Reading: [Simon] p. 313-333, 833-836.

0.1 The Chain Rule

Partial derivatives describe how a function changes in directions parallel to the coordinate axes. Now we shall demonstrate how the partial derivatives can be used to describe how a function changes in any direction. More generally, we are often interested in how a function changes as we move along a curve in its domain. For example, if inputs are changing with time, we may want to know how the corresponding outputs are changing with time. Or, if the input obeys budget restriction which is a line, then how the output changes along the budget line.

0.1.1 Curves

A *curve* in R^n is a continuous function $R \to R^n$

$$x(t) = (x_1(t),\ \dots, x_n(t))$$

here each $x_i(t)$ is a continuous function $R \to R$.

Particular case: A straight line which passes trough a point $x^* = (x_1^*,\ \dots, x_n^*)$ in direction of a vector $v = (v_1,\ \dots, v_n)$

$$x(t) = (x_1(t),\ \dots, x_n(t)) = (x_1^* + v_1 t,\ \dots, x_n^* + v_n t).$$

Note, that $x(0) = x^*$ and $x_k'(t) = v_k$.

The *tangent vector* for a curve $x(t)$ is the vector $(x_1'(t),\ \dots, x_n'(t))$.

Example. The curve $x(t) = (cos\ t, sin\ t),\quad t \in [0, 2\pi]$ is the unit circle.
The tangent vector at $t = 0$ is $(-sin\ 0, cos\ 0) = (0, 1)$ vertical.
The tangent vector at $t = \pi/2$ is $(-sin\ \pi/2, cos\ \pi/2) = (-1, 0)$ horizontal.

0.1.2 Differentiating Along a Curve - Chain Rule

Let $x(t)$ be a curve in R^n and $f : R^n \to R$. Consider the composition

$$g : R \xrightarrow{x} R^n \xrightarrow{f} R,$$

$g(t) = f(x_1(t),\ \dots, x_n(t))$.

The chain rule gives the expression for the derivative of this function

$$\frac{dg}{dt}(t_0) = \frac{\partial f}{\partial x_1}(x(t_0))x_1'(t_0) + \ \dots\ + \frac{\partial f}{\partial x_n}(x(t_0))x_n'(t_0).$$

or, shortly

$$\frac{dg}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \ \dots\ + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.$$

Example. Consider the Cobb-Douglas production function

$$Q = 4K^{\frac{3}{4}}L^{\frac{1}{4}}. \tag{1}$$

Suppose K and L vary with time t and the interest rate r by the rule

$$K(t,r) = \frac{10t^2}{r}, \qquad L = 6t^2 + 250r \tag{2}$$

(notice, as r increases, then K decreases and L increases, why?).

If we substitute in Q these expressions for K and L then Q becomes a two variable function of t and r:

$$Q(t,r) = 4 \cdot (\frac{10t^2}{r})^{\frac{3}{4}} \cdot (6t^2 + 250r)^{\frac{1}{4}}. \tag{3}$$

How to calculate the rate of change of Q with respect to t when $t = 10$ and $r = 0.1$?

There are three possibilities:

(i) Just calculate the partial derivative with respect to t of (3);

(ii) Use Chain Rule for (1) and (2):

$$\frac{\partial Q}{\partial t} = \frac{\partial Q}{\partial K} \cdot \frac{\partial K}{\partial t} + \frac{\partial Q}{\partial L} \cdot \frac{\partial L}{\partial t}.$$

(iii) Look in [SB], pp. 318.

Example. (14.12 from [SB]) At a given moment in time, the marginal product of labor is 2.5 and the marginal product of capital is 3, the amount of capital is increasing by 2 each unit of time and the rate of change of labor is 0.5. What is the rate of change of output at this time?

Solution. So what is given? $\frac{\partial Q}{\partial L} = 2.5$, $\frac{\partial Q}{\partial K} = 3$, $\frac{\partial K}{\partial t} = 2$, $\frac{\partial L}{\partial t} = 0.5$.

What we need to calculate?

$$\frac{\partial Q}{\partial t} = \frac{\partial Q}{\partial K} \cdot \frac{\partial K}{\partial t} + \frac{\partial Q}{\partial L} \cdot \frac{\partial L}{\partial t} = 3 \cdot 2 + 2.5 \cdot 0.5 = 7.25.$$

0.2 Gradient

For a function $F : R^n \to R$ the *gradient* at point $x^* = (x_1^*, ..., x_n^*)$ is the vector

$$\nabla F(x^*) = (\frac{\partial F}{\partial x_1}(x^*), \ \ldots \ , \frac{\partial F}{\partial x_1}(x^*)).$$

Sometimes gradient is denoted as $D_1F(x^*)$.

