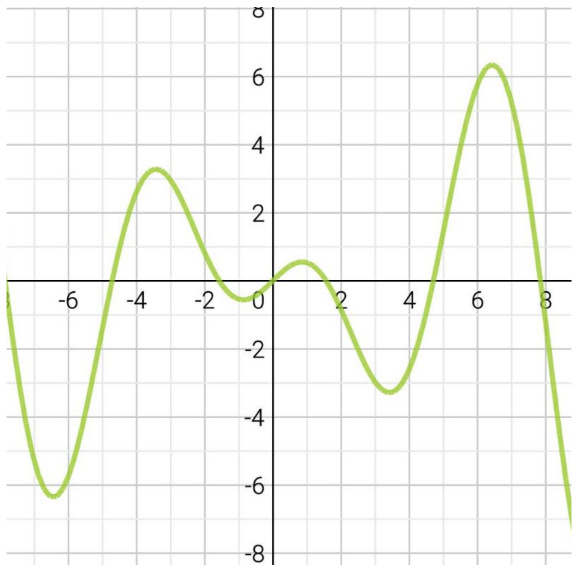
- a local minimum if $\frac{\partial^2 f}{\partial x_1^2}\big|_\xi > 0$ and $\Delta > 0$,
- a local maximum if $\frac{\partial^2 f}{\partial x_1^2}\big|_\xi < 0$ and $\Delta > 0$,
- no extremum if $\Delta < 0$.

Note that if $\Delta = 0$, Corollary 22.13 yields no information.

Finding Extrema

In searching for extrema of functions $f \in C^2(\Omega, \mathbb{R})$, we follow a four-step process:

1. Check for critical points $\xi \in \operatorname{int} \Omega$, i.e., those were $Df|_\xi = 0$.
2. Use Theorem 22.12 or Corollary 22.13 to check which of the critical points is an extremum.
3. Check the boundary $\partial\Omega$ separately for local extrema.
4. Identify the global extrema. Any finite global extremum must also be a local extremum, so it will be included among those found above.



Exercise

Let $f(x, y) = x^3 + y^3 - 3xy$, find the extrema of this function.

22.14. Example. Let $f(x, y) = x^3 + y^3 - 3xy$ on $\mathbb{R}^2$. Then $\nabla f = 0$ gives

$$\frac{\partial f}{\partial x} = 3x^2 - 3y = 0, \qquad \frac{\partial f}{\partial y} = 3y^2 - 3x = 0$$

or $x^2 = y$ and $y^2 = x$. The only two solutions to these equations are $(0, 0)$ and $(1, 1)$.

We have

$$\Delta(x, y) = f_{xx}f_{yy} - f_{xy}^2 = (6x)(6y) - (-3)^2 = 36xy - 9.$$

At $(0, 0)$, $\Delta < 0$ so there is no extremum at this point. At $(1, 1)$, $\Delta > 0$ and $f_{xx} = 6 > 0$, so this point corresponds to local minimum. Since $\Omega = \mathbb{R}^2$ is open, there are no other extrema.

Taylor Series

Higher-Order Derivatives

In this section we will suppose $(X, \|\cdot\|_X)$ and $(V, \|\cdot\|_z v)$ to be normed vector spaces, $\Omega \subset X$ an open set and we will consider functions $f\colon \Omega \to V$.

We may extend the strategy of Definition 21.1 to define derivatives of higher than second order inductively by setting

$$D^k f|_x = D(D^{k-1} f)|_x \in \mathscr{L}^{(k)}(\underbrace{X \times \cdots \times X}_{k \text{ times}}, V).$$

for $k = 2, 3, 4 \ldots$. Here, we again identify

$$\mathscr{L}^{(k)}(\underbrace{X \times \cdots \times X}_{k \text{ times}}, V) \cong \mathscr{L}(X, \mathscr{L}^{(k-1)}(\underbrace{X \times \cdots \times X}_{k-1 \text{ times}}, V))$$

We denote by $\boldsymbol{C^k(\Omega, V)}$ the set of those functions whose kth derivative is continuous and by $\boldsymbol{C^\infty(\Omega, V)}$ the intersection of all sets $C^k(\Omega, V)$, $k \in \mathbb{N}$.

Multi-Index Notation

For maps $f \in C^k(\mathbb{R}^n, \mathbb{R})$ the following ***multi-index notation*** for partial derivatives has been developed. This notation depends essentially on the interchangeability of partial derivatives.

