MATH 2010E ADVANCED CALCULUS I
LECTURE 13

WING HONG TONY WONG

14.9 — Taylor's formula

When we were estimating the error for the linearization of a function (Section 14.6), or when we were classifying a critical point by second derivative test (Section 14.7), we mentioned that Taylor's formula is the supporting reason behind the actions.

Before we proceed, recall the Taylor's formula for a single-variable function f that has $k+1$ continuous derivatives.

$$\begin{aligned} f(x) &= f(a) + \sum_{i=1}^{k} \frac{f^{(i)}(a)}{i!}(x-a)^i + \frac{f^{(k+1)}\big(a+c(x-a)\big)}{(k+1)!}(x-a)^{k+1} \\ &= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(k+1)}(\xi)}{(k+1)!}(x-a)^{k+1}, \end{aligned}$$

where $0 < c < 1$, and ξ is a number between x and a.

Lemma 1. *Let $D \subseteq \mathbb{R}^n$ be an open domain. Let $f : D \to \mathbb{R}$ be a C^k real-valued function. Let $\mathbf{a} + t\mathbf{h} \in D$. Let $F(t) = f(\mathbf{a} + t\mathbf{h})$. Then*

$$F^{(k)}(t) = \left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\right)^k f\Bigg|_{\mathbf{a}+t\mathbf{h}}.$$

Proof. By chain rule,

$$\begin{aligned} F'(t) &= \nabla f(\mathbf{a}+t\mathbf{h}) \cdot \mathbf{h} \\ &= \left(\frac{\partial}{\partial x_1} f\Bigg|_{\mathbf{a}+t\mathbf{h}}, \ldots, \frac{\partial}{\partial x_n} f\Bigg|_{\mathbf{a}+t\mathbf{h}} \right) \cdot (h_1, \ldots, h_n) \\ &= \left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\right) f\Bigg|_{\mathbf{a}+t\mathbf{h}}. \end{aligned}$$

In general, let $G(t) = F^{(j)}(t)$ for some $j \geq 1$. If $F^{(j)}(t) = \left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\right)^j f\Bigg|_{\mathbf{a}+t\mathbf{h}}$, then

$$\begin{aligned} G'(t) = \frac{d}{dt} F^{(j)}(t) &= \nabla \left[\left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\right)^j f \right]_{\mathbf{a}+t\mathbf{h}} \cdot \mathbf{h} \\ &= \left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n}\right)^{j+1} f\Bigg|_{\mathbf{a}+t\mathbf{h}}. \end{aligned}$$

By the principal of mathematical induction, we are done. □

Date: Tuesday, 23$^{\text{th}}$ June, 2015.

Theorem 2 (Taylor's formula)**.** *Let $D \subseteq \mathbb{R}^n$ be an open domain. Let $f : D \to \mathbb{R}$ be a C^{k+1} real-valued function. Let $\mathbf{a} + t\mathbf{h} \in D$ for all $t \in [0, 1]$. Then*

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \sum_{i=1}^{k} \frac{1}{i!} \left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n} \right)^i f \Bigg|_{\mathbf{a}} + \frac{1}{(k+1)!} \left(h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n} \right)^{k+1} f \Bigg|_{\mathbf{a}+c\mathbf{h}},$$

where $0 < c < 1$.

Proof. Let $F(t) = f(\mathbf{a} + t\mathbf{h})$. Since F is a composite function of infinitely differentiable functions, it is also an infinitely differentiable function. Hence, by Taylor's formula for single-variable functions, we have

$$F(1) = F(0) + \sum_{i=1}^{k} \frac{1}{i!} F^{(i)}(0) + \frac{1}{(k+1)!} F^{(k+1)}(c),$$

where $0 < c < 1$. By Lemma 1, we are done.

□

Another form of Taylor's formula is

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^{k} \frac{1}{i!} \left((x_1 - a_1) \frac{\partial}{\partial x_1} + \cdots + (x_n - a_n) \frac{\partial}{\partial x_n} \right)^i f \Bigg|_{\mathbf{a}} + \frac{1}{(k+1)!} \left((x_1 - a_1) \frac{\partial}{\partial x_1} + \cdots + (x_n - a_n) \frac{\partial}{\partial x_n} \right)^{k+1} f \Bigg|_{\mathbf{a}+c(\mathbf{x}-\mathbf{a})}.$$

This is obtained by substituting that $\mathbf{x} = \mathbf{a} + \mathbf{h}$.