Example. Consider again the Cobb-Douglas production function

$$Q = 4K^{\frac{3}{4}}L^{\frac{1}{4}}. \tag{4}$$

What is the gradient of this function at the point $(K = 10000, L = 625)$?

$$D_1Q(10000, 625) = (\tfrac{\partial Q}{\partial K}(10000, 625), \tfrac{\partial Q}{\partial L}(10000, 625)) =$$
$$(3K^{-1/4}L^{1/4}(10000, 625), K^{3/4}L^{-3/4}(10000, 625)) = (1.5, 8).$$

MAPLE calculates the gradient vector by
$> with(VectorCalculus):$
$> G := Gradient(4 * x^(3/4) * y^(1/4), [x, y]);$

0.2.1 Directional Derivative and Gradient

Directional derivative enables to compute the rate of change of function $F(x_1, ..., x_n)$ at a given point x^* and in the direction given by a *unit or normalized* vector $v = (v_1, ..., v_n)$.

This directional derivative is denoted as

$$\frac{\partial F}{\partial v}(x^*) \quad or \quad D_vF(x^*).$$

How can one calculate the Dir-Der? Take a straight line which passes trough the point x^* in the direction of a given vector $v = (v_1, \ ... \ , v_n)$

$$x(t) = (x_1(t), \ ... \ , x_n(t)) = (x_1^* + v_1t, \ ... \ , x_n^* + v_nt).$$

Note, that $x(0) = x^*$ and $x_k'(t) = v_k$.

The directional derivative $D_vF(x^*)$ is the **derivative of F along the straight line** $x(t)$

$$\frac{\partial F}{\partial x_1}(x(0))x_1'(0) + \ ... \ + \frac{\partial F}{\partial x_n}(x(0))x_n'(0) = \frac{\partial F}{\partial x_1}(x^*)v_1 + \ ... \ + \frac{\partial F}{\partial x_n}(x^*)v_n'.$$

Thus directional derivative equals to the inner product of gradient vector at x^* and the (*normalized*) vector $v = (v_1, \ ... \ , v_n)$

$$D_vF(x^*) = \nabla F(x^*) \cdot v.$$

Note that changing v proportionally (keeping the direction unchanged) changes the inner product in the same proportion.

Example. What is the derivative of our Cobb-Douglas production function $Q = 4K^{\frac{3}{4}}L^{\frac{1}{4}}$ in direction $v = (1, 1)$ at the point $(10000, 625)$?

First normalize v:

$$\hat{v} = \frac{v}{||v||} = \frac{v}{\sqrt{2}} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}),$$

then

$$D_{\hat{v}}Q((10000, 625) = D_1Q((10000, 625) \cdot (\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}) =$$
$$(1.5, 8) \cdot (\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}) = 1.5 \cdot \tfrac{1}{\sqrt{2}} + 8 \cdot \tfrac{1}{\sqrt{2}} = \tfrac{9.5}{\sqrt{2}} = 6.717514420.$$

And what is the rate of increasing of Q in direction of the *gradient vector* at the same point $(10000, 625)$? As we already know the gradient vector at this point is $(1.5, 8)$. Its normalization gives

$$\frac{(1.5, 8)}{||(1.5, 8)||} = \frac{(1.5, 8)}{8.139410298} = (\frac{1.5}{8.139410298}, \frac{8}{8.139410298}).$$

Then the derivative of Q in direction of the normalized gradient vector is

$$(1.5, 8) \cdot (\frac{1.5}{8.139410298}, \frac{8}{8.139410298}) = 8.13941.$$

Much faster than in *direction* of $v = (1, 1)$ which vas 6.7175!

0.2.2 Direction of Maximal Rate of Increasing

Recall that the inner product of two vectors w and v can be expressed as $w \cdot v = ||w|| \cdot ||v|| \cdot cos\ \alpha$, so if the lengths $||w||$ and $||v||$ are fixed the maximal value of the inner product is achieved when $\alpha = 0$. Thus the directional derivative

$$D_vF(x^*) = \nabla F(x^*) \cdot v = ||\nabla F(x^*)|| \cdot ||v|| \cdot cos\alpha$$

is maximal if the angle α between the gradient $\nabla F(x^*)$ and the direction v is 0, so

If the gradient is nonzero vector at x^* then it points into the direction in which F increases most rapidly.