The tuple $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ is called a ***multi-index*** of ***degree*** $|\alpha| = \alpha_1 + \cdots + \alpha_n$. We also define

$$\alpha! := \alpha_1!\alpha_2!...\alpha_n! = \prod_{i=1}^{n} \alpha_i!$$

For $f \in C^k(\mathbb{R}^n, \mathbb{R})$ we define

$$\partial^\alpha f := \frac{\partial^\alpha f}{\partial x^\alpha} := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_n}}.$$

For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we define

$$x^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n} = \prod_{i=1}^{n} x_i^{\alpha_i}.$$

Taylor's Theorem

23.1. Taylor's Theorem. Suppose that $f \in C^k(\Omega, V)$. Let $x \in \Omega$ and $h \in X$ such that the line $\gamma(t) = x + th$, $0 \leq t \leq 1$, is wholly contained within Ω.

Denote by $h^{(k)}$ a k-tupel as in (23.2). Then for all $p \leq k$,

$$f(x+h) = f(x) + \frac{1}{1!} Df|_x h + \cdots + \frac{1}{(p-1)!} D^{p-1} f|_x h^{(p-1)} + R_p(x,h) \tag{23.3}$$

with the remainder term

$$R_p(x,h) = \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p f|_{x+th} h^{(p)} \, dt.$$

Example

Find the Taylor expansion of

$$f(x_1, x_2) = \cos(x_1 + 2x_2)$$

at the origin point.

23.3. Example. The Taylor polynomial of degree 2 of the function $f(x_1, x_2) = \cos(x_1 + 2x_2)$ at $x_0 \in \mathbb{R}^2$ is given by

$$\begin{aligned}
& t_{f;x_0;3}(x_1, x_2) \\
&:= f(x_0 + x) - R_3 \\
&= \frac{1}{(0,0)!}\partial^{(0,0)} f(x_0) x^{(0,0)} + \frac{1}{(1,0)!}\partial^{(1,0)} f(x_0) x^{(1,0)} + \frac{1}{(0,1)!}\partial^{(0,1)} f(x_0) x^{(0,1)} \\
&\quad + \frac{1}{(2,0)!}\partial^{(2,0)} f(x_0) x^{(2,0)} + \frac{1}{(1,1)!}\partial^{(1,1)} f(x_0) x^{(1,1)} + \frac{1}{(0,2)!}\partial^{(0,2)} f(x_0) x^{(0,2
\end{aligned}$$

To find the Taylor polynomial at $x_0 = 0$ we have

$$\begin{aligned}
& t_{f;0;3}(x_1, x_2) \\
&= \frac{1}{0!0!} f(0) x_1^0 x_2^0 + \frac{1}{1!0!}\partial_{x_1} f|_{x_0=0} x_1^1 x_2^0 + \frac{1}{0!1!}\partial_{x_2} f|_{x_0=0} x_1^0 x_2^1 \\
&\quad + \frac{1}{2!0!}\partial_{x_1}^2 f|_{x_0=0} x_1^2 x_2^0 + \frac{1}{1!1!}\partial_{x_1 x_2} f|_{x_0=0} x_1^1 x_2^1 + \frac{1}{0!2!}\partial_{x_2}^2 f|_{x_0=0} x_1^0 x_2^2 \\
&= f(0,0) + \partial_{x_1} f|_{x_0=0} x_1 + \partial_{x_2} f|_{x_0=0} x_2 + \frac{1}{2}\partial_{x_1}^2 f|_{x_0=0} x_1^2 \\
&\quad + \partial_{x_1 x_2} f|_{x_0=0} x_1 x_2 + \frac{1}{2}\partial_{x_2}^2 f|_{x_0=0} x_2^2 \\
&= 1 - \frac{1}{2}x_1^2 - 2x_1 x_2 - 2x_2^2 \\
&= \cos(x_1 + 2x_2) - R_3.
\end{aligned}$$

We could also have obtained this result in an easier way by using the Taylor formula for functions of a single variable:

$$\begin{aligned}\cos(x_1 + 2x_2) &= 1 - \frac{1}{2}(x_1 + 2x_2)^2 + O(x^4) \\ &= 1 - \frac{1}{2}x_1^2 - 2x_1x_2 - 2x_2^2 + O(x^4).\end{aligned}$$

In cases where this (quick) strategy will not easily work, the full formula (23.5) needs to be evaluated.

Tips

- Parametrization is important! Do include the domain of a function and do the right exercise!
- Follow processes on Horst's slides and sample exam.
- For proofs, maybe Green's Identity will be tested, please do review it.
- Taylor series is a new part, which TAs talked about it quite little as we don't know how it will be tested. But you should master the basic formulas.
- Do make sure you know how to do surface integration, know the important theorems, find derivatives and find extrema,.

Reference

- VV285, slide. Horst Hohberger
- Sample Exam. Horst Hohberger
- RC, slides. TA-Pingbang Hu
- RC, slides. TA-Chen Yuxiang
- RC, slides. Yahoo

Thank You!