In particular, in $\mathbb{R}^2$,

$$\begin{aligned} f(x, y) &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \\ &+ \frac{1}{2}\big(f_{xx}(a, b)(x - a)^2 + 2f_{xy}(a, b)(x - a)(y - b) + f_{yy}(y - b)^2\big) \\ &+ \sum_{i=3}^{k} \frac{1}{i!} \left((x - a) \frac{\partial}{\partial x} + (y - b) \frac{\partial}{\partial y} \right)^i f \Bigg|_{(a,b)} \\ &+ \frac{1}{(k+1)!} \left((x - a) \frac{\partial}{\partial x} + (y - b) \frac{\partial}{\partial y} \right)^{k+1} f \Bigg|_{(a,b)+c(x-a,y-b)}. \end{aligned}$$

Recall from Section 14.6 in Lecture 10 that the error of the linear approximation of f at $\mathbf{x} = \mathbf{a}$ is

$$E(\mathbf{x}) = |f(\mathbf{x}) - L(\mathbf{x})| \leq \frac{1}{2} M(|x_1 - a_1| + \cdots + |x_n - a_n|)^2,$$

where $L(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$, and M is an upper bound of all $|f_{x_i x_j}|$ over D, $1 \leq i, j \leq n$, if it exists. This can be deduced directly from the Taylor's formula as follows.

$$
\begin{aligned}
f(\mathbf{x}) &= f(\mathbf{a}) + \left((x_1 - a_1)\frac{\partial}{\partial x_1} + \cdots + (x_n - a_n)\frac{\partial}{\partial x_n} \right) f \Bigg|_{\mathbf{a}} \\
&\quad + \frac{1}{2!}\left((x_1 - a_1)\frac{\partial}{\partial x_1} + \cdots + (x_n - a_n)\frac{\partial}{\partial x_n} \right)^2 f \Bigg|_{\mathbf{a}+c(\mathbf{x}-\mathbf{a})} \\
&= f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2} \sum_{1 \le i,j \le n} f_{x_i x_j}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big)(x_i - a_i)(x_j - a_j)
\end{aligned}
$$

$$
\begin{aligned}
|f(\mathbf{x}) - L(\mathbf{x})| &= \frac{1}{2} \left| \sum_{1 \le i,j \le n} f_{x_i x_j}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big)(x_i - a_i)(x_j - a_j) \right| \\
&\le \frac{1}{2} \sum_{1 \le i,j \le n} \left| f_{x_i x_j}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big) \right| |x_i - a_i| \, |x_j - a_j| \\
&\le \frac{1}{2} \sum_{1 \le i,j \le n} M |x_i - a_i| \, |x_j - a_j| = \frac{1}{2} M (|x_1 - a_1| + \cdots + |x_n - a_n|)^2.
\end{aligned}
$$

Also recall from Section 14.7 in Lecture 11 that the second derivative test says, if $\mathbf{a}$ is a critical point of f, i.e. $\nabla f(\mathbf{a}) = \mathbf{0}$, then let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the n eigenvalues of $D^2 f(\mathbf{a})$, and

- if $\lambda_1, \lambda_2, \ldots, \lambda_n < 0$, then $f(\mathbf{a})$ is a local maximum.
- if $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$, then $f(\mathbf{a})$ is a local minimum.
- if $\lambda_i < 0$ and $\lambda_j > 0$, then $f(\mathbf{a})$ is a saddle point.
- if $\lambda_i = 0$, and all λ's are of the same sign, then there is no conclusion.

To see this, since $\nabla f(\mathbf{a}) = \mathbf{0}$, we have

$$
\begin{aligned}
f(\mathbf{x}) - f(\mathbf{a}) &= \frac{1}{2} \sum_{1 \le i,j \le n} f_{x_i x_j}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big)(x_i - a_i)(x_j - a_j) \\
&= \frac{1}{2} \begin{pmatrix} x_1 - a_1 & \cdots & x_n - a_n \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} f_{x_1 x_1}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big) & \cdots & f_{x_1 x_n}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big) \\ \vdots & \ddots & \vdots \\ f_{x_n x_1}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big) & \cdots & f_{x_n x_n}\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big) \end{pmatrix} \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{pmatrix} \\
&= \frac{1}{2} (\mathbf{x} - \mathbf{a})^\top \cdot D^2 f\big(\mathbf{a} + c(\mathbf{x} - \mathbf{a})\big) \cdot (\mathbf{x} - \mathbf{a}).
\end{aligned}
$$