Example 14.2 from [SB]. Consider the Cobb-Douglas production function $Q = 4K^{\frac{3}{4}}L^{\frac{1}{4}}$, and, let the current input bundle be $(K = 10000,\ L = 625)$. In what proportion one should add K and L to $(10000, 625)$ to increase the production most rapidly? The gradient in this point is

$$\nabla Q(10000, 625) = (\frac{\partial Q}{\partial K}(1000, 625), \frac{\partial Q}{\partial L}(1000, 625)) = (1.5, 8).$$

So this function increases most rapidly in this direction thus one should add K and L in the same proportion $1.5:8$.

Nope! Suppose we want to allocate 95. In ratio $1.5:8$ it is

$$95 = 15 + 80$$

and calculation shows

$$F(10015, 705) = 20634.57297.$$

Now take different allocation

$$95 = 0 + 95$$

then

$$F(10000, 720) = 20720.16051,$$

it is more than in ration $1.5:8$!

0.2.3 Gradient and Level Curves

A level set for a function $f(x_1, \ldots, x_n)$ is a set on which the function is constant. For a function of two variables often this set is a curve.

Example. For the function $f(x_1, x_2) = x_1^2 + x_2^2$ the level curve corresponding to the value $f(x_1, x_2) = 4$ is the set

$$x_1^2 + x_2^2 = 4,$$

and this is the circle with the center in the origin and radius $r = 2$. Similarly, the level curve corresponding to value $f(x_1, x_2) = 5$ is the circle with the same center and radius $r = \sqrt{5}$, etc. So the level curves for $f(x_1, x_2) = x_1^2 + x_2^2$ are concentric circles (except $f(x_1, x_2) = 0$ which is a point).

Suppose $x(t) = (x_1(t), x_2(t))$ is a level curve for a function $f(x_1, x_2)$ corresponding to a value c, that is

$$f(x_1(t), x_2(t)) = c. \tag{5}$$

As we know the tangent vector for a curve $x(t)$ is the vector $(x_1'(t), x_2'(t))$. Differentiating the equation (5) by t we obtain

$$\frac{\partial f}{\partial x_1} x_1'(t) + \frac{\partial f}{\partial x_2} x_2'(t) = 0,$$

this means $\nabla(f) \cdot (x_1'(t), x_2'(t)) = 0$, i.e

If the gradient is nonzero vector, then it is orthogonal to (tangent of) the level curve.

0.3 Jacobian

Let us consider a function $F : R^n \to R^m$, which, as we know, is a collection of $R^n \to R$ functions

$$F(x) = F\begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix} = \begin{pmatrix} f_1(x_1, ..., x_n) \\ .... \\ f_m(x_1, ..., x_n) \end{pmatrix}.$$

The Jacobian of F is defined as the matrix

$$DF(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & ... & \frac{\partial f_1}{\partial x_n}(x) \\ ... & ... & ... \\ \frac{\partial f_m}{\partial x_1}(x) & ... & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}.$$

As we see the Jacobian is a matrix whose rows are gradients of the functions $f_1,\ ...\ , f_n$:

$$DF(x) = \begin{pmatrix} Df_1(x) \\ ... \\ Df_m(x) \end{pmatrix}.$$

Other notations for Jacobian

$$DF(x) = J_F(x) = \frac{\partial(f_1,\ ...\ , f_m)}{\partial(x_1,\ ...\ , x_n)} = \frac{\partial F}{\partial x}(x).$$

Slang: The Jacobian $\frac{\partial F}{\partial x}$ is differentiation of a **vector** $F(x) = \begin{pmatrix} f_1 \\ ... \\ f_m \end{pmatrix}$ by **vector** $x = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}$, and the result is a matrix.

Particular cases:

$(n = 1,\ m = 1)$ $F : R \to R$, Jacobian $\frac{\partial F}{\partial x}$ is ordinary derivative (scalar by scalar, result is a scalar);

$(\ m = 1)$ $F : R^n \to R$, Jacobian $\frac{\partial F}{\partial x}$ is ordinary Gradient DF (scalar by vector, result is a vector);

$(n = 1)$ $F : R \to R^m$, Jacobian $\frac{\partial F}{\partial x}$ is the tangent vector of the curve $F(t) = \begin{pmatrix} f_1(t) \\ ... \\ f_n(t) \end{pmatrix}$ (vector by scalar, result is a vectors).