To determine the nature of the critical point $\mathbf{a}$, we need to study the sign of $f(\mathbf{x}) - f(\mathbf{a})$ for all $\mathbf{x}$ in a small neighbourhood of $\mathbf{a}$. Let $\mathbf{x} = \mathbf{a} + t\mathbf{u}$, where $t > 0$ and $\|\mathbf{u}\| = 1$. Then

$$
f(\mathbf{x}) - f(\mathbf{a}) = \frac{1}{2}(t\mathbf{u})^\top \cdot D^2 f\big(\mathbf{a} + ct\mathbf{u}\big) \cdot (t\mathbf{u}) = \frac{t^2}{2} \mathbf{u}^\top \cdot D^2 f\big(\mathbf{a} + ct\mathbf{u}\big) \cdot \mathbf{u},
$$

so the sign of $f(\mathbf{x}) - f(\mathbf{a})$ depends only on $Q_{\mathbf{u}}(t) = \mathbf{u}^\top \cdot D^2 f\big(\mathbf{a} + ct\mathbf{u}\big) \cdot \mathbf{u}$.

Note that if f is a C^2 function, then all entries in $D^2 f$ are continuous, implying that $Q_{\mathbf{u}}(t)$ is continuous in t. Now, we are going to conduct the following analysis.

- If $\lambda_1, \lambda_2, \ldots, \lambda_n < 0$, then $D^2 f(\mathbf{a})$ is a negative-definite matrix. In other words, $Q_{\mathbf{u}}(0) < 0$ for all $\mathbf{u}$ such that $\|\mathbf{u}\| = 1$. Since $Q_{\mathbf{u}}(t)$ is continuous, there exists $\epsilon_{\mathbf{u}} > 0$ such that for all $0 < t < \epsilon_{\mathbf{u}}$, $Q_{\mathbf{u}}(t)$ share the same sign as $Q_{\mathbf{u}}(0)$.

Let $\epsilon = \min\{\epsilon_{\mathbf{u}} : \|\mathbf{u}\| = 1\} > 0$ (the minimum exists since we are taking the minimum over a compact set). For all $\mathbf{x} = \mathbf{a} + t\mathbf{u} \in B_\epsilon(\mathbf{a})$, $f(\mathbf{x}) - f(\mathbf{a})$ have the same sign as $Q_{\mathbf{u}}(t)$, which has the same sign as $Q_{\mathbf{u}}(0) < 0$.

Therefore, $f(\mathbf{a})$ is a local maximum.

- If $\lambda_1, \lambda_2, \ldots, \lambda_n > 0$, then $D^2 f(\mathbf{a})$ is a positive-definite matrix. In other words, $Q_{\mathbf{u}}(0) > 0$ for all $\mathbf{u}$ such that $\|\mathbf{u}\| = 1$. With the same analysis as above, there exists $\epsilon > 0$ such that for all $\mathbf{x} = \mathbf{a} + t\mathbf{u} \in B_\epsilon(\mathbf{a})$, $f(\mathbf{x}) - f(\mathbf{a})$ have the same sign as $Q_{\mathbf{u}}(t)$, which has the same sign as $Q_{\mathbf{u}}(0) > 0$.

  Therefore, $f(\mathbf{a})$ is a local minimum.

- If $\lambda_i < 0$ and $\lambda_j > 0$, then there exist $\mathbf{u}$ and $\mathbf{u}'$ such that $Q_{\mathbf{u}}(0) < 0$ and $Q_{\mathbf{u}'}(0) > 0$. Hence, there exists $\epsilon_{\mathbf{u}} > 0$ such that for all $0 < t < \epsilon_{\mathbf{u}}$, $Q_{\mathbf{u}}(t)$ share the same sign as $Q_{\mathbf{u}}(0) < 0$, and there exists $\epsilon_{\mathbf{u}'} > 0$ such that for all $0 < t < \epsilon_{\mathbf{u}'}$, $Q_{\mathbf{u}'}(t)$ share the same sign as $Q_{\mathbf{u}'}(0) > 0$. In other words, for all $\epsilon > 0$, there exists $\mathbf{x} = \mathbf{a} + t\mathbf{u} \in B_\epsilon(\mathbf{a})$ such that $f(\mathbf{x}) - f(\mathbf{a})$ have the same sign as $Q_{\mathbf{u}}(t)$, which has the same sign as $Q_{\mathbf{u}}(0) < 0$, and there exists $\mathbf{x}' = \mathbf{a} + t\mathbf{u}' \in B_\epsilon(\mathbf{a})$ such that $f(\mathbf{x}) - f(\mathbf{a})$ have the same sign as $Q_{\mathbf{u}'}(t)$, which has the same sign as $Q_{\mathbf{u}'}(0) > 0$.