Examples

1. Differentiation of scalar by vector. Consider a *linear function* $F : R^n \to R$, given by $F(x_1, \ ... \ , x_n) = a_1 x_1 + ... \ + a_n x_n$. In matrix form $F(x) = a^T \cdot_M x$ where $x = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}$ and $a = \begin{pmatrix} a_1 \\ ... \\ a_n \end{pmatrix}$ and $a^T \cdot_M x$ is the matrix product (which equals to $a \cdot_i x$ where this means the inner product). In this case we have

$$DF(x) = \frac{\partial (a^T \cdot x)}{\partial x} = a.$$

2. Differentiation of vector by vector. Consider a *linear function* $F : R^n \to R^m$, given by $F(x) = A \cdot x$ where

$$x = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & ... & a_{1n} \\ ... & ... & ... \\ a_{m1} & ... & a_{mn} \end{pmatrix}.$$

Thus

$$F \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix} = \begin{pmatrix} f_1(x) = a_{11} x_1 + \ ... \ + a_{1n} x_n \\ ... \\ f_m(x) = a_{m1} x_1 + \ ... \ + a_{mn} x_n \end{pmatrix}$$

and

$$DF(x) = \frac{\partial (A \cdot x)}{\partial x} = A.$$

3. Differentiation of scalar by vector. Consider a *Quadratic form* $Q : R^n \to R$, given by

$$Q(x_1, \ ... \ , x_n) = \sum_{i=1,...,n, \ j=1,...,n} a_{ij} x_i x_j,$$

in matrix form $Q(x) = x^T \cdot A \cdot x$ where

$$x = \begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & ... & a_{1n} \\ ... & ... & ... \\ a_{n1} & ... & a_{nn} \end{pmatrix}.$$

Calculation shows that (try!)

$$DQ(x) = \frac{\partial (x^T \cdot A \cdot x)}{\partial x} = x^T (A + A^T).$$

If A is symmetric then $DQ(x) = \frac{\partial (x^T \cdot A \cdot x)}{\partial x} = 2x^T A$.

0.4 Linear Approximation

Start with a function $F : R^n \to R^m$, which, as we know, is a collection of $R^n \to R$ functions

$$F(x) = F\begin{pmatrix} x_1 \\ ... \\ x_n \end{pmatrix} = \begin{pmatrix} f_1(x_1, ..., x_n) \\ .... \\ f_m(x_1, ..., x_n) \end{pmatrix}.$$

Our aim is to approximate $F(x^* + \Delta x)$ in terms of the value $F(x^*)$, the vector $\Delta x = (\Delta x_1, \ ... \ , \Delta x_n)$ and the value of Jacobian $DF(x^*)$. Linear approximation of each function $f_i : R^n \to R$ gives

$$\begin{array}{l} f_1(x^* + \Delta x) - f_1(x^*) = \frac{\partial f_1}{\partial x_1}(x^*)\Delta x_1 + \ ... \ + \frac{\partial f_1}{\partial x_n}(x^*)\Delta x_n \\ f_2(x^* + \Delta x) - f_2(x^*) = \frac{\partial f_2}{\partial x_1}(x^*)\Delta x_1 + \ ... \ + \frac{\partial f_2}{\partial x_n}(x^*)\Delta x_n \\ ... \\ f_m(x^* + \Delta x) - f_m(x^*) = \frac{\partial f_m}{\partial x_1}(x^*)\Delta x_1 + \ ... \ + \frac{\partial f_m}{\partial x_n}(x^*)\Delta x_n. \end{array}$$

Shortly this can be written as

$$F(x^* + \Delta x) - F(x^*) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x^*) & ... & \frac{\partial f_1}{\partial x_n}(x^*) \\ ... & ... & ... \\ \frac{\partial f_m}{\partial x_1}(x^*) & ... & \frac{\partial f_m}{\partial x_n}(x*) \end{pmatrix} \cdot \begin{pmatrix} \Delta x_1 \\ ... \\ \Delta x_n \end{pmatrix}.$$

Even more shortly it looks as

$$F(x^* + \Delta x) - F(x^*) = DF(x^*) \cdot (\Delta x)^T.$$

0.5 Jacobian of Composite function

Consider the composition

$$R^3 \xrightarrow{G} R^3 \xrightarrow{F} R$$

with

$$G\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u(x, y, z) \\ v(x, y, z) \\ w(x, y, z) \end{pmatrix}.$$

The composition is

$$(F \circ G)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = F(u(x, y, z), v(x, y, z), w(x, y, z)).$$

By chain rule

$$\begin{array}{l} (F \circ G)_x = F_u u_x + F_v v_x + F_w w_x \\ (F \circ G)_y = F_u u_y + F_v v_y + F_w w_y \\ (F \circ G)_x = F_u u_z + F_v v_z + F_w w_z. \end{array}$$

That is the Jacobian (gradient in this case) of the composite function is

$$D(F\circ G) = (F_u u_x + F_v v_x + F_w w_x,\ F_u u_y + F_v v_y + F_w w_y,\ F_u u_z + F_v v_z + F_w w_z).$$

Note that the Jacobian of G is

$$DG = \begin{pmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{pmatrix}$$

and the Jacobian (gradient) of F is

$$DF = (F_u, F_v, F_w).$$

Not hard to mention that

$$(F_u, F_v, F_w) \cdot \begin{pmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{pmatrix} =$$
$$(F_u u_x + F_v v_x + F_w w_x,\ F_u u_y + F_v v_y + F_w w_y,\ F_u u_z + F_v v_z + F_w w_z).$$

Thus

$$D(F\circ G) = DF \cdot DG.$$

This is a particular case of the following

Theorem 1 *Let*

$$G = \begin{pmatrix} g_1 \\ .. \\ g_n \end{pmatrix} : R^m \to R^n, \quad F = \begin{pmatrix} f_1 \\ .. \\ f_k \end{pmatrix} : R^n \to R^k,$$

and let

$$R^m \xrightarrow{G} R^n \xrightarrow{F} R^k$$

be their composition.

Then the Jacobian of the composition $(F\circ G)(x^*)$ is the product of Jacobian matrices

$$D(F\circ G)(x^*) = DF(G(x^*)) \cdot DG(x^*).$$

Example

Let $F(u,v,w) = u^2 + v + w$ where $u = x + 2yz,\ v = x^2 + y,\ w = z^2 + x$. Find $\frac{\partial F}{\partial x},\ \frac{\partial F}{\partial y},\ \frac{\partial F}{\partial z}$.

Note that in fact we have the composition

$$R^3 \xrightarrow{G} R^3 \xrightarrow{F} R$$

with

$$G\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u = x + 2yz \\ v = x^2 + y \\ w = z^2 + x \end{pmatrix}$$

and

$$F\begin{pmatrix} u \\ v \\ w \end{pmatrix} = u^2 + v + w.$$

We want to find partial derivatives of the composite function

$$(F \circ G)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (x+2yz)^2+(x^2+y)+(z^2+x) = 2x^2+4xyz+4y^2z^2+y+z^2+x.$$

Three ways:

1. The direct calculation of patrial derivatives show

$$\frac{\partial F}{\partial x} = 4x + 4yz + 1, \ \frac{\partial F}{\partial y} = 4zx + 8yz^2 + 1, \ \frac{\partial F}{\partial z} = 4yx + 8y^2z + 2z.$$

2. By chain rule

$$(F \circ G)_x = F_u u_x + F_v v_x + F_w w_x = 2u \cdot 1 + 1 \cdot 2x + 1 \cdot 1 =$$
$$2(x + 2yz) \cdot 1 + 1 \cdot 2x + 1 \cdot 1 = 4x + 4yz + 1$$

$$(F \circ G)_y = F_u u_y + F_v v_y + F_w w_y = 2u \cdot 2z + 1 \cdot 1 + 1 \cdot 0 =$$
$$2(x + 2yz) \cdot 2z + 1 \cdot 1 + 1 \cdot 0 = 4xz + 8yz^2 + 1$$

$$(F \circ G)_x = F_u u_z + F_v v_z + F_w w_z = 2u \cdot 2y + 1 \cdot 0 + 1 \cdot 2z =$$
$$2(x + 2yz) \cdot 2y + 1 \cdot 0 + 1 \cdot 2z = 4xy + 8y^2z + 2z.$$