  Therefore, $f(\mathbf{a})$ is a saddle point.

- If $\lambda_i = 0$, then there exists $\mathbf{u}$ such that $Q_{\mathbf{u}}(0) = 0$. As a result, we cannot determine the sign of $Q_{\mathbf{u}}(t)$. Hence, there is no conclusion.

Example 3. Find a quadratic approximation to $f(x, y) = \sin x \sin y$ near the origin. Find the error of the approximation if $|x| \le 0.1$ and $|y| \le 0.1$.

Solution. By Taylor's formula,

$$\begin{aligned}
f(x,y) &= f(0,0) + f_x(0,0)x + f_y(0,0)y + \frac{1}{2!}\big(f_{xx}(0,0)x^2 + 2f_{xy}(0,0)xy + f_{yy}(0,0)y^2\big) \\
&\quad + \frac{1}{3!}\big(f_{xxx}(h,k)x^3 + 3f_{xxy}(h,k)x^2y + 3f_{xyy}(h,k)xy^2 + f_{yyy}(h,k)y^3\big) \\
&= 0 + 0 + 0 + \frac{1}{2}(0 + 2xy + 0) \\
&\quad + \frac{1}{6}\big((-\cos h \sin k)x^3 - 3(\sin h \cos k)x^2y - 3(\cos h \sin k)xy^2 - (\sin h \cos k)y^3\big),
\end{aligned}$$

where h is between 0 and x, and k is between 0 and y. The quadratic approximation of $f(x, y)$ is $Q(x, y) = xy$, and the error is

$$\begin{aligned}
|E(x,y)| &= |f(x,y) - Q(x,y)| \\
&= \frac{1}{6}\left|(-\cos h \sin k)x^3 - 3(\sin h \cos k)x^2y - 3(\cos h \sin k)xy^2 - (\sin h \cos k)y^3\right| \\
&\le \frac{1}{6}|\cos h \sin k||x|^3 + 3|\sin h \cos k||x|^2|y| + 3|\cos h \sin k||x||y|^2 + |\sin h \cos k||y|^3 \\
&\le \frac{1}{6}(|x|^3 + 3|x|^2|y| + 3|x||y|^2 + |y|^3) \\
&= \frac{1}{6}(|x| + |y|)^3 \\
&\le \frac{1}{6}(0.1 + 0.1)^3 = \frac{1}{6} \times 0.008 = \frac{1}{750}.
\end{aligned}$$

□

14.10 — Partial derivatives with different independent and dependent variables

As mentioned in Section 14.3 in Lecture 8 that when we take partial derivatives, we may treat all independent variables as constants. However, all dependent variables must not be treated as constants. Hence, using different independent and dependent variables will lead to drastically different results.

Example 4. Let $w = x^2 + y^2 + z^2$ and $z = x^2 + y^2$. Find $\dfrac{\partial w}{\partial x}$ if
(a) x and y are independent variables.
(b) x and z are independent variables.

Solution. (a) We can substitute $z = x^2 + y^2$ into w to get

$$w = x^2 + y^2 + (x^2 + y^2)^2$$
$$\frac{\partial w}{\partial x} = 2x + 2(x^2 + y^2)(2x) = 2x + 4x^3 + 4xy^2.$$

(b) We can substitute $y^2 = z - x^2$ into w to get

$$w = x^2 + (z - x^2) + z^2$$
$$w = z + z^2$$
$$\frac{\partial w}{\partial x} = 0.$$

□

Note that in Example 4, the answers for $\dfrac{\partial w}{\partial x}$ are drastically different, and we cannot obtain one from the other using the relation $z = x^2 + y^2$. We can try to understand this through the geometry.

w is the square of the distance from the point (x, y, z) to the origin. When x and y are independent variables, $\dfrac{\partial w}{\partial x}$ fixes y unchanged. The path traced when x moves is a parabola parallel to the xz-plane, so w changes. However, when x and z are independent variables, $\dfrac{\partial w}{\partial x}$ fixes z unchanged. The path traced when x moves is a circle parallel to the xy-plane, so w does not change.

Example 5. Let $\ln wx + \sin \dfrac{yz}{x^2} = 0$ and $x^y + \dfrac{z^2}{\cos w} = 0$. Find $\dfrac{\partial w}{\partial x}$ if
(a) x and y are independent variables.
(b) x and z are independent variables.