That is the gradient (Jacobian in this case) of the composite function is

$$D(F \circ G) = (F_u u_x + F_v v_x + F_w w_x, \ F_u u_y + F_v v_y + F_w w_y, \ F_u u_z + F_v v_z + F_w w_z).$$

3. The Jacobian of G is

$$DG = \begin{pmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{pmatrix} = \begin{pmatrix} 1 & 2z & 2y \\ 2x & 1 & 0 \\ 1 & 0 & 2z \end{pmatrix}$$

and the Jacobian (gradient) of F is

$$DF = (F(u, F_v, F_w) = (2u, 1, 1) = (2(x + 2yz, 1, 1).$$

Then the Jacobian (the gradient in this case) is

$$D(F \circ G) = DF \cdot DG = (2(x+2yz), 1, 1) \cdot \begin{pmatrix} 1 & 2z & 2y \\ 2x & 1 & 0 \\ 1 & 0 & 2z \end{pmatrix} =$$
$$\begin{pmatrix} 2(x+2yz) + 2x + 1 \\ 2(x+2yz) \cdot 2z + 1 \\ 2(x+2yz) \cdot 2y + 2z \end{pmatrix} = \begin{pmatrix} 4x + 4yz + 1 \\ 4xz + 8yz^2 + 1 \\ 4xy + 8y^2 z + 2z \end{pmatrix}.$$

1 Higher Order Derivatives

The second order derivative of $f : R^n \to R$ can be calculated as

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

Sometimes the following notation is used $\frac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_j x_i}$. Also the notation $\frac{\partial^2 f}{\partial x_1 \partial x_1} = \frac{\partial^2 f}{\partial x_1^2}$ is used.

Similarly are determined higher order partial derivatives

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \, ... \, \partial x_{i_k}}.$$

All n^2 second order partial derivatives form the *Hessian matrix*

$$D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

Young's Theorem. $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$.

This implies that the Hessian matrix is *symmetric*.

Using this theorem one can prove similar fact for third order partials

$$\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k} = \frac{\partial^3 f}{\partial x_i \partial x_k \partial x_j} = \frac{\partial^3 f}{\partial x_j \partial x_i \partial x_k} = \frac{\partial^3 f}{\partial x_j \partial x_k \partial x_i} = \frac{\partial^3 f}{\partial x_k \partial x_i \partial x_j} = \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i} \; .$$

2 Taylor polynomials for a function of two variables

We already know the linear approximation

$$F(x_1 + h_1, x_2 + h_2) \approx P_2(h_1, h_2) =$$
$$F(x_1, x_2) + \frac{\partial F}{\partial x_1}(x_1, x_2) \cdot h_1 + \frac{\partial F}{\partial x_2}(x_1, x_2) \cdot h_2 =$$

$$F(x_1, x_2) + \nabla F(x_1, x_2) \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

The second order partial derivatives allow to construct better approximation by *second order Taylor polynomial* $P_2(h_1, h_2)$

$$F(x_1 + h_1, x_2 + h_2) \approx P_2(h_1, h_2) =$$
$$F(x_1, x_2) + \frac{\partial F}{\partial x_1}(x_1, x_2) \cdot h_1 + \frac{\partial F}{\partial x_2}(x_1, x_2) \cdot h_2 +$$
$$\frac{1}{2}\frac{\partial^2 F}{\partial x_1^2}(x_1, x_2) \cdot h_1^2 + \frac{\partial^2 F}{\partial x_1 \partial x_2}(x_1, x_2) \cdot h_1 h_2 + \frac{1}{2}\frac{\partial^2 F}{\partial x_2^2}(x_1, x_2) \cdot h_2^2 =$$

$$F(x_1, x_2) + \frac{\partial F}{\partial x_1}(x_1, x_2) \cdot h_1 + \frac{\partial F}{\partial x_2}(x_1, x_2) \cdot h_2 +$$
$$\frac{1}{2}[\frac{\partial^2 F}{\partial x_1^2}(x_1, x_2) \cdot h_1^2 + 2\frac{\partial^2 F}{\partial x_1 \partial x_2}(x_1, x_2) \cdot h_1 h_2 + \frac{1}{2}\frac{\partial^2 F}{\partial x_2^2}(x_1, x_2) \cdot h_2^2].$$

Note that last three (quadratic) terms in fact represent $\frac{1}{2}$ of the quadratic form determined by Hessian matrix. So

$$F(x_1 + h_1, x_2 + h_2) \approx P_2(h_1, h_2) =$$
$$F(x_1, x_2) + D^1 F(x_1, x_2) \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \frac{1}{2}(h_1, h_2) \cdot D^2 F(x_1, x_2) \cdot \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

where $D^1 F(x_1, x_2)$ is the gradient vector at (x_1, x_2) and $D^2 F(x_1, x_2)$ is the Hessian matrix at (x_1, x_2).

For a function $F(x_1, ..., x_n)$ the second order Taylor polynomial looks as

$$F(x_1 + h_1, \ \dots \ , x_n + h_n) \approx P_2(h_1, \ \dots \ , h_n) =$$
$$F(x_1, \ , \dots \ , x_n) +$$
$$\sum_{k=1}^n \frac{\partial F}{\partial x_k}(x_1, , \ \dots \ , x_n) \cdot h_k + \frac{1}{2}\sum_{k=1}^n \sum_{s=1}^n \frac{\partial^2 F}{\partial x_k \partial x_s}(x_1, \ , \dots \ , x_n) \cdot h_k h_s.$$

General Taylor polynomial

$$P_n(h_1, \ \dots \ , h_n) = F(x_1, \ , \dots \ , x_n) +$$
$$\sum_{k=1}^n \sum_{i_1 + \dots + i_n = k} \frac{1}{i_1! \cdot \dots \cdot i_n!} \frac{\partial^k F}{\partial x_1^{i_1} \partial x_2^{i_2} \ \dots \ \partial x_n^{i_n}}(x_1, \ , \dots \ , x_n) h_1^{i_1} \cdot \ \dots \ \cdot h_n^{i_n}.$$

Example. Let $f(x) = ax^2 + bx + c$ be a 1-variable quadratic polynomial. Let us show that the Taylor polynomial of f about the point $x^* = 0$ (i.e. the MacLaurin polynomial) of degree ≥ 2 is f itself.

Indeed, let us calculate the coefficients of Taylor polynomial

$$P_n(x) = f(0) + f'(0) \cdot x + \frac{1}{2!} f''(0) \cdot x^2 + \frac{1}{3!} f''(0) \cdot x^3 + \ \dots \ + \frac{1}{n!} f^{[n]}(0) \cdot x^n$$

of f:

$f(0) = c;$
$f'(0) = (2ax + b)_{x=0} = b;$
$f''(0) = 2a;$
$f''''(0) = 0;$
...
$f^{[n]}(0) = 0.$

Thus

$$P(x) = c + bx + \frac{1}{2} 2ax^2 + 0 + \ \dots \ + 0 = f(x).$$

Try to prove* the same for a polynomial of degree 2 of two variables: If

$$F(x, y) = a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + b_1x + b_2y + c$$

then its Taylor polynomial about the point $(0, 0)$ of degree $n \geq 2$ is F itself.

Example. For a given function (enough differentiable) $f(x)$ find a polynomial $P(x)$ of degree 3 such that

$$f(0) = P(0), \ f'(0) = P'(0), \ f''(0) = P''(0), \ f'''(0) = P'''(0).$$

Solution. Suppose this polynomial is

$$P(x) = d + cx + bx^2 + ax^3,$$

let us find the coefficients a, b, c, d.
$P(0) = d$, so $P(0) = f(0)$ gives $d = f(0)$.
$P'(0) = (c + 2bx + 3ax^2)_{x=0} = c$, so $P'(0) = f'(0)$ gives $c = f'(0)$.
$P''(0) = (2b + 6ax)_{x=0} = 2b$, so $P''(0) = f''(0)$ gives $b = \frac{1}{2}f''(0)$.
$P'''(0) = 6a$, so $P'''(0) = f'''(0)$ gives $a = \frac{1}{6}f'''(0)$.
Finally we get

$$P(x) = f(0) + f'(0)x + \frac{1}{2} f''(0)x^2 + \frac{1}{6} f'''(0)x^3.$$

Try to solve*: For a given function (enough differentiable) $f(x)$ find a polynomial $P(x)$ of degree n such that $f^{[k]}(0) = P^{[k]}(0)$ for $k = 0, 1, 2, \ \dots \ , n$.
Good lack Mr MacLaurin!

Formulate* the similar statement for 2-variable functions, and prove it at last for $n = 2$.