Solution. (a)

$$\frac{1}{wx}\left(w + x\frac{\partial w}{\partial x}\right) + \cos\frac{yz}{x^2} \cdot \frac{x^2 y \frac{\partial z}{\partial x} - 2xyz}{x^4} = 0$$
$$x^{y-1} + \frac{\cos w \cdot 2z\frac{\partial z}{\partial x} + z^2 \sin w \cdot \frac{\partial w}{\partial x}}{\cos^2 w} = 0$$

From the first equation, we have

$$\frac{\partial z}{\partial x} = \frac{1}{x^2 y}\left[-\frac{x^4}{wx\cos\frac{yz}{x^2}}\left(w + x\frac{\partial w}{\partial x}\right) + 2xyz\right] = -\frac{x}{wy\cos\frac{yz}{x^2}}\left(w + x\frac{\partial w}{\partial x}\right) + \frac{2z}{x}.$$

Substituting this into the second equation, we have

$$\begin{aligned}
x^{y-1} + \frac{\cos w \cdot 2z \left[-\frac{x}{wy\cos\frac{yz}{x^2}} \left(w + x\frac{\partial w}{\partial x} \right) + \frac{2z}{x} \right] + z^2 \sin w \cdot \frac{\partial w}{\partial x}}{\cos^2 w} &= 0 \\
-\frac{2xz\cos w}{wy\cos\frac{yz}{x^2}} \left(w + x\frac{\partial w}{\partial x} \right) + \frac{4z^2\cos w}{x} + z^2 \sin w \cdot \frac{\partial w}{\partial x} &= -x^{y-1}\cos^2 w \\
-\frac{2xz\cos w}{y\cos\frac{yz}{x^2}} - \frac{2x^2z\cos w}{wy\cos\frac{yz}{x^2}} \cdot \frac{\partial w}{\partial x} + \frac{4z^2\cos w}{x} + z^2 \sin w \cdot \frac{\partial w}{\partial x} &= -x^{y-1}\cos^2 w \\
\frac{x^{y-1}\cos^2 w - \frac{2xz\cos w}{y\cos\frac{yz}{x^2}} + \frac{4z^2\cos w}{x}}{\frac{2x^2z\cos w}{wy\cos\frac{yz}{x^2}} - z^2\sin w} &= \frac{\partial w}{\partial x}
\end{aligned}$$

(b)

$$\begin{aligned}
\frac{1}{wx}\left(w + x\frac{\partial w}{\partial x} \right) + \cos\frac{yz}{x^2} \cdot \frac{x^2 z\frac{\partial y}{\partial x} - 2xyz}{x^4} &= 0 \\
x^y \left(\ln x \cdot \frac{\partial y}{\partial x} + \frac{y}{x} \right) + \frac{z^2 \sin w}{\cos^2 w} \cdot \frac{\partial w}{\partial x} &= 0
\end{aligned}$$

From the second equation, we have

$$\frac{\partial y}{\partial x} = -\frac{z^2\sin w}{x^y\cos^2 w\ln x} \cdot \frac{\partial w}{\partial x} - \frac{y}{x\ln x}.$$

Substituting this into the first equation, we have

$$\begin{aligned}
\frac{1}{wx}\left(w + x\frac{\partial w}{\partial x} \right) + \cos\frac{yz}{x^2} \cdot \frac{x^2 z\left[-\frac{z^2\sin w}{x^y\cos^2 w\ln x} \cdot \frac{\partial w}{\partial x} - \frac{y}{x\ln x} \right] - 2xyz}{x^4} &= 0 \\
\frac{1}{x} + \frac{1}{w} \cdot \frac{\partial w}{\partial x} - \cos\frac{yz}{x^2} \cdot \left[\frac{z^3\sin w}{x^{y+2}\cos^2 w\ln x} \cdot \frac{\partial w}{\partial x} + \frac{yz}{x^3\ln x} + \frac{2yz}{x^3} \right] &= 0 \\
\frac{\frac{1}{w} - \frac{z^3\sin w\cos\frac{yz}{x^2}}{x^{y+2}\cos^2 w\ln x}}{\frac{yz\cos\frac{yz}{x^2}}{x^3\ln x} + \frac{2yz\cos\frac{yz}{x^2}}{x^3} - \frac{1}{x}} &= \frac{\partial w}{\partial x}
\end{aligned}$$

□