Exercises

1. Sketch (well, first in positive orthant $x \geq 0,\ y \geq 0$) level curves for the following functions

$(a)\ f(x,y) = x^2 + y^2, \quad (b)\ f(x,y) = |x| + |y|, \quad (c)\ f(x,y) = x \cdot y,$
$(d) f(x,y) = max(x,y), \quad (e)\ f(x,y) = min(x,y).$

2. Let $f(x_1, x_2) = 3x_1x_2^2 + 2x_1$ and $x(t) = (x_1(t), x_2(t))$ be a curve given by $x_1(t) = -3t^2,\ x_2(t) = 4t^3 + t$.

(a) Use the substitution and direct differentiation to compute the rate of change of the composite $f(x_1(t), x_2(t))$.

(b) Use the chain rule to compute the same rate. Compare the answers of (a) and (b).

3. Find a point on the curve $x(t) = (e^t + 5t^2, t^4 - 4t)$ where the tangent vector is parallel to x axis.

4. In what direction should one move from the point $(2,3)$ to increase $4x^2y$ most rapidly? Present the answer as a vector of length 1.

5. Consider the function y^2e^{3x}. In which direction should one move from the point (0,3) to increase most rapidly. Present the answer as a vector of length 1.

6. Compute the directional derivative of $f(x,y) = xy^2 + x^3y$ at the point $(4,-2)$ in the direction $(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}})$.

7. A production function $Q = F(K,L)$ obeys the *law of diminishing marginal productivity* if $\frac{\partial F}{\partial K} > 0$ but $\frac{\partial^2 F}{\partial K^2} < 0$ and $\frac{\partial F}{\partial L} > 0$ but $\frac{\partial^2 F}{\partial L^2} < 0$.

For what values of parameters the Cobb-Douglas function $AK^\alpha L^\beta$ obeys this law?

8. Write third order Taylor polynomial of a function $F(x,y,z)$.

9. Compute the Taylor approximation of order two of the Cobb-Douglas function $F(x,y) = x^{1/4}y^{3/4}$ at $(1,1)$. Estimate the value $F(1.1, 0.9)$ with order one and order two Taylor approximations.

Exercises 14.11-14.17, 14.18-14.20, 14.23-14.27, 30.11-30.15.

Homework

Exercises 14.17, 14.19, 14.20, 14.27, 30.13 from [SB]

Short Summary
Gradient

The **tangent vector** for a curve $x(t) = (x_1(t),\ ...\ , x_n(t))$ is the vector $(x_1'(t),\ ...\ , x_n'(t))$.

Chain rule

$$\frac{df(x_1(t),\ ...\ , x_n(t))}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \ ...\ + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.$$

Gradient of a function $F(x_1,\ ...\ , x_n)$:

$$\nabla F(x_1,\ ...\ , x_n) = (\frac{\partial F}{\partial x_1}(x_1,\ ...\ , x_n),\ ...\ , \frac{\partial F}{\partial x_1}(x_1,\ ...\ , x_n)).$$

Directional derivative of F in direction of a unit vector v at a point x^*

$$D_v F(x^*) = \nabla F(x^*) \cdot v = \frac{\partial F}{\partial x_1}(x^*) \cdot v_1 + ... + \frac{\partial F}{\partial x_n}(x^*) \cdot v_n.$$

The **gradient** vector at x^* points into the direction in which F increases most rapidly, and is orthogonal to the (tangent of) level curve.

Jacobian of a function $F : R^n \to R^m$: If $F(x) = \begin{pmatrix} f_1(x_1, ..., x_n) \\ .... \\ f_m(x_1, ..., x_n) \end{pmatrix}$, its Jacobian is

$$DF(x) = J_F(x) = \frac{\partial(f_1,\ ...\ , f_m}{x_1,\ ...\ , x_n} = \frac{\partial F}{\partial x}(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & ... & \frac{\partial f_1}{\partial x_n}(x) \\ ... & ... & ... \\ \frac{\partial f_m}{\partial x_1}(x) & ... & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}.$$

Second order Taylor

$F(x_1 + h_1,\ ...\ , x_n + h_n) \approx F(x_1,\ , ...\ , x_n) + \sum_{k=1}^n \frac{\partial F}{\partial x_k}(x_1, ,\ ...\ , x_n) \cdot h_k +$
$\frac{1}{2} \sum_{k=1}^n \sum_{s=1}^n \frac{\partial^2 F}{\partial x_k \partial x_s}(x_1, ,\ ...\ , x_n) \cdot h_k h_s.$